INTEGRATION OF SOCIAL NETWORKS IN A LARGE-SCALE TRAVEL BEHAVIOR MICROSIMULATION

A dissertation submitted to
ETH ZURICH
for the degree of
Doctor of Sciences

presented by

JEREMY HACKNEY

Master of Science in Technology and Policy,
Massachusetts Institute of Technology

Prof. Kay W. Axhausen, examiner
Prof. Eric J. Miller, co-examiner

2009
Acknowledgements, dedication, and thanks

In memory of my grandfather, Robert A. Kloss, who inspired by his example of hard work, community service, humor and modesty.

Swiss National Fund support, "Travel behaviour in a dynamic spatial and social context," SNF Project Number 105212-112482.

VW Stiftung support, "Travel impacts of social networks and networking tools".

I am indebted to Antonio Paez, Juan Antonio Carrasco, Elenna Dugundji, Theo Arentze, and Fabrice Marchal for discussions and international collaboration supporting the notion of integrating simulations of social networks with transportation behavior, for co-organizing the Frontiers in Transportation workshop in 2005 and 2007 at which my work first took shape, and for promoting discussions in the wider transportation planning field on social interactions and transportation behavior.

Lars-Erik Cederman's courses on multi-agent simulation of social systems provided the impulse for the project and a rich introduction to the history and use of agent-based systems.

I thank Marcel Rieser and Michael Balmer for supporting the coding development within MATSim and, along with David Charypar, for discussions on the meaning of agent behavior and complex adaptive systems.
# Table of Contents

Acknowledgements, dedication, and thanks ................................................................. 3

Abstract .......................................................................................................................... 15

Kurzfassung .................................................................................................................. 17

1 Introduction ............................................................................................................. 19
   1.1 Independent versus networked travellers ....................................................... 19
   1.2 Social network ................................................................................................. 19
   1.3 The challenges of using social networks ...................................................... 20
   1.4 The use of microsimulation ........................................................................... 21
   1.5 The goal: exploratory toolbox and verification ............................................ 22
   1.6 Organization of the dissertation .................................................................... 23

2 Social networks and transportation ................................................................... 24
   2.1 Definition of social network ........................................................................... 24
      2.1.1 What is a relationship? ........................................................................... 25
      2.1.2 What is meant by "social network dynamics"? ..................................... 26
      2.1.3 The networked traveller ....................................................................... 26
   2.2 Properties of real social networks ................................................................. 27
      2.2.1 Topology .................................................................................................. 28
      2.2.2 The geography of social travel ............................................................... 31
      2.2.3 Social network degree in the travel context ......................................... 37
      2.2.4 Other measures of social networks in the travel context .................... 38
   2.3 Microsimulations of social networks and travel ............................................ 43

3 Microsimulation of social systems and social networks ................................... 44
   3.1 Explaining spatial processes .......................................................................... 45
   3.2 Using multi-agent models ............................................................................... 46
      3.2.1 What is an agent? .................................................................................. 46
      3.2.2 Choosing what an agent represents ...................................................... 46
      3.2.3 Complex systems ............................................................................... 47
      3.2.4 Bracketing the system's response ......................................................... 48
   3.3 The simulation of social networks ................................................................. 48
5.6.2 Mathematical formulation ............................................................................. 83
5.6.3 The scope of the decision being represented ............................................. 84
5.6.4 Form of the social interaction utility term ................................................... 85
5.7 Replanning strategy for cooperative agents .................................................... 86
5.8 Output and analysis tools .................................................................................. 87
  5.8.1 Statistics package ..................................................................................... 87
  5.8.2 Pajek output .............................................................................................. 91
  5.8.3 R statnet and geographic analysis code .................................................... 91
5.9 Sample implementation ..................................................................................... 91
6 Verification experiments .................................................................................... 96
  6.1 Assumptions and fundamental structure of the experiments ......................... 96
  6.2 Hypotheses .................................................................................................... 97
    6.2.1 Social ties ............................................................................................... 97
    6.2.2 Insights into transportation behavior microsimulation ............................ 99
    6.2.3 Social geography ................................................................................... 100
  6.3 Experiments .................................................................................................... 101
  6.4 Details of the Experiments ............................................................................. 105
    6.4.1 SocialnetInit ............................................................................................ 110
    6.4.2 SocialnetEvolve ...................................................................................... 112
    6.4.3 Social exchange ...................................................................................... 112
    6.4.4 Scoring ..................................................................................................... 113
    6.4.5 Replanning ............................................................................................... 117
  6.5 Scenario used for the experiments .................................................................. 118
  6.6 Hypothesis testing by comparing the verification experiments ..................... 119
    6.6.1 Baseline sensitivity: random seed effects ................................................. 120
    6.6.2 Baseline social contact: mapping of face to face meetings in a model without 
a social network ................................................................................................. 120
    6.6.3 Time synchronization with a static social network and a utility reward for 
socializing ............................................................................................................ 121
    6.6.4 Time synchronization with a socially dense static social network and utility 
reward for socializing ........................................................................................... 121
    6.6.5 Baseline location choice: unconstrained secondary location choice of all 
available facilities ............................................................................................... 122
6.6.6 Secondary location choice with exchange of location knowledge on a static social network and standard utility ............................................................. 122

6.6.7 Secondary location choice with a spatially contracted static social network and standard utility ..................................................................................... 123

6.6.8 Secondary location choice with exchange of location knowledge on a static social network and a utility reward for socializing ...................................... 123

6.6.9 Secondary location choice with an evolving social network and no utility reward for socializing .................................................................................. 124

6.6.10 Secondary location choice with an evolving social network and a utility reward for socializing .................................................................................. 124

6.6.11 Boundary conditions of the utility function and the facility opening times.. 124

6.6.12 Verification experiments not carried out ..................................................... 126

7 Results of the experiments .............................................................................128

7.1 Establishing the base sensitivity of the MATSim system ................................. 128

7.2 Descriptive summary of the geographically embedded social networks........... 131

7.2.1 Social networks summary ........................................................................ 131

7.2.2 Social network degree distributions.......................................................... 136

7.2.3 The spatial extent of ego nets ................................................................... 139

7.2.4 Spatial density of social networks: dependence of ego net radius on agent degree ........................................................................................................ 141

7.2.5 Population density effects on social network ........................................... 144

7.2.6 The social context of establishing ties........................................................ 154

7.2.7 Agent degree versus number of activities in plan ...................................... 156

7.2.8 The effect of the degree on knowledge ...................................................... 158

7.3 Travel behavior ........................................................................................... 161

7.3.1 Average utility score of the experiment .................................................. 162

7.3.2 Chain length, trip distance, distance to acts .............................................. 163

7.3.3 Trip duration, speed ................................................................................ 163

7.3.4 Time profiles of agents travelling............................................................. 163

7.3.5 Time profiles of agents at activities ......................................................... 174

7.3.6 Distribution of plan distances (sum of trip distances).............................. 180

7.3.7 Distribution of the average distance to activities ....................................... 182

7.3.8 Average distance to activities versus number of activities in plan .......... 183

7.3.9 Summary of the consumption of space and travel .................................... 184

7.3.10 Quantification of facility popularity ......................................................... 185
7.3.11 Activity duration ................................................................. 189

7.4 Social travel and face to face interactions .............................................. 191
  7.4.1 Summary of the number of alters encountered at leisure activities .... 194
  7.4.2 Summary of the duration of face-to-face leisure encounters .......... 194
  7.4.3 Scoring the social interactions ...................................................... 194
  7.4.4 Number of agents encountered by activity type .......................... 195
  7.4.5 Number of alters encountered by activity .................................. 197
  7.4.6 Duration of face to face social interactions by activity type ........... 202
  7.4.7 Face to face group size and time of day ..................................... 208
  7.4.8 Distances travelled to participate in the different activities .......... 222
  7.4.9 Number of agents meeting: group size and distance travelled ....... 223
  7.4.10 Interdependence of distance travelled and agent degree: location choice 230

7.5 Performance ............................................................................. 234

8 Discussion of experimental results in the light of hypotheses and model verification ............................................................................... 236
  8.1 Comparison of the experimental output .......................................... 236
    8.1.1 Baseline sensitivity: random seed effects .................................. 236
    8.1.2 Baseline social contact: mapping of face to face meetings in a model without a social network ..................................................... 237
    8.1.3 Time synchronization with a static social network and a utility reward for socializing ................................................................. 238
    8.1.4 Time synchronization with a socially dense static social network and utility reward for socializing .............................................. 239
    8.1.5 Baseline location choice: unconstrained secondary location choice of all available facilities ......................................................... 240
    8.1.6 Secondary location choice with exchange of location knowledge on a static social network and standard utility .................................. 241
    8.1.7 Secondary location choice with a spatially contracted static social network and standard utility ...................................................... 243
    8.1.8 Secondary location choice with exchange of location knowledge on a static social network and a utility reward for socializing .................. 246
    8.1.9 Secondary location choice with an evolving social network and no utility reward for socializing .................................................... 248
    8.1.10 Secondary location choice with an evolving social network and a utility reward for socializing .................................................... 251
    8.1.11 Boundary conditions of the utility function and the facility opening times . 252
List of Figures

Figure 1  Long and short-term decisions of the networked traveller ..................27
Figure 2  Inverse cube dyadic tie probability as a function of distance between zones of telephone calls .................................................................33
Figure 3  Distribution of distance between alters’ home locations in a survey of real social relationships (up to 100km) .................................................36
Figure 4  The tie probability with distance, population distribution with distance, and ratio in a survey of real social relationships .........................................37
Figure 5  The degree distribution of a survey of personal trust relationships ..........38
Figure 6  Inverted encounter frequency of personal trust relationships ..................41
Figure 7  Encounter frequency and type of relationships from a survey of personal trust relationships .............................................................42
Figure 8  Approaches for representing social relationships in a network ............52
Figure 9  Structure of the MATSim Controller with iterative evolutionary algorithm for utility maximization ............................................................57
Figure 10 Structure of MATSim with integrated social networks module ..........68
Figure 11 Pseudocode for pruning facility knowledge ......................................76
Figure 12 A TimeWindowMap for activities of a single type at a given facility, and the spatial interaction statistics: Statistics are summarized for D’s act. C and D are friends .................................................................78
Figure 13 The distance-dependent probability weighting of a social tie ...............111
Figure 14 Marginal rates of substitution for additional hours of activity duration versus an additional friend present at the activity ...............................116
Figure 15 Isouility surface of the hours of activity duration versus the number of friends present at the activity .........................................................117
Figure 16 Number of agents travelling for the ensemble travel models ............130
Figure 17 Final social network degree distributions by experiment ..................137
Figure 18  Distributions of the average distance between the home location of an ego and that of all of its alters for six social network experiments ...........................................140

Figure 19  Boxplots of the average distance between the home location of an ego and that of all of its alters versus the agent’s degree .........................................................143

Figure 20  Spatial distribution of population density in 3km x 3km square zones in the Zurich scenario ..................................................................................................................145

Figure 21  Personal social ties embedded in space and geographically aggregated to 3km x 3km grid: Experiment 5 ..................................................................................................148

Figure 22  Personal social ties embedded in space and geographically aggregated to 3km x 3km grid: Experiment 2 ..................................................................................................149

Figure 23  Personal social ties embedded in space and geographically aggregated to 3km x 3km grid: Experiment 6 ..................................................................................................150

Figure 24  Personal social ties embedded in space and geographically aggregated to 3km x 3km grid: random12 social network ..................................................................................151

Figure 25  Personal social ties embedded in space and geographically aggregated to 3km x 3km grid: Experiment 102 ..........................................................................................152

Figure 26  Personal social ties embedded in space and geographically aggregated to 3km x 3km grid: Experiment 104 ..........................................................................................153

Figure 27  Context of the social ties .................................................................................................................................156

Figure 28  Boxplots of the agent degree versus the number of activities in the agent's plan 157

Figure 29  The number of activities known about versus number of alters in ego net 159

Figure 30  Time profiles of the number of agents travelling .................................................................................................166

Figure 31  Time profile of the number of agents at activities .................................................................................................176

Figure 32  Activity chain length distribution .......................................................................................................................181

Figure 33  Average trip length (distance between activities, km) per number of activities .................................................................................................................................182

Figure 34  Distribution of Euclidean distance to activities .................................................................................................183
Figure 35   Distribution of distance to activities versus the number of activities per plan

Figure 36   Number of agents choosing each leisure facility versus the facility ID for models 1, 7, 5, and 6........................................................................................................188

Figure 37   Average duration of activities by experiment (home activity not shown for clarity). The duration for peak marginal utility is shown at left for comparison.................................................................191

Figure 38   Average number of agents other than the ego co-present at the different activities (leisure on the scale at right) .................................................................196

Figure 39   Average number of an ego's alters co-present at the different activities .................................................................198

Figure 40   The number of alters encountered at all activities ................................................201

Figure 41   The total duration of encounters with alters at all activities .................................204

Figure 42   Distribution of face to face duration with friends (friend-hrs) at leisure activities:................................................................................................................206

Figure 43   Distribution of face to face utility reward at leisure activities: ..........................208

Figure 44   Number of other agents encountered at all activities versus time of day 211

Figure 45   The number of alters encountered at all activities versus time of day ........216

Figure 46   Number of alters present at the same time and leisure activity versus the time of day of the activity ..........................................................219

Figure 47   Number of non-alter agents present at the same time and place versus the average distance the agents travelled to the activity .................................225

Figure 48   Number of alters present at the same time, activity, and place versus average distance the alters travelled to the activity ........................................227

Figure 49   Boxplots of the average distance between the home location of an ego and all of its activities versus the agent's degree ........................................232
### List of Tables

**Table 1**  Summary statistics of graphs that emerge from social processes (social networks) ..........................................................30

**Table 2**  Model fits relating dyadic tie probability to distance, various datasets ........34

**Table 3**  Model fits by Butts (2000) of the data in Latane *et al.* (1995) relating dyadic activity to distance, various datasets ..........................................................35

**Table 4**  MATSim objects relevant to the social networks simulation ......................58

**Table 5**  MATSim social networks objects ................................................................70

**Table 6**  Contents of the file "edge.txt" ..................................................................89

**Table 7**  Contents of the file "agent.txt" ..................................................................89

**Table 8**  Contents of the file "graph.txt" ..................................................................90

**Table 9**  Simple social network configuration file ..................................................92

**Table 10**  Experimental test battery ....................................................................103

**Table 11**  Comparisons between the experiments ................................................104

**Table 12**  Configurations of the verification experiments for the social network interactions and the joint utility function .................................................................106

**Table 13**  Configurations of the verification experiments that test the assumptions of boundary conditions which affect the utility valuations of activities ........108

**Table 14**  Configurations of the verification experiments that use knowledge of friends as the choice set for location choice ..................................................109

**Table 15**  Settings for the standard experiment ....................................................114

**Table 16**  Possible socializing parameters for tests ...........................................115

**Table 17**  Travel behavior summary of seven social network travel models (standard deviation of ensemble in parentheses) ..................................................129

**Table 18**  Social network statistics .....................................................................132

**Table 19**  Summary of the geographical embedding of the social networks; values averaged over the population .................................................................134
<table>
<thead>
<tr>
<th>Table 20</th>
<th>Summary travel statistics averaged over all agents's active plans and/or all trips</th>
<th>162</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 21</td>
<td>Number of different facilities used in the final plans in the models with secondary location choice and the variance in the number of agents choosing the facilities in their plans</td>
<td>186</td>
</tr>
<tr>
<td>Table 22</td>
<td>The aggregate statistics of social travel indicating the effect of the algorithms on spatial (face to face) agent interaction</td>
<td>193</td>
</tr>
<tr>
<td>Table 23</td>
<td>The average distance from home of all agents co-present at the different activities (km)</td>
<td>223</td>
</tr>
<tr>
<td>Table 24</td>
<td>Average run times of the iteration steps and total run time for 500 iterations for representative runs (seconds)</td>
<td>235</td>
</tr>
</tbody>
</table>
Abstract

This dissertation presents a general multi-agent microsimulation of activity-travel coupled with a simulation of agent socializing mediated by social networks, to be used to evaluate hypotheses about coupled social influences and travel behavior. Simulation is used because data about social interactions and activity-travel behavior is too scarce to enable studies of social/travel interdependence. Instead, a world is built in-silica in which experimental activity-travel output is generated, or grown, using behavioral hypotheses as input. The system can also be used in the opposite way to construct social network topologies that may lie at the basis of observed activity-travel behavior. The social networks which are used or which are generated by the system are different from those used in previous work because they are embedded in geography and can co-evolve with the particular activity-travel patterns of the agents.

The social networks are used to guide agent interactions in space (face to face meetings) or via other communication media (exchange of information). Examples are illustrated for verification purposes in which agents may exchange of information about locations where activities may take place, where they receive different utility rewards for face to face meeting, where they can forge new social relationships and break others, and combinations of these phenomena. Emphasis is placed on measures which may be useful to detect social influences on activity-travel decisions.

The social network mechanism does provide measurable emergent coordination of activity travel in time and space, which is not possible with a centrally-governing mechanism. Information sharing mechanisms about activity locations results in heterogeneous clumping of demand centered on different activity locations, i.e. an emergent activity attribute can be generated that is based on social interaction. The model system, with its focus on calculating short-term traffic flows, is however not well-suited to constructing comprehensive geographic social networks, which would require a model system that incorporates the passage of time or some other method of appropriately allocating the value of maintaining social connections and socializing, versus other aspects of geographic location and participating in activities. However the issues involved in valuating face to face social contact or the value of a social relationship, versus travel disutility for meeting face-to-face, is not entirely different from that of valuations of any activity. The short versus long term gains, costs, and consequences of these decisions, as well as the time horizon of dynamics like learning, adapting, forging new friendships and forgetting old ones, learning about new activities and locations, are similar modelling problems whether it regards activity travel habits or social circles.
The calculation overhead in the results presented is approximately double the memory used in standard MATSim experiments using independent agents, depending on the average number of social relationships per agent, and run times of up to twice as long, depending on the number of social relationships and what interactions are to be simulated.

Rather than validation of particular emergent social networks or travel patterns against observations, the emphasis of the methodology is to establish measures with which to study the dynamics of the model system's components as social interactions are integrated with the travel microsimulation. The output is analyzed statistically, using a model without socialized agents as a reference. Information exchanges between agents, representing the result of social interactions long-term relationships, utility rewards for face-to-face socializing, and dynamic social ties which evolve with activity-travel habits, all have ranges in which their effects on the travel demand simulation can be detected and characterized according to the social mechanism that is in place in the model. However, it would not always be obvious from the results that a social mechanism was at work producing the outcomes, if this was not known in advance of making the analysis. This points to the need for more study of social interactions and activity travel behavior in order to be able to correctly estimate travel models while accounting for decision maker interdependencies.
Kurzfassung


Das soziale Verhalten richtet sich nach den sozialen Netzen entweder im Raum (face-to-face meetings) oder via anderen Kommunikationsmitteln (Informationaustausch). Beispiele für Verifikationszwecke sind erläutert, in denen die Agenten Informationen über Standorte für die Ausführung von Aktivitäten miteinander austauschen dürfen, in denen sie verschiedene Nutzengewinne für "face to face" Treffen erzielen, in denen sie neue Beziehungen schliessen und alte auflösen können, und Kombinationen dieser Ansätze. Der Schwerpunkt liegt bei der Evaluierung von Kenngrössen, die für das Aufspüren und die Erkennung von sozialen Einflüsse auf Aktivitäts- und Verkehrsverhalten nützlich sein könnten.

Der soziale Netzwerkmechanismus fördert in der Tat messbare Koordination zwischen den Agenten, welches in dieser Ausprägung nicht durch ein zentralgesteuerten Mechanismus entstehen würde. Mechanismen, die den Austausch von Informationen über Standorte ermöglichen, ergeben heterogenes "Klumpen" von Nachfrage für Aktivitätsstandorte sowie - Zeiten, die in wenigen Standorten sowie –Zeiten konzentriert ist: ein Attribut der Standorte (Beliebtheit) ergibt sich dabei auf dynamischer und nicht zentralgesteueter Weise in der Simulation. Das Modellsystem ist jedoch nicht geeignet für die allgemeine Konstruktion von geographischen sozialen Netzen aufgrund seines kurzen Zeithorizonts. Längere Zeithorizonte (Lebensgeschichten) für die Agenten und deren Lebensplanung wären in der Modellierung nötig, oder sonst wäre eine Methode angesagt, um der Nutzen langfristiger Anliegen, wie z.B. die Aufrechterhaltung sozialer Beziehungen, in einem kurzfristigen Zeithorizont, in dem die Aktivitätsstandortwahl, die Routenwahl, sowie die Aktivitätsdauer bestimmt werden, bewerten zu können. Nichtdestotrotz unterscheiden sich diese kurz- und langfristigen

Der zusätzliche Rechenaufwand in den hier vorgeführten Resultaten der sozial vernetzten Agenten ist bis auf eine quasi Verdoppelung der virtuellen Speicher eines Standard-MATSim-Experiments mit unabhängigen Agenten, abhängig von der Anzahl sozialer Beziehungen pro Agent, sowie eine Laufzeit von bis zu zwei Mal so lang, abhängig von der Anzahl und Typ der sozialen Interaktionen, die möglich sind zwischen den Agenten.

Anstatt die sich ergebenden sozialen Netze oder Verhaltensmuster gegenüber echten Daten (Beobachtungen) zu validieren, liegt der Schwerpunkt der Dissertation auf der Festlegung von Kenngrössen, mit denen die Dynamik der in ein einziges Modellsystem zusammengekoppelten Sozial- und Verkehrsmikrosimulationen komponentenweise studiert werden können. Das Output wird statistisch analysiert mit als Maßstab einem Model ohne sozialisierten Agenten (unabhängigen Agenten). Die Modelle für Informationenaustausche zwischen Agenten, die das Ergebnis sozialer Interaktionen über längere Zeithorizonte repräsentieren, für Nutzenzusätze für "face to face" Treffen, sowie für dynamische soziale Beziehungen, die sich mit den Aktivitäts- und Verkehrsverhaltensmustern ändern, haben gewisse und eigenartige Auswirkungen auf die Verkehrsnachfrage, die erkannt und erläutert werden können, und auf den sozialen Mechanismen in dem Modell zurückgeführt werden können. Diese Hinweise deuten auf die wichtige Rolle von weiterer Forschung im Bereich sozialer Einflüsse auf Aktivitäts- und Reiseverhalten, sowie umgekehrt, um Verkehrsmodelle richtig bestimmen zu können um die relevanten sozialen Einflüsse zu berücksichtigen.
1 Introduction

1.1 Independent versus networked travellers

Travel utilities (and decisions) are almost always modelled as individually independent across decision makers: a function of the traveller's own (or household) characteristics and the attributes of modes, routes, and destinations. The fact that people plan trips and activities jointly, depending on the trips and activities of other people, has been ignored with few exceptions (Kurani and Kitamura, 1996). This compromise has been reached consciously and with good reason, including convenient econometric model estimation tools (e.g. Ben Akiva and Lehrman, 1985) and the lack of datasets detailing decision-maker interactions (Manski, 2000).

Representing behavior as entirely independent decisions succeeds as long as it considers regularly occurring, relatively inelastic equilibrium behavior, like work (non-discretionary) trips, where the boundary conditions of the problem dominate the solution. Travel preferences in this case can be assumed to be explained by the sociodemographic characteristics of the traveller and by generalized travel costs; any deeper processes of orchestrating the trip take either a long-term character exogenous to the model (Hensher, 2003), i.e. choice of home/work location, or they play an insignificant role due to the relatively limited flexibility of the traveller. More flexible (discretionary) behaviors are explained by adding variables like taste heterogeneity and cohort or habitual behavior effects, either as random coefficients or based on panel datasets or assumed distributions. While such extensions improve the statistical specification and the conceptual validity of the models somewhat, they neglect accounting for traveller/decision maker interactions. This rules them out for studying important processes of real behavior, and may limit their usefulness in prediction, as well. That is, the model output without interactions may be plausible, but it may be for the wrong reason (Shy, 2001). The distinction is important because it may have complex, non-linear consequences in the society as a whole, as a result of cascading influence from one social group across individuals to their social groups, and so on. Thus the traveller, and the transportation system, is not just affected by immediate friends and family, but through those who have influenced them, in turn: a wider network of influences.

1.2 Social network

A social network (graph) is an abstract mapping used to visualize and to calculate the effects of such relationships. It is a collection of dots (nodes or vertices), each representing an entity
like a person, household, activity, etc., which are connected by lines (edges or links) indicating a cause, effect, or other relationship between the dots. The important feature of a graph as a modelling tool is that the behavior and salient features of the graph as a whole can be quite different from the behavior or the microstructure around any node, and can therefore be used to describe many emergent (non-centrally driven) phenomena: for example, in a Potts or spin-glass model, a dot represents a particle which can take on one of two "spin" values: "up" or "down" (1 or -1), depending on the average spin of dots connected to it by lines. Depending on the topology (arrangement and number) of the lines, the entire graph, representing a crystal, may exhibit a majority of stably aligned spins (a permanent magnet), randomly aligned spins (non-magnetic material), or regionally aligned spins which are unstable (non-permanent magnet). Graphs are used increasingly in social science in similar models when group behavior can be directed by individuals' behavior or the behavior of small regions of influence, as opposed to a centralized global controlling function, such as studying the changing trends in attitudes toward public policy issues. The patterns of lines and dots can be mathematically analyzed to yield descriptive statistics of the patterns, which have correlations to expected emergent characteristics of the collective of individuals, or used in models explicitly depicting flows or neighbor-to-neighbor influences, as above, to study the complex emergent behaviors of the social group.

1.3 The challenges of using social networks

A reciprocal interaction between travel/communication and social networks is postulated: Both communication and person-to-person encounters are crucial for the maintenance of social networks and the social capital (the realizable value in the relationships, Freeman, 1977) that they enable (Larsen et al., 2006). Social networks generate communication/travel, limit the individual's travel choices by constraining schedules, and drive habits or "sunk" investments like the choice of residential, and work locations, which are (socially, professionally) expensive to change once established. Who is participating in the activity-travel, and the type of relationship, can influence trip timing, destination, expenditures, mode, and cost sharing of the activity and/or trip. On the other hand, communication/travel opportunities open up with the discovery of new locations (expanding the choice set), experiencing alternative time slots for activities (where other people might be encountered), and can alter the spatial and topological character of the social network, leading via social exchanges to yet other new experiences, knowledge, and habits (or, reduced mobility can cause the opposite (e.g. Grieco, 1995).

Endogenizing interactions in conceptualizations of transportation behavior amounts to explicitly accounting for who influences whom, when and how much. This new dimension may help resolve spatio-temporal behaviors that would otherwise be masked by
sociodemographic explanations, leading to better conceptual and predictive models. But there is a major drawback to methods using social interactions, because it becomes necessary to know the specific social connections (social network), as well as the direction and strength of influence on an individual's decision making that is communicated by this network, and finally to know all of these relationships. This places high demands on data collection and imputation models, as well as behavioral assumptions.

For example, a sociologist observing individuals interacting is faced with the problem of determining to what extent the correlated decisions between an individual and a group is a result of a process of choosing peers with the same attitudes (homophily = liking like, McPherson et al., 2001), versus a process by which members of a group influence one another to conform to some group norm: the so-called reflection problem (Bramoullé et al., 2009; Manski, 1993) This metaphor refers to the difficulty an observer has in determining whether a person moves his image in a mirror or vice versa: without knowing the relationship between the two people in view, i.e. which one is a reflection and which is real, the observer cannot separate cause and effect. In order to correctly represent how the society functions, one needs precise knowledge about, or hypotheses of, social topologies as well as the dynamic processes at work in the system. The nature of these relationships are rarely observed (Manski, 2000).

1.4 The use of microsimulation

Study of interdependent travellers is therefore desirable, but spatial observations of social interactions in the geographical or transportation context are uncommon, because they are expensive to make, burdensome on participants, and the cooperation rates of subjects seem to be lower than for conventional travel surveys (Kowald, 2009). Compromises have been made in scope and extent with respect to geography, activities, sociodemographics, geodesic (graph) distances, and number of people studied. Studies tend to be limited to either small social circles, very specific kinds of relationships, or to neglect specifics about travel behavior.

Surveys have an indispensable place in establishing empirical cornerstones, but are clearly inefficient in gathering large samples with rich content, or observing system responses in time. The effort lends insight only into the questions asked, and for inference or studies of complex emergent behavior, acquired samples need scaling up to a full population. In the case of travel behavior and social influences, imputation may imply a complex model of geography and time-space behavior.

Because the phenomena of travel behavior are emergent: the group result is a result of actions which are not centrally directed, a "whole" is needed for study of networked, i.e. complex
propagating interdependent, decision making. Given the difficulty finding suitable data, and
the need for precise and accurate descriptions of the social networks, the controlled
evironment of microsimulation is chosen as the tool best suited for experimentation with
coupled social- and activity travel behavior. A coupled social-travel demand simulation
environment enabling "what if" queries is a way to complete the information missing from
surveys, and to more generally observe the effect of assumptions about socializing behavior
on travel behavior and vice versa.

The micro-simulation is a modular object-oriented structure using the Java-based MATSim
Toolbox (summarized in Rieser et al., 2007) and it enables the researcher to conduct and
document experiments by combining desired algorithms and/or datasets for the three main
elements of study: socializing, geography, and travel behavior, and to adjust the strength of
their coupling.

Exploratory agent modeling allows researchers to control and experiment with microscopic
behavior and observe the emergent macroscopic system (Axtell, 2000b; Bankes, 1993). The
method is a combination of deductive and inductive methods which has been called
“generative science” (Sawyer, 2004) or "consolidative modeling" (Bankes, 1993). The
usefulness of the results depends on clear explanation and validity of the assumptions and
thorough model verification.

1.5 The goal: exploratory toolbox and verification

The approach taken here is not to write a data-validated simulation representing the actual
behavior of people, their interactions, and their response to certain activity-travel situations.
Instead, this effort encompasses several goals intended to evaluate the general
microsimulation approach to the problem of interactive travel behavior. The first is to write a
computer simulation toolbox with the ability to grow geographically-embedded social
networks with any combination of topology, population density, world dimension, set of
activity locations, and transportation network. Second, to embed this social network within a
set of activity plans in the world, such that the social networks can take on the meaning of
mapping face-to-face and/or long-term contacts, or a mixture of the two. This coupling of the
social network with the daily activity plan of the agents enables two kinds of experiments:
first, comparisons of the mobility and activities of the agents within their social context (how
many friends are encountered each day?) and second, the tuning of the social network to
activity-travel behavior by mixing a component of hypothesized (or measured, or inferred)
long-term social relationships with a component of daily activity travel. Third, this work
investigates the suitability of the MATSim iterative directed relaxation framework for
incorporating social interactions. The stability, complexity, sensitivity, reproducibility, and
computing resource requirements of the simulation package with social networks are evaluated by examining the effect on the simulation of integrating the social interactions in the evaluation of plan utility, in information exchanges about activity locations, and in response to a social network which evolves based on face-to-face contacts of the agents. To these ends, the work finally introduces an initial set of measures to help understand the state of a social network embedded in a geographical and activity-travel environment.

This thesis studies the interdependence of social network properties on geography for basic social science understanding and to help design better transportation models, policy recommendations, and travel surveys. The results of the hypothesis tests might be applied to the activity spaces associated with suburban versus urban lifestyles, mobility and aging, and the change in the nature of social contacts with decreasing (increasing) transportation costs.

1.6 Organization of the dissertation

Thus the work presents several different primary research threads woven together to make a research tool, all of which will be described in turn. First, the state of the art of work done on social influences in transportation are summarized, followed by an introduction into social networks and some of the findings from the data on global and ego networks that can be used to support hypotheses for modelling. The generative modelling approach in computer simulations of social systems is explained as a way to grow large datasets of collective behavior based on initial well-founded hypotheses. The specific computing tool used for this system, MATSim, is described next, with the social behavior module and the statistical module that were added for the experiments. A test battery to verify the individual and combined effects of the social interactions on a MATSim activity-travel demand simulation is defined, with the results and their interpretation. The dissertation concludes with an assessment of the usefulness of the modelling approach and on further developments in the topic of social networks and transportation and the contribution of the microsimulation approach to understanding socially-mediated activity-travel.
2 Social networks and transportation

While there is no large dataset on activity travel and social interactions, there is an emerging literature on social networks in the context of geography which can provide sufficient cornerstones for developing models of social travel behavior.

2.1 Definition of social network

A social network (graph) is a mathematical expression referring to a set of nodes (vertices), representing people, and links (edges) representing well-defined relationships between the people. In this work, the term "global" network refers to the collection of all nodes and edges in the system, whether they are all connected together or not. The links can be valued and directed (arcs) to represent relationship strengths, and these values can change in time. The 2-person subset of a social network consisting of a node-link-node subgraph is called a "dyad". "Geodesic distance", or "shortest path", is a definition from graph theory referring to the smallest number of edges that must be crossed to jump from one node to another node: thus dyads have geodesic distance 1. In a social network in which the only metric of distance is the number of links between individuals, geodesic distance is the same as social distance. The subject being studied is called an "ego" and those he is linked to, "alters". The locus of alters and the links from the ego constitutes an "ego network" or "personal network". This work uses "friend" loosely to mean "alter", or those nodes a geodesic distance 1 from an ego.

Network statistics help discern paths of information flow and give some indication of the network's efficiency, resilience, and resistance to disruptions. The measures are counting procedures focused on identifying certain constellations of links and nodes. Some definitions will help the following discussion:

- geodesic distance or shortest path (the minimum number of edges in a chain between two nodes),
- clustering coefficient (the proportion of instances in which 3 agents are connected in a triangles relative to the number of instances in which they are connected by fewer than 3 edges),
- average shortest path length (average for each pair of nodes of the minimum number of edges connecting the two nodes),
- degree (number of links entering or exiting a node),
- component (a collection of nodes which can reach one another by edges; a network can consist of multiple components),
• main component (the largest component of a network),
• diameter (the maximum geodesic distance between all pairs of nodes which are reachable along the edges of the graph, for practical purposes limited to nodes within the same component)

The statistics defined above will be referred to here because they are the most commonly observable statistics in real social networks (Dorogovtsev and Mendes, 2003). Other commonly reported metrics include "density", the number of realized edges relative to the number of possible edges in the graph (which is \((N \times (N-1))/2\) in a non-directed graph).  

While the term “social network” will be used throughout this work to distinguish these graphs, which represent social interactions, from other graphs representing biological, chemical, physical, economic, or other processes, a more precise term in view of the hypothesis-oriented methodology might be “relational econometric” networks (Bidart and Degenne, 2005). The difference (and the specificity of the terminology) lies in the definition of relationships.

### 2.1.1 What is a relationship?

The social ties between the individuals in a social network (edges between nodes of a social network) must be well-defined for an analysis or model to have meaning. Often a relationship is defined as an observed set of circumstances between individuals: who spoke to whom, who owes whom money, which people encountered one another at a party. In the sense of observing (or surveying) behavior, the relationships are a map of past social interaction, an existing obligation or other potential avenue open for further future interactions. Thus the social tie, the relationship, indicates a gateway for the flow of some tangible or intangible good between people. For the utilitarian purposes of the social networks defined and used here, a social network is a mapping of possible agent interactions: if agents are connected by a tie, they can interact socially, and otherwise they cannot.

A graph theoretical treatment of a social network would study only the topology, or the way in which the nodes are connected by edges, usually treated as binary variables (they exist or they do not). The direction of connections and any weights they have to modify flows across them would be ignored. Social network analysis uses graph theoretical foundations to calculate topological statistics, but attaches social meaning to the statistics. It is thinkable in sociological (social) networks that a social tie, or relationship, could have describing features: a measure of strength amplifying or limiting the flow from one individual to the next; a

---

1 This unitless metric in graph space is to be distinguished from "spatial density", which is the number of individuals living in a geographical area, agents/km².
direction of who may influence (pay, promote, inform) whom, in which case the edge is called an arc; a type, such as family, work colleague, or friendship; an age or a date at which the relationship began; a distance across which the relationship is maintained, etc. These additional attributes add social explanatory value to the edge and the configuration of the edges as a whole, but do not figure in graph statistics, which are topologically-based.

2.1.2 What is meant by "social network dynamics"?

A distinction is made in this work between "static" and "dynamic" social networks. Both refer here to topology. Static means that the number and connections of the edges and nodes remains the same. Dynamic means that these structures can change. Information can be passed along the social network edges in either static or dynamic graphs. The attributes of the graph edges and nodes can also be updated in either type of graph.

2.1.3 The networked traveller

While it has long been apparent to travel demand researchers that social contacts can both constrain and induce travel (Axhausen et al., 2007), activity-based tools for understanding and modelling travel demand have not systematically incorporated social network approaches.

A framework for incorporating the social context of activity-based tools is proposed by Axhausen, Larsen, and Urry (2007), summarized in Figure 1. Friends, family, work colleagues, etc. constitute a social network of acquaintances ("alters"). In the short term, these relationships with the decision maker ("ego") conduct information and obligations, which can motivate, constrain, or substitute for travel and activities. The social network influences the allocation of time between travelling versus planning and participation in activities (see e.g. Miller, 2005). Meanwhile, the long-term effectiveness of the upkeep of social contacts determines how social influences evolve for the future. Though much social interaction occurs via synchronous and asynchronous electronic media, making geography practically moot, co-presence is vital to certain social processes and activities and thus relationships (Urry, 2003).
Figure 1 Long and short-term decisions of the networked traveller

That activities and/or travel are undertaken jointly leads to the less obvious converse: the time budgets left over for each of the involved individuals is also influenced by the joint events. The displacement of other activities in time and in space is different than if each person were acting independently instead of tied to the activities of another person. The first joint plan results in a cascade of modifying or constraining other joint plans, not just because of who is present, but who can no longer adjust his or her schedule because of prior commitments. Joint decisions cascade in time and beyond the ego network to the next ego network, across geodesic distance in the global social network.

2.2 Properties of real social networks

Modelling travel behavior based on influences of extra-household individuals in a social network has a first hurdle in the identification of a plausible set of social contacts, and if possible, their integration in geography. Available datasets on the social network structures are not directly helpful in this area. The researcher is challenged to find a social network in which the topology, geography, demographics, and travel choices of the individuals are simultaneously consistent, along with the interactions that take place on the social ties that relate to travel. Most surveys of “global” networks are understandably constrained to cases of
small groups with very specific characteristics that are not generalizable (Valente, 2006). Panel observations of social network dynamics are rarer still.

2.2.1 Topology

Graphs are complicated (also complex) structures and an orientation of theoretical graph constructs can help as a reference point for understanding what plausible.

In a "classical random", Bernoulli, or Erdős/Renyi network (synonyms), relationships are equally likely between all individuals, independent of physical distance. Random networks have Poisson distributed degree, low clustering, and short average path lengths which let information diffuse across the networks quickly and homogeneously. Connections are by definition not preferential and are equally likely to be found between any two agents. The macroscopic character of a random network is not changed by randomly or systematically perturbing the structure (removal of links or nodes).

Lattices are highly ordered structures in which every node's link pattern is identical. Depending on the average degree, lattices can have very long path lengths in which information travels less efficiently, and because of which localized heterogeneity can emerge in models of information percolation on the network. The macroscopic character of the network can be highly sensitive to local imperfections: a semiconducting material is an example of a nonconducting lattice that is interrupted with a trace impurity to become semiconducting. Perfect lattices have a single degree with no variance.

Real human social networks have properties in between these two well-studied theoretical structures. However, there is no work systematically summarizing the characteristics of socially-generated networks (Jackson and Rogers, 2005).

They seem to be locally organized into groups which share similar information, world views, needs and resources, with occasional ties to other groups which provide conduits to different characteristics, leading to exchanges (Burt, 1992; Granovetter, 1973).

Jackson and Rogers (2005) summarize three commonalities of real human social networks in the sense of graph metrics from the findings in the literature: First, society in general is a “small world”, in which a few individuals belong to a local structural order, but also to other distant ordered structures, thus providing shortcuts through society (Newman et al., 2002; Watts, 1999). Small world network statistics are characterized by a higher clustering coefficient than a classical random graph, and a comparably small average shortest path length (log (number of nodes)). While the average shortest path length has been measured repeatedly through snowball sampling (Liben-Nowell and Kleinberg, 2008; Travers and
Milgram, 1969), the clustering characteristics of real relationship networks is expensive to study and has been assumed to hold based on specific studies of small groups. These network structures are consistently present in large associations of scientists, actors, or smaller groups of powerful CEOs, for which digital databases are available (summaries in Barabasi, 2002; Kleinberg, 2000; Kleinberg, 2003; Strogatz and Watts, 1998; Watts, 1999). Second, in addition, many examples of real social networks exhibit a tendency for social links to form preferentially around nodes which already are well-connected, i.e. popular people become more popular (Barabasi, 2002), exhibiting a degree distribution between exponential (the number of nodes with degree $k$ is a function of $e^{-k}$) and scale-free (the number of nodes with degree $k$ is a function of $1/k$) (Dorogovtsev and Mendes, 2003). The frequency of observations of high degree nodes is higher than in a Poisson-distributed degree of a classical random or the small world graph constructed by Watts and Strogatz (1998). This characteristic is seen in neither a random network nor a lattice, and can be a result of networks adding nodes (growing), which add links to an existing network (such as a literature citation index), but can also result from other processes (Dorogovtsev and Mendes, 2003). In preferentially attached networks, linked nodes tend to have positively correlated degree (assortativity). Third and finally, the highly skewed distribution of degree means that clustering of the neighbors of high-degree nodes is lower than for neighbors of lower-degree nodes (sometimes referred to as a “core-periphery” structure).

Few "global" social networks have been studied because of the difficulty of observing the large number of variably interpretable edges (relationships) between the actors. Often observations are biased by practical limitations placed on the survey, such as recording only a certain maximum number of associations in order to reduce the workload on the respondent and the survey team, or limiting the survey to a certain language, activity, self- or otherwise identified social group, or spatial region. Most complete networks resulting from human interactions that have been quantitatively observed are the kind which lend themselves to being electronically monitored and stored. While they emerge from social interactions, they are also all special processes that may not relate to the kind of social networks one imagines would be significant for influencing travel decisions. Co-authorship activities, for example, depend on the size of research groups, the feasibility of collaboration on the particular topic, the freedom of the researchers to collaborate, and the reward structure and rules of the discipline and the universities or laboratories where the authorship occurs, often in relation to the stage of the researcher's career. Obviously the social behavior leading to sexual contacts among people at risk for HIV, recorded by hospital visits, is another completely different dynamic. Nevertheless, to provide an orientation of what quantities have been observed, some summary network statistics of prominent samples of social networks compiled from the literature by Newman (2003) are reprinted here in Table 1. The broad indications summarized by Jackson and Rogers (2005), above, are evident in the wide range of parameters.
Table 1 Summary statistics of graphs that emerge from social processes (social networks)

<table>
<thead>
<tr>
<th>Network Type</th>
<th>Type</th>
<th>N</th>
<th>Average Degree</th>
<th>Average shortest path length</th>
<th>Exponential parameter if exponential</th>
<th>Clustering coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sexual contacts</td>
<td>undirected</td>
<td>2,810</td>
<td>-</td>
<td>-</td>
<td>3.2</td>
<td>-</td>
</tr>
<tr>
<td>Email messages</td>
<td>directed</td>
<td>59,912</td>
<td>1.44</td>
<td>4.95</td>
<td>1.5/2.0</td>
<td>-</td>
</tr>
<tr>
<td>Student relationships</td>
<td>undirected</td>
<td>573</td>
<td>1.66</td>
<td>16.01</td>
<td>-</td>
<td>0.005</td>
</tr>
<tr>
<td>Telephone call graph</td>
<td>undirected</td>
<td>47,000,000</td>
<td>3.16</td>
<td>-</td>
<td>2.1</td>
<td>-</td>
</tr>
<tr>
<td>Email address books</td>
<td>directed</td>
<td>16,881</td>
<td>3.38</td>
<td>5.22</td>
<td>-</td>
<td>0.17</td>
</tr>
<tr>
<td>Math coauthorship</td>
<td>undirected</td>
<td>253,339</td>
<td>3.92</td>
<td>7.57</td>
<td>-</td>
<td>0.15</td>
</tr>
<tr>
<td>Physics coauthorship</td>
<td>undirected</td>
<td>52,909</td>
<td>9.27</td>
<td>6.19</td>
<td>-</td>
<td>0.45</td>
</tr>
<tr>
<td>Company directors</td>
<td>undirected</td>
<td>7,673</td>
<td>14.44</td>
<td>4.60</td>
<td>-</td>
<td>0.59</td>
</tr>
<tr>
<td>Biology coauthorship</td>
<td>undirected</td>
<td>1,520,251</td>
<td>15.53</td>
<td>4.92</td>
<td>-</td>
<td>0.088</td>
</tr>
<tr>
<td>Film actors</td>
<td>undirected</td>
<td>449,913</td>
<td>113.43</td>
<td>3.48</td>
<td>2.3</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Newman (2003) cites the sources of these datasets and analyses.

Social networks used in/emerging from transportation studies should exhibit these realistic characteristics, or else network information flows, influences, etc., will not have the proper heterogeneity or spread at plausible rates. This does not narrow the choices appreciably: In short, many social networks may be considered "plausible", only classical random networks and perfect lattices can be ruled out. Furthermore, how the clustering, geodesic distances, degree, and other characteristics of social networks relate to geography and travel behavior must still be understood.

Other social networks, summarized in the next section, have given information about the physical distance between actors' home base locations, physical distance distributions of tie- and interaction probabilities, and indications of encounter frequencies.
2.2.2 The geography of social travel

The datasets summarized above are not observations of relationships that are relevant for everyday travel. It is not clear what their relationship to travel behavior over longer horizons is, either. Neither the topology of the networks that are relevant for influencing travel decisions, nor how they evolve, are well-understood. Indeed, the set and strength of relationships that are relevant to transportation demand over different distances and time horizons is yet to be defined.

There are few datasets available to study the geography of social interactions. Besides the social ties, corresponding observations are needed of the frequency and location of face-to-face meetings, the activity types associated with these meetings, the number of participants, their relationship, the distances travelled to meet, the activity planning process, and so on. A partial survey illustrates some consistencies in the findings. The common measure of geographic embedding used in existing studies is the straight-line, or Euclidean, distance between residential locations of the alters.

Christakis and Fowler (2007) and (2008) have constructed a geocoded dynamic (panel) social network of spatial interactions of 5124 people over 35 years. The data been yielded significant parameters to describe the group context of the spread of obesity and smoking (or quitting smoking). A geographical analysis is pending. The study is limited to the alters listed as "emergency contacts" on patients admission sheets, and so the number of alters is very small and usually a family member within the household, or of course an individual in a reasonably close physical proximity who could react in the case of emergency. Rothenberg et al. (2005) is another study from epidemiology reporting distances between alters. In the survey of 595 HIV-risk persons, fifty-two percent of all dyads were separated by a distance of 4 km or less. In the main component of 348 connected respondents, almost half the individuals were between 3 and 6 steps from each other in the social network and were separated by a distance of 2 to 8 km. The mean distance between all people in the survey area (Colorado Springs) at a certain level of spatial aggregation was 12.4 km, compared with a mean distance of 5.4 km between all dyads in the study.

Axhausen and Frei (2007) summarize a wide range of studies about the geographical distribution of social contacts. Universally, face-to-face contact frequency falls with an inverse function of distance. Liben-Nowell et al. (2005) use the hometowns of participants in an online social network of bloggers to derive a small world model in which the probability of befriending a person is inversely proportional to the number of physically closer people.

Butts (2000) presents Bayesian tie-probability models to generate spatially-embedded social networks based on link probabilities determined by the joint distributions of characteristics of observed social ties. The author notes that "very little of the available data relating ties to
distance is suitable for modeling in this fashion." (page 34). Various classic spatial social network studies are summarized, and then the methods are demonstrated on real datasets from Hägerstrand (phone calls and migration datasets), Zipf (assorted datasets), Latane et al. (last memorable encounter), Bossard (marriages), Stewart (students' residential locations), and Festinger et al. (friendship in dormitory housing). The specific works are cited below. An inverse-square relationship of tie probability with distance is expected, based on long-standing arguments in sociology of the cost of searching in two dimensional space.

The definition of "social tie" is intentionally left ambiguous in the analysis because the datasets all measure a different social interaction, and the point of the study is to identify a common relationship between tie probability and spatial distance across the datasets. A best-fit model for Hägerstrand's entire set of phone call data, for example, is a General Attenuation Model in which proximate ties are roughly equally likely for a short radius, before the probability of ties drops off as a power law (however a cubic function) with distance (Figure 2).
Figure 2  Inverse cube dyadic tie probability as a function of distance between zones of telephone calls

Source: Butts (2000) Figure 6, based on the data in Hägerstrand (1967)

Simple power law models (functions of $1/r^β$) fit the other datasets better, i.e. without the initial radius of constant tie probability. Table 2 summarizes Table 10 in Butts (2000) showing the exponent for simple power law regressions of the datasets. The distance effect is both strong and consistently present. All fits are power laws in distance, with different base probabilities of a social tie (not shown in the table). Telephone contacts follow approximately an inverse cube, migration approximately an inverse square, and marriage/student ties follow a slightly less than inverse square relationship.
Table 2 Model fits relating dyadic tie probability to distance, various datasets

<table>
<thead>
<tr>
<th>Source</th>
<th>Data</th>
<th>Distance exponent $\beta$</th>
<th>N distance bins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hägerstrand (1967)</td>
<td>Total person to person calls</td>
<td>-3.514</td>
<td>24</td>
</tr>
<tr>
<td>Hägerstrand (1967)</td>
<td>Median analysis of all calls</td>
<td>-2.230</td>
<td>11</td>
</tr>
<tr>
<td>Bossard (1932)</td>
<td>Origin of marriage partner</td>
<td>-1.844</td>
<td>12</td>
</tr>
<tr>
<td>Hägerstrand (1967)</td>
<td>Total migration</td>
<td>-1.692</td>
<td>32</td>
</tr>
<tr>
<td>Stewart (1941)</td>
<td>Student residential distance from University</td>
<td>-1.659</td>
<td>6</td>
</tr>
<tr>
<td>Festinger et al. (1950)</td>
<td>Same floor</td>
<td>-0.962</td>
<td>4</td>
</tr>
<tr>
<td>Festinger et al. (1950)</td>
<td>Different floors</td>
<td>-0.914</td>
<td>4</td>
</tr>
</tbody>
</table>

Summary of Butts (2000), Table 10, page 35.
All models are of the form $y = ax^\beta + \varepsilon$

As mentioned, these samples are the best available classic sociological studies, but not ideal for the task of fitting distance models. Festinger et al.'s study limited the number of alters to a maximum of three. The Hägerstrand phone calls are aggregated to zones and each zone-zone phone call is treated as a new dyad pair in this analysis, which it certainly is not. Repeat communications between pairs of people is highly likely, but is not recorded in the data. This may be one cause of the model form as well as the unexpected inverse cube relationship with distance. Bossard's data only considers the marriage partner, which is a special kind of social search. Stewart's data is also geographically aggregated to only two dozen zones in the U.S.

Table 11 in Butts (2000) summarizes total quantities within distance rings to distance, i.e. an integral of the quantity over distance. These studies, summarized here in Table 3, tend to show strong $1/d$ relationships. Assuming a uniform population distribution, this implies inverse square laws for tie probability.
Table 3 Model fits by Butts (2000) of the data in Latane et al. (1995) relating dyadic activity to distance, various datasets

<table>
<thead>
<tr>
<th>Data</th>
<th>Exponent of distance, $\beta$</th>
<th>N distance bins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encounters/distance in Florida</td>
<td>-1.01</td>
<td>11</td>
</tr>
<tr>
<td>Encounters/distance in China</td>
<td>-1.05</td>
<td>10</td>
</tr>
<tr>
<td>Encounters/distance between academics (sociology)</td>
<td>-0.93</td>
<td>8</td>
</tr>
</tbody>
</table>

Summary of Butts (2000), Table 11, page 37

All models are of the form $y = \alpha x^\beta + \varepsilon$

The fit is of encounters/unit distance versus distance and so the exponent of the number of encounters (probability of encounter) is decreased by 1.0, indicating an inverse square relationship between encounter probability and distance.

Frei and Axhausen's (2008) survey of personal contacts (ego networks) cannot let us construct a graph, but it does provide new information on the distance between individuals engaged in a trusting relationship, their modes of communication, and the frequency with which they meet face to face. The name generator attempted to elicit biographically significant, very long-term trusted friendships around the world, and respondents could list an unlimited number. The survey captured responses from 332 egos who named 4100 alters with whom they maintain a trusting relationship. The median distance between alters is 9km and the mean is 109km (the relationships extend around the world). Figure 3 and Figure 4 show the distance distribution of the distance between alters in the study.
Figure 3  Distribution of distance between alters' home locations in a survey of real social relationships (up to 100km)

A re-analysis of the data up to 100km distance shows that the exponential decay with distance of the likelihood of a tie between two people per unit distance is -1.04 ($R^2 = 0.96$). This is entirely consistent with the fits in Table 3 and Table 2, and indicates that the name generator which tries to solicit the names of trusted individuals results in similar distance probability distribution as those in the studies re-analyzed by Butts (e.g. Latane, Zipf, Bossard, Festinger, and Hägerstrand). The exponent of $\sim$-1 indicates an inverse square probability of a friendship tie with distance. The effect of heterogenous (clumpy) population distributions in the Zürich region is not important here: the same inverse-square distance relationship results from the Frei and Axhausen (2008) data if the incidence of a social ties is normalized for the population density in rings of distance (Figure 4).
2.2.3 Social network degree in the travel context

Datasets focusing on the social influences on travel behavior and telecommunications (Carrasco and Miller, 2006; Carrasco et al., 2008a; Carrasco et al., 2008c; Frei and Axhausen, 2008; Mok et al., 2007; Ohnmacht, 2006; Schlich, 2004; Silvis et al., 2006) are characterized by detailed questions about behavior posed to small numbers of travellers. Thus by going into detail, they sacrifice breadth and data volume. They are useful for indications of ego net size and the number of other people present at an activity, but not as a basis for constructing a global social network or a quantitative model of social influence on travel behavior. Key quantities are not present all together in a single study, such as the attributes of alters, distances travelled, or the coordinates of the meeting place. Furthermore, attempting to solicit social network structure from the ego shows that, when an ego is asked to draw the relationships between his alters, they are all drawn as if they know each other, resulting in a clustering coefficient = 1 and a constant degree. Surveying each individual would likely reveal more variance in the mutuality of these relationships, and fewer strong cliques (Carrasco et al., 2008b; Kowald, 2009). Also, the small sample sizes do not allow reliable extrapolation (Silvis et al., 2006).

However, some consistent results emerge. Axhausen and Frei (Axhausen and Frei, 2007) established approximately 12 important egos in their study of biographical (long-term)
significant friendships, summarized in the positive-skewed distribution shown in Figure 5 (Frei and Axhausen, 2008).

Figure 5 The degree distribution of a survey of personal trust relationships

Source: Frei and Axhausen (2007)

Carrasco et al. (2008b) indicate that 9-12 others are named as meaningful individuals in the ego network, using a slightly different name generator and interview process. Kowald et al. (2009) are conducting a snowball survey to assess the connections beyond the ego network to begin to grapple with relationships between spatial embedding, global social network topology and short-term travel (the latter may not be completed in the study). The preliminary results also indicate that the name generator prompts recall of between 10 and 20 individuals.

2.2.4 Other measures of social networks in the travel context

The frequency of face-to-face encounters and the type of activity involved in the meeting are addressed in the studies by Silvis et al. (2006) and Frei and Axhausen (2008) mentioned above, as well as a leisure time survey by Schlich (2004).

Schlich's (2004) 12-week leisure travel diary of Swiss travellers illustrates the strong role that social contact plays in mobility habits. Socializing was the motive for 43% of leisure trips,
while a person travelled with someone else to 72% of the activities. Face to face contact was also a central motive for activities which had other specific purposes aside from socializing: for 58% of political activities and volunteer work; and 40% of visits to museums, cinema, and the theater. This represents a profound component of travel that is decentrally coordinated in time and space. Additionally, these datasets hint at marginal distributions we might expect for the number of other people encountered at daily (and other) activities; a reference value for the simulation of social behavior.

Three cases of joint activity-travel are evident in the data: one in which a person is an addition or an enhancement to the activity (cinema, museum); a case in which one of the people is a necessary tool (accompaniment to school); and a case where a person is an activity destination. Inasmuch as a person is mobile, this latter case means that a destination to a fulfilling "socializing" activity can even be without a concrete geographic location; that is, the location (or mode of joint travel) may be very elastic and maybe even meaningless to the participants. An example of the latter is a group of teenagers just as happy "hanging out" in a shopping mall, in a tram, or at home playing music. The activity, and the travel behavior associated with it, are not focused on the context (type of activity, location, accessibility) but on the person-specific attributes of the occasion and the personal needs that are fulfilled by it. This "functional approach" toward activity-travel behavior is described by Stauffacher et al. (2005).

Silvis et al. (2006) used a diary of 2 months’ travel to measure the effort to travel to meet social contacts: distance, time, and frequency of encounters, taking into account the type of social contact involved: duration of friendship, number of friends present at the encounter, and whether the contact is a relative or not. The iterative, or snowball, recruitment elicited 24 respondents reporting a total of 505 trips (441 non-work trips), and 972 social interactions spanning two months.

The average duration of the trips by type show that trips for the purpose of socializing are 3x as long as the work commute.

The trip duration correlates slightly with the number of friends present at a location, but this correlation is not significant if only socializing trips are considered. This may be because the relationship is poorly represented in the small sample. Social activities with 2, 4, 5, 6, or 8 friends lasted 2.5-3.5 times as long as trips performed alone. Only activities where 1 or 7 other friends were present did not correlate with trip duration. There is therefore an indication that the presence of friends at social activities prolongs the activity; that the friends themselves are the feature of the activity that makes it valuable, i.e. adds utility.
The size of the ego network (degree) correlates at 0.490 with the number of different locations visited, i.e. “spatial scope” of activities. As far as trip frequency, the degree, or total number of friends, correlates weaker with the number of social trips than the proportion of social trips which repeatedly occur between two individuals; thus socially "close" friends who visit each other frequently account for more trips than visits to diverse friends. Interestingly however, repeatedly socializing with the same person does not correlate with repeatedly visiting the same locations. These meetings takes place in different locations. Silvis et al. summarize, “However, the best predictor of the scope of a respondent’s activity space was social network size, with an $R^2$ of 0.490. In summary, we saw that a higher number of repeated interactions led to greater number of trips, while a larger social network size led to visiting more locations. ... Two different socio-mobility styles stood out: one group (Style 1, N=14) made many shorter trips to see a large number of people individually, while the other (Style 2, N=5) made fewer longer trips to see many people simultaneously at each destination.” Five of the subjects had socializing patterns in-between these poles of behavior.

Frei and Axhausen (2008), too, permit insight into the frequency of socializing. Figure 6 indicates the number of personal encounters to be expected per day by inverting the reported frequency of encounters.
The individuals in each dyadic relationship are encountered in median 12 (average 49) times per year. The encounter frequencies in the survey were converted to continuous values and inverted to yield an estimate of the number of very significant people encountered per day by the respondents. The peak is just under 2 and the distribution is positive-skewed with no respondent having more than 9 significant person-encounters per day. The indication is that this measure reflects household members above all (see Figure 7).

Figure 7 plots the discrete breakdown of the percent of social encounters made with very significant alters, by type of encounter and frequency.
The daily encounters are dominated by contact with the partner and other household members or close family, and trusted colleagues from work. In less frequent encounters, it is clear that the proportion of distant and out-of-household relatives, work colleagues, and school friends increases, while the categorization "other" makes up the plurality.

Finally, Mok et al. (2007) offer a detailed dataset and analysis of different kinds of social interactions, from support to kinship, fitted to functions of distance between dyad members. The models show depth and detail in the specific elasticities of the frequency of travelling for these kinds of interactions with respect to increased distance. The work shows strongly heterogeneous results in the fit parameters depending on the type of interaction, and highlights the importance of differentiating the type of social tie and the type of social interactions being modelled in order to obtain appropriate model behavior.
2.3 Microsimulations of social networks and travel

Microsimulations offer a glimpse not only of mean field effects but true agent-level effects of networked decision-making. Recent coupled models include work from Arentze and Timmermans (2008), Hackney and Axhausen (2006), and Marchal and Nagel (2005). Arentze and Timmermans (2008) present a fully developed concept for social interactions and activity patterns based on the ego-centric (personal) network, including abstractions of homophily (McPherson et al., 2001), social need, and satisfaction, i.e. the progression of time. Their utility functions maximize the value of ego networks within the total discretionary time budget of the agent. Tests were limited to four social groups of five agents each, so there are no summary statistics of the social networks, and it is noted that the complex results of even the small population tested are difficult to summarize and understand.

In Hackney and Axhausen (2006), social networks of circa 1000 agents evolve with activity spaces by weighing travel cost against participating in social activities in the utility function, while agents exchange information with their affiliates about where other socializing opportunities exist, exploring space to maximize the utility of socializing. It is a rudimentary activity travel model with homogenous agents and no explicit valuation of a social network, intended as a template for realistic models with estimable utility functions. Despite its simple logic, the geographic provenience of the agents adds substantial complexity to existing network generation algorithms. Statistical analysis of the networks indicates exponential degree distributions and probabilities of affiliation proportional to an inverse function of distance between alters' home adresses. This model's algorithms are not scalable to models larger than several thousand agents because of high demands on computer memory.

Marchal and Nagel (2005) have modeled the spread of information about secondary location choice (shopping) along the affiliation network of co-located coworkers to accelerate the learning curve of agents. The work served above all to illustrate the feasibility of the approach in a large-scale microsimulation, but does not have an extensible API or easily exchanged "world" scenarios, and is a fixed classical random graph topology (except for geographical effects), and does not output socializing statistics.
3 Microsimulation of social systems and social networks

Using multi-agent simulations to model inter-actor relationships that lead to macro-scale network properties is a combination of deductive and inductive methods, sometimes called “generative science” (Sawyer 2004). Agent modeling allows researchers to control and experiment with microscopic behavior and observe the emergent macroscopic system: the formation and enforcement of emergent phenomena like social norms, popularity, polarization, and behavioral tipping.

Axtell (2000b, in the paper's abstract) argues for three situations in which multi-agent models can be useful:

"One such use — the simplest — is conceptually quite close to traditional simulation in operations research. This use arises when equations can be formulated that completely describe a social process, and these equations are explicitly soluble, either analytically or numerically. In the former case, the agent model is merely a tool for presenting results, while in the latter it is a novel kind of Monte Carlo analysis. A second, more commonplace usage of computational agent models arises when mathematical models can be written down but not completely solved. In this case the agent-based model can shed significant light on the solution structure, illustrate dynamical properties of the model, serve to test the dependence of results on parameters and assumptions, and be a source of counter-examples. Finally, there are important classes of problems for which writing down equations is not a useful activity. In such circumstances, resort to agent-based computational models may be the only way available to explore such processes systematically, and constitute a third distinct usage of such models."

The last two uses are of interest in this case. The simulated society is permitted to evolve with simple rules, valid at microscopic level, that are agreed upon by experts in the respective fields of behavior; in this case, the disciplines of complex networks and transportation science. Random numbers are used to simulate that part of behavior that is not understood. A range of random seeds and input configurations are then run, often thousands of times, to bracket the possible outcomes of the macroscopic phenomena and to quantify the effects of random influences. The precise microscopic results are considered to be statistical representations of many possible microscopic ensembles (final states). Thus, an individual agent is not showing what a real entity will do, but what it might do.

The large volumes of output of multi-agent model systems are subjected to statistical tests to verify the model function ("does the model respond as intended?")), and, perhaps but not
always, to validate results (to confirm realism). The goal is rarely to program a crystal ball of outcomes, rather, the models serve the purpose of a laboratory experiment to observe how a result develops. The models are generally loosely constrained within a controlled set of boundaries, and are therefore well-suited to experiments where exogenous boundary conditions are systematically modified, for example policy interventions, different agent reward structures, different rules for agent engagement, different agent characteristics, etc. Outcomes are qualified, and sometimes quantified (e.g. response elasticities) by clear and repeatable input assumptions.

3.1 Explaining spatial processes

Examples of spatial agent models abound. Moeckel, et al. (2007, page 2) summarize two early models of spatial agent interactions to motivate enthusiasm for advantages of microsimulation in urban modelling:

"A famous early example explaining the advantages of microsimulation was developed by T. Hägerstrand (1967). He studied the distribution of tuberculosis (TB) control in agriculture in a rural Swedish region. His model of spatial diffusion shows that farms that are located close to other farms that apply TB control are more likely to adapt this control. Every farm that “received” the innovation TB control becomes a “sender” itself, spreading this innovation to other nearby farms. Thus, microsimulation allows representing spatial diffusion. Another prominent example is the self-forming neighbourhood model developed by T. A. Schelling (1978). In this model individuals of two distinguishable groups, such as rich and poor, black and white, or students and professors, select a location on a checkerboard. Each individual accepts a certain number of individuals of the other group. If the number of neighbours of the other group is too high, the individual will decide to relocate. After several iterations an equilibrium is reached, every individual has no more neighbours of the other group than it was willing to accept. Due to a microscopic chain reaction leading individuals that are unsatisfied with their neighbourhood towards clusters of one’s own kind, the resulting segregation is much higher than the individuals' preferences would indicate, taken alone."

These agent models provide spatial inhomogeneity which is different with each random number stream, i.e. each ensemble run. But the important observation is that the qualitative conclusions do not change, and the models enable worthwhile study of behavior and policy implications that would not be tractable with analytical systems (such as differential equations).
3.2 Using multi-agent models

3.2.1 What is an agent?

An agent is an object in a computing language which represents a real-life entity. Most simply, an agent has states and rules of behavior (Axtell, 2000b). These can be private, or known to other agents. However agents also typically have additional properties:

1. The behavior rules usually incorporate a goal to reach, strategies for reaching the goal, and reinforcing rewards/punishments from the environment.

2. The ability to interact with their environment, including other agents, whereby they may receive feedback from the actions of other agents

3. Rationality or bounded rationality.

4. Some sort of adaptation. Reaction to the environment according to behavior rules is not adaptation but following rules; adaptation occurs by changing strategies, and is usually simulated using some kind of strategy mutation (genetic algorithm that mixes current strategies to make new ones, or random mutation) followed by a culling or survival-of-the-fittest algorithm. Adaptation changes the strategies randomly, however learning is a special kind of adaptation in which the new strategy is changed specifically in response to experiences from the environment.

5. Many more complicated interactions with other agents.

3.2.2 Choosing what an agent represents

An agent can represent anything with a state and a set of rules for behaving: a traffic light, a bureaucracy, a tree. In social science models, it is generally a person or a group (e.g. household).

The behavior rules (goals, strategies, rewards) also define the agent and must be chosen appropriately to the modelling goals. The decision horizon is important: short term (i.e. a constant environmental context) or long term (changing contexts). If the agents focus on a short term decision horizon, and the model lasts through time, is the model still useful? Are the short term decisions consistent to yield plausible long-term decisions? If the agents can use strategies for both long and short-term, what triggers agents to shift horizons from the short-term to the long-term?
Are the feedbacks from the environment and the other agents consistent with the state changes ("decisions") permitted within the agents? Does the reward make sense with respect to the strategy such that the agent in fact makes progress toward its goal? As will be seen later, these considerations are important in the context of simulating social contacts within a single-day's travel behavior model. Failure to consider the meaning of the time horizons of agent motivations can make a model difficult to defend.

### 3.2.3 Complex systems

Agent-based models are often complex systems. Complex systems is a term referring to various types of models with many components which may interact and adapt, and even change behavior rules with time. They differ from "complicated" systems in being self-directed. A precise definition is elusive. Durlauf (1997, page 1) writes,

"For our purposes, a system is said to be complex when it exhibits some type of order as a result of the interactions of many heterogeneous objects."

The systems typically exhibit self-organization and emergence. These are aggregate characteristics that are not recognizeable from the actions of a single component of the system, or that depend on the group of components to exist at all, such as the flocking of birds and the "behavior" of the "flock". New states of the system are not simply a summation of the states of the individuals making up the system. Durlauf (1997) describes this as the situation in which the patterns of a system occur at a different level of description than that at which the interactions occur. Stochastic interactions and external noise typically also act on the system. "The common characteristic of all complex systems is that they display organization without any external organizing principle being applied." (Amaral and Ottino, 2004).

The flow collapse on a roadway with high traffic density is an example of nonlinearities in a complex system. After traffic volumes exceed a certain number of vehicles per unit time, the speed on the roadway could remain constant or it could break down to extremely low speeds. The nonlinear result is emergent because no participant intends to cause traffic breakdown, and indeed cannot cause it on his own without the other participants, but it happens as a result of the interactions between the individuals with one another and the environment.

Studying complexity is an effort to understand how a system's microscopic structures determine its aggregate form, such as booms and busts in the broad stock or housing markets due to investors' individual actions rather than a centrally-controlling governing body, or geographic segregation resulting from racist tendencies in individuals rather than from policy.
3.2.4 Bracketing the system's response

The experimental design consists of verification, validation, and hypothesis testing of model results at each step, on the basis of ensemble and sweep runs. The core task is the application of social network analysis techniques to compare social network properties.

Verification consists of making sure that the model is carrying out the desired algorithm correctly (Macal, 2005):

- The model is programmed correctly
- The algorithms have bee implemented properly
- The model does not contain errors, oversights, or bugs

Verification does not ensure:

- Meeting specified performance requirements
- That the model is mature for application ro real-world problems
- That the model correctly represents real world processes

Clearly, not all contingencies can be tested and a model cannot be guaranteed to be perfectly verified. The goal is a model which passes a broad set of tests that are anticipated to cover a space of values that will be commonly used in later application, and perhaps validation, of the model (U.S. Department of Transportation Federal Highway Administration, 1997).

For example, the use of a probabilistic link-addition algorithm enables the emergent social network to be evaluated within analytical bounds of well-known random networks, on the one hand, and regular lattices, on the other. In addition, the distribution of link attachments per agent (degree distribution) will also be bounded and well-predicted by this network growth algorithm.

Validation is the comparison of the model with reality (Box et al. 1978). The properties of real social networks are not well-known. However the properties of the emergent social network can be compared with what has been observed empirically (see Section 2.2). Hypothesis testing involves statistical comparisons based on the variance of the emergent parameters in ensemble runs or within a known sensitivity of the model system.

3.3 The simulation of social networks

Given aggregate characteristics of social networks, statistical or econometric methods have been used to generate sets of social networks that simulate ties in an artificial society. Models of social behavior have also been proposed which attempt to explain what is known about
social network topology and relationship dynamics (e.g. Bidart and Degenne, 2005 provide a review). The modelling choices can be reduced to: probabilistic social tie generation (tie existence), microeconomic tie generation (tie processes), and hybrid models (process feedbacks into tie existence). Some of the many approaches that have been developed to generate social networks are summarized here.

### 3.3.1 Existence versus process

The distinction between social network processes and outcomes is unresolved, even in sociology (Goodreau et al., 2009):

"Clustering at the population level can arise from a range of different underlying processes at the micro level: - transitivity (triangle closure bias) - homophily (preference for forming ties within group)- sociality (degree heterogeneity). In general, there is no 1-to-1 mapping of process to outcome in networks, so it is worth recognizing them as separate features. The literature is not very clear about this, so you will find terms like homophily sometimes used to describe the underlying process, and other times used to describe the outcome. Ditto ' assortative mixing'. Assortative is only one type of mixing -- the general term here is selective (or biased) mixing, and it can be disassortative (as with heterosexual preferences for sexual partners), or idiosyncratic (preferences for specific outgroups over others)."

Schweitzer et al. (2009 (forthcoming)) make the point that aggregate graph characteristics like scale-free degree distributions may have multiple explanations and result from different dynamic processes.

The relationships taking place on the network ties are as important as the topology. Giuraniuc et al. (2005) apply a renormalization technique, a procedure to reduce the dimensionality of problems in statistical physics, to transform the equal interaction strengths between nodes (the binary Ising model) connected in a scale-free graph to a distribution of interaction strengths which result in percolation statistics on a scale-free graph that are equivalent to those resulting from equal (binary Ising) interactions on a classical random graph. In other words, the researchers show that a transformation of interaction strengths is equivalent to a different graph topology. Network researchers should be aware of this substitutability and choose

---

2 Martina Morris, Professor of Sociology and Statistics, University of Washington, in a contribution to the SOCNET email list on 21 May, 2009.
appropriate interaction and network forms which are compatible with the behavioral theory being modelled.

The recommendation from well-established social network researchers in sociology and physics is, then, to accept tie existence and the processes leading to its existence as separate features. This has important consequences for what type of model to choose to represent static social networks (i.e. a result of past interactions and current relationships) versus social network evolution (i.e. resulting from social processes that would have to be represented explicitly, i.e. microeconomically (Jackson and Watts, 2002)).

For the former, a link-generation model would place the links where desired, and social interactions would take place across them. For the latter, it is important whether the process of socializing which leads to social tie formation is the focus of study, or the simple fact that a link might exist between two agents as a matter of probability based on co-presence or some other geographic and activity-travel context.

The difference is whether the model is feasible or makes more sense when using an internally consistent microeconomic strategy-and-reward system for both socializing, long-term valuation of social ties, and the costs of socializing, as well as the usual activity-travel valuations, or by using a link creation mechanism that is independent of the microeconomic modelling within the activity-travel system.

The following sections outline the different approaches.

3.3.2 Network generation by a growth rule

These algorithms use a central social network governor that dictates which two agents become linked together with a social tie. The agents do not establish ties based on their own experiences with their environment, but are instructed to do so based on system rules at a higher level.

Analytic graphs have a strong basis in statistical physics and the asymptotic behavior is well understood. But translating these models into motivations and actions of people is not straightforward. Link attachment procedures to construct canonical graphs are analytically well-described and possibly useful in the context of the activity-travel problem (Dorogovtsev and Mendes, 2003):

- Erdös/Renyi classical random graph generator: an average graph degree is given and the number of pairs of nodes to link together is calculated. Pairs of nodes are chosen at random and linked together until the desired number of pairs is reached. The graphs have a Poisson degree distribution, low to zero clustering, and the lowest
average shortest-path length (order ln(N)). The low clustering makes these graphs unrealistic representations of social networks.

• Equilibrium random graphs with given degree distribution: The degree distribution is allowed to depart from Poisson. Each node is assigned a degree, drawn from the desired degree distribution. Edges are added between the nodes until all nodes have their pre-assigned number of connections. Apart from the degree distribution, these graphs are similar to classical random graphs. Algorithms have also been published with tunable degree and clustering distributions (e.g. Volz, 2004).

• Small World Networks: a regular lattice with very few global imperfections (long links short-circuiting the lattice) is constructed by removing a link from each node and replacing it randomly. This algorithm results in average shortest path lengths nearly as short as in random graphs and a Poisson degree distribution, but it retains high clustering.

• Preferential attachment graphs make agents with more links attract new links with higher probability, resulting in correlated degree that is distributed like a power law (Barabasi 2002). The original algorithm has low clustering like a random graph. Graphs with scale-free degree distributions, embedded in space, were also published concurrently by Warren et al. (2002) and ben-Avraham et al. (2003).

• Barabasi and Bonabeau (2003) propose another preferential attachment algorithm which also creates small-world clustering.

• A static small-world generator using only homophilic classifications is described by Watts et al. (2002).

• Jin et al. (2001) construct a small world graph in the opposite direction from Watts: they begin with a random graph and add local structure to it by introducing friends of friends to one another (triad closure). The graph is constructed iteratively and the average degree is held constant by a random social tie removal algorithm.

• A purely probabilistic treatment of social tie formation as a function of distance between alters is presented in Wong et al. (2006).

• A family of network generation models based on predicting ties between nodes as a function of the existing ties between nodes. Starting from a seed graph, or a sample of relationships, algorithms project where ties would most likely form with a set of non-connected nodes. Work in this topic is represented by Bayesian or Markov simulation processes, e.g. the Exponential Random Graph Model (ERGM) described in Robins et al. (2005) and Liben-Nowell and Kleinberg’s (2003) work on proximity in social space.

Figure 8 illustrates a space of link-addition approaches to modelling social networks in geography.
3.3.3 Network generation based on microscopic behavior (network economics)

The notion that market forces may rely not just on supply and demand, but also on what other consumers or suppliers are supplying and consuming, i.e. that the decisions of these actors are not independent, but connected, is called network economics. Positive feedback, in which more participation induces even more participation, is now thought to be a partial explanation of bubble behavior and other apparently non-rational behavior in economic systems, as well as driving complex social phenomena like collective action (a kind of system "tipping" or, in a physical analogy, like magnetizing a ferric material) (Durlauf, 1997; Krugman, 1996).

Theories and models can be based on networks or on a mean-field model of feedback, in which who affects whom is not as important as a net population effect. For example, a rising stock market index means that more people are buying stocks, and this may induce others to follow the trend and invest, as well. The social networks may not be appropriate at all, or taken to be fixed, based on assumptions or on other models. A general utility model for mean-field feedback is presented in Manski (2000). Extension of the linear model to a discrete
choice model form was made by Brock and Durlauf (2001) and Durlauf and Cohen-Cole (2004). Ioannides extends the Brock and Durlauf model to arbitrary network topologies and illustrates the model on a star, wheel, and path (line) networks (2006).

Dugundji and Walker (2005), Dugundji and Gulyas (2008), and Paez and Scott (2004) have applied the mean-field theory of Manski (1993) and the socially interdependent discrete choice model of Durlauf and Cohen-Cole (2004) to a global social network to iteratively estimate Logit choice models with mean field feedback. In this vein of work, social networks were generated on the basis of a number of factors (common zip code, common residential or work zone, common workplace, sociodemographic categories) and used to estimate econometric discrete choice models on revealed preference data in which the decisions of a relevant peer group entered into the utility equation of each ego. The studies show that the endogenous normative opinion of a peer group in certain social networks can have significant explanatory power for mode choice (bus/car) and trip generation (whether to telecommute).

Theoretical models representing individuals' contributions to network formation are only beginning to be developed (Jackson and Watts, 2002). At the basis of a microeconomic model is myopic behavior of the agents, who do not know the network structure, some cost to the agent for making a social tie, and some benefit to the agent for having it. Currarini et al. (2007) present a microeconomic model of friendship based on the utility of social ties as a function of homophily, race, and separation from other races as motivations for making friends, and compare the results to data.

An alternative microscopic model of network formation emerges from a discipline called "econophysics", which has united the science of complex networks and statistical mechanics from the physics discipline with agent modelling. Physical concepts like Brownian motion, entropy, and temperature are applied to artificial societies to construct aspects of self organization and networking (Gonzalez et al., 2006; Schweitzer and Tilch, 2002). There are many more examples of physical principles applied in disciplines from finance to cell biology, ecological food chains and foreign policy.

Game theory has also been combined with social networks. A summary of approaches for transportation models is presented by Hollander (2006). Wilhite (2001) reviews the literature and presents a game-theoretic representation of bilateral trade on a social network. A fixed network of relationships would be used to represent either a priori interaction rules or an exogenous history of the simulation, to determine which agents can interact with one another. The use of a simple, well-studied game theoretic interaction with different network topologies (as well as turn order of the agents) permits study of the role of network constraints in determining the macroscopic character of a multi-agent game’s outcome (Axtell, 2000a). Random, small-world, or lattice interactions can change certain key properties of a game’s
macroscopic outcomes. Game theoretical approaches are not limited to fixed networks however, and can be the basis for generating social ties. For example, Jackson (2005) demonstrates a method for fitting social network models to data using a simulation that adds social network ties based on game theoretic utility payoffs.

Purely microscopic models, giving the agents no information beyond that which they can experience by direct interaction, provide the most transparent systems with the most straightforward and easily explained agent motives. However, the goal is to observe this emergent behavior of the aggregate body of agents. The result is not easy to tune. The emergent social network topology is difficult to control, and its feedbacks can be unpredictable. There are problems of scoring, distribution of reward, strategies, turn order, equilibrium, etc. that complicate their use in tools which hope to eventually have standing in the world of real data.

In the sense of modelling short-term travel, only the "use" of a tie can generate benefits for being attached to the network (Hackney and Axhausen, 2006). The long-term value of a social tie cannot be explicitly captured in the social network. This necessitates a delicate balance which unifies or makes consistent the utility rewards of the social interaction and the costs for adding and maintaining, as well as the conditions for removing, social ties. The challenge is similar to the resolution of the problem of valuing the long-term expenditures made on a motor vehicle purchase or a bus pass in short-term mode choice.

### 3.3.4 Hybrid models between microeconomics and probabilistic tie generation

Social networks could be generated and regulated using behavioral tendencies, as summarized by Bidart and Degenne (Bidart and Degenne, 2005). Perhaps the most certain sociological hypotheses that might serve as building blocks for the generation of a base social network are: “homophily”, which is the tendency for people to associate with people who are like themselves, or who are friends of their friends (McPherson et al., 2001); “bridging social capital” (Putnam, 1999), which are the associations of a person with those who are like himself in only one way but different in other ways; and notions that people can only maintain a maximum number of relationships, implying a saturation point (Barrett et al., 2002). In terms of dynamics and link removal, studies in sociology indicate that the strength of different relationships changes with time depending on a number of factors which are also not trivial and have to do with the history, activity, and spatial context of the relationships (Burt, 2000; Reagans, 2005).

Note that implementing these hypotheses in computer models necessitates either that agents know more about each other and their social network than just what they can experience from
their environment (they have perfect knowledge) or a governor role must be introduced into the model, as above. The latter is necessary for the model to be easily controllable. While the hybridization of microscopic behavior and a centralized social network governing function can help infuse sociological processes into the generation of social links, the mechanism uncouples the utility of socializing with the longer-term costs of social search and maintenance of social ties.
4 The Microsimulation Framework

The social network module extends the Multi-Agent Transportation Simulation Toolbox (MATSim, Rieser et al., 2007). The MATSim program system is written in Java and has the flexibility to simulate activity-travel of agents representing individual people with arbitrary systems of transportation networks and time-space geographies of activities. It cannot however simulate the passage of medium- or long-term time beyond the extent of the activity plan being simulated. Such passage of time must be handled external to the travel optimization iterations, or external to the MATSim system entirely.

This chapter describes the relevant parts of the MATSim toolbox on which the social network module builds: the Java objects, the iteration procedure, the meaning of the iterations, the stopping criterion, and the limitations of the system for simulating geographic and temporal social interactions. The social network module is introduced in the following chapter.

4.1 The MATSim toolkit

MATSim is an activity-based traffic flow simulation in which agents maximize their own utility by carrying out an activity-travel plan, including simulated travel on a transportation network. The common implementation awards positive utility for participating in activities throughout the time period of an activity plan, and disutility for travel and delay.

A MATSim scenario consists of a geographic world, a transportation network, a set of facilities (activity locations), and a set of agents with plans to be carried out (Balmer, 2007). The "travel demand" consists of "plans", which are a set of activity schedules for each agent, including the locations and routes for the activities. The plans are representative of short-term travel behavior and do not correspond to a particular time period. An evolutionary procedure called a "Controller" (Figure 9) improves the agents' plans iteratively. A queueing algorithm simulates traffic flow in the world for all agents (this has also been called, "execution", the "mobility simulation", "dynamic assignment" or, the name of the specific algorithm used in the examples here, "QueueSim" (Charypar et al., 2007). A scoring function evaluates their utility (Scoring, 4.2.4); a replanning strategy randomly modifies a portion of the activity-plans (Replanning, 4.2.4); and a selection procedure favors the better plans by deleting the worst ones according to their score. The iterative process of utility maximization should be understood to be an optimization algorithm seeking system relaxation given a certain utility function, rather than a learning process or a sequence of time steps.
The Controller is described more in depth in Section 4.3 because the details are important for the social network modifications. The specific fundamental MATSim objects are described first in Section 4.2 to aid understanding of the iterations. Time is distinguished from iterations in Section 4.4. The stopping criterion for the iterations is discussed in Section 4.5 because it is important for social networks and for the statistical representation of simulations "in silica".

### 4.2 Basic data structures (objects) in MATSim

The data structures, or objects in Java, which are relevant to the social networks module and the subsequent discussion, are listed in Table 4. The important features to note are the division between supply and demand objects, and planned versus executed stages of the iteration. The standard version of MATSim has no data structures or objects for joint behavior of any kind apart from queueing conflicts on the transportation network in the traffic flow simulation. The basic objects in Table 4 are extended (Chapter 5) to accommodate agent interactions.
Table 4 MATSim objects relevant to the social networks simulation

<table>
<thead>
<tr>
<th>Object</th>
<th>Contains</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>All Persons</td>
<td>Objects related to an agent and its planned activity-travel</td>
</tr>
<tr>
<td>Person</td>
<td>Plans</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sociodemographic Characteristics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mobility tool ownership</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Knowledge</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Household ID***</td>
<td></td>
</tr>
<tr>
<td>Plan</td>
<td>Acts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Legs (travel leg)</td>
<td></td>
</tr>
<tr>
<td>Act</td>
<td>Location**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Times (start, end)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Activity type</td>
<td></td>
</tr>
<tr>
<td>Leg</td>
<td>Route</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Travel duration</td>
<td></td>
</tr>
<tr>
<td>Route</td>
<td>Links</td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td>Activities</td>
<td></td>
</tr>
<tr>
<td>Facility</td>
<td>X, Y Coordinates</td>
<td>Objects related to fixed geography</td>
</tr>
<tr>
<td></td>
<td>Activity Types</td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>Location**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Type</td>
<td></td>
</tr>
<tr>
<td>Road Network</td>
<td>Links</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nodes</td>
<td></td>
</tr>
<tr>
<td>Event</td>
<td>Time (arrival, departure)</td>
<td>Objects related to realized activity-travel</td>
</tr>
<tr>
<td></td>
<td>Link, Facility**, Node</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Act*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leg*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Person</td>
<td></td>
</tr>
<tr>
<td>Score (Utility)</td>
<td>Scoring function (Events)</td>
<td></td>
</tr>
</tbody>
</table>

* This information is not currently written out and is lost when the run is stopped.
** A Location is a Link or a Facility.
***The Households were added late while this dissertation was in preparation, independently of Social Networks.
4.2.1 Plans and Persons: objects related to intended activity-travel

A Person has identifying sociodemographic and mobility characteristics (sex, age, % employment, transit pass and driver's license ownership), some "Knowledge" about the geographic opportunities for performing activities, and one or more Plans. Knowledge is a list of Activities (Section 4.2.2). A Plan consists specifically of Acts connected by Legs, and involves time, geography, and the person. Acts are elements of plans and consist of the location, activity type, and intended participation schedule for the person owning the Plan. Legs are combinations of transportation mode and sequences of links on the transportation network from Act to Act, and the times during which the agent passes over them.

4.2.2 Geographic objects

The "Location" at which Acts take place are fixed objects in geography with X,Y coordinates and a unique ID. Presently they are either "Facilities" or the "Links" in the road network model. Facilities represent buildings or other infrastructure. They can have opening times and capacities, and specific "Activities" can be carried out at them. To distinguish them from "Act", which is time- and agent (plan)-dependent and represents demand, the term "Activity" is used on the supply side for the time use opportunities at the facility. It is a combination only of the Location (X, Y coordinates and an ID number of the Facility) and the type of activity.

Legs and Links are related in a similar way. Links are part of the road network (supply), and legs are the combination of links that satisfy the agent's travel demand. A "Link" has a length, capacity, and free speed. It is important for the structure of the social networks module (Section 5.4) that the agent is routed to the link, but not to the doorstep of the building (Facility). Thus changing the facility in which an Act takes place to another facility on the same link is not a different plan in terms of the traffic flow simulation; any difference in Utility between two such plans (all else equal) would be due to the facility attributes and not the accessibility (travel to it).

4.2.3 Events: realized travel

While the "plan" is the agent's intention to participate in activities and to travel between them, the experience of having done so after the traffic of all agents is assigned to the transportation network is called the Events object. The Events indicate the results of the traffic flow simulation and yield the movements of each agent on the road network and the time of arrival/depature at the Facilities. The Events contain pointers to the Acts and Legs of each Person and so the experimenter can access both planned and actual movements from the events object. The events stream to the scoring function as they exit the traffic flow.
simulation, and are not ordered geographically or by agent. This is one reason for the high speed of MATSim's calculations.

Thus events do not "belong" to an agent, link or node of the transportation network, or to a location (facility). This means that there are no standard containers that aggregate which agents were on what links or in what facilities (performing what activities) at what time. In order to make sense of Events in terms of a particular agent or location requires accumulating the stream into containers for later analysis, adding a level of overhead to models which incorporate interdependencies between individuals or the places they frequent (Replanning in Section 4.2.4 and Social network scoring functions in Section 5.6).

4.2.4 Utility score

The utility, or "Score", function calculates a double-valued number for each agent's active plan from the events stream (or from the result of aggregating the events stream). The form of the utility function is not restricted. In a standard scenario, agents gain utility for participating in activities and disutility for travel and delay cumulative over the time horizon of the plan.

The utility for plan \( i \) is defined according to the function published in Charypar and Nagel (2005). This function is explicited here because it is also used at the core of the joint utility valuations with social networks for the sake of comparison with the case of non-socially interdependent agents (Section 5.6):

\[
U_i = \sum_{i=1}^{n} U_{act}(type_i, start_i, dur_i) + \sum_{i=2}^{n} U_{trav}(loc_{i-1}, loc_i),
\]

where the utility of each activity in plan \( i \) consists of five components:

\[
U_{act,i} = U_{dur,i} + U_{wait,i} + U_{late,arrival,i} + U_{early.departure,i} + U_{short.duration,i}
\]

and \( U_{trav} \) is simply the travel time from the previous location to the current location, multiplied by a constant \( \beta_{trav} \) (less than 0);

\[
U_{dur,i} = \beta_{dur} t^* \ln \left( \frac{t_{dur}}{t_0} \right)
\]

is the utility of the activity lasting a certain duration, which is maximal at \( t^* \) and zero at and below \( t_0 \), with decreasing returns to additional duration;

\[
U_{wait,i} = \beta_{wait} t_{wait}
\]

is the utility of waiting for the activity to start, subject to the arrival times and the opening times of the facilities where the activities take place;
\[
U_{\text{late.arrival}} = \begin{cases} 
\beta_{\text{late.arrival}} (t_{\text{start}} - t_{\text{latest.arrival}}), & t_{\text{start}} > t_{\text{latest.arrival}} \\
0, & \text{else}
\end{cases}
\]

and \[
U_{\text{early.departure}} = \begin{cases} 
\beta_{\text{early.departure}} (t_{\text{earliest.departure}} - t_{\text{end}}), & t_{\text{end}} < t_{\text{earliest.departure}} \\
0, & \text{else}
\end{cases}
\]

are the utilities of arriving too late or of having to leave the activity too early; and

\[
U_{\text{short.duration}} = \begin{cases} 
\beta_{\text{short.duration}} (t_{\text{short.duration}} - (t_{\text{end}} - t_{\text{start}})), & t_{\text{end}} < t_{\text{short.duration}} \\
0, & \text{else}
\end{cases}
\]

is an additional penalty if the activity duration is shorter than a minimum duration.

The arrival and departure times (activity durations) and travel times result from the traffic flow simulation, expressed in the events. The $\beta$ values and the minimum duration, earliest start time, latest end time, and ideal duration of activities are set in the input configuration file that initializes the Controller. These values have not been estimated for any population yet and the values used in the demonstration experiments will be given for each run. This utility function has never been estimated on data and the values of the $\beta$ are generally chosen to have the values $\beta_{\text{wait}}, \beta_{\text{trav}}, \beta_{\text{late.arrival}} = -6/h, -12/h$, and $-18/h$ to match the ratios used in Vickrey's (1969) illustrative model for the equilibrium pricing of road services in a bottleneck.

### 4.3 The MATSim iteration

The MATSim iteration is an implementation of an evolutionary algorithm to improve the activity plans of the agents. The plans are altered and subjected to a survival test (selection). The Controller is a flexible implementation enabling any number of algorithms to alter plans, score them, and to select them. The specific algorithms activated by the Controller, and their initializing parameters, including those of the social networks and social interactions, are stipulated in a configuration file in XML format.

Referring to Figure 9, the plans and the geographic world are initialized first. At the beginning of each iteration, a particular percentage of agents is randomly chosen to alter their plans in one of several "replanning" algorithms. The percentage of the population using each algorithm is determined by the experimenter at the outset of the run. A "replanning" algorithm is an operator which is sent a plan which it copies and alters in a particular way that is not related to how well the plan performed: either the route, the departure time, the secondary (shopping, leisure) location, or mode. The precise ways in which these components of the plan can be altered will be discussed. The solution space of plan configurations for each agent is therefore potentially completely searchable by not incorporating any information about the plan's performance in the alteration of the plan (a non-learning algorithm). This is intended to
eliminate path-dependence in the solution of the traffic flow problem. Altering the order, type, and number of activities is not currently allowed in replanning, due to the unresolved combinatorial complexity of the calculation, and a lack of mid- and long-term motives for the agents (for an attempt at adapting MATSim to be able to adjust the number of activities see Feil et al. (2009)).

Thus MATSim is suited for the simulation of generated traffic: "Additional vehicle trips on a particular roadway or area that occur when roadway capacity is increased, or travel conditions are improved in other ways. This may consist of shifts in travel time, route, mode, destination and frequency" (Victoria Transport Policy Institute, 2009). Induced travel, "an increase in total vehicle mileage due to increased motor vehicle trip frequency, longer trip distances or shifts from other modes, but excluding travel shifted from other times and routes", is possible only with the secondary location choice replanning module and/or mode choice module, both new additions to the toolbox. Latent demand, "additional trips that would be made if travel conditions improved (less congested, higher design speeds, lower vehicle costs or tolls)," would require the ability to change the activity order and number for each agent, which is not currently possible in MATSim.

The altered plans are executed, or assigned, to the transportation network in the traffic flow simulation. The other, non-chosen agents, change nothing at all in their intentions and attempt to execute a previous plan they have stored. At the end of each iteration, each agent's executed plan, whether freshly modified by replanning or not, receives the score of the the events that point to the acts and legs in the plan. The potential for complex interaction arises in the plan execution in that the actions of the agents competing for slots in the network queues can influence one another's score and subsequent "best" plan. The utility of each plan is accumulated event-by-event as these stream through the utility function object. Each agent can remember a number of different plans according to a user-defined fixed allowance (usually 3 or 4). The plans evolve in that the worst-scoring plan of each person is removed from its memory after the utility scoring step. Subsequent alterations to plans occur based on the plans remaining in memory. The plan to be executed can be chosen by various algorithms (Best, Logistic, Random, etc.). This mutation-and-culling process ensures that the utility space is searched with equal probability and that the system is not confined by search parameters to a local maximum.

There is no connection between the state of an agent's plans and the likelihood the agent is chosen to replan. The percentage of agents which replan is also fixed (experiments that reduce this proportion with the number of iterations have reduced the number of iterations required to reach stable utilities, but the optimal rate of changing the share of agents who replan each iteration depends on the scenario (Charypar, 2008). There is also no connection between the changes made to the plan and the plan; i.e. the replanning strategy is a potentially complete
search of the utility space. This is a critical contrast to "learning": incorporating past experience into intentions attempts to mimic real behavior, but risks arriving at path-dependent solutions which have an unknown relationship to optimal behavior. The random search (in time) is unrealistic, but less likely to become trapped in a local maximum of the search space. The optimum of the system is not known, however, and solutions are not sought to any predefined tolerance, so the MATSim solution method cannot be called optimization.

4.4 Time in MATSim

The only passing of time in the simulation is limited to the events occurring between the earliest and latest time of the agents' plans. The exogenous world of a road network, geography (planning zones, municipal boundaries), and locations (facilities) remain constant, as do the population's sociodemographics and any medium and long-term projects of the agents which their short-term activities may be trying to achieve. Most MATSim experiments are based on the schedule of a single representative weekday, but the system's structure does not place any limit on the duration of the input plans; thus medium term goals may indeed be incorporated into (very long) activity plans, if computer memory suffices.

The iteration in MATSim does not represent time, but is a search through a multi-dimensional space of activity-travel trajectories; indeed as far as the timing of the activities is concerned, the search is random. Since the agents' plan evolution differs from learning algorithms in being a random mutation (at least in scheduling) rather than applying experiences from a previous iteration, a search through space for a utility-maximizing secondary location should not be confused with "variety seeking" or an "activity space", which would reflect learning over a longer biographical timeline of an agent's experience. The dissertation returns to this issue in Section 5.3.3.

4.5 MATSim stopping criterion and equilibrium

The stopping criterion for the iterations is a difficult technical issue. The MATSim user community has agreed that a qualitatively stable average utility score with increasingly small differences between the score of plans within each agents' memory indicates a stable solution, which suffices to assume an optimum distribution of activity-travel. At this stage, the marginal utility of the day's entire plan with respect to a change in the scored variables as a result of exploration in the strategy dimension (route, departure time to first activity) should be equal for all the plans retained in each agent's memory. However there are several unsolved questions about what this equilibrium means. Randomly chosen agents continue to alter their plans until the run is stopped. This has several consequences. It means that the
agent's "best" plan may not correspond to the current state of the system, since a portion of the other plans have been changed since it was scored: only the "executed" plan has a score current with the system state. In fact an agent may end up retaining copies of the same plan with several different scores, each depending on what the rest of the system was doing, whereas in actuality only one of those scores would be valid at the present time.

Unless the point is reached in which agent behavior does not change in any sense from one iteration to the next (activity choices, route and time choices, delays experienced, etc., indicating that all plans in each agent's memory are identical), the equilibrium is only statistically stable. The steady temperature of a perfectly insulated ideal gas is a statistical equilibrium: the average kinetic energy of the molecules is constant, like the average utility of the executed plans. The speeds of the molecules are also distributed in a constant distribution, according to how they collide with each other. Which molecules are going fast and which are going slow does not matter in maintaining the constant temperature. In fact, it changes as the molecules collide. But this doesn't matter either, since the molecules are otherwise all identical. A statistical model is therefore perfectly acceptable and very useful, since if we cannot distinguish between them, we don't care which molecule is behaving a certain way.

However the MATSim agents differ socio-economically and in spatial distribution. We are interested in the particular trajectory in time and space of different agents, and we are interested in the distributions of more than just one parameter. Timing, location, route, delays are all important in the geographic embedding of the traffic problem. It matters which agent has high utility in which iteration, and if this changes in the next iteration, even if the population average remains the same.

In the statistical physics model of an ideal gas, only the kinetic energy (speed) of the molecules is important. Differences in the system state of the activity-travel microsimulation from one iteration to another might have meaningful consequences for the components of utility of certain agents, as they continue to re-plan their activities.

Returning to the MATSim "equilibrium", even if precisely stable with no random variation across iterations, the stopping criterion is no guarantee of a Nash equilibrium, in which each agent would have to try to maximize utility while others keep their intentions fixed. As far as recognizing the deviation from optimality for each agent's executed plan, the "best" score cannot be used as a reference unless all plans for each agent are identical (from the agent's best to worst plan), in which case they are all optimal. Indeed the MATSim stable state has never been compared to the Nash solution. The stable result is also not guaranteed to be a Wardrop equilibrium, since it optimizes activities alongside travel and not just the route assignment, itself. The precise characterization of what is now called the MATSim equilibrium is only beginning.
The social networks module

The social network module is an API in the Java language that extends the Multi-Agent Transportation Simulation Toolbox (MATSim, Rieser et al., 2007) and generalizes the small-scale multi-agent computer simulations of Hackney and Axhausen (2006) and Marchal and Nagel (2005). These implemented elements of joint travel behavior, but were limited by fixed agent interaction rules, fixed geographies, fixed social network topologies, and no traffic flow simulation. The new system adds the capability of representing the socially-mediated travel outlined in Section 2.1.3.

The social networks extension to MATSim provides building blocks to construct simulations of traveller interdependencies. Both the activity-travel intentions of agents and the valuation of the traffic flow simulation can be socially influenced, letting the experimenter construct interaction rules and observe the outcomes. At the same time, the MATSim toolbox provides a geographic context and the detailed population data structures to enable models of social networks which grow exogenously to the traffic flow simulation, at longer time scales and with a broader-brush, if the researcher has suitable models to apply or to test.

5.1 Changes to the structure of the MATSim iteration

The iterative evolution of individual plans is maintained, with an envelope of social interactions, statistical analysis, and output occurring around the traffic flow simulation.

5.1.1 Basic elements of social interactions

Utilities are provided for:

- Initializing a social network;

- Spatial interactions in Facilities (or during Activities) to make new social associations and to strengthen (or otherwise alter) existing ones, actually a time-space interaction;

- Non-spatial interactions (electronic communications, unobserved communications including biography) to share information about geography and social networks, actually also independent of time in that these interactions do not relate to the events in the traffic flow simulation (and are dependent only on social network topology and the interaction rules set out in the interactor);

- Interdependent utility of activities, travel, social network ties, and entire plans
5.1.2 Approach to joint plan intent and joint valuation

Some objects which fit conceptually into the notion of joint activity travel are missing which would require major changes to the fundamentals of MATSim. These are simulated abstractly in combinations of the above building blocks:

- Acts and plans are still individual. There is no object for a genuine joint "act" and "leg" which would represent appointments (intended face-to-face meetings or accompaniment), nor is there a coding workaround to coordinate intent in plans, which would take the form of an "appointment manager" utility that allocates and tracks acts and social groupings. There is no way to divide costs among participants in a joint activity or travel leg. Currently the only indicator in the data structures of a desire for agents to meet face to face is acts (legs) which overlap in time, space, and type, and an existing social relationship. This coincidence is not a good model of the intent to meet: four acquainted people who happen to be planning to do the same activity at the same time and place do not necessarily have to have planned to meet each other, nor will they necessarily interact with one another. However, while conceptually incomplete, the building blocks in place enable the emergence of joint acts and legs. Joint alterations to plans and joint scoring are possible which couple agents' plans to simulate the joint intent, and the implementation of the spatial interactor allows for a probabilistic treatment of whether co-present agents interact, to account for the fact that some co-presence is inconsequential. Likewise, joint activities can be strongly implied if the social network is defined as "people who have appointments with one another" and the relationship type is the type of activity; respectively, a comprehensive social network can be defined in which the edges represent the type of interactions, including face-to-face, that are to take place between agents during the period of the plans. But no explicit joint intent objects have been defined.

- The plans evolve and are optimized for each agent. Joint plan evolution or joint utility maximization per se have not been implemented. There are two reasons which are related. First, the Controller iteration deletes the worst plan of each individual, not considering whether this worst plan belongs to a joint plan that involves other agents (in part since joint plans or acts are not defined). Second, the plan promoted to be "active" for one agent should be activated in all other involved agents if it is a joint plan, but this is also not available in MATSim. To simulate evolving joint plans, the corresponding plans in all involved parties should be deleted or made active, at the same time. If the joint plan is optimally formed for all involved, then it won't matter in the end result which evolutionary algorithm, individual or joint, is used, since all those
involved in the joint plan will eventually find each other through individual optimization. Indeed this is the current assumption. It requires carefully accounting for every element of joint utility, including accurately distributing the marginal costs and benefits due to joint behavior, in the utility function. If the individual agent has a higher-utility plan that is not the joint plan, this plan will be more likely to be active instead, and the agent will have "decided" not to take part in the joint plan, whose utility would then change as a result, and the other affected agents would "decide" how to adapt appropriately. But "joint plan deletion" and activation would be necessary to be able to model joint plans in which not every agent in the joint plan has an optimal utility through joint action (i.e. he would opt out of the joint plan if some other obligation wasn't forcing him to remain in the coalition). In contrast, estimating the opportunity costs due to obligations are only possible in the current implementation of individual utility optimization with formulations of very high utility for upholding the obligation or very large penalties for neglecting them. This is less a shortcoming as it is a particular operationalization of networked utility, in which the group decision making dynamics is incorporated into the utility function and is not explicitly modelled as a behavior.

This chapter describes the new logical flow for long- and short-term socializing, the modified data structures (objects) necessary for the modelling, and examples of replanning and scoring which simulate specific interdependent activity travel decision making processes outlined in Section 1.1.

5.1.3 The Social Networks Controller

Figure 10 illustrates the modifications to the iteration scheme of Figure 9 to introduce the social interactions. The simulation begins as usual with plans and a geographic scenario. At this point, a set of social relationships between agents is generated, based on a model or data that the researcher either believes in or wants to test. The agents' "Knowledge" is adapted to track the social exchange of information. It may be important at this stage to initialize the agents' spatial knowledge (about "Activities", see 4.2) so that it is consistent with the social contacts the agent has: knowing a person may imply knowing something about that person's activity space, as well. Note that this is the first break with conventional implementations of MATSim, in that the "Knowledge" object is used here as a historical record, i.e. memory, of agent behavior in a time that passed before the period in which plans to be optimized take place; an activity space of sorts.

Before the traffic flow simulation is called, containers are prepared to aggregate the events that stream out of the simulation to track which agents were at which facilities during which times, and which social relationships they have with one another. A container to track
route/link information (for joint travel, for example) would have an identical structure but has not been written, yet. Socially networked (joint) utility functions can now use the events containers. Face-to-face encounters and other socializing interactions are simulated in the scoring method of the Controller immediately after the traffic flow simulation, and are rewarded or penalized in the utility function, as needed (strictly, these should occur as "joint events" in the traffic flow simulation, but the traffic flow simulation does not use facilities, see 4.2.2). The Replanning stage now may use the socially exchanged knowledge and/or access the social network of agents to alter plans. New socially-moderated replanning strategies are described in Section 5.6.3.

5.1.4 Social network iteration

Social interactions and the modification of social ties take place during a social iteration that envelopes the MATSim iteration to relax the travel behavior. The reason for this step is explained as follows: Ideally, the evolution of relationships across iterations should also conform to MATSim's evolutionary simulation in which an iteration is not time progression.
and time mutations are random, changing the current state of the system, but not dependent on it. An agent should not use information from a previous iteration in changing its current situation. But some of the hypotheses of social network geography that are desirable to test, for example incorporating knowledge about locations learned about from alters into the selection of new destinations, involve notions of memory. Thus, if MATSim is used in this way, the meaning of one iteration of social network evolution versus one iteration of activity-travel optimization is very different.

The solution employed is to decouple the rate of change in the two systems by defining a "social network iteration" that consists of one or more "MATSim" iterations. The MATSim iteration step is the activity-travel optimization. This can yield variable results initially from step to step as the random changes to plan schedules cause large displacements in travel times. The perturbations take from several to several dozen iterations to dampen. The social network iteration represents social network dynamics (changes to the edges of the social network) and information exchanges. Though the social activities are geographically embedded, they need not respond to each rapid change in the MATSim activity-travel optimization. The new agent actions, "spatially interact", "nonspatially interact", "forget knowledge", and "dissolve social ties" occur only during social network iterations. All other actions (scoring, replanning, statistics, write) occur each MATSim iteration.

The new social network state should be viewed as determining a new set of travel demand whose routes, modes, times, and locations need to be optimized by the MATSim iterations. The researcher can choose how quickly the social interactions and social networks evolve relative to the activity-travel optimization, both in terms of the magnitude of the iteration-to-iteration changes, and the amount that the traffic flow simulation relaxes between social iterations. Running MATSim to a stable state between each evolutionary step of the social network would take a tremendous amount of time, and dwells on a degree of precision that is not necessary for influencing the social network evolution. The goal is to change the social network at a rate that allows the utility function to steadily increase, as in a non-socially-interacting MATSim run. As it turns out, even configurations in which the social network evolves each iteration of the traffic flow simulation are stable in this respect.

5.2 Key Objects in the social nets module

Table 5 outlines the main new structures and the changes made to existing structures which were described in Table 4 above. As experimental structures, their precise interfaces have not been fixed. Likewise, once finalized, their permanent incorporation into MATSim will be accompanied by appropriate read/write utilities and document type definitions. The objects and new classes are discussed one by one below.
### Table 5 MATSim social networks objects

<table>
<thead>
<tr>
<th>Object</th>
<th>Contains</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>EgoNet ManageKnowledge</td>
<td>List of Persons*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cull excess knowledge (Facilities)</td>
</tr>
<tr>
<td>SocialNetwork</td>
<td>Type</td>
<td>Directed/Undirected</td>
</tr>
<tr>
<td></td>
<td>Saturation model</td>
<td>For each agent</td>
</tr>
<tr>
<td></td>
<td>Link dissolution model</td>
<td>Choice of algorithms</td>
</tr>
<tr>
<td></td>
<td>EdgeList</td>
<td>SocialNetEdges</td>
</tr>
<tr>
<td>SocialNetEdge</td>
<td>PersonFrom</td>
<td>Person object*</td>
</tr>
<tr>
<td></td>
<td>PersonTo</td>
<td>Person object*</td>
</tr>
<tr>
<td></td>
<td>Strength</td>
<td>double</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>EuclideanDistance</td>
</tr>
<tr>
<td></td>
<td>TimeFirstMet</td>
<td>Iteration number</td>
</tr>
<tr>
<td></td>
<td>TimeLastMet</td>
<td>Iteration number</td>
</tr>
<tr>
<td>TimeWindow</td>
<td>Person, Act, StartTime, EndTime</td>
<td>Mapped in TimeWindowMap: Facility&lt;&gt;[TimeWindowMaps]</td>
</tr>
<tr>
<td>SpatialInteractor</td>
<td>MeetFaceToFace</td>
<td>Choice of algorithms</td>
</tr>
<tr>
<td>NonSpatialInteractor</td>
<td>ExchangeGeographicInformation</td>
<td>Facility locations</td>
</tr>
<tr>
<td></td>
<td>IntroduceFriendsOfFriends</td>
<td>Closes open triangles</td>
</tr>
<tr>
<td>RePlanning</td>
<td>SecondaryLocationChoice</td>
<td>Choice of algorithms</td>
</tr>
<tr>
<td></td>
<td>JointArrivalTime</td>
<td>Experimental</td>
</tr>
<tr>
<td>Utility</td>
<td>JointActivityValuation</td>
<td>Choice of parameters</td>
</tr>
<tr>
<td>Statistics</td>
<td>Graph</td>
<td>Over SocialNetwork</td>
</tr>
<tr>
<td></td>
<td>Edge</td>
<td>Over SocialNetEdges</td>
</tr>
<tr>
<td></td>
<td>Agent</td>
<td>Over Persons</td>
</tr>
<tr>
<td></td>
<td>Confidence Ellipses</td>
<td>Postprocessing</td>
</tr>
<tr>
<td>PajekWriter</td>
<td>Net Graphics File</td>
<td>Entire social network</td>
</tr>
<tr>
<td></td>
<td>Geo Graphics File</td>
<td>Geographically aggregated</td>
</tr>
<tr>
<td>KML Writer</td>
<td>Knowledge Ellipse</td>
<td>Post process geographic output</td>
</tr>
</tbody>
</table>

*Person = Agent

### 5.3 Social network: EgoNet, SocialNetwork, SocialNetEdge

Social interactions and interdependent activity-travel are simulated at the agent, or ego-, level. A global social network is of interest either as: 1) an initial condition; or 2) as an emergent property of the ego-centered interactions. An initial social network is a statistical entity constructed by some algorithm the researcher trusts or wants to test with respect to its effects.
on (representative) short-term activity travel behavior, when coupled with some kind of social interaction rules. A social network as an emergent property of the system is an object to analyze in terms of topology, geography, and activity-travel patterns as a result of the assumed socializing rules. Here, the intent is to make a comparison to a data sample, to verify the functioning of model components, or to compare an ensemble of runs to investigate the stability of the social-traffic flow simulation system.

5.3.1 Social network

The social network class is a List object of all the social relationships (specifically, SocialNetEdges) and a set of methods for managing them. It defines whether social links are directed or undirected, and contains interfaces to initiate, strengthen, and remove links, and a saturation function to simulate the cognitive limitations that inhibit humans from maintaining very many links. Whether and when all of these features are necessary is not yet clear, but for the moment they are available to be used.

5.3.2 EgoNet

While the SocialNetwork manages the relationships between agents (but does not store the pointers to agents), the agents keep track of the alters in their own ego networks (but do not store the edges). Specifically, the MATSim agent's "Knowledge" is extended by a list of its acquaintances in a List object called an "EgoNet". The alters in the ego net are always on the receiving end of a social net edge ("PersonTo" object). In this way, both the ego networks and the global social network are stored for quick searching. There is some overlap here (it is not necessary to store the alters in an "EgoNet" since they can be accessed with some searching by the edges in the social network) but this dual-level organization at ego- and global levels speeds access to the social network elements. For global processes, which are usually statistical calculations, the researcher can access any and all social network edges using the edge list of the social network, and for local search and local behavior, the alters are accessed directly via the ego net, as the ego would do in real life (without using a global social network).

5.3.3 Social relationships: SocialNetEdge

The social net edges have attributes type, Euclidean length, strength, to/from agent, and the iteration. The strength can be generated by any algorithm and is set to "1.0" for the demonstrations here.

The condition for agents to add a social relationship is established in the Interactors (Section 5.5). Once the condition is met, a request is sent to the social network object to update the
social net edge. If the edge already exists between the ego and the alter, the iteration at which it was last "used" ("activated") is updated with the current iteration. The "type" may be a constant for the model, or it may be updated at this point with the context of the latest interaction (note: the latter is the current configuration but it may be desirable to set the "type" constant and have another variable to track the context of the last interactions). The strength may be a function of this meeting, as well, and be updated. If the edge exists, the "EgoNet" object is not altered.

If the edge does not exist (no social relationship between the agents), a new edge may be established from the ego to the alter. The addition of the link is mediated by the saturation function, in which the probability of adding an additional link decreases according to how many links the ego already has:

\[ p_{ij} = \exp(\beta_{sat} \ast z_i) , \]

where \( p_{ij} \) is the probability for adding the new link from \( i \) to \( j \), \( \beta_{sat} \) is a constant < 0 chosen by the experimenter, and \( z \) is the number of social contacts of ego, \( i \). If a randomly drawn number between \([0,1]\) < \( p_{ij} \), the new link is added between the ego and the alter with the "time" of establishment equal to the current iteration, a "type" according to the researcher's needs, and a "strength" set to some functional value (usually a constant = 1.0). The edge is added to the edge list of the social network, and the alter is added to the ego's EgoNet.

In a "directed" network, each edge can have a different "to" and "from" Person, with a different saturation probability according to the degree of the "from" Person, and each edge established is recorded in the edge list of the social network. Directed networks are not commonly used however, because they have the disadvantage that many graph statistics are not well-defined for them. In an "undirected" graph, the more common object of study in graph theory, only one edge is recorded in the social network between any two agents. This gives the correct edge count for the statistical calculations, whereby it is understood that the edges permit influence between nodes to flow in both directions. The edge is added in both directions if it satisfies the saturation condition of the ego. Whether an edge is directed or not, the agents of each added edge are entered accordingly into each others' ego nets.

### 5.3.4 Dissolution of social relationships

Edge removal is not determined by an Interactor, but is an algorithm in the social network and is in this sense a (exogenously set) social standard for ending relationships. While a specific social interaction may lie at the dissolution of some social relationships in reality (e.g. a "breakup", a project end, graduation from school), it is more likely that relationships lie fallow, become less useful, or are gradually forgotten (Burt, 2000). Placing the link removal
utility in the social network is an attempt to assert that a universal rule for dissolving social relationships holds for the experimental society. The algorithms consist of a set of rules, or tests, run for each edge during each social network iteration.

The link removal algorithms have non-trivial effects on the social network topology (as in the Jin et al. (2001) model in Section 3.3.2). The algorithms implemented have a minimum social link "age", equal to the number of social network iterations that have passed since the edge was first created, before which removal is not possible. This permits the social network to evolve some before links are removed. The choice of removal conditions applicable each iteration after that include:

- "none": no removal;
- "random": iterates through edges once and removes each with a probability "remove_p", set by the experimenter;
- "random_node_degree" iterates through edges once and removes each with probability proportional to the normalized degree of the "From" person times the probability "remove_p". Degree is normalized by dividing by the graph maximum degree;
- "random_link_age" iterates through edges and removes each with probability proportional to the normalized link age times probability remove_p. Link age is normalized by dividing by the iteration number;
- "random_constant_kbar" keeps average degree constant. It first calculates the number of edges to remove, then randomly picks this number of random edge indices and removes these edges.

When an edge is removed from the social network edge list, the corresponding alters are removed from the ego nets in the respective persons' knowledge.

5.3.5 Social network initialization

While the link removal algorithm and the spatial and non-spatial interactors are used in conjunction to iteratively evolve the social network, the network's initial state is set in a generation algorithm which may or may not incorporate travel plans and geography. With appropriate Java implementations of the interfaces, social networks can be generated initially with arbitrary geography, topology, and coupling with activity plans. The initial social network can also be read in from a previous MATSim run or another program generating social networks. The initial state to use is up to the research question and the type of model being posed: investigative, verification and validation, or predictive. The social relationships
relevant to the behavior being studied have to be carefully defined in order for a model to be useful (see the introduction of this chapter, for instance), and any iterative evolution procedure of the social network has to be consistent with this definition. This is up to the researcher to establish and to defend (see Section 3.2).

Even if it were realistic to expect useful results from a model attempting to capture all socializing explicitly, MATSim is not a suitable tool for representing time dynamics at the scope of a lifetime and the scale of the world, which would be the necessary scope for such an endeavor. In this respect, it is more useful to consider representing only certain social interactions relevant to high-frequency (e.g. daily or weekly) mobility patterns that may be simulated in MATSim.

Various algorithms from the literature are pre-installed to initialize the social network and which can be chosen from the input configuration file include: small world (Newman et al., 2002; Watts, 1999), scale-free (Barabasi et al., 1999), random (e.g. Dorogovtsev and Mendes, 2003), as well as spatially embedded versions of these (Andrade et al., 2005; Penrose, 1991; Wong et al., 2006).

### 5.4 MentalMap

The "Knowledge" object that each person has normally contains the activities (i.e. type and facility, Section 4.2) it knows about. Knowledge is extended in two ways; it now contains the EgoNet, described already in Section 5.3, and the "MentalMap", described here. The mental map was initially needed to map the activities in the knowledge to the acts in the plans, and vice-versa. This was because the acts in the plans had only a "Link" as a location, exactly as they are modelled in the traffic flow simulation. In this superseded (< mid-2007) version of MATSim, changing the facility in which an act took place did not change the plan unless the facility was on a different link. The additional information about the facilities that was needed to distinguish two plans with different (secondary) locations was stored in the "MentalMap". This mapping was written out (ActivityActMap.txt) to be associated with social networks from the same iteration upon restarting experiments, to keep the mental map consistent with the social network. Now that acts can contain pointers to facilities, this mapping is superfluous since the location of acts is recorded in the plans file.

Note that the acts having facility-precision in the plans object does not change the traffic flow simulation: Agent travel and Events are still not resolved any higher than the link on which the act takes place. The significance of having different facilities in the acts only regards the spatial interactions possible for the agent during the act, and the utility valuations of the facilities, but not the results of the traffic flow simulation.
The "MentalMap" now primarily provides tools to manage the activities in knowledge. The first tool is a method to initialize the knowledge of the agents with facility locations. This is first uses the facilities pointed to by the acts of the plans. If no facilities are assigned to the acts, which is possible, since facilities are not required and the acts could be assigned to links instead of facilities, the facilities are either randomly assigned according to the link, or they are assigned by reading in a user-generated file (no support is provided by the module for this option). Then, additional activities can be initialized in knowledge, using a random process based on the link, or a user-generated file (like the "ActivityActMap"). Part of the procedure to prepare newly initialized knowledge for the simulation is to reconcile the names of the activity types between the types listed in the facilities read in, the plans read in, and the expected names for them in the social networks interactors. This activity-type-naming section is hard-coded and needs to become integrated with the XML/DTD's of the supporting data files.

The second tool is a method to give the activities a score. This score is used only within MentalMap and has no bearing on the Plan score, though it itself could be a function of the Plan score, if desired. It helps each person rank his geographic knowledge. The score can be useful for experiments in the percolation of information (deciding which information to share with friends, see Altenhoff (2003), for use in choosing new locations for activities (weighting the probability of choosing a particular new location), or in algorithms that forget, or erase, less-valuable activities from the person's memory. The sample simulations illustrated in Chapter 6 use as a score the number of alters passing ("recommending") the activity to the ego, and the score is cumulative over social network iterations. But the score could be any real number: the number of other agents at the activity, a cost for participating in the activity at that facility (such as how crowded it is or how far from a parking space), the score of the plan with an act using that activity, the distance to the activity from the home of the agent, etc.

The final tool prunes the activities in the knowledge of each person. When agents can learn about activities from other agents, very many activities can accumulate in knowledge which are not assigned to an act in a plan. The tool allows reducing this knowledge to simulate a model of a mental map. Many conceivable algorithms could be written to cull the excess knowledge, with different effects on the mental map. In the examples shown in this work, each social network iteration, the mental map sorts the activities by score and erases the lowest-scoring ones in excess of a certain threshold number of allowable activities per agent. The activities that are assigned to acts in the active and inactive plans in the agent's memory cannot be erased. Common settings for this algorithm are to allow 50% more activities than the total number of acts in the person's plans. Since inactive plans can be chosen for re-
execution in the traffic flow simulation, those activities which are mapped to the acts in the inactive plans must be maintained at minimum in knowledge (Figure 11).

Figure 11    Pseudocode for pruning facility knowledge

Sort FacilitiesInKnowledge by facility score (lowest to highest)
MaxNumToKeep = k * Number of Acts/Plan * Number of Plans in memory
NumInExcess = NumberOfFacilitiesInKnowledge - MaxNumToKeep

FacilitiesToForget = subset of FacilitiesInKnowledge[0:NumInExcess]

Do i=0,NumInExcess
  If(Facility(i) is not in plans){
    mark Facility for deletion from memory
  }
End do
Forget marked Facilities

Note: facility score is any double; k >1.0

The activities ultimately contained in knowledge, which may be considered a "mental map", depend then on the initialization of knowledge, the exchange of geographic information between agents, its scoring, the success of the plans that use the activity, and the deletion procedures for excess knowledge.

5.5 Interactors

An "Interactor" is a mechanism to simulate social processes, specifically, those prompting the generation or updating of a "SocialNetEdge" or the passing of "Knowledge" across the social network. Other kinds of explicit social interaction are currently subsumed in utility. In an attempt to give the modeller as many tools as possible for flexibly generating social networks and social travel dynamics, the interactors are divided into "spatial" (spatio-temporal) and "nonspatial" (nonspatial, nontemporal) types, with different socializing activities possible in each case.

5.5.1 Spatial interactions

The agents spending time at a facility for a certain reason ("Activity") can establish or alter (say, strengthen or reaffirm) a social association with other agents with activities at the facility. The spatial interaction introduced by Marchal and Nagel (2005) use a queueing
model to simulate spatial encounters: the next agent entering a facility encounters the previous one who has entered, if it is still there. If the previous agent has left already, the queue at the activity is empty and the newly arriving agent encounters no one. Upon encountering, the agents make "friends" and exchange knowledge of a location for carrying out an activity. The strength of the relationship wanes exponentially with the number of iterations since the last encounter, until the relationship dissolves at a certain minimum strength threshold. The mechanism is a fast way to propagate socially-held knowledge, but it is a non-intuitive representation of spatial interactions, and is the only interaction possible in that model.

The social networks module for MATSim enables the incorporation of time and multiple agents at once in the encounters. A container for collecting the "Act Events" was mentioned in Section 4.2. As MATSim evolves, so does the role of this container. It is necessary at the moment to map "acts" back to "persons"; for the sole reason that MATSim does not contain a mapping from an act back to the plan or person to which it belongs. This container is named "TimeWindow". Each "TimeWindow" is constructed from the Events stream out of the traffic flow simulation and consists of (points to) the start time, end time, person, and the act in the plan. (Note that the start time and the end time are carried over from older code versions. They are now redundant and could just as easily be extracted from the act).

In a map of "TimeWindows" (TimeWindowMap), called TimeWindowMap, the "TimeWindows" are associated with each facility, and stored as a List: for each facility that is visited by agents in the iteration, a list is available of which agents were there, when, and for what reason.
Figure 12 A TimeWindowMap for activities of a single type at a given facility, and the spatial interaction statistics: Statistics are summarized for D’s act. C and D are friends.

<table>
<thead>
<tr>
<th>actStats for Person D</th>
<th>Total for act</th>
</tr>
</thead>
<tbody>
<tr>
<td>nFriends</td>
<td>0 0 1 1 0 0 0 2</td>
</tr>
<tr>
<td>nNonFriends</td>
<td>2 1 1 2 2 1 1 2</td>
</tr>
<tr>
<td>Friend/NonFriend</td>
<td>- - - - - - 0.67</td>
</tr>
<tr>
<td>FriendTimeOverlap (minutes)</td>
<td>0 0 40 55 0 0 0 95</td>
</tr>
</tbody>
</table>

The total value for the act is the statistic used to calculate socializing utility. For nFriends, nNonFriends, and FriendTimeOverlap, the total is simply the sum over all co-present friends during the period in which the ego is present and participating.

The Friend/NonFriend ratio = friend/(foe+0.1*(friend+foe)) is calculated from the total. Its form was chosen mathematically to give a maximum of 10.0 and to avoid division by zero.

The TimeWindowMap structure is a time-map of movements based on the streaming events, and it is useful for facility-specific calculations in the spatial interactor. A set of "TimeWindow" tools called "compareTimeWindows" lets the interactor determine which agents overlapped in time as well as space and activity type by sending the appropriate attributes to compare in the argument list: one or more of time overlap, activity type, and location. The object "actStats", for example, keeps a count of the number of time-overlapping friends of the ego, non-friends of the ego, and the ratio of friends to non-friends co-present at facility and engaged in the same activity type (Figure 12). This aggregate statistic is used in a demonstration social utility function (Section 5.6).

With the "TimeWindow" information available, the interactor carries out instructions to do something with the agents that have been identified as having a chance to encounter one
another. If agents participate in the same activity\textsuperscript{3} during the same time window, it is a face-to-face meeting and particular social processes depending on proximity can take place: droplet infection, violence, material transactions, etc. These detailed interactions are not implemented in the demonstration experiments here, but this is the interactor where they would be realized.

However, we may be interested in modelling not only face-to-face meetings, but plausible social connections in plausible activity spaces, in a representative sense without regard to actual subsecond precision afforded by the MATSim traffic flow simulation. In particular, since the MATSim plans do not represent actual behavior on an actual day, but a sample of "representative" behavior on a "representative" day, the precision of the iterative mobility calculations is more a confounding attribute than a helpful one as far as representing plausible geographic social relationships that could have developed over a longer period than the simulation horizon, but would have been tied to the same general representative activity travel patterns. The researcher may believe that the circumstantial evidence that two agents frequent the same place for the same activity may imply a social association between them that could not be observed in the particular iteration or the particular sample of plans. The "TimeWindow" tools can be used in the spatial interactor for spatial meetings to occur (have occurred) even if the agents missed each other in time during the MATSim iteration. This allows for different kinds of interaction (ant pheromone models, for example) and analyses of, for example, overlapping social and spatial knowledge.

Specific spatial interactors that have been implemented on the basis of "TimeWindows" and which can be chosen in the input configuration file include: agents interact with a certain probability with all agents present (per social network iteration) in overlapping time windows at the same facility; agents interact with a certain probability with one single agent (per social network iteration) present in overlapping time windows at the same facility; agents interact with a certain probability with all other agents (per social network iteration) who used the same facility during the duration of the plan; and agents interact with a certain probability with one other agent (per social network iteration) who used the same facility during the duration of the plan.

5.5.2 Non-spatial interactions

Agents can also activate their social networks independent of spatial proximity or time overlap in order to exchange the information stored in knowledge. The configuration that will

\textsuperscript{3} This same mechanism could be applied to shared travel legs ("vehicles"). The procedure would be more complicated because routes consist of multiple contiguous legs, and utility rewards for shared in-vehicle time would have to be appropriately allocated.
be illustrated permits the exchange of information about activities (facilities) and about the "EgoNet", i.e. other agents, which amounts to a simulation of personal introductions (closure of open social triangles). These two exchanges are chosen because of the clean structure of the utilities for storing and tracking knowledge. Of course it would be desirable for agents to exchange their desired activity start times, for example, or their desired mode of travel. However, the agents have no other place to store this information besides in the Plan object itself, and therefore exchanges of information of this type can only be simulated by changing a plan. However changing elements of Plans is reserved for the Replanning function (covered in Section 5.7).

The exchange of location (facility) information involves setting a pointer in the relevant data structure and is currently implemented such that each agent may make N exchanges of geographic knowledge of a random "activity type" per social network edge per social network iteration. The relative likelihood of exchanging knowledge about locations for each type of activity can be set by the researcher and is set to 1.0 for all types in the demonstrations here (exchanges for all types of activities are equally likely).

As described in Section 5.4, a score is accorded to every facility in agent knowledge to facilitate subsequent simulations of agent cognition, memory, and/or judgment. In the examples that will be shown, each time the same activity (facility) is named to an agent, the score of that location in his knowledge is increased by one. If the memory management is actively culling agents' knowledge of locations on the basis of this score, it simulates reinforcement of common activity spaces within a community or clique. (Note from Section 5.4 that the facilities actually used in the Plans in agent memory cannot be removed from knowledge, so that if a location which is associated with a high-scoring plan is not popular among an agent's friends, it will not be forgotten as long as the plan is not erased in the evolution phase of the iterative loop).

Many other means of exchange and knowledge scoring are imaginable besides random exchange and popularity of the location. For example, each agent could choose to pass the best-scoring, or most valuable, location, to its alters, rather than a random choice. The value of the location (its score) could be based on the plan score associated with the location, or the distance from the agent's home to the location. This may not be the best activity location for the alter receiving the information, however, and using it might result in a lower-scoring plan for this agent. Because of this, the information would not be passed further. Each means of permitting information exchange results in a different percolation of information as each agent acts as an active filter of what information is good or bad (Altenhoff, 2003).

The second currently implemented non-spatial interaction is a way to represent the increased likelihood that friends of friends are also friends, which leads to the observed high clustering
coefficients of real social networks (forming transitive triads from intransitive triads, i.e. closing open triangles, an adaptation of the algorithm described in Jin et al. (2001), see Section 3.3.2). Here, agents associated with each other in a social network can introduce two of their friends to one another, choosing pairs of friends randomly N times per social network iteration.

5.6 SocialScore function

The technical issues of joint plan evolution in MATSim were covered in Section 5.1.2. This section details the formulation of socially-networked utility specifications and illustrates sample implementations which will be demonstrated in chapter 6.

A joint utility function enables the attributes, utility, or attitudes of one or more agents to be used as decision variables in the utility of another agent, in addition to the attributes of the choice alternatives and the characteristics of that agent. These influences may depend on the explicit social relationships, or on average (mean field) influences arising from relevant groups (or groupings, depending on whether the researcher knows the "true" group associations or has to assume them, himself) of individuals. While some forms of utility could be divided across individuals in a group (e.g. splitting the price of a road toll or of the fuel used among vehicle passengers), others cannot (the allergic reaction of one group member to seafood cannot be divided among the members of a group which decided to go to a seafood restaurant together). For this reason and for the technical reasons outlined in Section 5.1.2, in the context of the social network module in MATSim, the utility function remains a property of the ego and its choice set: utility is neither divided across nor maximized for a specific group.

5.6.1 Reasoning behind joint utility and examples

Joint utility functions are explanatory enrichments. Explicitly representing each relationship's or group's contribution to decision making, rather than using a convenient socio-demographic variable (e.g. age, sex, income, race), may seem to add unneeded complications and data requirements to otherwise acceptable models based on aggregate indicators like sociodemographics, cohorts, or hedonics, but the idea is to provide a more accurate and higher-resolved model response to changed conditions. Since the mid-1990's, joint utilities are being increasingly used in network economics (Shy, 2001) to explain paradoxes like the preference of hungry people to wait in line for a table at a popular restaurant rather than satisfy their hunger at an available table around the corner, or using unsatisfactory and expensive computer software rather than choosing a better niche product. In the first instance, waiting produces negative utility, so the queue should warn new arrivals to go someplace else. But in some cases, the presence of the queue is a signal that the restaurant is popular: "the" place to
be. Thus the decision (co-presence) of other decision makers generates positive utility which outweighs the disutilities of both waiting and hunger. In the second instance, the utility of being able to exchange data files with colleagues using the same unsatisfactory software outweighs the disutility of learning to convert files from other formats, and to large and stable market shares for certain software products.

The goal, then, is to represent in a function not only the direct value of an alternative to an individual, but also the extent to which an individual's valuation is influenced by others' characteristics and previous decisions. Some key considerations in the formulation of the function are illustrated by example:

- Social norms may emerge to influence behaviors as people compare themselves to one another in a kind of self-regulation. For example, a person making six single-occupant home-based automobile trips per day may begin to feel awkward if he notices his neighbors only taking one or two chained trips, and reduce his demand for trip-making (by arranging trips into chains) in order to conform. Conversely, all the neighbors may find reinforcement for making six trips per day if their neighbors do it, too. What keeps a positive feedback model like this from spiraling out of control are the other decision constraints which limit how much of the good can be consumed by each individual.

- The reference group may be spatially proximate or diffuse, and described by specified or unspecified relationships. The choice of car to buy (big car or little car) might be influenced by how threatened the buyer feels driving a small car if others are driving a big car. A positive feedback of the proportion of big cars on the road leading to a higher likelihood of purchasing a big car is an example of a mean field effect where only the number of "others" and some aggregate measure of their behavior matters, in contrast to the example above which used immediate neighbors as a reference group.

- The type of relationship might be important, and expressed as a dummy variable in specifying the influence of independent variables on ego's utility. For example, the income of a child's parents may be an important factor in the child's decision of which brand of clothing to wear. However, the choice of clothing brand made by the child's peer group may be just as important in the child's decision. The first "social" influence model is a trivial example that could easily be subsumed by a family income variable, instead, without explicit modelling of social connections between the child and its parents. The second model would be difficult to simplify into demographic indicators without losing precision, though this is what is done in econometric analysis when the social networks are not known. Note that the second model risks correlation with other things that the friends have in common, like income and exposure to the same broader
cultural influences. This hints at the problem of homophily defined in Section 1.2 and formalized in Section 5.6.2.

- The relevant reference characteristic to include in the utility is important. In the above example, the income of the parents is a sociodemographic characteristic, and the choice of clothing made by peers is a decision or a utility valuation.

A model neglecting this "community" or extra-ego component of utility may be tuned or fitted to yield the correct result for a static context, but if the reality of the decision involves interpersonal influence, this model will not respond correctly to a changed scenario if it does not simulate how the decision making of agents are tied together.

5.6.2 Mathematical formulation

Manski (1993) pioneered the formalization of peer effects on decision making by proposing the linear-in-means model that uses terms to describe group-averaged influences on individual utility. This requires groupings of individuals made by the researcher. As Bramoullé et al. (2009) summarize, Manski distinguishes between endogenous effects, exogenous or contextual effects, and correlated effects on individual utility. The first is the influence of peer outcomes (decisions or utility), the second is the influence of peer characteristics, and the third is the confounding problem that individuals in the same reference group tend to behave similarly because they are alike or face a common environment: homophily (Section 3.3.4).

Manski's task was to formalize the mathematical components of joint utility and evaluate their observability. Bramoullé et al. (2009, page 1) write:

"Manski shows that two main identification problems arise in the context of a linear-in-means model. First, it is difficult to distinguish real social effects (endogenous + exogenous) from correlated effects. Second, even in the absence of correlated effects, simultaneity in behavior of interacting agents introduces a perfect collinearity between the expected mean outcome of the group and its mean characteristics. This ‘reflection’ problem hinders the identification of the endogenous effect from the exogenous effects."

The practical implication is that the researcher needs prior knowledge of the groupings of agents, or else such models are only identifiable under certain conditions.

Bramoullé, et al (2009) provide a general formulation of joint utility on a social network:

\[ y = \alpha_0 l + \alpha_1 x + \alpha_2 gy + \alpha_3 gx + \epsilon, \]

83
where $y$ is the vector of individual utilities $y_i$, $1$ is a vector of ones, $x$ is a vector of individual characteristics (only one attribute is shown, for clarity, but the model is valid for any number of them: $x_1, x_2, \ldots$), and $g$ is the general interaction matrix for existing interactions, which may be row-normalized (with $g_{ij} = 1/n_i$ if $j \in N_i$) or not. Each row is a group of peers who influence the decisions of ego $i$, and which does not include $i$. $E(\varepsilon|x) = 0$ and $E(\varepsilon\varepsilon'|x) = \Sigma$ where $\Sigma$ is not restricted but is symmetric and positive definite. The deterministic portion is analagous to spatial autoregression functions (with exogenous effects). The formulation is shown to generalize earlier utility functions which include decision-maker interactions, like the linear-in-means model (Manski, 1993).

The network utility formulation can also be used under assumptions of independent and identically distributed errors in discrete choice models. The approach and model estimation are described by Durlauf and Cohen-Cole (2004), with a notable implementation in transportation behavior by Dugundji and Walker (2005) which estimates mean-field effects of postal code groupings on mode choice in the Dutch Randstadt.

It should be clear from the examples above and from the general discussion that the interaction matrix of alters to use in ego's decision, and which attributes to include about their decisions, person, or judgments, depends on the type of decision making being modelled. The arguments about the identifiability of the function are moot in the agent-based implementation of the joint utility function, since the generative approach allows the researcher to hypothesize the interactions (and correlations). However if the hope is to calibrate an agent model with an estimable joint utility function, its form should be carefully considered based on the data that is anticipated to be collected.

5.6.3 The scope of the decision being represented

Clearly, one could formulate microeconomic utility functions to value social relationships, as well as activities, with which social networks could be formed (Section 3.3.3). The utility would somehow rate the longer-term value of the social connection. A model of this form is a true model of social (or economic, or business) fabric and would be expected to reflect such processes as aging, learning, long- and short-term priorities, and social profit or "capital", etc. Normally, a utility valuation of long-term phenomena is beyond the scope of the scenarios typically treated in a transportation model. Their dynamics are of little interest in the short term in which travel planning takes place, and their representation would require more information than is available anywhere. An additional complication, if explicit valuations of social relationships is desired, is that some of the value of the social relationship might already be realized by the information flowing across the social network described in the
previous section. If the information itself results in decisions of higher value, its utility contribution should not be counted again as part of the value of the social relationship.

An example of a model attempting to incorporate a measure of social capital as a microeconomic behavior into mobility and location decisions was described by Hackney and Axhausen (2006). This very abstract agent model represented the allocation of travel budgets over long-term horizons, allowing them to be spread between travelling and socializing, with explicit social capital rewards for maintaining strategic social connections, which had to be reaffirmed through face-to-face meetings or else they dissolved over time. The agents received a high reward for obtaining "betweenness centrality", and negative utility for travelling to make social contacts. Information was not explicitly exchanged and valued, but the potential for agents to obtain information from social connections was valued as part of "social capital" of the agents' relationships through the betweenness centrality measure of the local social network structure (structural analysis, Wasserman and Faust (1994)). While toy models of this nature may be useful to describe network formation processes given stylized geographic and mobility constraints, they are too detailed (socially) for use in mobility microsimulations like MATSim.

Nevertheless, evolving social networks are desired in this module, which therefore employs the method of random establishment and dissolution of social connections as described in Section 5.3, instead of a utility function valuing social links, in order to avoid the untenable requirement to simulate a lifetime of socializing in order to build social networks for a travel microsimulation.

The simple social scoring function used in the social networks module in MATSim is the standard MATSim scoring function described in Section 4.2.4, plus a socializing bonus for making face to face contact at leisure activities.

### 5.6.4 Form of the social interaction utility term

Four different social interaction functions are used (see Figure 12 for the derivation of the input variables):

1. \[ U = U_{\text{std}} + \beta_{1_{\text{soc}}}*(\text{FriendFoeRatio} \times \delta_{\text{leisure}}) \]
2. \[ U = U_{\text{std}} + \beta_{2_{\text{soc}}}*(\text{Nfriends} \times \delta_{\text{leisure}}) \]
3. \[ U = U_{\text{std}} + \beta_{3_{\text{soc}}}*(\ln(\text{Nfriends}+1) \times \delta_{\text{leisure}}) \]
4. \[ U = U_{\text{std}} + \beta_{4_{\text{soc}}}*(\ln(\text{TotalTimeWithFriends}(hr)) \times \delta_{\text{leisure}}) \]

The first equation simulates an agent's preference for places and times of day in which familiar people are co-present. Equations 2-4 represent different valuations of time spent face
to face with friends: one formulation with constant marginal utility and two with decreasing marginal utility. The socializing statistic, and therefore the bonus, is only nonzero for the relevant activity type(s). The Kronecker $\delta_{\text{leisure}}$ function is illustrative here. It is implicit in the TimeWindow statistical summary and is not part of the utility function.

A positive parameter indicates that the activity has added value for the agent if friends are present to enjoy the activity with the agent; respectively, that those activities with a higher proportion of friends are favored. It is intended to represent many imagineable social interactions, for example the requirement of face-to-face presence in leisure-time meetings without tying agents to an exogenous schedule (for example a book club or planning a hike), the added fun of a party if friends rather than unknown people are there, or the ability to get work done more quickly if others are co-present (also a valid scenario for "leisure time", which could very well include household chores, gardening, etc. for which help is appreciated).

However, it also implies a substitution of socializing valuations for activity or travel time (Figure 15): utility in this system may increase over the case with independent agents (ceteris paribus if agents' ideal plans happen to already overlap), but since ceteris paribus assumptions will likely not hold in the mobility dynamics with this added utility reward unless exogenously constrained, utility may also remain the same or decrease, and be redistributed across activity duration and travel decisions relative to the case without socializing valuations. In addition to social activities with many people present, these utility functions could very well produce longer activity durations, or longer travel distances at highly congested peak periods to enable short encounters between two individuals.

These functions implicitly value the endogenous utility of other agents as manifested in their choice of location and timing of the activity. The functions do not explicitly incorporate endogenous utility or exogenous characteristics of peers. The specific reward structure and the determination of useful parameter values is discussed in depth in chapter 6 and the configuration file for using the function can be is presented as an example in Table 9.

### 5.7 Replanning strategy for cooperative agents

The standard replanning suite available in MATSim was outlined in Section 4.3. Choosing a new location for the secondary activities (shopping and leisure) was not part of MATSim at the time this research began. A new Replanning-strategy for changing the location of secondary activities was implemented for social networks to demonstrate the spread of geographic information in social networks.
Secondary location choice was implemented in the earlier simulation of cooperative mobile agents in the work of Marchal and Nagel (2005), however their spatial encounter algorithm was coupled with information exchanges. The replanning algorithm implemented here is not dependent on the way in which agents know about alternative locations.

A character string of activity types where locations may be freely chosen is loaded from the experiment's configuration file: for example, "leisure", "shop" (see Section 5.9). The agents are chosen randomly, as usual, to replan with the secondary location choice strategy. For each of the activity types, the agents may replace their location with one randomly chosen from a Knowledge resource. Currently there are three implementations of the algorithm.

In the first, the agents draw on their own Knowledge of facilities. This Knowledge may have been already iteratively augmented by social exchanges and mediated by the MentalMap memory manager (forgetfulness). This model simulates the tendency for people looking for new locations to first turn to trusted information sources, in this case, the memory of conversations with friends.

In the second implementation, the agents draw on all facilities in the World. This simulates a person who looks in the telephone book or online, and trusts all information equally, objectively seeking an optimum for his utility function. It is used as a reference case for the optimal search for an activity location.

In the third implementation, agents may choose locations randomly from their and all their friends' Knowledge. This implies agents asking friends directly for advice, rather than relying on memories of conversations with them. This has a similar function to the first model but it occurs each MATSim iteration (i.e. decoupling the social and traffic flow simulation iterations is not possible). The goal of this exchange is that it is simpler in that the information exchange between agents about locations is no longer necessary (the MentalMap memory manager is not used).

The experiments demonstrating these replanning strategies are illustrated in Section 6.3.

### 5.8 Output and analysis tools

#### 5.8.1 Statistics package

MATSim uses varied inputs and produces a complex output landscape in many dimensions. The hypotheses posed about the program's behavior, the tests carried out on it, and the tools available to examine output, are sparse and not standardized. Users are encouraged and indeed required to develop their own tools for their own questions about what the simulation is
producing. The event-based activity travel available in MATSim Events, combined with the Plans file, enable nearly any imaginable transportation behavior query to be posed in post-processing: flows, loads, and time profiles can be summarized for the road network, facilities, households, and agents; trip-, tour-, and plan-based summaries can be analyzed with respect to travel times, distances, routes, and utility components. The MATSim file readers can be used to read in outputs so that postprocessing can be done in Java with all its advantages of speed and flexibility/specificity of the calculation over commercial or non-compiled packages.

The common analyses for which tools have been contributed to MATSim include time profiles of trip begin- and end times, giving an idea of the traffic flows throughout the day ("leg histograms"); average trip travel times, giving an idea of congestion (but also trip distance == route); the development of the average utilities of agents' plans over the iterations, giving an idea of the success of the replanning algorithms at locating a utility optimum and an evaluation of the choice of stopping point; the average trip distances each iteration, giving an idea of the route choice, and the run time analysis for the components of the simulation.

Specific new measures of the traffic flow simulation, some of which include the social networks, include: location distributions, location popularity, activity durations, duration of face-to-face contact, group size at activities (both friends and non-friends), distances travelled to activities, activity start times and participation profiles in time (similar to the leg histograms), utilities by plan component, and marginal utilities by plan component (both per plan and aggregate).

Social networks increase the scope of the analysis manifold. The statistics of the social network are output periodically (the interval is set by the researcher in the configuration file) and include delimited text files of graph average statistics ("graph.txt"), all the social network edges and their characteristics (Section 5.3.2) ("edge.txt"), and all the agents and the summary of their EgoNets ("agent.txt").
Table 6  Contents of the file "edge.txt"

<table>
<thead>
<tr>
<th>iter</th>
<th>MATSim iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>tlast</td>
<td>Iteration of last encounter between ego and alter</td>
</tr>
<tr>
<td>tfirst</td>
<td>Iteration of first encounter between ego and alter</td>
</tr>
<tr>
<td>dist</td>
<td>Euclidean distance between home locations of ego and alter</td>
</tr>
<tr>
<td>egoid</td>
<td>ID number of ego</td>
</tr>
<tr>
<td>alterid</td>
<td>ID number of alter</td>
</tr>
<tr>
<td>purpose</td>
<td>Context of last social engagement (user-definable)</td>
</tr>
<tr>
<td>timesmet</td>
<td>Number of encounters through iterations, since first meeting</td>
</tr>
</tbody>
</table>

Note: the variable, "strength" is currently defined as the geographic distance between alters

Table 7  Contents of the file "agent.txt"

<table>
<thead>
<tr>
<th>iter</th>
<th>MATSim iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>Agent (ego) ID</td>
</tr>
<tr>
<td>homeid</td>
<td>Agent home facility number</td>
</tr>
<tr>
<td>deg</td>
<td>Number of social connections (degree)</td>
</tr>
<tr>
<td>asd1</td>
<td>Activity Space measure 1: Average Euclidean distance to all alters' home locations: geographic radius of the ego net.</td>
</tr>
<tr>
<td>asd2</td>
<td>Activity Space measure 2: Average Euclidean distance to all visited locations</td>
</tr>
<tr>
<td>asd3</td>
<td>Activity Space measure 3: Plan length = sum of Euclidean distances between activity locations in activity chain order</td>
</tr>
<tr>
<td>clust</td>
<td>Clustering coefficient of agent's social relationships (Section 7.2.1)</td>
</tr>
<tr>
<td>plantype</td>
<td>String of first character of each activity type in plan, e.g. &quot;hwh&quot; == home-work-home</td>
</tr>
<tr>
<td>placesknown</td>
<td>Activity Space measure 4: Number of locations in agent's Knowledge</td>
</tr>
<tr>
<td>pop</td>
<td>Population in the zone of the agent's residence, used if desired to normalize. Zones are defined as a MATSim input layer.</td>
</tr>
</tbody>
</table>
Table 8  Contents of the file "graph.txt"

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>iter</td>
<td>MATSim iteration</td>
</tr>
<tr>
<td>deg</td>
<td>Graph average degree (=number of social connections per agent): incoming + outgoing links</td>
</tr>
<tr>
<td>clust</td>
<td>Graph average clustering (Section 7.2.1)</td>
</tr>
<tr>
<td>clustratio</td>
<td>Ratio of graph average clustering to that of an Erdos/Renyi random graph of same number of nodes and average degree</td>
</tr>
<tr>
<td>asd1</td>
<td>Activity Space measure 1: Graph average Euclidean distance from all egos to all alters' home locations: average geographic radius of ego nets</td>
</tr>
<tr>
<td>asd2</td>
<td>Activity Space measure 2: Graph average Euclidean distance from all egos to all visited locations</td>
</tr>
<tr>
<td>asd3</td>
<td>Activity Space measure 3: Average plan length = average over all agents of the sum of Euclidean distances between activity locations in the activity chain order</td>
</tr>
<tr>
<td>dyad_dist</td>
<td>Graph average of all distances between home locations of ego-alter pairs</td>
</tr>
<tr>
<td>link_age</td>
<td>Graph average number of iterations since last encounter between all ego-alter pairs</td>
</tr>
<tr>
<td>meet_freq</td>
<td>Graph average frequency of face to face encounters between agent pairs (average over iterations).</td>
</tr>
</tbody>
</table>

The statistics are currently calculated using the JUNG statistics library in Java (O'Madadhain et al., 2005). This software was originally chosen for its flexibility as long as the exact suite of statistics needed was not known, yet. It is not optimized to calculate specific statistical queries of the MATSim social networks and as such it proves to be burdened by large networks. It is additional overhead in the MATSim release library, and may be replaced by self-written code optimized to the queries used in MATSim runs. Note that the number of graph components, maximum shortest path length, average path length, etc. are very time consuming calculations and are not performed during a MATSim run; these may be performed in postprocessing. Note also that the "edge.txt" file is the only one needed to perform graph statistics in postprocessing. Together with a "plans.xml" file and the geographic initialization files of the experiment ("facilities.xml", "world.xml"), even the geographic measures of social networks can be post-processed. The best way to perform analyses involving calculations with events, plans, and/or social networks is to use the MATSim I/O tools and Java. However two other tools are also used for simpler postprocessing tasks and for graphical analysis.
5.8.2 Pajek output

The graph is output at the same intervals in Pajek format for graphical analysis in two forms: the entire social network and a geographically aggregated representation of social connections. The former is difficult to work on graphically due to its large size. However Pajek can produce additional graph statistics quickly. The latter geographic summary is more useful for graphical depictions. It is a simplified network in which agents are allocated to zones (a "layer" in MATSIM, Section 4.2.2), which become the nodes of the output graph, and the number of interzonal ingoing and outgoing personal relationships are recorded as the edge strengths between the zones (see Figure 21 for an example). This graph's nodes (the zones) can additionally be weighted by the population in the zone, which is also written out, and the graphics can be tailored to suit the weights. The entire Pajek graph, or the geographical graph, can be exported to ESRI GIS products and depicted there as geographically embedded networks for various effects and additional calculations.

The graphical analysis of social networks is time-consuming, and often does not bring insights because its qualitative nature makes comparison between experiments difficult to repeat. Pajek is most useful for its quantitative methods that enable post-processing of statistics which were not foreseen to be needed in the MATSim social networks statistics package.

5.8.3 R statnet and geographic analysis code

While Pajek handles special cases and can be used to make images for presentation, brute-force statistical post-processing and graphical output are run instead in batch using R.

One function produces analyses of the agent, edge, and graph files, including changes occurring over the iterations of the MATSim optimization.

The other function requires an intermediate MATSim postprocessing run of an EventHandler to produce a table of the numbers of agents at facilities at different times and a time profile of the number of agents engaged in each activity.

5.9 Sample implementation

An annotated excerpt from a simple configuration file in Table 9 illustrates how an experiment is carried out. The social networks configuration is added as a module to the standard MATSim configuration file. The section headings are comments only and are divided up to correspond to the steps in the flowchart in Figure 10. Only the relevant sections (replanning strategy and social networks) are shown.
Table 9  Simple social network configuration file

```xml
<module name="strategy">
  <!-- Add Modules here, Probability values get all summed up and normalized to 1.0 -->
  <param name="Module_1" value="SelectExpBeta" />
  <param name="ModuleProbability_1" value="0.7" />
  <param name="Module_2" value="ReRoute" />
  <param name="ModuleProbability_2" value="0.1" />
  <param name="Module_3" value="TimeAllocationMutator7200_ReRouteLandmarks" />
  <param name="ModuleProbability_3" value="0.1" />
  <param name="Module_4" value="KSecLoc" />
  <param name="ModuleProbability_4" value="0.1" />
</module> <!-- strategy -->

Replanning strategy
Rule for selecting plans not changed this iteration.

> Standard rerouting
Standard activity time mutator
Change secondary location by selecting from own knowledge

... <module name="socialnetwork" >

  <!-- How many replanning iterations occur before letting agents socially interact again -->
  <param name="replanning_interval" value="1" />
  <param name="reporting_interval" value="50" />

  <!-- OUTPUT DIRECTORY -->
  <param name="outputDirSocialNets" value="&OUTPUTBASE;/socialnets" />

  <!-- INITIALIZE THE SOCIAL TIES -->
  <param name="socnetalgorithm" value="euclidrandom" />
  <param name="kbar" value="12" />
  <param name="edge_type" value="UNDIRECTED" />
  <param name="inputSocNetDir" value="&INPUTBASE;/socialnets" />

Social network loop
Number of MATSim iterations per social iteration
How frequently to write out social network statistics

Initial social network
Construction rule for initial social network
Average degree of initial social network
Link type
If social net is read in ("edge.txt")
```

...
Table 9  Simple social network configuration file

<param name="inputIter" value = "0" />
<param name="euclid_alpha" value ="-1.5" />
<param name="euclid_rmin" value ="1000" />

<!-- SOCIAL NETWORK DYNAMICS -->
<param name="socnetlinkremovalage" value="0" />
<param name="socnetlinkremovalalgorithm" value="random_constant_kbar" />
<param name="socnetlinkremovalp" value="0.05" />
<param name="socnetlinkstrengthalgorithm" value="constant" />
<param name="prob_befriend" value="1." />
<param name="degree_saturation_rate" value="0" />

<!-- INFORMATION EXCHANGES IN SOCIAL SPACE -->
<param name="factype_ns" value="home,leisure,shop,education,work" />
<param name="fract_introduce_friends" value="0.0" />
<param name="fract_ns_interact" value="1.0" />
<param name="num_ns_interactions" value="1!" />
<param name="memSize" value="1.5" />

<!-- INFORMATION EXCHANGES IN GEOGRAPHICAL SPACE -->
<param name="spatial_interactor_type" value="timewindowrandom" />
<param name="s_weights" value="1.,1.,1.,1.,1." />
<param name="act_types" value="home,work,shop,education,leisure" />

<!-- SCORING -->
<param name="betafriendfoe" value = "0" />
<param name="betanfriends" value = "0" />

Iteration to read in for initializing from an "edge.txt" file
Exponent for Euclidrandom network construction algorithm
Min. distance for Euclidrandom network construction algorithm

Social network evolution
First MATSim iteration to attempt social link removal
Algorithm for removing social links
Parameter for random link removal
Algorithm for modifying social link strength

Probability of making a friend given an encounter
Exponent of degree saturation

Nonspatial information exchanges
Type of location information to exchange
Fraction of agents closing intransitive triangles
Number of actions each agent may take this social iteration
Ratio of number of locations in Knowledge vs. number in plans

Face to face information exchanges/interactions
Time/space conditions in which encounters may take place
Probability of encounters corresponding to the following activities:
... the corresponding activities

Social scoring function (leisure only)
Utility parameter for ratio of friends to non-friends at activity
Utility parameter for number of friends at activity
Table 9  Simple social network configuration file

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;param name=&quot;betalognfriends&quot; value = &quot;10&quot; /&gt;</td>
<td>Utility parameter for log of number of friends at activity</td>
<td></td>
</tr>
<tr>
<td>&lt;param name=&quot;betatimewithfriends&quot; value = &quot;0&quot; /&gt;</td>
<td>Utility parameter for total time spent with friends at activity</td>
<td></td>
</tr>
<tr>
<td>&lt;!-- REPLANNING --&gt;</td>
<td></td>
<td>Replanning parameters</td>
</tr>
<tr>
<td>&lt;param name=&quot;switch_weights&quot; value=&quot;0.0,0.1.0,0.1.0&quot; /&gt;</td>
<td>Probability of switching this location this MATSim iteration:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;home&quot;,&quot;work&quot;,&quot;shop&quot;,&quot;education&quot;,&quot;leisure&quot;</td>
<td></td>
</tr>
</tbody>
</table>

</module>
The experiment directed by this configuration file initializes a social network with $k_{bar} \cdot N/2$ random associations whose likelihood diminishes with the distance between home locations of the agents, making a network with shorter geographic distances between egos and alters than purely random links would imply. The maximum probability is constant and scaled to 1.0 for agent pairs chosen which have a separation between 0 and 1000m, and it decreases with separation to the exponent $-1.5$. The ties are undirected, i.e. communication is possible in both directions, while the edge is counted only once in the statistics (this is a definition of "graph" that is consistent with traditional graph theory). The agents interact socially each MATSim iteration, and social nets results are written out every 50 iterations. The social interactions consist first of all in making friends with other agents encountered face to face. This occurs with 100% probability, in all types of activities. The average social network degree is kept at 12 by deleting excess relationships randomly. Finally, of each agent has a chance to exchange knowledge of the location of one of any type of facility with each of its friends. A maximum of 1.5 times as many facilities as the agent has in its plans can be retained in its memory (according to the "popularity" algorithm of facility scoring). The replanning strategies are: rerouting, shifting activities in time, and replacing the activity location of shopping and leisure with locations from the agent's own knowledge. The agents also receive a scoring bonus of 10 times the natural logarithm of the number of friends encountered at leisure activities.

Table 9 will be referred to frequently in chapter 6, which illustrates the separate effects of each of these settings on the MATSim output.


6 Verification experiments

In view of the lack of a comprehensive and large dataset of observations for comparison, validation of the model system, i.e. precise reproduction of real reference values, is not the goal. While qualitative similarities in the emergent agent behavior to reality certainly reinforce the credibility of any model, there are several other goals to the verification procedure pursued here:

- To develop appropriate tools and measures for probing, aggregating, simplifying, and representing the complex spatiotemporal social-travel output;
- To use these tools to describe the emergent behavior of the agents;
- To learn to distinguish, recognize and reliably detect signatures of given algorithms and/or their parameter values in the output; and
- To evaluate which algorithms may or may not be useful for simulating socially-interactive travel behavior in MATSim.

To this end, hypotheses are posed about the model system's response when various algorithms of agent social behavior are incorporated into the iterative directed relaxation cycle. The experimental plan consists of a broad battery of simulation experiments oriented toward efficiently addressing the hypotheses. It is unavoidable given the many dimensions of the output that multiple experiments touch upon more than one hypothesis, and vice-versa. The results enable characterization of the model responses to the specific implementations of the social interaction models, and evaluation of whether the emergent result qualitatively or quantitatively resembles expected and/or observed relationships (chapter 3). This is an indirect verification process (Nicholson, 1995).\(^4\) It is hoped that the conclusions enable qualitative generalizations to be made about coupling agent interactions in an iterative directed system relaxation process such as that used in MATSim.

This chapter delineates the goals of the experiments and their configuration, outlines the expected outcomes, and links the experiments to the hypotheses about model behavior.

6.1 Assumptions and fundamental structure of the experiments

Travel, waiting, and late arrival, are generalized costs, or disutility, for individual agents (chapter 4.) Also, for each agent, participation in activities produces positive utility with

\(^4\) The real processes are unknown and so the abstract behavior models could be assumed to be valid if the result is plausible, thus this process is also a kind of validation.
decreasing returns to the time spent participating. The directed system relaxation is effectuated by iterative improvements to individual plan utility, as it is in the standard MATSim simulation.

In the experiments with social networks, long-term social ties or contexts are not valued in utility, but the social network will have an influence on utility, as it is assumed to be able to affect the travel patterns in three ways: 1) in experiments allowing the exchange of knowledge, it may have bearing on an agent's choice sets for locations, endogenizing the information in social clusters., 2) in experiments rewarding face-to-face encounters with alters directly in the utility function (chapter 5), the utility reward is a time-and-space varying location attribute, endogenized by agents' time allocation and location choice, and 3) social networks are assumed in some experiments to evolve in conformity with the travel patterns of the agents through a reinforcement process in series with the MATSim iteration, which affects the information flow between agents. Alone or together, these mechanisms provide a feedback to socially moderated activity-travel behavior, and that coupling is the focus of study.

6.2 Hypotheses

The verification experiments enable linking the cause and effect between algorithms and model output. As such, the hypotheses do not describe agent- as much as model behavior. Three basic effects are expected in each model (experiment) of activity travel and social interaction. The first type of hypotheses tests the approach to social network modelling by investigating the topological properties of the emergent social network; the second type evaluates changes to the travel behavior model caused by the social networks assumptions; and the third questions what the coupled model can reveal about the embedding of social networks in geography. In a way, a final hypothesis is that these effects can be detected and discerned from other processes in the model. The hypotheses overlap in some cases and the three groupings are not necessarily exclusive from one another.

6.2.1 Social ties

The module offers a general method to generate social networks that are embedded in space and/or activity-travel patterns.

Real social networks often exhibit a "fat-tailed" (positive skew) degree distribution with higher-than-Normal (or Poisson) probability of high-degree nodes. It is explained as a characteristic of systems exhibiting growth (addition of new nodes to an existing graph). It is also often observed in real social networks that clusters form in which the frequency of
membership in three-party triangles is high, and between which infrequent single ties connect the clusters, forming "long distance relationships" (in the geodesic, not geographic, sense).

A series of hypotheses ask the question, what social networks emerge from the specific mechanisms?

They test whether such distributions can be generated in social networks by geographic and activity-travel considerations (face to face meeting within the framework of maximizing activity travel utility). This would not rule out that non-explicitly modelled systematic social processes might underly the geographic effects, but it would point to a dominant role of geography and activity-travel at least as latent processes (variables) in contributing to observed statistics of social networks.

1) A basic MATSim scenario has no social network, randomly allocates closest shopping and leisure locations, and a realistic distribution of home, work, and education locations. Three bipartite mappings can be generated to quantify a bounding social network of spatio-temporal overlap of agents crossing paths at activities:

- The mapping of face to face meetings of agents in the relaxed state of a basic scenario;
- The mapping of locations visited and activities pursued during the day (bipartite network of activity/locations and agents);
- The mapping of locations visited during the day (bipartite mapping of locations and agents).

The activity types, schedules, and locations connect certain agents and are mechanisms to cluster their ties. High clustering is a characteristic of bipartite graphs, as associating any three nodes of the first type to one node of the second type forces the three to form a triangle in the transformation of the bipartite graph to nodes of the first type. Furthermore, the tie between the two different types of nodes in a bipartite network is a classification and thus aggregation of characteristics, which further clusters nodes of the other type. This is necessarily mono-variate and does not account for joint distributions of group membership. Grouping the agents in a bipartite network is not a question of being in this group AND that group, but this group OR that group, further reducing the precision of associations. Thus ties between agents and multiple nodes which represent categories of characteristics will associate many agents together in clusters when the network is represented with a single node type (Dorogovtsev and Mendes, 2003).

Therefore, these social networks are hypothesized to be small world social networks with higher clustering and longer shortest paths than are expected for a classical random graph of the same degree and number of nodes. A classical random graph of average degree K results
from this generation algorithm if, over K iterations, each agent is instructed to add one social contact at an activity where all activities were of the same type, the number of activities as well as facilities in the agents' plans would equal the number of agents, these facilities were equally accessible to all agents from all locations, and the allocation of the locations to each agent's activities were random. Any perturbation to this perfect homogeneity would form a complex graph, i.e. a small world. As an initial reference, the activity types/locations/schedules are expected to be highly atomized in (a) with little time overlap of agents at the activity. Thus this social network will have low degree, but the clustering coefficient will be very sensitive to time overlaps at activities and could be high or low. The selectiveness of the social ties decreases in the order listed (a-c); i.e. the probability of agents to be included in one another's ego nets increases. One would expect corresponding increasing average degree and clustering.

2) An equilibrium dynamic social network can be generated within the MATSim iterative directed system relaxation framework using a combination of meeting rules and link removal rules (where equilibrium is defined as a constant clustering coefficient, distance distribution to friends, and degree (Hackney and Marchal, 2009)). Such an equilibrium was previously sought unsuccessfully, with less restrictive mechanisms (Hackney and Axhausen, 2006).

3) Including the mechanism which introduces friends of friends to produce transitive triangles, independent of space-time, increases clustering and the incidence of long-distance connections.

6.2.2 Insights into transportation behavior microsimulation

Ideally, generalizations would be hoped for about agent interactions in agent-based activity traffic flow simulations. But the iterative directed utility maximization algorithm of MATSim limits the planning time horizon, and the range of social interactions in terms of the cost/benefit models of investments in socializing, that other model forms may be able to incorporate better. The insights about the influence of travel on the evolution of social relationships are limited, and the focus of the research is on what the models of social interactions reveal about this transportation behavior model:

1) Evidence can be found of coordination of activity schedules by a social interaction bonus term in the utility function

2) A utility bonus for face to face socializing means that agents will trade off longer distance trips or longer travel times for the opportunity to meet with friends.
3) Information exchange on a social network about secondary activity locations that reinforces choice sets results in secondary location choice that is better than a random allocation of locations (as measured by utility) but sub-optimal compared to free secondary location choice that is not socially-mediated.

4) Rewarding face to face social interactions in conjunction with free choice of secondary locations results in a different distribution of destination choice relative to the same secondary location choice algorithm without extra social utility rewards (exchange of location information alone).

5) Rewarding face to face socializing at activities results in fewer facilities being chosen, and in these being more popular, than in a scenario without this reward.

6) In models rewarding face-to-face co-presence, these activities will be of shorter duration (other activities or travel duration can be longer).

6.2.3 Social geography

These hypotheses are posed to deepen the understanding of geographically-embedded social networks which also depend on activity-travel. How are social networks embedded in space given coupling of social links with activity travel demand?

1) Social network effects on activity travel behavior can be distinguished from the effects of the geographic framework like population density, density of facilities, and accessibility measures.

2) In a social network which evolves with the activity-travel plan, or which mediates secondary location choices, the geographic extent of the ego networks is larger than the geographic extent of the locations used by the agent in the agent's final location choice.

3) In a static social network, the geographic extent of the ego networks may be larger or smaller than the geographic extent of the locations used by the agent who has free location choice. Agents which are free to choose locations with the standard utility function will attempt to reduce the travel times. This either means choosing locations nearby, choosing locations on high-speed, high-capacity roadways which may be farther from home, or choosing activity times during the off-peak travel periods (if possible). If agents are maximizing utility independently, they thus may well increase their travel distances beyond the geographic extent of their ego network, even if they learn about activity locations through friends. Feedback which rewards face-to-face meetings links the geographic size of the ego network with the geographic extent of the activities. This will tend to concentrate friends in the same locations (at the same times), according to how much utility the meeting brings to
each friend's total day's utility. This reward structure is expected to lead to location choice in which the travel cost is essentially split across the friends' ability to reach the location in relation to its other time obligations. Rewarding the participation by the number of other agents present is a way to weight this activity utility against the travel disincentive such that even agents living farther from the location may choose it over a closer location, as a result of the presence of friends (i.e. the combination of the size of the ego network and the ability of the alters to travel to the location at the given time). The physical extent of the activity locations is likely to be smaller than that of the ego network. Finally, if new friends are made at activities, the geographic extent of the ego network versus the activity locations will be very similar.

4) In a social network which evolves with the activity-travel plan, an agent's spatial knowledge covers a much larger geographic area than the area subtended by the ego nets.

5) In a social network which evolves with the activity-travel plan, the link probability distribution vs. distance between agent home locations is exponentially decreasing.

6) Social network effects can substitute for exogenously fixed variables and endogenize behavior in the model: social norms can replace exogenously established desired start time; desired duration; activity location.

7) In a social network which evolves with the activity-travel plan, the agent's degree scales with the number of activities (will depend on the rules for making friendships face to face).

8) In a social network which evolves with the activity-travel plan, the agent's degree scales with the size of the facilities visited (number of other agents at the facilities) it visits (related to weighted accessibility measures of facilities)

### 6.3 Experiments

The influence of the social interactions module is investigated one behavioral or model dimension at a time in the experiments. The model results, i.e. the effects of the behavioral assumptions, are compared to each other and to reference model configurations, rather than to observations. Evidence relevant to the hypotheses is produced in each experiment. Conclusions regarding the hypotheses will be derived from comparisons between the experiments in Section 8.1.

The experiments systematically alter the social network topology, its geography, the utility reward, information exchange in the form of knowledge about secondary locations, and
observe social network evolution. Table 10 presents a summary of the experiments, which are explained in detail in Section 6.4
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Secondary Location Choice</th>
<th>Social Network</th>
<th>Utility Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All locations available</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social knowledge exchange; use own knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use alters' knowledge of locations directly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>6</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>7</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>12</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>22</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>102</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>104</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>4_4</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>13_4</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>23_4</td>
<td>X</td>
<td>None</td>
<td>X</td>
</tr>
</tbody>
</table>
The experiments are compared with one another and evaluated in support or refutation of hypotheses in Section 6.6, as follows:

<table>
<thead>
<tr>
<th>Experiments</th>
<th>Section</th>
<th>Main Category</th>
<th>Subcategory</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>6.6.3</td>
<td>Time rescheduling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reference socializing utility reward</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>6.6.3</td>
<td>Time rescheduling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Two socializing utility functions</td>
</tr>
<tr>
<td>12</td>
<td>22</td>
<td>6.6.3</td>
<td>Time rescheduling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Socializing utility parameter values</td>
</tr>
<tr>
<td>2</td>
<td>102</td>
<td>6.6.4</td>
<td>Time rescheduling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Social network density</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>6.6.5</td>
<td>Secondary location choice</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reference unconstrained location choice</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>6.6.6</td>
<td>Secondary location choice</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Constrained location choice, socializing utility reward</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6.6.8</td>
<td>Secondary location choice</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Constrained location choice, socializing utility reward</td>
</tr>
<tr>
<td>4</td>
<td>104</td>
<td>6.6.7</td>
<td>Secondary location choice</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Social network spatial contraction</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6.6.9</td>
<td>Social network evolution</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Without socializing utility</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>6.6.10</td>
<td>Social network evolution</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>With socializing utility</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>6.6.8</td>
<td>Utility reward</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Location choice, static social network</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6.6.10</td>
<td>Utility reward</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Location choice, evolving social network</td>
</tr>
<tr>
<td>4</td>
<td>4_4</td>
<td>6.6.6</td>
<td>Type of social interaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Without socializing utility</td>
</tr>
<tr>
<td>3</td>
<td>23_4</td>
<td>6.6.8</td>
<td>Type of social interaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>With socializing utility</td>
</tr>
<tr>
<td>4_4</td>
<td>23_4</td>
<td>6.6.8</td>
<td>Utility reward</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Location choice, static social network</td>
</tr>
<tr>
<td>13_4</td>
<td>23_4</td>
<td>6.6.8</td>
<td>Utility reward</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Socializing utility parameter values</td>
</tr>
</tbody>
</table>

Additional experiments examine the effect of constraints imposed by standard MATSim model settings on the scoring function and the facility opening times, which might affect the
socializing behavior. All of the experiments and their components are described in detail in Section 6.4.

### 6.4 Details of the Experiments

Table 12 summarizes 9 configurations of the socializing dynamics, including with and without social networks (Sections 6.4.1 and 6.4.2), exchange of location information (Section 6.4.3), and utility rewards for socializing (Section 6.4.4). Table 13 summarizes six additional configurations which evaluate the effects of common MATSim assumptions about the utility function, facility opening times, and the desired activity durations and start times. Finally, the four models in Table 14 use a different Controller for an alternate means to model the sharing of location information (Section 6.4.5).

The meaning of the table columns is described immediately after the tables without delving into the specific expectations of the experiments, which follows in Section 6.6.
<table>
<thead>
<tr>
<th>Name</th>
<th>SocialnetInit</th>
<th>SocialnetEvolve</th>
<th>Social exchange</th>
<th>Scoring</th>
<th>Replanning</th>
<th>Test Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>independent agents</td>
<td>none</td>
<td>none</td>
<td>standard</td>
<td>Time, Route</td>
<td>Standard reference without social networks and ensemble</td>
</tr>
<tr>
<td>2</td>
<td>Euclid, deg 12</td>
<td>none</td>
<td>none</td>
<td>leisure bonus $10*\ln(nf)$</td>
<td>Time, Route</td>
<td>Face to face utility reward</td>
</tr>
<tr>
<td>3</td>
<td>Euclid, deg 12</td>
<td>none</td>
<td>location knowledge 1</td>
<td>leisure bonus $10*\ln(nf)$</td>
<td>Time, Route, Knowledge 2ndary Loc</td>
<td>Location choice with face to face utility reward</td>
</tr>
<tr>
<td>4</td>
<td>Euclid, deg 12</td>
<td>none</td>
<td>location knowledge 1</td>
<td>standard</td>
<td>Time, Route, Knowledge 2ndary Loc</td>
<td>Location choice with information exchange and standard utility</td>
</tr>
<tr>
<td>5</td>
<td>Euclid, deg 12</td>
<td>deg 12, &lt;= 1 new friend/act</td>
<td>location knowledge 1</td>
<td>standard</td>
<td>Time, Route, Knowledge 2ndary Loc</td>
<td>Location choice with information exchange and standard utility and social network evolution</td>
</tr>
<tr>
<td>6</td>
<td>Euclid, deg 12</td>
<td>deg 12, &lt;= 1 new friend/act</td>
<td>location knowledge 1</td>
<td>leisure bonus $10*\ln(nf)$</td>
<td>Time, Route, Knowledge 2ndary Loc</td>
<td>Location choice with face to face utility reward and social network evolution</td>
</tr>
<tr>
<td>7</td>
<td>independent agents</td>
<td>none</td>
<td>none</td>
<td>standard</td>
<td>Time, Route, Omniscient 2ndary Loc</td>
<td>Location choice reference with standard utility and unlimited choice set</td>
</tr>
<tr>
<td>12</td>
<td>Euclid, deg 12</td>
<td>none</td>
<td>none</td>
<td>leisure bonus $10*\ln(tf)$</td>
<td>Time, Route</td>
<td>Face to face duration utility reward</td>
</tr>
<tr>
<td></td>
<td>Configurations of the verification experiments for the social network interactions and the joint utility function (continued)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>Euclid, deg 24, none, leisure bonus $10*\ln(nf)$, Time, Route, Face to face utility reward, Location choice with information exchange and standard utility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>Euclid dense, deg 12, location knowledge 1, standard, Time, Route, Knowledge 2ndary Loc, Location choice with information exchange and standard utility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1_501_F2F</td>
<td>Experiment 1 plans, Befriend All, none, none, none, Map of all face to face meetings Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7_501_F2F</td>
<td>Experiment 7 plans, Befriend All, none, none, none, Map of all face to face meetings Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 13 Configurations of the verification experiments that test the assumptions of boundary conditions which affect the utility valuations of activities

<table>
<thead>
<tr>
<th>Name</th>
<th>SocialnetInit</th>
<th>SocialnetEvolve</th>
<th>Social exchange</th>
<th>Scoring</th>
<th>Replanning</th>
<th>Test Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>Euclid, deg 12</td>
<td>none</td>
<td>none</td>
<td>leisure bonus 24*ln(tf)</td>
<td>Time, Route</td>
<td>Face to face duration utility reward with high parameter value</td>
</tr>
<tr>
<td>32</td>
<td>Euclid, deg 12</td>
<td>none</td>
<td>none</td>
<td>leisure bonus 24*ln(tf)</td>
<td>Time, Route</td>
<td>Face to face duration utility reward with high parameter value. No minimum durations for activities.</td>
</tr>
<tr>
<td>42</td>
<td>Euclid, deg 12</td>
<td>none</td>
<td>none</td>
<td>leisure bonus 24*ln(tf)</td>
<td>Time, Route</td>
<td>Face to face duration utility reward with high parameter value. All opening times and desired activity start times are undefined.</td>
</tr>
<tr>
<td>51</td>
<td>independent agents</td>
<td>none</td>
<td>none</td>
<td>standard</td>
<td>Time, Route</td>
<td>No minimum durations for activities. No late penalty.</td>
</tr>
<tr>
<td>52</td>
<td>Euclid, deg 12</td>
<td>none</td>
<td>none</td>
<td>leisure bonus 24*ln(tf)</td>
<td>Time, Route</td>
<td>Face to face duration utility reward with high parameter value. No minimum durations. No opening and closing times. No late penalty.</td>
</tr>
<tr>
<td>72</td>
<td>Euclid, deg 12</td>
<td>none</td>
<td>none</td>
<td>leisure bonus 24*ln(tf)</td>
<td>Time, Route</td>
<td>Face to face duration utility reward with high parameter value. No minimum durations for activities. Facilities all open 24hrs.</td>
</tr>
</tbody>
</table>
Table 14 Configurations of the verification experiments that use knowledge of friends as the choice set for location choice

<table>
<thead>
<tr>
<th>Name</th>
<th>SocialnetInit</th>
<th>SocialnetEvolve</th>
<th>Social exchange</th>
<th>Scoring</th>
<th>Replanning</th>
<th>Test Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>4_4</td>
<td>Euclid, deg 12</td>
<td>none</td>
<td>location knowledge 2</td>
<td>standard</td>
<td>Time, Route, SNPick</td>
<td>Choose secondary locations from friends' knowledge.</td>
</tr>
<tr>
<td>13_4</td>
<td>Euclid, deg 12</td>
<td>none</td>
<td>none</td>
<td>leisure bonus 24*ln(tf)</td>
<td>Time, Route, SNPick</td>
<td>Choose secondary locations from friends' knowledge with face to face utility reward. Minimum duration =0.</td>
</tr>
<tr>
<td>23_4</td>
<td>Euclid, deg 12</td>
<td>none</td>
<td>location knowledge 2</td>
<td>leisure bonus 10*ln(nf)</td>
<td>Time, Route, SNPick</td>
<td>Choose secondary locations from friends' knowledge with face to face utility reward. Compare to 3</td>
</tr>
</tbody>
</table>
6.4.1 SocialnetInit

This lists the initial social network used. All social networks used in the experiments are undirected (symmetric); i.e. each edge permits flows of influence in both directions and the relationship is counted only once per agent, despite there being two-way communication. "Independent agents" means the standard MATSim run in which no information can be shared. "Euclid, deg 12" is the standard social network for the experiments, used in 2, 3, 4, 5, 6, 12, 22, 32, 42, 52, 72, 4_4, 13_4, and 23_4. It and "Euclid, deg 24" (used in 102) is a random social network generation algorithm which distributes social links randomly with a spatial weighting until the average graph degree is 12 or 24, respectively. "Euclid dense, deg 12" (104) is a spatially contracted version of the first spatial network of average degree 12. "Random12" is a reference classical random network that has no spatial weighting for friendship (spatial characteristics of the graph depend entirely on the geometry of the World and the spatial distribution of the population).

The spatial embedding generation algorithm proceeds by choosing two agents at random. If they are not already linked socially, the algorithm tries to associate them by establishing a link. This is the basis for a random graph generation algorithm and it results in a Poisson degree distribution. To influence the spatial form of the random network, the probability that the tie will be made from agent $i$ to agent $j$ is made to be proportional to the spatial distance between the agents' home locations:

$$p_{ij} = \left( \frac{r_{ij} + r_{\min}}{r_{\min}} \right)^{-\alpha}, \alpha > 0$$

where $r_{ij}$ is the (Euclidean planar) distance between home locations of the agents, and $\alpha$ is typically between 1.5 and 2.5 (Section 2.2.2) ($r_{\min} = 1000$m and $\alpha = 1.5$ for the verification tests, here). Thus the distance-based probability is 1.0 if the agents share the same home location, and the probability of becoming friends decreases exponentially the farther away they live from one another to approximate $1/\text{distance}^\alpha$ at the long-distance limit.

The social network is random with respect to person attributes or attributes of the social network itself: there are no edge correlations or other personal linking preferences. But it concentrates friendships into spatially compact areas by preferentially linking spatially proximate agents. Because some agents are favored over others, indeed spatially co-located ones, the social network is by construction spatially clustered.

Returning only to the distance parameter used in network construction, the distance-dependent probability of each ego-alter tie is illustrated in Figure 13 for three values of $\alpha$ and
r_{min}, and for the distance-independent probability as in a classical random graph. This model is similar to the model used in Butts and Carley (1999) and directly reflects the fits in Table 2 and Table 3, as well as the qualitative features of the distance decay in the model in Figure 2. The final distribution of the dyad separations in the social network is a convolution integral of the spatial distribution of agents within the scenario and this distance-dependent function for making friends, given the edge geometry of the scenario as a boundary condition of the integral.

A distribution concentrating relationships physically proximate to an ego may represent poor mobility or strong neighborhood identity. A flatter distribution with distance may represent high mobility and active search for optimal social partners. In a modelling sense, the former has the benefit of reducing the confounding effects of the geographic boundary of the scenario, while the distribution of ego-alter distances in the latter case is determined strongly by this geometry. In a more realistic model, it would be expected that the parameters $\alpha$ and $r_{min}$ would be themselves distributed across the population and would be derived from person-related characteristics.

Figure 13 The distance-dependent probability weighting of a social tie.

![Graph showing the distance-dependent probability weighting of a social tie.](image)

Given a potential dyad of agents $i$ and $j$. The probability weighting with distance for a non-spatial classical random network (shown for reference with $\alpha$ and $r_{min} = 0$) is 1.0.
6.4.2 SocialnetEvolve

The evolution algorithm for the social network proceeds in series with the MATSim directed system relaxation iteration. The abbreviation "deg 12, meet 1/act" (5,6) means that the agents are permitted to add up to one new friendship tie to their egonet per activity within the time window of co-presence, per iteration (making a new friend as a result of face-to-face contact). In the experiments illustrated here, the determination of which agent to add to the egonet is made randomly. If the social tie which the agent wishes to initiate exists already, this tie is reinforced by adding 1 to its strength. The context of the social encounter is recorded as an attribute of the social edge (e.g., "renew work", "new leisure", etc.). The iteration that the social network edge was established is also recorded as an edge attribute. Immediately after all the agents have had the opportunity to create new social edges based on their face-to-face encounters, the average number of alters (degree) is maintained constant at 12 through random removal of the requisite number of social network edges (as in Jin et al., 2001). This culling is an abstract way to represent cognitive constraints on the agents of maintaining social ties. Since the algorithm chooses edges between agents randomly with equal weighting, all agents lose the same percentage of their social edges, and those agents with more edges (high degree) will tend to lose a higher absolute number of their connections. This mechanism reduces the positive skew of the degree distribution (Jin et al., 2001).

6.4.3 Social exchange

The type of social exchange that is permitted is closely related to the replanning algorithms (Section 6.4.5). The exchange occurs along the intact social edges in the social network, regardless of the face-to-face encounters this iteration. Any algorithm can be implemented in this fashion. Two algorithms for the exchange of location knowledge were tested here. The other currently implemented "interaction" is the introduction of friends of friends to one another, to increase local social clustering. This was not tested on this scenario with 8760 agents but in a dataset of 1008 agents used for code development, it was found to have the expected effect of quickly raising the clustering coefficient.

Social knowledge exchange: Each iteration, for each alter, each ego chooses a random location from its Knowledge (of any type, though the experimenter can choose to only allow certain types to be exchanged) and gives it to the alter. Each time an agent receives a location from an alter in its ego network, the "score" for that location increases by one. This score has nothing to do with the plan score (utility), though in the general way that the model is programmed, the plan score associated with the location could certainly be used instead. The "location score" is a generic double valued parameter intended to ascribe some measure of value the location has to an agent. After each agent has played the role of offering one location, the locations of all agents are sorted in order of decreasing score. The lowest-scoring
(least frequently mentioned) locations are erased from each agent's memory by the Memory Manager within the Mental Map structure. The number of locations to erase for each agent is determined by the locations needed for the plans in memory (generally 4 plans), and a user-entered parameter. The locations needed for the plans in memory may not be forgotten until the plan is erased from memory. This occurs in the scoring phase of the iteration and will almost always cause superfluous locations to be retained in memory (and passed on to alters) for at least one iteration, until they may or may not be removed by the Memory Manager. The user-entered parameter for memory management indicates a multiple of the number of locations to be kept in excess of the number of locations needed for all plans in the agent's memory. The value used is 1.5, or 50% more locations than each agent needs for the plans in memory (Cases 3, 4, 5, 6, 104).

Social knowledge exchange 2: This is an attempt to model the same exchange above in a simpler form. It requires a new replanning strategy. The simplification is that there is no "nonspatial interaction" class required; in fact, and the exchange takes place during the replanning phase of the iteration. Each iteration, each ego chooses to learn a new location randomly from all the knowledge of the alters in its ego net. Thus the ego has access to the Knowledge objects of its alters. Knowledge percolates through the social network as the chosen location is immediately incorporated into the plan being modified, and added to the ego's own knowledge. The choice of location is made with a Logit function in which the alternative probabilities are equal to the frequency with which the locations appear in the alters' knowledge. Because only locations which are permitted to be switched (secondary locations: shop, leisure), much less information is exchanged (home, work, and education facilities are not exchanged), and memory management is not used. Cases (4_4, 13_4, 23_4).

6.4.4 Scoring

The scoring function is enhanced in some models with two different socializing functions: "leisure bonus $\beta\ln(nf)$" (case 2,3,6,102, 23_4) and "leisure bonus $\beta\ln(tf)$" (case 12), with $\beta=10$. The latter is also tested with $\beta=24$ (cases 22, 32, 42, 52, 13_4). The first function weights only the number of alters present. It is expected to have less precise influence on activity timing decisions. The second is a function weighting the total time (hrs) that the ego spends face to face with its alters at each leisure activity. See Figure 12.

All configurations use the following generalized cost parameters ($\beta$, cost/hr) unless otherwise noted:

- earlyDeparture = -0.0
- lateArrival = -18.0
• performing activity = 6.0
• traveling = -6.0
• waiting = -0.0

**Activity and facility boundary condition parameters**

All configurations use the activity parameters (per activity) in Table 15 unless otherwise noted.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Desired duration (hh:mm:ss)</th>
<th>Latest start time</th>
<th>Minimum duration (hh:mm:ss)</th>
<th>Opening time</th>
<th>Closing time</th>
</tr>
</thead>
<tbody>
<tr>
<td>home</td>
<td>12:00:00</td>
<td>-</td>
<td>08:00:00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>work</td>
<td>08:00:00</td>
<td>09:00:00</td>
<td>06:00:00</td>
<td>07:00:00</td>
<td>18:00:00</td>
</tr>
<tr>
<td>education</td>
<td>06:00:00</td>
<td>09:00:00</td>
<td>04:00:00</td>
<td>07:00:00</td>
<td>18:00:00</td>
</tr>
<tr>
<td>shop</td>
<td>02:00:00</td>
<td>-</td>
<td>01:00:00</td>
<td>08:00:00</td>
<td>20:00:00</td>
</tr>
<tr>
<td>leisure</td>
<td>02:00:00</td>
<td>-</td>
<td>01:00:00</td>
<td>06:00:00</td>
<td>24:00:00</td>
</tr>
</tbody>
</table>

The desired duration defines the "working point" of the utility function at which the marginal returns to longer duration are maximum and after which they decrease (chapter 4), and are determined to the nearest hour based on the actual census dataset (Meister et al., 2008). The latest start time defines the point after which being late is penalized. The minimum duration defines the point after which utility accrues (it is zero until this minimum duration is reached). Activities commencing before the facility opening time are penalized with a waiting disutility. Activities in progress cannot proceed beyond the facility closing time and will cease accruing utility at that point.

**Parameter values for the social interaction utility term**

The socializing parameter values for the utility functions in Section 5.6.4 were determined by a marginal utility analysis with respect to activity duration. Taking the second expression in Section 5.6.4 as an example, which values the number of friends present with decreasing marginal utility,

\[
U_{act} = \beta_{dur} t^* \left[ \frac{\ln(t)}{t^*} + \frac{10}{\beta_{dur} t^*} \right] + \beta_{soc} * \ln(N + 1)
\]
where $N$ is the number of friends present at the activity and $t$, $t^*$ and $\beta$ are as defined in chapter 4, yields the marginal rate of substitution (MRS) between meeting the next friend and spending the next hour at the activity with the current number of friends:

$$dU_{act} = 0 = \frac{dU_{dur}}{dt} dt + \frac{dU_{soc}}{dN} dN$$

$$-\frac{dN}{dt} = \frac{\beta_{dur} t^* (N + 1)}{t \beta_N} = \frac{6t^* (N + 1)}{t \beta_N}$$

to give the number of friends that can replace an hour of activity time, or, inversely:

$$-\frac{dt}{dN} = \frac{t \beta_N}{\beta_{dur} t^* (N + 1)} = \frac{t \beta_N}{6t^* (N + 1)}$$

for the hours of activity time an agent would exchange to have an additional friend at the activity.

Assuming that the agent, acting alone ($N=0$), has an optimized time allocation such that the activity duration = $t^*$, $t$ and $t^*$ cancel each other, sample values can be tabulated for the tradeoff the agent would make to meet with another friend.

Table 16 Possible socializing parameters for tests

<table>
<thead>
<tr>
<th>MRS = $-\frac{dt}{dN} = \beta_N/6$ (additional hr of activity / additional friends co-present)</th>
<th>$\beta_N$, $N=0$, $t=t^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.5</td>
<td>3</td>
</tr>
<tr>
<td>-1.0</td>
<td>6</td>
</tr>
<tr>
<td>-2.0</td>
<td>12</td>
</tr>
</tbody>
</table>

A standard "$\beta$" for socializing was chosen to be 10. This value is useful both for the case in which the number of friends is valued in utility, as well as when the number of hours of overlapping time with friends or the friend/foe ratio is valued, since these all have roughly the same range of values. Some experiments use the value 24 to increase the incentive to socialize rather than allocate longer time at the activity. Figure 14 illustrates the marginal rates of substitution (the willingness to trade off hours of leisure activity duration for one additional friend being present) at different points on the utility surface. An agent spending 2 hours alone pursuing a leisure activity would be willing to give up 1.67 hours of the activity if a friend were to co-participate.
Figure 14  Marginal rates of substitution for additional hours of activity duration versus an additional friend present at the activity

For the logarithmic function of face-to-face co-presence, $\beta_{soc}=10$, $t^*=2h$, $\beta_{dur}=6$.

The indifference curves (Figure 15) between the number of friends meeting face to face and the value of the time spent at the activity emphasize the duration/socializing tradeoff.
Figure 15  Isouitility surface of the hours of activity duration versus the number of friends present at the activity

Logarithmic function, $\beta_{soc} = 10$, $t^* = 2h$, $\beta_{dur} = 6$.

An agent participating in a leisure activity alone for 3 hours and 40 minutes can increase the utility of the activity from ca. 12 to ca. 18 if one additional friend were to be co-present. Alternatively, the agent can receive the same utility of ca. 12 if one new friend took part in the activity, but the activity lasted only 1 hour and 45 minutes.

Increasing the reward for socializing should encourage agents to make longer trips for socializing, as observed by Silvis et al. (2006) (Section 2.2.4).

6.4.5 Replanning

Five replanning strategies were tested: The first two are the standard Time (2 hour time mutation window) and Route algorithms (1, 2, 12, 102, 22, 32, 42, 51, 52), standard in the MATSim toolbox, to establish baseline measures for the subsequent experiments with interactive replanning algorithms.
Secondary location choice has been recently added to the MATSim toolbox by Horni, et al. (2008) to utilize the Potential Path Area for allocating locations, with a utility penalty for attending overfilled facilities (using maximum capacities available in the "facilities" dataset), leading to competition for slots at activities. While this is still an individual-based strategy and the location attributes are related entirely to accessibility measures, the capacity constraint and associated facility service penalty are compatible with any social networks replanning algorithm.

The three secondary location choice algorithms implemented here use neither of these mechanisms (time budget nor facility capacities) to redistribute demand in space, instead using only the standard or social reward-utility function and the information about locations available to agents to study knowledge propagation and how it influences location choice. The first implementation (see Section 6.4.3) uses the constrained location choice sets of the agents and randomly switches one randomly chosen "shop" and "leisure" location in each agent's plan with another location of the same type from the agent's own knowledge. If the location randomly chosen is the same as the current one, the agent simply does not change location. The choice set is limited by the social network topology, the locations passed on by the alters, and the memory management scheme: used for cases 3, 4, 5, 6, 104.

The second type of location choice also replaces a randomly chosen secondary location of each agent's plan with a new one. The algorithm allows the agent to randomly choose secondary locations from all facilities available in the scenario (the reference case 7). This is intended to provide a basis of reference for how the particular plan utility function would allocate facility locations if perfect information were available (barring facility capacity).

Finally, each agent can choose to change one shopping or leisure location to one chosen directly from the knowledge of friends, and immediately use the new locations in its plan, noting the location in its memory for possible further propagation (Section 6.4.3, experiments 4_4, 13_4, 23_4).

6.5 Scenario used for the experiments

A "scenario" in MATSim refers to the transportation network, the population of agents, their day plans, and the layers of geographic zones in the World. The scenario is introduced at this point because certain of the experiments in the following discussion, like geographic density, refer to its characteristics.

To cover the range of effects to be tested for, a scenario is needed with sufficient resolution, sample size, agent- and geographic heterogeneity. Any a-priori scenario will leave traces of its geographic, transportation, or social landscape on the final result. Therefore, an arbitrary
available real-world scenario was chosen for the verification experiments, and comparisons are drawn only across these cases, to control for the scenario's own effects. All model configurations have the following in common:

1. 1% sample of the population in the Zurich region (car mode only) = 8760 agents and 6438 geocoded facilities with opening times (Meister et al., 2008). Multiple activities are possible at each facility. The number of unique facilities at which the following activities are possible are: home: 4917; work: 1521; leisure: 114; shopping: 183; education: 71.

2. 5 aggregate activity types: home, work, education, shopping, leisure, with activity chains drawn from micro-census travel data (Swiss Federal Statistical Office, 2001). The population makes 3.38 trips per day.

3. 100km x 100km region surrounding Zurich (agents' initial plans were chosen such that the entire day plan takes place within this region).

4. Swiss National Road Network modified to 3% of capacity (Vrtic et al., 2003)

The agents have 6 identifying characteristics, based on distributions of characteristics from the census data: ID number, age, sex, percent of employment or education status, car ownership, and public transportation pass ownership. Their home and work (or education) locations (primary locations) correspond to municipally-aggregated distributions from the commuting matrix in the 2000 census (Swiss Federal Statistical Office, 2001). The set of primary locations for each agent are sampled from the matrix according to the sociodemographic characteristics of the agents, and allocated randomly within the municipality while upholding the constraint of the facility capacities. The activity chains used are sampled from the activity chains in the microcensus and are also allocated according to the sociodemographic characteristics of the agents. The leisure and shopping locations (secondary locations) are allocated randomly within a radius to home or workplace (Meister et al., 2008).

6.6 Hypothesis testing by comparing the verification experiments

Only the most important experiments testing the core behavioral elements were carried out. This section outlines the experimental design that was possible, and describes the experimental context of each model configuration listed in Table 12, Table 13, and Table 14. It concludes with verification experiments that were curtailed or not performed up to this point.
The experiments are grouped by algorithmic families to provide component-wise evidence for or against the hypotheses in Section 6.2.

6.6.1 Baseline sensitivity: random seed effects

Configuration 1 with no social networks was repeated five times with different random seeds to assess MATSim's variability with respect to random influences. The output of the other experiments can only be considered relevant if a signal is detected above the background noise inherent in the MATSim model system.

This small ensemble was run with lowered road capacity (1% instead of 3% used for the other experiments) and generated considerable traffic queueing as a result. The model relaxed more slowly to a less satisfactory convergence between the best and worst agent scores. The variability is thus taken to be higher than it would be had the models either run longer or been run at a higher capacity (with less queueing) as all the rest of the experiments presented here were. Therefore it is asserted for the experiments on the network with 3% capacity that the variability of the average quantities in response to random variation is less than that of this ensemble. Values in excess of these sensitivity bounds in each experiment can be associated with the mechanisms at work in the experiment.

6.6.2 Baseline social contact: mapping of face to face meetings in a model without a social network

The relaxed activity travel of the standard configuration 1 is used to make a map of all face to face encounters, which is then treated statistically as a social network. The statistics of this social network illustrate one extreme of geographic determinants on socializing (1_501_F2F). The exercise is repeated to provide a reference for face-to-face opportunities with optimized secondary locations in 7_501_F2F. The individually-optimized secondary location choices which did not respect facility capacities are hypothesized to have increased the number of co-present agents at highly accessible locations, and to change the statistics of face-to-face meetings relative to configuration 1.

Random_12 is an experiment with a randomly generated non-spatial classical social network whose statistics are used as a reference for these and all experiments using social networks. There is no travel behavior generated and no travel statistics for this experiment.
6.6.3  Time synchronization with a static social network and a utility reward for socializing

Experiments 2 and 12 pertain to time synchronization (with route choice) and use the static spatially-embedded social network described in Section 6.4.1. They each use a different utility reward for socializing. The experiments do not allow location change and therefore do not need information exchange; the agents can optimize face to face meetings in time only. (Social network evolution with time synchronization was not run but would be desirable to complete the test suite).

The experiments test time and route planning when encountering friends is valued as an activity attribute in the utility. Agents can schedule their travel more freely because socializing utility can compensate for higher travel cost or shorter activity durations. There may be evidence of agents coordinating their schedules.

Experiment 2: Utility reward for socializing of $10 \times \log(\text{Number of friends})$. This experiment allows less feedback of the social network into the duration of the activities: even one second of co-presence suffices to increase utility appreciably and yet permit the agent to continue adjusting the remainder of its plan unburdened by the opportunity costs of missing continued socializing.

Experiment 12: Utility reward for socializing of $10 \times \log(\text{Time overlap with friends})$. This experiment explicitly couples the duration of face-to-face time with friends with the duration of an activity. Leaving an activity at which one or many friends are present, in order to commence the next activity, carries with it opportunity costs of not having spent more time with friends. The experiments may have more pronounced influence on scheduling than experiment 2.

Experiment 22: As Experiment 12 with utility parameter = 24. The higher parameter should exhibit more easily recognizable evidence of schedule coordination.

6.6.4  Time synchronization with a socially dense static social network and utility reward for socializing

Experiment 102 is identical to experiment 2, but it doubles the number of agents in the social network. The spatial embedding parameters for the social network are unchanged. This results in roughly twice the number of friends per unit area, provided there are enough friends available: if all possible agents are linked to an ego's net within a given radius, the agent is of course forced to make the next friend further out, which would change the spatial embedding statistics slightly. The experiment is intended to increase the utility reward for co-presence at the activities that the agents have in common, and to make any resulting shifts in schedules
more obvious in the output, in case they could not be detected with the "standard" social network of average degree 12. Roughly doubling the number of friends at leisure activities roughly doubles the utility reward for leisure-time overlaps with friends of the same duration. The friends are an additional source of utility, like increasing a utility parameter or the road capacity; and they increase the agents' flexibility in adjusting their plans to locate a utility maximum. It is not clear exactly what will happen. The agents will either shorten their leisure activities and gain the same utility values for the activities, in order to take advantage of traffic conditions to spend longer durations at other activities; or they will spend the same amount of time at leisure activities but profit from their friends' co-presence, and increase utility; or they will spend longer times at leisure activities because the socializing reward is so high that it is worthwhile to sacrifice utility in shorter times spent at other activities or in time spent in traffic.

6.6.5 Baseline location choice: unconstrained secondary location choice of all available facilities

Experiment 7 is the reference case for location choice. Socializing plays no role and there is no social network. Agents receive standard utility but are free to choose any alternative location (direct from the database of "facilities") for their shopping and leisure. Agents are expected to redistribute their secondary activities to many distinct locations. They may accept longer trips to avoid congestion and seek closer locations to reduce travel, in relationship with the penalties in the utility. The choice of location should follow the accessibility accorded by the transportation network. Because they can choose from all available locations, the stable solution will represent the best possible location choices for secondary activities for independently acting agents with the standard utility function. (It would be interesting to have another experiment with a social network and a face-to-face socializing reward, but with free choice of locations for secondary activities).

6.6.6 Secondary location choice with exchange of location knowledge on a static social network and standard utility

The models 4 and 4_4 use the standard utility function and allow secondary locations to be changed with bounded knowledge in the social exchange and replanning algorithms described in Sections 6.4.3 and 6.4.5. The behavior represented is "asking friends for advice" about locations. The social network is the static initialization in 6.4.1.

Experiment 4 uses the standard spatially-embedded social network, the non-spatial interactor implementation to exchange location knowledge, the replanning module which chooses locations randomly from an agent's own knowledge, and a memory manager to erase less-popular locations, forgetting the advice that is heard less frequently. Agents can retain a
number of locations in their memory up to 1.5 x the number of plans in memory x the number of activities per plan.

Experiment 4_4 uses the replanning module which allows agents to choose secondary locations from their friends knowledge and to learn the location, themselves. The standard spatially-embedded static social network is used.

The completeness of knowledge percolation through the social network is expected to limit the learning about space, relative to unlimited location choice (experiment 7). The number of distinct secondary locations and other activity-travel characteristics is expected to be between that of experiment 1 and experiment 7, with more agents co-present at the activities due to social reinforcement of knowledge.

6.6.7 Secondary location choice with a spatially contracted static social network and standard utility

Experiment 104 uses the same configuration as experiment 4, but with a static social network that is spatially contracted, (see Figure 13 with $\alpha=2.5$, $r_{min}=500$). That is, the average distance to alters is shorter such that the number of social relationships in a local spatial neighborhood is higher and the number of social relationships to farther locations is lower.

The social network is expected to cause the knowledge to be spatially as well as socially localized and thus the choice of secondary locations is expected to be limited in both these senses.

6.6.8 Secondary location choice with exchange of location knowledge on a static social network and a utility reward for socializing

Experiments 3 and 23_4 allow secondary location choice with knowledge bounded by a static social network, plus positive utility for socializing. Thus two positive feedbacks (information exchange from experiment 4 (4_4) and socializing score from experiment 2) reinforce the socializing. This model reflects the endogeneity of location choice and the pool of potential social opportunities by implicitly building "socializing opportunity" into the utility as a location attribute. The secondary activities of an agent are expected to be spatially more concentrated than in experiment 4 (4_4): agents' leisure locations are expected to converge to a pattern according to the social network topology such that more friends are encountered. Willingness to travel or to shorten activity durations should also be observed, as these are compensated by the socializing score.
6.6.9 Secondary location choice with an evolving social network and no utility reward for socializing

The issues of treating phenomena subject to evolution in time, such as social network development, within a MATSim iterative loop have been detailed in Section 5.1.4. Experiments 5 and 6 attempt social network evolution in series with the iterative activity travel relaxation to observe the stability of the system and the emergent topology of the social network and its geographic embedding. Both models permit new friendships to be made face to face, location knowledge to be exchanged, and activity location choice to be optimized, as described above in this chapter. Friendships are dissolved each iteration, as above, to maintain a constant average degree. Experiment 6 has socializing utility and model 5 does not.

Experiment 5 permits information exchange of location and social network evolution, but no additional socializing score. The activity travel has the standard utility. The emergent social network is therefore a strong function of the location choice optimization, and are likely to be comparable with the results of the experiments in Sections 6.6.2 and 6.6.6, above.

6.6.10 Secondary location choice with an evolving social network and a utility reward for socializing

Experiment 6 permits information exchange of location, social evolution, and additional socializing score in the form of activity travel having utility for the total number of friends co-present. The results are influenced by the location choice optimization, the link removal algorithms, and the face-to-face rule for making new friends, as in model 5, but the incentive to return to a location at a time when friends were present is a difference which manifests itself as the tradeoffs illustrated in Figure 15. The results will be comparable with the results of the experiments in Sections 6.6.2 and 6.6.8, above.

6.6.11 Boundary conditions of the utility function and the facility opening times

Tests were performed in experiments 22, 32, 42, 52, and 72 of the "time synchronization" configuration of Section 6.6.2, in experiment 51 of the standard base configuration with no social network, and in experiment 13_4 of the location choice replanning algorithm in which location knowledge is taken directly from friends.

These experiments systematically eliminate boundary conditions and assumptions about the facility opening times and the standard form of the utility function, common to all the experiments, to discover ways to relax the MATSim system to allow more endogeneity to be directed by the social interactions.
Experiment 51: the standard MATSim configuration with no social interactions and no social network. The penalty for being late is 0 and the minimum duration of activities is also 0. Some users of MATSim specify the utility function this way. The first change to the scoring is an attempt to measure the relative influence of the "late penalty" to the ideal durations at activities, the argument being that all the positive utility subsumes opportunity costs of not spending time elsewhere, and additional late costs are not necessary for simulating time allocations. The second test was made because it has been remarked (Log Warn message in the code by Kai Nagel, 2008) that setting minimum durations > 0 is not advisable because, once an agent has shortened its activity duration such that it is below the minimum duration for the activity, it is in a window of nonzero durations during which it can either increase or decrease its duration and not get any additional utility. Since accruing no utility for spending time at the activity means that it can only lose utility elsewhere in its plan by not immediately leaving the activity for the next one, the agent will always spend zero time at the activity, making the trip for no benefit. If, however, the minimum activity duration is zero, the agent spending zero time at the activity can stand to gain utility by increasing the duration of its stay.

Experiment 32 repeats experiment 22 with the minimum activity durations set to zero in order to test the hypothesis that the agents could be wasting trips to activities of zero-length if the minimum duration is >0.

Experiment 52 repeats experiment 32 with no penalty for arriving late, and undefined facility opening and closing times.

Experiment 42 is similar to experiment 52 except that late penalties are assessed and that the desired start times for the activities are not defined.

Experiment 72 repeats experiment 32 but the facilities are open 24 hours (this experiment supercedes 52 because undefined times, as in experiment 52, may have caused strange results)

The two experiments 42 and 52 without defined facility opening and closing times or latest start/end times were an attempt to simulate the agents endogenously scheduling their activities by comparing schedule outcomes with one another (essentially) and adjusting their plans accordingly. However there is no substitute in the MATSim scoring function to the exogenous (user-entered) desired activity start and end time, -duration, and the facility opening and closing times. Code has not been written to re-set the desired start/end times of the utility function of each agent to correspond to social norms. This would be required in order to relax the system endogenously without exogenously set opening times and activity start/end times.

Experiment 23_4 is the experiment analogous to experiment 3, but with a different implementation of the replanning module.
Experiment 23_4 is complemented by experiment 13_4, which uses a scoring function based on the total time face-to-face with friends and the high parameter (=24), initiated to help detect signals of social interaction in the output analogous to experiment 22.

6.6.12 Verification experiments not carried out

The initial scope of model verification was clearly too ambitious at the scale of scenarios necessary for 1) statistical stability and statistical analysis of social networks; 2) spatial variation in activity plans with simultaneously large activity attendance to allow for social interactions and to illustrate location choice behavior. This does not mean that too little was learned however: some verification tests were not possible because it was quickly evident that the model was poorly conceived. In these cases, insight was gained in failing to be able to carry out the test.

Among the curtailed plans for verification experiments were:

- The ensemble analyses was cut to five experiments rather than 10 and constrained to the standard case with independent agents rather than applied to all social interaction models. This was an efficient use of computing resources and sufficient to illustrate the very small variability of the model system.

- With link removal algorithms using fixed ("remove 10,000 links per iteration) or fixed proportional ("remove 5% of links per iteration) rates of link removal, no range of parameters could be found that led to an equilibrium social network, i.e. low variation in the average degree and clustering coefficient of the graph. Only a link removal algorithm that was tuned to the rate of adding new social relationships could maintain equilibrium without eventually removing all the social links, or permitting the social network to grow beyond the memory limits of the computer. In most social interactions proposed here, a positive feedback develops between the agent degree and the rate of adding new social ties. For instance, information exchange causing more agents to visit the same location, utility reward enticing agents to remain at a location, or friend-of-friend introductions. Iteration for iteration, these add a variable but increasing number of new edges to the social network. A link removal algorithm with a fixed functional form, for instance removing either a fixed number of edges or a fixed fraction of edges, cannot respond to the complex dynamics of the social network and cannot be tuned to a stable equilibrium. Only a link removal algorithm which incorporates information about the state of the social network can succeed in establishing an equilibrium. The link removal algorithm that constrains the social network to a mean degree was useful. Others may be, as well, for instance one which enforces a target clustering coefficient.
The scaling experiments to a 10x bigger population (87,600) ran for 8 days each and produced 18.5GB of output per experiment, which has been prohibitively large to analyze in the postprocessing software (R) written for this project. The data has not been analyzed yet. This branch of the research was therefore postponed.

Scaling to smaller populations was not attempted. Scenarios that are too small risk wild fluctuations in social network statistics, too few opportunities for socializing, and too much spatial homogeneity to observe systematic deterministic or complex tradeoffs between socializing and location choice. The average household size, size of groups at activities, and the incidence of encounters was already very low in the 1% random sample of 8670 individuals, and it was anticipated that further dilution of population and facilities with seldom social encounters would render the models very unstable. Sparse social ties would become such a dominant influence; changes to a single social tie could completely alter the character of the social graph, rendering the drawing of conclusions from interaction modelling fruitless. Scaling tests will have to be performed using another scenario that samples households and other social groups rather than individuals, so that there is at least a core of consistent social structure.

The socializing utility terms for "Friend/Foe ratio" was implemented but not verified in the 1% scenario. The function of this variable is extremely nonlinear in small groups and may be destabilizing rather than useful.

The intended parameter sweeps of all algorithms were not performed, with the exception of comparing two values for the socializing utility, two social network densities, and two spatial models of social networks.

Social interactions based on relationship type, such as "family/household" and "work colleague" were never implemented.

Social exchanges consisting of closing transitive triads: "introducing friends of friends" were implemented but not verified. Other explicit feedbacks of social network topology (popularity, etc.) were not implemented.

A sociodemographic dependence for making social connections, explicitly simulating "homophily", or the reflection problem (Manski, 1993), i.e. the tendency for people with similar characteristics to share a correlated interest, was never implemented. The MATSim variables "sex", "age", and "employment status" might however be useful in such a model that establishes a priori homophilic associations, see e.g. Mossong, et al. (2008).
7 Results of the experiments

The analysis examines the links between social interaction mechanisms and the observed travel patterns. The simulation was run once for each of the described models. Each experiment iterates 500 times, requires between 10 and 20 hours to run, and produces approximately 3 GB of output. User equilibrium conditions are assumed by confirming that the executed score (i.e. utility) remained constant within a 1% variation over the last 100 iterations.

This chapter is organized as follows:

- Results illustrating the baseline sensitivity of MATSim
- Summary of social network topology and geography
- Summary of travel behavior results
- Summary of socializing travel behavior results
- Synthesis of results experiment-by-experiment

The results are summarized and discussed in chapter 8.

7.1 Establishing the base sensitivity of the MATSim system

This section summarizes the model results from the point of view of social networks, i.e. the representation of long-term social ties that lie at the basis of the integrated social-traffic flow simulation. There is no clean way to isolate these from travel behavior and face-to-face socializing, so reference will be made to results that appear later in the chapter, and, respectively from later sections to this section.

Table 17 presents the average spatiotemporal characteristics of the plans resulting from the directed system relaxation of the first 7 scenarios in Table 12, with a 1% road network capacity scaled 1:1 with the population sample. The first row of the table shows the values for the ensemble experiments of experiment 1, a MATSim model without social interactions. The standard deviation of the average results establish the basic variation of the model. Figure 16 shows the time profiles of traffic for the ensemble model configurations, averaged in 5 minute bins.
Table 17: Travel behavior summary of seven social network travel models (standard deviation of ensemble in parentheses)

<table>
<thead>
<tr>
<th>Ensemble Experiment Name</th>
<th>Avg. trip distance (km)</th>
<th>Activity radius (km)</th>
<th>Avg. score (utils)</th>
<th>Avg. social score</th>
<th>Avg # friends per leisure</th>
<th>Avg # non-friends per leisure</th>
<th>Avg trip duration (min)</th>
<th>Trip speed (km/h)</th>
<th>Avg distance to friends (km)</th>
<th>Dyad distance (km)</th>
<th># graph components</th>
<th>Clust ratio</th>
<th>Avg. link age (iter)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>9.47 (0.0)</td>
<td>7.18</td>
<td>165 (0.9)</td>
<td>0.0</td>
<td>0.00</td>
<td>10.2 (0.1)</td>
<td>18.1 (0.4)</td>
<td>31.5 (0.7)</td>
<td>-</td>
<td>-</td>
<td>8760</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.2</td>
<td>9.48</td>
<td>7.18</td>
<td>172 (0.9)</td>
<td>0.9</td>
<td>0.14</td>
<td>11.9</td>
<td>17.9</td>
<td>31.9</td>
<td>17.6</td>
<td>15.4</td>
<td>11</td>
<td>3.21</td>
<td>500</td>
</tr>
<tr>
<td>0.3</td>
<td>12.49</td>
<td>9.04</td>
<td>172 (1.3)</td>
<td>1.3</td>
<td>0.20</td>
<td>53.0</td>
<td>24.7</td>
<td>30.4</td>
<td>17.6</td>
<td>15.4</td>
<td>11</td>
<td>3.21</td>
<td>500</td>
</tr>
<tr>
<td>0.4</td>
<td>10.54</td>
<td>7.85</td>
<td>163 (0.0)</td>
<td>0.0</td>
<td>0.11</td>
<td>13.6</td>
<td>18.2</td>
<td>34.9</td>
<td>17.6</td>
<td>15.4</td>
<td>11</td>
<td>3.21</td>
<td>500</td>
</tr>
<tr>
<td>0.5</td>
<td>10.10</td>
<td>7.74</td>
<td>163 (0.0)</td>
<td>0.0</td>
<td>2.93</td>
<td>13.3</td>
<td>18.6</td>
<td>32.5</td>
<td>10.1</td>
<td>11.3</td>
<td>202</td>
<td>135</td>
<td>249</td>
</tr>
<tr>
<td>0.6</td>
<td>11.83</td>
<td>8.77</td>
<td>189 (12.8)</td>
<td>3.25</td>
<td>42.3</td>
<td>22.5</td>
<td>31.5</td>
<td>12.1</td>
<td>13.9</td>
<td>212</td>
<td>114</td>
<td>114</td>
<td>280</td>
</tr>
<tr>
<td>0.7</td>
<td>11.25</td>
<td>8.30</td>
<td>166 (0.0)</td>
<td>0.0</td>
<td>0.00</td>
<td>9.1</td>
<td>17.5</td>
<td>38.6</td>
<td>-</td>
<td>-</td>
<td>8760</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

1 Average distance from agent home location to all visited locations in the plan.
2 Average distance from agent home location to home locations of all friends (radius of the ego network).
3 Average distance between each pair of friends.
4 Ratio of clustering coefficient to that of a classical random network of the same degree.
5 Number of other agents an ego encounters at a leisure activity who are (are not) in the ego net
As noted in Section 6.6.1, these experiments did not relax within 500 iterations to the extent that the experiments with a 3% road capacity did, and the difference in the final state between the experiments could have to do with incomplete relaxation. Nevertheless, there is very little variation in the average statistics of the system across the experiments. Referring to Table 12, the trip distance (routing replanning module) and final score vary by approximately 0.5%. The trip duration (time allocation mutation replanning module) varies just 2% across models, also resulting in a 2% variation in trip speed.

The standard deviation of total travel volume in Figure 16 is an average of 15 agents per bin (depending on the bin, this ranges from 1-27%) and is a maximum of 30 agents (out of 8760), which occurs over mid-day (ca. 12%).

These baseline results will be used in two ways: First, values in excess of these sensitivity bounds in each experiment can be associated with the mechanisms at work in the experiment. Second, they provide a reference for estimating the effect of changing the road network...
capacity parameter on social network coupling, since the remaining experiments use a capacity three times higher. All following sections will occasionally refer to these results.

The experiment group 11 (introduced in Section 6.6.11) questions why this result is so stable by perturbing the model's boundary conditions and utility assumptions. The insights of these experiments are summarized in Section 8.1.11.

7.2 Descriptive summary of the geographically embedded social networks

This section summarizes the statistics of the social network topology and geography of all the experiments. The results trace the effects of spatial and social tie density assumptions, spatial embedding versus spatial independence, and of the coupled evolution of social networks and travel plans.

7.2.1 Social networks summary

Table 18 summarizes basic nonspatial statistics of the social network of each experiment. The statistics are defined as follows:

- Degree: the average number of social edges (friends) per agent
- Clustering coefficient: the proportion of closed triads, averaged over all agents: \(C = \frac{3 \times \text{number of triangles on the graph}}{\text{number of connected triples of vertices}}\) (Strogatz and Watts, 1998):
- Clustering ratio: the ratio of the clustering coefficient to that expected in a randomly generated reference graph.
- Diameter: the longest of all the shortest paths between all 2 pairs of agents (except infinity)
- Average path length: the average length of the shortest geodesic paths through the social network between all agents.
- Components: the number of isolated social groups with no social connection to any other social group
- Main component: the size of the largest contiguous group of agents
- Link age: average over all dyads of the last iteration since the pair of agents who are connected by a social link met face to face.
- Number of encounters: average over all dyads of the number of times these two agents have met (as friends) throughout the iterations. Thus this statistic is reset to 0 if a friendship is dissolved and re-made. In the experiments in which the agents did not socially interact, no statistic could be computed ("-").
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Deg</th>
<th>Clust. coeff.</th>
<th>Clust. ratio</th>
<th>Dia.</th>
<th>Avg. path length</th>
<th>Main component</th>
<th># components</th>
<th>Link age</th>
<th># encounters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>0.0000</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>0</td>
<td>8760</td>
<td>NaN</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>12.0</td>
<td>0.0044</td>
<td>3.2</td>
<td>10</td>
<td>4.0</td>
<td>8752</td>
<td>9</td>
<td>500</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>12.0</td>
<td>0.0044</td>
<td>3.2</td>
<td>10</td>
<td>4.0</td>
<td>8752</td>
<td>9</td>
<td>500</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>12.0</td>
<td>0.0044</td>
<td>3.2</td>
<td>10</td>
<td>4.0</td>
<td>8752</td>
<td>9</td>
<td>500</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>12.0</td>
<td>0.1860</td>
<td>135.8</td>
<td>12</td>
<td>4.8</td>
<td>8572</td>
<td>173</td>
<td>3.6</td>
<td>1.98</td>
</tr>
<tr>
<td>6</td>
<td>12.0</td>
<td>0.1357</td>
<td>99.1</td>
<td>12</td>
<td>4.4</td>
<td>8535</td>
<td>204</td>
<td>3.5</td>
<td>1.65</td>
</tr>
<tr>
<td>7</td>
<td>0.0</td>
<td>0.0000</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>0</td>
<td>8760</td>
<td>NaN</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>12.0</td>
<td>0.0044</td>
<td>3.2</td>
<td>10</td>
<td>4.0</td>
<td>8752</td>
<td>9</td>
<td>500</td>
<td>-</td>
</tr>
<tr>
<td>102</td>
<td>24.0</td>
<td>0.0088</td>
<td>6.4</td>
<td>6</td>
<td>3.3</td>
<td>8760</td>
<td>1</td>
<td>500</td>
<td>-</td>
</tr>
<tr>
<td>103</td>
<td>12.0</td>
<td>0.1862</td>
<td>135.9</td>
<td>17</td>
<td>5.4</td>
<td>634</td>
<td>20</td>
<td>500</td>
<td>1.39</td>
</tr>
<tr>
<td>1_501F2F</td>
<td>27.2</td>
<td>0.5376</td>
<td>173.2</td>
<td>13</td>
<td>4.9</td>
<td>8627</td>
<td>120</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>7_501F2F</td>
<td>26.4</td>
<td>0.5200</td>
<td>172.6</td>
<td>10</td>
<td>4.3</td>
<td>8633</td>
<td>115</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>12.0</td>
<td>0.0044</td>
<td>3.2</td>
<td>10</td>
<td>4.0</td>
<td>8752</td>
<td>9</td>
<td>500</td>
<td>-</td>
</tr>
<tr>
<td>32</td>
<td>12.0</td>
<td>0.0044</td>
<td>3.2</td>
<td>10</td>
<td>4.0</td>
<td>8752</td>
<td>9</td>
<td>500</td>
<td>-</td>
</tr>
<tr>
<td>42</td>
<td>12.0</td>
<td>0.0044</td>
<td>3.2</td>
<td>10</td>
<td>4.0</td>
<td>8752</td>
<td>9</td>
<td>500</td>
<td>-</td>
</tr>
<tr>
<td>51</td>
<td>0.0</td>
<td>0.0000</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>0</td>
<td>8760</td>
<td>NaN</td>
<td>0</td>
</tr>
<tr>
<td>52</td>
<td>12.0</td>
<td>0.0044</td>
<td>3.2</td>
<td>10</td>
<td>4.0</td>
<td>8752</td>
<td>9</td>
<td>500</td>
<td>-</td>
</tr>
<tr>
<td>72</td>
<td>12.0</td>
<td>0.0044</td>
<td>3.2</td>
<td>10</td>
<td>4.0</td>
<td>8752</td>
<td>9</td>
<td>500</td>
<td>-</td>
</tr>
<tr>
<td>4_4</td>
<td>12.0</td>
<td>0.0044</td>
<td>3.2</td>
<td>10</td>
<td>4.0</td>
<td>8752</td>
<td>9</td>
<td>500</td>
<td>-</td>
</tr>
<tr>
<td>13_4</td>
<td>12.0</td>
<td>0.0044</td>
<td>3.2</td>
<td>10</td>
<td>4.0</td>
<td>8752</td>
<td>9</td>
<td>500</td>
<td>-</td>
</tr>
<tr>
<td>23_4</td>
<td>12.0</td>
<td>0.0044</td>
<td>3.2</td>
<td>10</td>
<td>4.0</td>
<td>8752</td>
<td>9</td>
<td>500</td>
<td>-</td>
</tr>
<tr>
<td>Random12</td>
<td>12.0</td>
<td>0.0013</td>
<td>1.0</td>
<td>6</td>
<td>3.9</td>
<td>8760</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- = not calculated because the statistics package is called during agent interactions
NaN= division by zero or solution does not exist

Table 19 summarizes the geographical embedding of the social networks and some measures of the geographic spread of activity locations and the distance travelled:

- Distance to all alters: the sum of the planar (Euclidean) distance from each ego to each of its alters, divided by the number of alters in the ego net.
• Distance to all acts: the sum of the planar (Euclidean) distance from each agent to each of its activities, divided by the number of activities in the agent's plan.

• Chain length: the sum of the planar (Euclidean) distance between the activities of each agent's plan, in the sequence they are carried out; thus an indicator of the distance travelled per agent to carry out its plan.

• Dyad separation: the average distance between each ego-alter pair.

• Various ratios of the distances for comparison of the distance of social links versus activity locations
### Table 19
Summary of the geographical embedding of the social networks; values averaged over the population

<table>
<thead>
<tr>
<th>Experiment ID</th>
<th>Dist. to all alters (km)</th>
<th>Dist. to all acts (km)</th>
<th>Chain length (km)</th>
<th>Dyad separation (km)</th>
<th>Dist to acts/dist. to alters</th>
<th>Dyad dist/dist to alters</th>
<th>Chain length/dist. to all acts</th>
<th>Chain length/dist. to alters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NaN</td>
<td>7.2</td>
<td>21.6</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>3.0</td>
<td>NaN</td>
</tr>
<tr>
<td>2</td>
<td>17.6</td>
<td>7.2</td>
<td>21.6</td>
<td>15.4</td>
<td>0.4</td>
<td>0.9</td>
<td>3.0</td>
<td>1.2</td>
</tr>
<tr>
<td>3</td>
<td>17.6</td>
<td>10.1</td>
<td>33.8</td>
<td>15.4</td>
<td>0.6</td>
<td>0.9</td>
<td>3.3</td>
<td>1.9</td>
</tr>
<tr>
<td>4</td>
<td>17.6</td>
<td>7.6</td>
<td>22.9</td>
<td>15.4</td>
<td>0.4</td>
<td>0.9</td>
<td>3.0</td>
<td>1.3</td>
</tr>
<tr>
<td>5</td>
<td>9.8</td>
<td>7.5</td>
<td>21.8</td>
<td>10.9</td>
<td>0.8</td>
<td>1.1</td>
<td>2.9</td>
<td>2.2</td>
</tr>
<tr>
<td>6</td>
<td>13.5</td>
<td>9.5</td>
<td>30.9</td>
<td>15.8</td>
<td>0.7</td>
<td>1.2</td>
<td>3.2</td>
<td>2.3</td>
</tr>
<tr>
<td>7</td>
<td>NaN</td>
<td>7.8</td>
<td>23.6</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>3.0</td>
<td>NaN</td>
</tr>
<tr>
<td>12</td>
<td>17.6</td>
<td>7.2</td>
<td>21.6</td>
<td>15.4</td>
<td>0.4</td>
<td>0.9</td>
<td>3.0</td>
<td>1.2</td>
</tr>
<tr>
<td>102</td>
<td>17.7</td>
<td>7.2</td>
<td>21.6</td>
<td>15.5</td>
<td>0.4</td>
<td>0.9</td>
<td>3.0</td>
<td>1.2</td>
</tr>
<tr>
<td>104</td>
<td>3.5</td>
<td>7.1</td>
<td>21.1</td>
<td>3.0</td>
<td>2.0</td>
<td>0.9</td>
<td>3.0</td>
<td>6.0</td>
</tr>
<tr>
<td>1_501F2F</td>
<td>9.0</td>
<td>7.2</td>
<td>21.6</td>
<td>10.0</td>
<td>0.8</td>
<td>1.1</td>
<td>3.0</td>
<td>2.4</td>
</tr>
<tr>
<td>7_501F2F</td>
<td>10.2</td>
<td>7.8</td>
<td>23.6</td>
<td>11.1</td>
<td>0.8</td>
<td>1.1</td>
<td>3.0</td>
<td>2.3</td>
</tr>
<tr>
<td>22</td>
<td>17.6</td>
<td>7.2</td>
<td>21.6</td>
<td>15.4</td>
<td>0.4</td>
<td>0.9</td>
<td>3.0</td>
<td>1.2</td>
</tr>
<tr>
<td>32</td>
<td>17.6</td>
<td>7.2</td>
<td>21.6</td>
<td>15.4</td>
<td>0.4</td>
<td>0.9</td>
<td>3.0</td>
<td>1.2</td>
</tr>
<tr>
<td>42</td>
<td>17.6</td>
<td>7.2</td>
<td>21.6</td>
<td>15.4</td>
<td>0.4</td>
<td>0.9</td>
<td>3.0</td>
<td>1.2</td>
</tr>
<tr>
<td>51</td>
<td>NaN</td>
<td>7.2</td>
<td>21.6</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>3.0</td>
<td>NaN</td>
</tr>
<tr>
<td>52</td>
<td>17.6</td>
<td>7.2</td>
<td>21.6</td>
<td>15.4</td>
<td>0.4</td>
<td>0.9</td>
<td>3.0</td>
<td>1.2</td>
</tr>
<tr>
<td>72</td>
<td>17.6</td>
<td>7.2</td>
<td>21.6</td>
<td>15.4</td>
<td>0.4</td>
<td>0.9</td>
<td>3.0</td>
<td>1.2</td>
</tr>
<tr>
<td>4_4</td>
<td>17.6</td>
<td>7.4</td>
<td>22.2</td>
<td>15.4</td>
<td>0.4</td>
<td>0.9</td>
<td>3.0</td>
<td>1.3</td>
</tr>
<tr>
<td>13_4</td>
<td>17.6</td>
<td>10.1</td>
<td>35.7</td>
<td>15.4</td>
<td>0.6</td>
<td>0.9</td>
<td>3.5</td>
<td>2.0</td>
</tr>
<tr>
<td>23_4</td>
<td>17.6</td>
<td>9.2</td>
<td>30.7</td>
<td>15.4</td>
<td>0.5</td>
<td>0.9</td>
<td>3.3</td>
<td>1.7</td>
</tr>
<tr>
<td>Random12</td>
<td>36.1</td>
<td>7.2</td>
<td>21.6</td>
<td>36.1</td>
<td>0.2</td>
<td>1.0</td>
<td>3.0</td>
<td>0.6</td>
</tr>
</tbody>
</table>

NaN cannot be calculated.

The values are significant to 3 digits after the decimal due to the large population but this significance does not shed light on the conclusions so the values are rounded, here.

A single non-spatial random social network of average degree 12 was generated as a reference ("Random12"). It has a single large component, as expected, which contains all members of the population, and a longest geodesic path length between any two agents of 6. Its degree distribution is Poisson (Figure 17). The expected average distance between egos' and alters'
home locations ("dyad distance") of randomly drawn pairs of homogenously distributed agents in a 100km x 100km square would be 31km (Dunbar, 1997), and is 36.1km in the heterogeneously distributed population used here. The average ego net radius is 36.1km, as well.

The spatially-embedded social network with parameters $\alpha=1.5$, $r_{\text{min}}=1000\text{m}$ (Section 6.4.1, used in Experiments 2, 3, 4, 12, 22, 32, 42, 52, 72, 4_4, 13_4, 23_4) is slightly more fractured and clustered. The ratio of its clustering coefficient to that of the non-spatial network is 3.2, meaning that it is 3.2 times more likely in this society that friends of friends are mutually friends (closed triads), versus if the friendships were made randomly without consideration of spatial proximity. The graph has 9 components, which is misleading, because the main component is very large and the other 8 components are all single agents with no social connections (the population is all essentially in one component). The diameter of the spatial social network is 10 with an average separation of 4.0, meaning that, on average, the connectivity resembles that of the random graph, but that some agents are socially connected via one-dimensional strings of agents, resembling tails in the social graph. This will slow the percolation of information to and from these agents, who are socially more distant from the core, which is slightly more tightly-knit (higher clustering) than in the non-spatial graph. The network has an average dyad distance of 15.4km, compared to 36.1km in the nonspatial reference case (ego net radius = 17.6km). Thus the social network used in the experiments has higher clustering and more fractured components than a random graph, and is a spatially embedded small world with highly clustered local (spatial, possible social) neighborhoods and longer (geographic) connections to more (geographically, possibly socially) distant neighborhoods (Wong et al., 2006).

Doubling the social density to an average degree of 24 in experiment 102 doubles the clustering coefficient relative to a graph of degree 12. The longest shortest path between any two agents is the same as in the random graph (6), but the average path length is much shorter at 3.3. There is once again a single main component. This suggests that information can flow across this social network in a manner similar to the non-spatial random network of degree 12, while the high clustering like a spatial network indicates mutual friendships where behavior may be locally reinforced.

Spatially contracting the social network in experiment 104 ($\alpha=2.5$, $r_{\text{min}}=500\text{m}$) dramatically increases the clustering and the number of components of the graph. The largest component is only 634 agents, indicating a fracturing of the population into similarly-sized isolated neighborhoods at the outset of the scenario (the remaining ca. 93% of the population is distributed into 19 components with no component having more than roughly 7% of the agents in it). The diameter and average path length refer to the value within the components, which are small relative to the population, and are both much longer than in any of the other
three social networks shown, and indicate a string-like social structure within the components. One imagines many peripheral members, but a highly clustered core, and potentially slow percolation of information to the social outskirts. Of course there is no social information exchange between components of the graph.

The evolving social networks (Experiments 5 and 6) reach equilibrium after ~80 iterations. A constant graph average degree and clustering coefficient are taken to indicate that the evolving graph is in equilibrium (personal communication with Kosinetts, 2007). The initial social network has also completely disappeared by this point. High clustering, short average distance to friends, and a very high number of components relative to the original spatial random graph are consistent social network characteristics across the two models (Table 18), and indicate highly spatially clustered structures. These structures resemble spatial neighborhoods, but are based on the simplistic face-to-face meeting criterion and lack the richer fabric of "neighborhoods" that incorporates demographics, heritage, life goals, etc. The degree distribution is similar in the two experiments (Figure 17) and favors low degrees. This degree distribution emerges from three influences: the social pool (a combination of the number of activities per agent, the number of other agents present at each activity, and the accessibility of the activity location), the rule for making friends, and the link removal algorithm. The rule that was used for making friends face to face means that the social network strongly reflects spatiotemporal movements of the agents. The link removal algorithm tends to isolate those agents which do not have many activities or which live in sparsely populated regions (the high number of components, yet a large main component). This effect of the mechanism was also observed by Jin et al. (2001), who noted first, that increased clustering and a shift of the degree distribution to smaller degrees was caused by iterative removal of links on the basis of node degree; and second, that edge removal algorithms which choose edges randomly will tend to remove edges from high-degree nodes and "trim" the right-hand tail of the degree distribution. The many components are mostly single agents because the main component is large in both experiments. These agents were likely cut off from the rest by this mechanism.

Like the spatial contraction in Experiment 104 before, these graphs exhibit very long diameter (average path length), double (25% longer than) that of a nonspatial reference social network. This indicator, together with the high clustering coefficient, is evidence of long chains of relationships radiating from a dense core of relationships. Information will probably be slow to percolate from and to these peripheral regions.

7.2.2 Social network degree distributions

The degree distributions of the social networks are plotted in Figure 17. The spatially embedded initial networks are by construction Poisson, modified by a dependence on the
spatial distribution of agents (Section 6.4.1). All of the spatial social networks exhibit positive-skewed degree distributions, including the social networks which evolve together with the activity-travel plans. Only the degree distribution of the non-spatial network is symmetric (Figure 17). The right-skew increases if the social network is denser or if the social network is spatially contracted (compare Figure 17 with Table 19). This fat tail was not fitted and so its exponent is not known. But the distributions compare qualitatively favorably to exponential distributions published in the literature (Section 2.2.3).

Figure 17 Final social network degree distributions by experiment

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Social Network Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Static degree 12 spatial social network (initialization)</td>
</tr>
<tr>
<td>24</td>
<td>Experiment 102: Static degree 24 spatial social network (initialization)</td>
</tr>
<tr>
<td>5</td>
<td>Experiment 5: Social network which evolved with activity plan optimization</td>
</tr>
<tr>
<td>6</td>
<td>Experiment 6: Social network which evolved with activity plan optimization</td>
</tr>
</tbody>
</table>
Three points are important to make, here. First, the random graph generation algorithm produces the expected Poisson-distributed degrees, low clustering, and short geodesic distances (ca. log(N)) when distance is not an additional factor in the probability of making a social link, so the graph construction algorithm is thus far verified (Figure 17). Second, the density of agents (number of agents in their home locations per unit area) determines the positive-skewness of the degree distribution when distance is used as a weight in establishing the social links. This property emerges in the initial static spatial social network, both in the degree-12 case and in the spatially contracted graph of experiment 104 (Figure 17). Third, the
spatial locations of the agents' activities combined with the rule for making up to one new friend per activity determines the skewness as well as in the evolving graphs of experiments 5 and 6 (Figure 17), where spatial distance between alters is governed by where they meet face-to-face, i.e. not necessarily their home locations, but their activity locations.

Thus, spatial embedding of a social network, whether based on home locations or on face-to-face meetings away from home, forces a positive skew onto the degree distribution. How much this depends on the geography versus the activities and/or the population distribution has not been investigated formally and should be approached analytically.

7.2.3 The spatial extent of ego nets

The radius of an ego net is defined here as the average distance between home locations of an ego and all of its alters. This distance depends on the rules for constructing the spatially embedded social network and/or for making and breaking social ties. The distributions associated with the average values in Table 19 are plotted in Figure 18.

The distributions for the six social networks plotted are all positive-skewed. The distributions in Experiments 2, 102 (average degree is doubled), and 5 are qualitatively similar. Experiment 6 has many more short-distance friendships (a wider peak) and experiment 104 (spatially contracted) has a very high, narrow peak at a short radius from the ego's home. The distance to alters in a non-spatial (classical random) social network is determined by the population density and the physical dimension as well as geometry of the landscape, which leaves a signature on the distribution (Section 6.2.3).
Figure 18  Distributions of the average distance between the home location of an ego and that of all of its alters for six social network experiments

- Static degree 12 spatial social network (initialization)
- Experiment 102: Static degree 24 spatial social network (initialization)
- Experiment 6: Social network which evolved with activity plan optimization
- Experiment 5: Social network which evolved with activity plan optimization
The distance distributions lack the high number of spatially proximate relationships and decay less slowly than real observations and theory ($r^{-2}$) hypothesize (Section 2.2.2). Experiment 104 exhibits a more severe tie probability decay with distance that is more consistent with expectation. The inverse distance function for the right tail of the distribution was not calculated however. The parameters in the initialization function for the marginal probability of a social tie with respect to distance have a good basis in literature and theory. Though it was not specifically tested, the particular distance distributions resulting here are likely influenced by the sparse population density in the geographic scenario (1% of real population), thus stretching the distance distributions of social ties over what would emerge in a scenario with denser populations.

### 7.2.4 Spatial density of social networks: dependence of ego net radius on agent degree

The ego net radius has a relationship to the agent's degree that depends either on the assumptions used to spatially embed the static social network, or on the rules for making and losing social ties. The latter of course depend on the travel behavior and all of its geographical complexity in the coupled models. Figure 19 illustrates these relationships for six models.

Three types of distributions emerge:

- Decreasing radius with increasing degree: Figure 19 (Experiments 2 [same as 3, 4, 12, 22, 32, 42, 52, 72, 13_4, and 23_4], and 102)
In reality, it would be expected that the distance between alters and the degree be some kind of equilibrium between the effort needed to maintain relationships over a distance, the benefit from doing so, and the budget available to meet new friends and travel enough to visit them (Section 2.2.2). Whether the slope of a plot of the number of alters versus the distance to the alters is positive, negative, or zero, depends on how the alters are defined (what kind of relationship) and how specialized (or far) a journey a person must make in order to maintain this particular relationship. In the absence of real time evolution in MATSim, the rules for making and dissolving friendships must substitute for a complete model of this equilibrium. The figure illustrates that the choice of social network construction algorithm can obtain a range of distance distributions with respect to degree. Though the correct (valid) model is not identified in this method, the entire range of positive and negative slopes can be generated in this modelling context, showing that some control of the experiment is possible.
Figure 19  Boxplots of the average distance between the home location of an ego and that of all of its alters versus the agent's degree

Static degree 12 spatial social network (initialization)

Static degree 24 spatial social network (initialization), experiment 102

Social network which evolved with activity plan optimization, experiment 6

Social network which evolved with activity plan optimization, experiment 5
7.2.5 Population density effects on social network

The MATSim structure enables the definition of "zones", much like a Geographical Information System (GIS). The zones are square and the other elements of the MATSim World ("layers") are mapped to the zones: road network and facilities. The definition of zones enables calculations of population densities and spatial aggregations of the social networks, which can be used to investigate the effect of the geographical scenario on the social network. A grid of zones of 3km x 3km dimension is overlaid in the scenario to create a matrix of ca. 1000 zones. The population density is illustrated in Figure 20.
The social network is aggregated in space by establishing a new network with a node in the center of each grid point. For each pair of zones, the number of social connections between agents residing within each zone is summed and attributed to the link between the nodes of the two zones. A network of 1000 nodes is large enough to enable studies of topological statistics, while being small enough to shown graphically and to be more likely to capture at least one link between zones (reducing the dimension of the zones increases the number of links between zones by capturing the very short connections, but increases the number of zones to the point that they can not be practically analyzed graphically). The degree of the
zone nodes is an indicator of both the popularity of agents in a zone, as well as the number of agents living in the zone (population density). The plots in Figure 21 through Figure 26 show these aggregated social networks embedded in space, with the node diameter proportional to the degree and the edge color proportional to the number of connections between agents in the zones (darker = more). The top graphic is the raw count, and the bottom graphic divides the degree by local population density to filter out this effect and to try to isolate the other effects of the rules for making social ties. All the graphs use the same scale for the node size and line color for comparison. Most relationships under 3km long are not captured in the line color because they are interzonal and do not cross zone boundaries. Dark lines emphasize relationships longer than about 3km.

In Experiment 2 with the spatially-embedded random social network (Figure 22), as in all experiments, the densely populated regions have the highest degree and this decreases gradually over several zones away from these regions (spatially proximate social ties are favored). The degree/population ratio is spatially homogeneous. The distribution of line color is slightly darker near the population centers, indicating more interzonal relationships here. These are indications that population density and the tie probability function that depends on distance together determine the spatial social network fabric.

In Experiment 5 (Figure 21), the regions surrounding the population centers and those at the edge of the scenario boundary have the highest degree/population ratio. Regions in densely populated centers have relatively low degree/population ratios. The darker lines show more relationships over long spatial distances than in Experiment 2.

Experiment 6 (Figure 23) exhibits especially non-homogeneous degree/population ratio in areas surrounding the dense centers. Here, as in Experiment 5, it is likely that the high-degree nodes from the zones at the city-center and border are divided by a settlement density which is artificially lowered by the zone incorporating part of an uninhabited lake or a zone in a mountain region. The dark lines are short and they link populous centers.

The randomly generated degree-12 social network without spatial embedding (Figure 24) exhibits very long spatial relationships in the dark lines traversing the graph. There is an even distribution of degree/population ratio with a slightly diminished value in the densest part of Zurich, as seen in the other experiments, but the values are larger than in the other experiments because the interzonal degree is higher in this model due to the number of social ties longer than 3km.

The socially dense social network with double average degree (24) in Experiment 102 (Figure 25) illustrates relative to Experiment 2 (Figure 22) slightly bigger nodes lines that are altogether somewhat darker, though they are not longer or transferred spatially. The lines are
still much lighter than Experiments 5, 6 or 12 because the longer spatial relationships are lacking in as high a number.

The spatially contracted social network in Experiment 104 (Figure 26) exhibits much smaller nodes due to lower interzonal degrees and lighter lines, due to fewer relationships at 3km or more distance.
Figure 21  Personal social ties embedded in space and geographically aggregated to 3km x 3km grid: Experiment 5

degree

degree/population
Figure 22  Personal social ties embedded in space and geographically aggregated to 3km x 3km grid: Experiment 2
Figure 23  Personal social ties embedded in space and geographically aggregated to 3km x 3km grid: Experiment 6

degree

degree/population
Figure 24  Personal social ties embedded in space and geographically aggregated to 3km x 3km grid: random12 social network

degree

degree/population
Figure 25  Personal social ties embedded in space and geographically aggregated to 3km x 3km grid: Experiment 102
Figure 26  Personal social ties embedded in space and geographically aggregated to 3km x 3km grid: Experiment 104.
The social context of establishing ties

In these experiments, the context of the social encounter is recorded in the social net edge as the "type" of relationship. The "type" is just a character string that can contain any information. In all but experiments 5 and 6, the context is "random" from the initialization of the static social network. Figure 27 shows the social context of the ties in final iteration of the experiments in which the social networks were allowed to evolve. For each ego in experiments 5 and 6, one social tie is either created ("new") or renewed ("renew") per activity per iteration, and the social ties are removed thereafter with equal probability until the average degree is restored to 12. If a social tie is made from agent A to agent B during an activity, and then agent B attempts to re-make the social relationship after that, it is "renewed", even though it did not exist last iteration.

Since each agent co-present will have a chance to make a social tie, this means that the probability of an ego making a new social tie at a given activity is strongly related to how many agents are co-present at the activity who are not already alters. This graphic is therefore a strong reflection of the relative number of agents at each activity, as well as how many social ties of this type have been erased last iteration, to make allowance for "new" rather than "renewed" ties of this type. There are two important effects of the link creation and removal algorithm illustrated in this figure. First, there is some indication of the rate of turnover of social ties (there is nothing in the socializing algorithms to secure a tie from agent A to agent B; only that agent A and agent B have a certain number of ties); and there is an indication that the type of activity at which social ties are made heavily favors leisure. Another important physical factor not represented in this graphic, and not considered in these experiments, is the effect of the size of the facility on the size of social group that can mix (see Section 7.3.10 for a summary of how many agents visit individual facilities if the capacity is not enforced).

The rate of link turnover is determined by the allowed rate of making new social ties and the cap on the average degree in the link removal algorithm. Coincidentally, permitting agents to make up to one new social tie per iteration slightly more than doubles the degree to 27 in the base case (see 1_501F2F in Table 18). So, in the activities "shop", "work", and "education", the agents approximately double, and then lose approximately half of their social ties in the link removal algorithm each iteration. These social ties are renewed each iteration and dissolved again. The height of the columns "renew" and "new" are about equal for these activities. Which of the agents are "new" to the ego network versus which have been "renewed" is purely random: continuity is not guaranteed by the algorithms investigated, here.

For "shop" and "leisure", many more "new" social ties are made per iteration than "renewed". A corresponding proportion of each are then dissolved. In experiment 5 without the socializing utility reward, 40% of the social edges are "new leisure". This is the case for 50%
of the social ties in experiment 6. This is because these locations can be changed, and new alters found, at the new location. The turnover at leisure and in shopping is higher because of this.
7.2.7 Agent degree versus number of activities in plan

Boxplots of the agent degree versus the number of activities in the agent's plan shows that more active individuals make more friends with the face-to-face rule for adding social ties. It
is a check on the influence of the socializing algorithm. There is no ascertainable trend of agent degree with the number of activities in the fixed (randomly generated) social networks.

Figure 28  Boxplots of the agent degree versus the number of activities in the agent's plan

Static degree 12 spatial social network  (initialization)  Experiment 102: Static degree 24 spatial social network (initialization)

Experiment 6: Social network which evolved with activity plan optimization  Experiment 5: Social network which evolved with activity plan optimization
7.2.8 The effect of the degree on knowledge

Degree is investigated as an influence on knowledge, with the hypothesis that more social connections give access to more information, giving choice sets with a higher likelihood of containing higher-utility locations. Only the experiments with social information exchanges are relevant (Experiments 3, 4, 5, 6, 104, 3_4, 4_4, and 23_4).

The number of locations known about per agent (see Section 5.4) does not depend on agent degree when the social network is fixed (Figure 29 Experiments 104, 3, and 4). This is not a function of culling the knowledge of the agents: the agents can retain a number of locations in memory up to the number of distinct activities in all of their plans (here, for example, 4 plans times the number of activities per plan). Additionally, the agents are allowed 1.5 times this number of locations in memory (Section 5.4).

The number of facilities that agents are aware of increases with degree in models with evolving social networks where new friends are made at new locations (Experiments 5 and 6). This concurs with the findings of Silvis et al. (2006) in Section 2.2.3. The number of locations in an agent's knowledge increases the size of the choice set for its secondary locations. Recall that the higher the number of activities in an agent's plan, the higher the number of agents it will encounter; with the simple proportional algorithm for friend-making used here, this means that more activities results in higher degree. Aside from the social exchanges enabled
by having many friends, more activities also means the agent has more knowledge about locations.

The number of locations known about in Experiment 23.4 is much lower, but superfluous information is not passed across the agents in these models.

Figure 29 The number of activities known about versus number of alters in ego net

Experiment 102: Static degree 24 spatial social network (initialization, without information exchange)  
Experiment 5: Social network which evolved with activity plan optimization

Experiment 6: Social network which evolved with activity plan optimization  
Experiment 104: Static degree 12 spatial social network spatially contracted
Experiments ending in "_4" are conceived as a more efficient representation of information exchange in avoiding storing data unnecessarily and avoiding an extra calculation step (1: exchange of location information with friends, 2: choose from knowledge, versus 1: choose from friends' knowledge). The agents in these experiments have fewer unique facilities in their knowledge than in the experiments with the first type of information exchange. The difference in the information stored by each agent is large in these models (the interquantile distance), and this will certainly influence the net flow of information across the social...
network. Agents with lowest and highest degree appear to be particularly disadvantaged by having a small choice set of locations. There could be a nonlinear influence of the social network topology, like betweenness centrality, which is a measure of the access to information flowing across a graph. No specific test (such as tracing which agents know about a particular location) was performed to qualify or quantify information flow across the social network however.

7.3 Travel behavior

This section summarizes the travel behavior of the agents in terms of travel times and the durations and locations of activities; i.e. behavior typically analyzed in a MATSim simulation.

Table 20 summarizes key statistics of the model output for comparison. The meaning of the columns is defined after the table.
Table 20  Summary travel statistics averaged over all agents's active plans and/or all trips

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Score</th>
<th>Chain length (km)</th>
<th>Trip distance</th>
<th>Dist. to all acts (km)</th>
<th>Trip dur. (min)</th>
<th>Trip speed (km/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>175</td>
<td>21.6</td>
<td>9.1</td>
<td>7.2</td>
<td>9.8</td>
<td>55.7</td>
</tr>
<tr>
<td>2</td>
<td>184</td>
<td>21.6</td>
<td>9.1</td>
<td>7.2</td>
<td>9.8</td>
<td>55.5</td>
</tr>
<tr>
<td>3</td>
<td>187</td>
<td>33.8</td>
<td>13.6</td>
<td>10.1</td>
<td>18.2</td>
<td>44.8</td>
</tr>
<tr>
<td>4</td>
<td>175</td>
<td>22.9</td>
<td>9.5</td>
<td>7.6</td>
<td>9.9</td>
<td>57.5</td>
</tr>
<tr>
<td>5</td>
<td>175</td>
<td>21.8</td>
<td>9.1</td>
<td>7.5</td>
<td>9.4</td>
<td>57.9</td>
</tr>
<tr>
<td>6</td>
<td>204</td>
<td>30.9</td>
<td>12.4</td>
<td>9.5</td>
<td>15.7</td>
<td>47.3</td>
</tr>
<tr>
<td>7</td>
<td>175</td>
<td>23.6</td>
<td>9.7</td>
<td>7.2</td>
<td>9.9</td>
<td>58.8</td>
</tr>
<tr>
<td>12</td>
<td>185</td>
<td>21.6</td>
<td>9.1</td>
<td>7.2</td>
<td>9.8</td>
<td>55.8</td>
</tr>
<tr>
<td>102</td>
<td>187</td>
<td>21.6</td>
<td>9.1</td>
<td>7.2</td>
<td>9.8</td>
<td>55.8</td>
</tr>
<tr>
<td>104</td>
<td>176</td>
<td>21.1</td>
<td>8.7</td>
<td>7.1</td>
<td>9.1</td>
<td>57.7</td>
</tr>
<tr>
<td>1_501F2F</td>
<td>175</td>
<td>21.6</td>
<td>9.1</td>
<td>7.2</td>
<td>9.8</td>
<td>55.7</td>
</tr>
<tr>
<td>7_501F2F</td>
<td>175</td>
<td>23.6</td>
<td>9.7</td>
<td>7.2</td>
<td>9.9</td>
<td>58.8</td>
</tr>
<tr>
<td>22</td>
<td>201</td>
<td>21.6</td>
<td>9.1</td>
<td>7.2</td>
<td>9.8</td>
<td>55.7</td>
</tr>
<tr>
<td>32</td>
<td>201</td>
<td>21.6</td>
<td>9.1</td>
<td>7.2</td>
<td>9.8</td>
<td>55.7</td>
</tr>
<tr>
<td>42</td>
<td>223</td>
<td>21.6</td>
<td>9.0</td>
<td>7.2</td>
<td>9.2</td>
<td>58.9</td>
</tr>
<tr>
<td>51</td>
<td>190</td>
<td>21.6</td>
<td>9.1</td>
<td>7.2</td>
<td>9.2</td>
<td>58.8</td>
</tr>
<tr>
<td>52</td>
<td>237</td>
<td>21.6</td>
<td>9.1</td>
<td>7.2</td>
<td>9.2</td>
<td>59.3</td>
</tr>
<tr>
<td>72</td>
<td>209</td>
<td>21.6</td>
<td>9.1</td>
<td>7.2</td>
<td>9.4</td>
<td>57.9</td>
</tr>
<tr>
<td>4_4</td>
<td>175</td>
<td>22.2</td>
<td>9.3</td>
<td>7.4</td>
<td>9.6</td>
<td>57.5</td>
</tr>
<tr>
<td>13_4</td>
<td>220</td>
<td>35.7</td>
<td>14.2</td>
<td>10.1</td>
<td>19.7</td>
<td>43.2</td>
</tr>
<tr>
<td>23_4</td>
<td>189</td>
<td>30.7</td>
<td>12.5</td>
<td>9.2</td>
<td>16.4</td>
<td>45.5</td>
</tr>
</tbody>
</table>

7.3.1 Average utility score of the experiment

The average executed plan score is influenced by socializing in some models. This reward is introduced in Table 12 through Table 14 and recapped in Table 22. The travel behavior in the models with and without socializing rewards are not comparable because the differing utility functions represent fundamentally different agent motives, and thus the behavior and the state of the traffic flow simulation are different as a result.
7.3.2 Chain length, trip distance, distance to acts

These different measures of geographic coverage, or "consumption" of space by the agents depend on the locations in the initial plans, the route choice, and/or the secondary location choice algorithm, and are a representation of an activity space in a single day. In the absence of socializing rewards, the only choice variable that can distinguish locations is accessibility (travel time, late penalty) so rewarding socializing means that utility can be increased even by choosing locations which have lower accessibility. Experiment 7 in which all facilities may be used provides a control on the "desired" travel distance the agents seek, given only the accessibility to distinguish the locations. The distances in this model are only slightly longer than those in the base case (initialization). The longest travel distances are realized with the fixed social network and the location choice and social reward mechanisms in the family of comparable experiments 3, 13_4 and 23_4, as well as Experiment 6, which additionally allows social network evolution as a function of face-to-face meetings. Not rewarding face-to-face socializing results in very small changes to travel distances (experiment 4, 4_4, 5). Different social network topologies do not seem to matter in determining the radius of activities or chain/trip length (experiments 102, 104).

7.3.3 Trip duration, speed

These measures reflect a mixture of the effects of route choice (higher-speed road infrastructure) and traffic congestion (time of day, location choice, as well as route choice). The reward for face-to-face meeting tends to increase travel times and cause congestion. Experiments 3, 6, 13_4 and 23_4 have much lower average speeds than the other experiments.

7.3.4 Time profiles of agents travelling

Throughout this section, which presents the time profiles of agents travelling, it is helpful to refer to the next section describing the time profiles of the agents' participation in activities.

The typical high travel demand peaks occur in the morning and evening work rush hours (experiment 1, Figure 30). Note that the plots generated by MATSim have different scales. Nearly no change in the time profile is qualitatively evident in experiments without utility rewards for face-to-face encounters (experiments 4, 5, 7, 104, 4_4). The changes to the secondary locations in these models enable the agents to avoid congestion costs (late penalties and excess travel time) but offer no other quantity to be evaluated in the utility, so the travel behavior in time is still tied closely to the desired schedules in the model boundary conditions, and the time profiles of travel volumes are not strongly affected.
Utility rewards for face-to-face socializing, without location choice replanning, cause peak spreading in the evening (2, 12, 102). The effect of valuating the total face-to-face time (Experiment 12) rather than the number of friends (Experiment 2) is to slightly widen the evening travel peak. The effect of doubling the number of friends (Experiment 102) also widens this evening peak (increased utility for socializing). Experiment 22 is identical to Experiment 12 except for a higher socializing parameter (rewarding the total time of face-to-face contact with alters at leisure activities). A separate, additional peak of travel is generated in the late evening as a result. Experiment 32 with no minimum activity durations enforced is identical to the result of 22.

The broadening of the travel volume peaks is stronger when location choice (3, 13_4, 23_4) and friend-making feedback is possible (experiment 6). In these experiments, the travel peaks are also 20-30% higher than in the other experiments. There is both more travel and it takes a longer time. The profiles support the observation in Table 20 that the agents trade off slower travel speed at peak times and remain on the road longer, in return for large rewards for spending time at the (socially) desirable locations, which become ever more attractive through the social reinforcement feedback of either the particular location choice algorithm used, or the ability to make new social ties there. The ability to make new social ties at the new locations in Experiment 6 is a positive feedback loop to make the location an even more fruitful source of social utility: where utility cannot be found it can be created by cultivating social connections at the location. The difference in these experiments between rewarding the number of friends encountered (23_4) versus the total time spent with friends (13_4) is consistent with the comparison between the time-only relaxation (Experiments 12 vs. 2).

Lifting or altering the facility constraints and/or the utility parameters for penalties alters the time profiles meaningfully because these penalties are potentially much higher than activity duration or social rewards. Experiments 42 and 52 have various social interaction mechanisms in play, but they have in common that the late penalty is not enforceable. The traffic volumes are very low relative to the other experiments and only a hint of morning and evening travel peaks are visible. A large peak of travel is attempted at the last hour. A similar pattern of travel volume with time emerges in experiment 51, with no social interaction mechanisms and no penalties for late arrival.

Experiment 72 has late penalties but all facilities are always open. There are three peaks of traffic volume, of low magnitude relative to the other experiments, with steadily climbing volume and a sudden reduction in the morning peak, and symmetric triangle-shaped peaks in the afternoon and late evening. Compared with Experiment 22 or 32, it shows that the facility opening times constrain the travel volume profiles. Some signatures of Experiment 72 are recognizeable in the output of Experiment 22, but it is obvious that the facility opening times
in Experiment 22 also have strong bearing on the travel volumes with respect to time, and thus the influence of the socializing utility reward in the models.

From the result of model 51 it is evident that some punishment for agents arriving late is necessary in order to resolve travel peaks seen in real behavior, whether this is an explicit penalty term in the utility function or implicit in a lowered realized utility for activities begun too late.

Experiments 42 and 52 attempt to induce schedule coordination without facility opening or closing times and without exogenously penalized tardiness to an arbitrary latest possible start time: solely through endogenous social penalization for missing social appointments. This mechanism fails to produce models with customary morning and evening travel peaks, because the socializing opportunity costs do not affect the plan utility in the same way that a specific penalty, applied to a specific "desired" arrival time of work and education activities has. This does not mean that endogenous schedule coordination using a method of social reward/penalization is not possible with a social network mechanism within MATSim. It means that the mechanism for endogenously determining the behavior norm and enforcing it has to be specific to the activity and of sufficiently high magnitude relative to the activity- and travel utility: i.e. the type of relationship ("work"), the strength of the penalty, and the reference point for determining the penalty need all align.
Figure 30  Time profiles of the number of agents travelling

Experiment 1

Experiment 2
Figure 30  Time profiles of the number of agents travelling (continued)

Experiment 3

Experiment 4
Figure 30  Time profiles of the number of agents travelling (continued)
Figure 30  Time profiles of the number of agents travelling (continued)
Figure 30  Time profiles of the number of agents travelling (continued)
Figure 30  Time profiles of the number of agents travelling (continued)

Experiment 22 and 32

Experiment 42
Figure 30  Time profiles of the number of agents travelling (continued)
Figure 30  Time profiles of the number of agents travelling (continued)

Experiment 72

Experiment 4_4
7.3.5 Time profiles of agents at activities

Leisure and shopping activities are permitted to be re-located in many experiments, and leisure encounters are rewarded in some. However, participation times in all activities are
affected by these changes. The shopping and leisure activities bring maximum marginal utility at 2 hours' duration and they are not subject to desired start and end times, being scheduled around the anchor activities. In most plans these activities take place before or after the main activity of "working" (desired duration 8hrs) or "education" (6hrs), and all plans begin and end at home (12 hrs duration).

Changes in both the duration and the schedule are visible in the plots of aggregate participation in activities in Figure 31. In the reference case 1, shopping peaks occur at 9:00 and at 17:30, and leisure at 7:30 and 18:00. In Experiments 2, 12, and 102 (face-to-face reward and time rescheduling) there are large displacements to the evening for leisure: both a later peak (19:00) and more agents/longer duration at this time. The slight lunchtime peak of shopping participation is eliminated. No difference can be detected between the algorithms rewarding face-to-face contact (2) versus the duration of face-to-face contact (12), or the model with higher social network density (102). In the shift of leisure activities to a later time, the time at which the agents return home is also later.

In Experiments 4 with secondary location choice and no social reward, 104 (spatially contracted social network) and 4_4, the time profiles are nearly identical to the reference experiment 1, as is the profile in Experiment 5 with social network evolution. This points to independence of social network topology (since 5 and 104 have different social networks and the result in 102 did not differ qualitatively from the result in 2).

In Experiments 3, 6, 23_4 the leisure activities shift to a later time (19:30) with a broad range of start- and end times and durations, and shopping is also distinctly later (19:00). The profile is similar for experiment 13_4, but participation in leisure activities peaks even later at 21:00.

Experiments 22 and 32 show very late leisure participation peaks (20:30-21:00) with large spread but without the early-evening participation seen in Experiments 3 and 6. Experiments 42 and 52 spread the activities throughout the day and though leisure takes place predominantly in the early morning and evening, there is hardly any variation throughout the day. Shopping activities are evenly distributed throughout the day. Only circa 5500-6000 agents are at home concurrently; thus there is clearly a need for the strong boundary conditions on the utility function and the participation times within MATSim. These cannot be simply lifted to make endogenous schedule planning possible, and alternative approaches have to be found to effectively allow agents to determine their own schedules based on endogenous as opposed to exogenous schedule constraints.

In Experiment 51 the double-peak shopping and leisure behavior disappears and shopping and leisure take place all day, but especially in the late afternoon/early evening. The double-peak returns in experiment 72 with morning peak for shopping and leisure at 7:00 and later peaks at
17:30 (shopping) and 22:00 (leisure). Again, the experimental settings of these two models are unlikely to be used in MATSim because the loose constraints do not lead to realistic behavior in time.

In these activity time shifts, work and education activities are affected in a marked way only in those models which relax the desire start times and/or late start penalties. The home and shopping activities are shifted in time to compensate for the leisure utility reward because they are not fixed in time in the boundary conditions (exogenous schedule) of the model.

Figure 31 Time profile of the number of agents at activities

![Experiment 1](image1)
![Experiment 2](image2)
![Experiment 3](image3)
![Experiment 4](image4)
Figure 31  Time profile of the number of agents at activities (continued)
Figure 31  Time profile of the number of agents at activities (continued)

Experiment 102

Experiment 104

Experiment 22

Experiment 32
Figure 31  Time profile of the number of agents at activities (continued)
7.3.6 Distribution of plan distances (sum of trip distances):

The distribution of the total length of the activity chain (average of 21.6km in the reference case in Table 20) is summarized in Figure 32. The distribution reflects not only travel distances between activities but also the number of activities per plan. Longer chains are expected to be associated with more activities. The activity chain length is a strongly positively-skewed distribution that is similar for all models and only two variations are qualitatively distinguishable: 1) Experiments 1 (shown), 4, 5, 7, 104, 4_4 as well as all time-only replanning models (2, 12, 102, 22, 32, 42, 51, 52, 72) which have either no socializing utility reward or no secondary location choice mechanism have a lower median and average chain
length; and 2) Experiments 3 (shown), 6, 13_4, and 23_4, which have secondary location choice replanning as well as face-to-face socializing utility rewards, have a right-shifted length distribution with more very long activity chains.

Figure 32 Activity chain length distribution

Experiment 1 (as well as 2, 12, 102, 22, 32, 42, 51, 52, 72; 4, 5, 7, 104, 4_4 similar)  Experiment 3 (6, 13_4, 23_4 similar)

Because the effect of the number of activities in the plan and the average trip length cannot be distinguished in Figure 32, the corresponding activity chain lengths are divided by the number of activities in each plan and plotted by number of activities. This yields a trip length distribution versus the number of activities in a plan, in order to evaluate to what extent the activity chain lengths are related to the number of activities. The result is shown in Figure 33.

The median and interquartile range of the trip length distribution in Figure 33 does not vary strongly across plans of different numbers of activities, for plans with 3-10 activities in those experiments with weak responses to the secondary location choice mechanisms (left plot), and 5-10 activities in the experiments with a strong response (right plot). Because the trip distance distribution is unchanging as the number of activities increases, it is clear that adding an activity is highly likely to add distance to the plan of the order of one more average trip distance. In the experiments with strong responses to secondary location choice (right plot), a large increase in total plan distance also occurs when the fourth activity is added to the plan. The longest median and interquartile trip lengths occur in plans with 11-14 activities in both groups of model results. Plans with this many activities become much longer by having both more activities and longer trip lengths. The longest trip lengths are the outliers in the plans with fewer activities (3-7 activities in the first group of experimental results on the left, and 3-11 activities in the second group of experimental results on the right).
The long activity chains respond to the secondary location replanning mechanism (with socializing reward) by becoming slightly shorter (right plot in Figure 33). The lack of a very strong response to location choice replanning (i.e., changing many of the locations to closer facilities) is a result of these plans not having as many re-locatable "leisure" and "shopping" activities as they do fixed "work" and "home" activities.

Allowing the secondary locations to be replanned (in combination with a utility reward for face-to-face socializing) shifts the trip length distribution to become longer and more positive-skewed throughout the range of plans with 3 or more activities.

So, the main difference in the activity chain length distribution between the two groups of results in Figure 32 is caused by a longer expected trip length in the plans with 3-11 activities, but not in the trip chains with more than 11 activities. No evidence is found for trip distances to decrease on a per-activity basis (which would be a strategy for increasing the duration of activities). Instead, each new activity comes at a constant increased travel cost. Though plans with fewer activities have a higher incidence of extremely long trips and could occasionally cover very long distances, more activities can be reliably expected to be longer in distance.

Figure 33  Average trip length (distance between activities, km) per number of activities

| Experiment 1 (as well as 2, 12, 102, 22, 32, 42, 51, 52, 72; 4, 5, 7, 104, 4_4 similar) | Experiment 3 (6, 13_4, 23_4 similar) |

7.3.7  Distribution of the average distance to activities

The average distance to all activities (in a plan) (7.2 km in Table 20) may or may not be function of the number of activities in the chain. The distribution is strongly positive-skewed.
with a higher-than-Normal incidence of longer distances. The plots in Figure 34 show as in the previous section that the scoring function (experiments 3, 6, 13_4, 23_4) has more to do with the distribution of the distance to activities than the alternative set of locations (Experiments 4, 7, 4_4) or the social network evolution (5) or the social network geography (104). The distribution is fixed for Experiments 2, 12, 102, 22, 32, 42, 51, 52, and 72, and equal to that in Experiment 1.

Figure 34 Distribution of Euclidean distance to activities

Experiment 1 (as well as 2, 12, 102, 22, 32, 42, 51, 52, 72; 4, 5, 7, 104, 4_4 similar)  Experiment 3 (6, 13_4, 23_4 similar)

7.3.8 Average distance to activities versus number of activities in plan

Do more activities mean travel to farther locations? Or are the activities concentrated together such that a geographic footprint of the activities (a kind of activity space for a single day's plan) is constant? Or are an agent's activities localized in space independently of how many there are in a plan? The average distance from an agent's home to its activities across all the experiments is plotted versus the average distance by the number of activities in a plan in Figure 35. As for the trip distance and chain lengths (km), the scoring function (experiments 3, 6, 13_4, 23_4) has more to do with the distribution of the average Euclidean distance to activities than the alternative set of locations (Experiments 4, 7, 4_4) or the social network evolution (5) or the social network geography (104). The distribution is fixed for Experiments 2, 12, 102, 22, 32, 42, 51, 52, and 72, and equal to that in Experiment 1.

There is higher variance the fewer the activities in the plan (as in the previous examples of chain- and trip length), and the median and interquartile ranges do not reveal any trends across the number of activities.
This means that agents locate their activities within similarly-dimensioned spaces, no matter how many activities are to be carried out. Summarizing the dimension into a single measure, or radius, is a first approximation of its size for comparison, and results in an interquartile range of roughly 3-10km for the initialization and 5-15km for the experiments permitting the shopping and leisure activities to be re-located in response to social utility reinforcement. The distance between activities that is charted in Figure 32 and Figure 33 is contained within this space, though the distance travelled (chain length) increases with the number of activities.

Figure 35 Distribution of distance to activities versus the number of activities per plan

Experiment 1 (as well as 2, 12, 102, 22, 32, 42, 51, 52, 72; 4, 5, 7, 104, 4_4 similar)  
Experiment 3 (6, 13_4, 23_4 similar)

7.3.9 Summary of the consumption of space and travel

With some variation and a positive-skewed distribution, the following points summarize the agents' use of space in carrying out their activity plans:

- Activity radius (Figure 34 and Figure 35): the footprint of geographic territory consumed for a given number of activities in a plan (a measure akin to activity space, but only for a day) is constant across the number of activities.
- Trip length (Figure 33): The distance between activities is constant across the number of activities, except for plans with a very high number of activities.
- Chain length (Figure 32): the distance travelled in the day plan is a linear function of the number of activities.

Neither quantity changes (within the model variation) with a replanning strategy that randomly alters the locations of shopping and leisure activities, unless there is a concurrent utility reward for meeting alters face to face at the leisure activity.
As far as spatial or geographical forms of the footprint, neither the measure "chain length", nor "average distance to activities" nor any product or quotient thereof can distinguish between activity plans which are spatially dispersed versus aligned on a corridor or an arc on which the agent travels back and forth. To measure this would require two indices, such as average variance in x and y of the locations of the activities for each plan (for example, a confidence ellipse, e.g. Schönfelder, 2006).

7.3.10 Quantification of facility popularity

The number of different facilities used by the agents in each model does not vary by more than a handful of facilities (Table 21). The education, work, and home locations were not permitted to change. But the allocation of shop and leisure locations (secondary locations) changes depend on the experiment. None of the models allowed searching through all possible locations except Experiment 7. Thus, the task of secondary location choice comprised finding a new allocation of locations among the initial set of locations, including the possibility that some locations may not be used at all in the resulting utility-maximized plans. This discussion begins by looking at the facilities that were used, before looking at the issue of how many agents use which facilities.

4989 unique facilities are used in the initial plans, e.g. in Experiment 1, the control in this experiment. The activities in all models of secondary location choice are also distributed among the 4989 facilities, except in Experiment 6, in which three of the original facilities are not used. Experiment 7 makes use of the two shopping facilities that were not frequented in the initialization, but also finds a solution in which the original 4989 facilities are the ones used (the new shopping locations are located in buildings already in use in the initialization for another activity purpose).

In the initialization, the agents' activities are distributed across all 114 distinct leisure and 181 of the 183 available distinct shopping locations. In Experiment 6 (utility socializing reward at leisure, friend-making at all activities, and information exchange) this activity is concentrated into only 104 and 176 locations, respectively. In models 3 (utility socializing reward at leisure, information exchange) and the slightly different formulation of the same model with a choice of secondary location from the ego net, 13_4 (higher reward) and 23_4, there are also one or two fewer unique leisure and shopping locations, showing a concentration of activities in fewer locations. In Experiment 7, where two shopping activities are added to the original set, the indication is that for this utility function and the flow capacities on this road network, the algorithm to allocate the initial activity locations does not minimize travel costs along the entire activity chain. However the trip-based algorithm that was used for the initialization results in an allocation of locations that is very similar to that which agents would choose were they choosing the locations in order to maximize the plan's utility.
Table 21 Number of different facilities used in the final plans in the models with secondary location choice and the variance in the number of agents choosing the facilities in their plans

<table>
<thead>
<tr>
<th>Experiment</th>
<th>education</th>
<th>home</th>
<th>leisure</th>
<th>shop</th>
<th>work</th>
<th>all</th>
<th>Shop</th>
<th>Leisure</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total available</td>
<td>71</td>
<td>4917</td>
<td>114</td>
<td>183</td>
<td>1521</td>
<td>6438</td>
<td>14</td>
<td>35</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>71</td>
<td>3591</td>
<td>114</td>
<td>181</td>
<td>1367</td>
<td>4989</td>
<td>14</td>
<td>35</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>71</td>
<td>3591</td>
<td>110</td>
<td>179</td>
<td>1367</td>
<td>4989</td>
<td>28</td>
<td>143</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>71</td>
<td>3591</td>
<td>114</td>
<td>181</td>
<td>1367</td>
<td>4989</td>
<td>22</td>
<td>49</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>71</td>
<td>3591</td>
<td>113</td>
<td>181</td>
<td>1367</td>
<td>4989</td>
<td>24</td>
<td>49</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>71</td>
<td>3591</td>
<td>104</td>
<td>176</td>
<td>1367</td>
<td>4986</td>
<td>28</td>
<td>142</td>
<td>24</td>
</tr>
<tr>
<td>7</td>
<td>71</td>
<td>3591</td>
<td>114</td>
<td>183</td>
<td>1367</td>
<td>4989</td>
<td>12</td>
<td>32</td>
<td>12</td>
</tr>
<tr>
<td>104</td>
<td>71</td>
<td>3591</td>
<td>113</td>
<td>180</td>
<td>1367</td>
<td>4989</td>
<td>18</td>
<td>40</td>
<td>13</td>
</tr>
<tr>
<td>4_4</td>
<td>71</td>
<td>3591</td>
<td>114</td>
<td>181</td>
<td>1367</td>
<td>4989</td>
<td>16</td>
<td>43</td>
<td>15</td>
</tr>
<tr>
<td>13_4</td>
<td>71</td>
<td>3591</td>
<td>109</td>
<td>179</td>
<td>1367</td>
<td>4988</td>
<td>17</td>
<td>135</td>
<td>23</td>
</tr>
<tr>
<td>23_4</td>
<td>71</td>
<td>3591</td>
<td>112</td>
<td>181</td>
<td>1367</td>
<td>4989</td>
<td>17</td>
<td>122</td>
<td>21</td>
</tr>
</tbody>
</table>

The column "all" in the table is less than the sum of the number of locations in the columns because multiple activities are possible at some facilities, and these facilities will be counted as "unique" for that activity type in each column.

The second part of Table 21 provides a measure of the distribution of the number of agents visiting each facility. It is the variance about the mean number of activities taking place at a facility as plotted in Figure 36. A lower variance in the number of agents choosing a facility indicates relatively constant distribution of agents across locations, i.e. no strong preferences. A higher variance indicates that some locations are chosen by very many agents compared to other locations, i.e. a measure that there are preferred locations.

In Experiments 1 and 7 without any agent interactions, the variance of the average number of agents choosing a facility is 12 if all locations, even those not changeable in the run, are considered. That the variance in Experiment 7 is lower for leisure and shopping than in Experiment 1 indicates that facility choice is spread out more in Experiment 7, which is consistent with the slightly higher number of unique facilities used and with the philosophy of
Experiment 7 to let agents relax the constraint on their locations that was enforced in the initializing algorithm.

The experiments can be divided into three types of outcome. Those with the lowest variance and thus most evenly spread visits to locations are the base cases of experiment 1 and 7. When the social network for information exchange is spatially localized (Experiment 104), this raises the variance in the number of agents choosing a facility over the base case, but due to the fractured social networks, information does not spread about popular facilities, so the variance (and concentration of activities) stays lower than for the next group of experiments. Experiment 4.4 also lacks the information exchange dynamic to concentrate agents into popular locations (discussed further, below) because there is no utility reward for socializing in these experiments that would cause feedback encouraging agents to only visit locations where their friends are. So, this lack of strong preference for certain facilities is a phenomenon of information exchange and the choice set, in combination with the accessibility of the facilities which is the sole driver of the location choice in experiment 7.

Experiments 4 and 5, also with no socializing rewards, show intermediate variances where the social networks and the agent interactions result in concentrations of demand at certain shopping as well as leisure activities. Thus similar information exchange and location choice algorithms result in similar popularity measures for facilities, with or without social network evolution.

Those experiments with socializing rewards for face-to-face meetings have the highest variance, whether the social network evolves with face-to-face meetings or not (Experiments 3 and 6, 13.4 and 23.4). Though Experiments 3 and 23.4 are identical except for the implementation of the social exchange algorithm, the variance of the number of agents choosing a shopping location is much higher in experiment 3 (and 6) than in 23.4 (or 13.4), suggesting as in Experiment 4.4 (as compared to Experiment 4) that the algorithm for exchanging information about locations affects the choice of locations: the "_4" experiments do not pass superfluous location information that is not used immediately in a plan change, whereas the other experiments do retain and pass information that has not necessarily been used. Thus this information can be passed on further to agents who might find it useful, and the information percolates through the social network better.
Figure 36  Number of agents choosing each leisure facility versus the facility ID for models 1, 7, 5, and 6.

Arranged in order of popularity; only the most popular facilities shown.

The facility number is shown in the graphic in Figure 36 for the facilities which were especially attractive in Experiment 6. For insight into why these particular locations emerged as most popular, it would be necessary to investigate the components of attractiveness that enable agents to distinguish them in their plans: accessibility measures versus their social attractiveness. The locations have been plotted in space but reflect an even distribution along the main transportation arteries as well as the population density. No conclusions can be drawn from the plot. Other measures of social popularity or accessibility of these particular locations have not been calculated yet, and the investigation remains open.

If one can (suspend algorithmic rigor and) imagine the MATSim social network iteration to represent the passage of time over a series of similar Tuesdays or Thursdays in which agents repeatedly meet face to face, then the results can be compared to the findings of Silvis et al. (2006) regarding where repeated social visits occur. Repeat visits do not mean visits to the same places; indeed, the models in which friends are drawn by a utility reward to return to the same location to meet again with other friends may not be producing desirable dynamics in this respect. Such reinforcement may more realistically be replaced by a module which permits social interaction to generate more social interaction. Such an implementation would
truly require a model of real time progression however, and not the "pseudo-time" interpretation of the MATSim interval.

### 7.3.11 Activity duration

The utility function encourages an agent to align the durations of its activities with a working point of duration at which a maximum marginal utility accrues. The standard model response in the case of no social interactions (Experiment 1) shows that the average duration of all activities, except "education" well exceeds the "working point" such that decreasing marginal returns to longer duration are realized (Figure 37). This is rational as the transportation losses that may be taken as a result are still less than or equal to the gains of longer duration activities, and the net utility is as high as can be for the agents. The response of the activity durations to the social network experiments in Figure 37 can be summarized as follows:

- All models relax with work, leisure, and shopping durations that are longer than the working point for that activity type. The education activity is often shorter (utility could still be gained by attending school longer).
- Although shopping and leisure both have the same working duration (per instance of the activity), the time spent per leisure activity is always higher than the time spent shopping.
- Free secondary location choice (Experiment 7) results in much longer leisure durations relative to Experiment 1, with little change to the durations of other activities.
- Removing late penalties or setting no working duration results in meaningless activity durations (models 42, 51, 52, 72)
- A utility reward for socializing face to face at leisure activities tends to lengthen the duration of leisure activities relative to the base case (Experiment 1); a scoring function rewarding the number of agents met, as opposed to the duration of the encounter, lengthens the activity even more. However this may be a question of the parameter value in the utility function: 10 * (number of friends) does not equal 10 * duration in hours (compare Experiments 2 and 12).
- The shortest shopping durations of roughly 3.2 hrs are the experiments with shopping/leisure location choice and no utility reward for socializing: Experiments 4, 104 (spatially contracted social network) and 4_4, as well as experiment 5 with social network evolution. Those (plausible) experiments with the longest shopping durations of 5.25 hours are 13_4 and 23_4 with socializing utility reward and less exchange of information about locations.
- Of the experiments with plausible travel behavior results (i.e. a penalty for arriving late), Experiment 2 with no secondary location choice has the longest leisure durations. Experiments 4, 5, and 104 with secondary location choice but no utility reward have the shortest leisure activity durations. No certain trends in leisure duration can be ascertained among the other experiments.
• The duration of the education activity is determined strongly by the boundary conditions of the model (Experiments 42, 51, 52, 72). If these are left as standard, models 22 and 32 with the high-parameter socializing reward for the duration of face-to-face contact at leisure activities increase the education duration to 7.5 hours (instead of 6 hours). These experiments use the random spatially embedded social network so it is not clear why education activities should last longer. Besides these models, only experiments 12, 13_4 and 23_4 have education durations that are equal to the working duration, though they are very different models. All other models have education durations that are shorter than the working duration, meaning that marginal utility for longer durations is still increasing. Most likely, the agents in these models have arrived at education late due to traffic congestion, and must leave when the facility closes, so they are not able to realize the next units of utility. There is no ascertainable trend between model specification and the education durations in these remaining models.

• The duration of work activities also has no ascertainable trend across model specifications except for the sensitivity to boundary conditions and to the utility parameter for the late arrival penalty.

With the exception of some extreme cases, the resulting activity durations do not seem to follow recognizable, much less predictable, trends as functions of the model specifications: scoring, information exchange, social network topology, location choice. Since the duration and location of activities are at the core of the activity-based planning approach, it would be important to more extensively probe the model responses with respect to activity duration and to perhaps set up more specific experiments that target producing detectable signals in the activity duration that can be attributed to the social network and social interaction mechanisms in the model.
7.4 Social travel and face to face interactions

This section summarizes the interplay between the social network and the travel behavior that was observed in the experiments. The socializing mechanisms are related to: the number of alters and non-alter agents encountered at the activities, the duration of the encounters, the distribution of the encounters in time, measures of the spatial expanse of the travel involved to meet, and the contribution of social encounters to the plan score.
Table 22 summarizes aggregate measures of agents overlapping in time at their activities and has a reference column recalling the scoring function used. Columns summarizing the social interactions at leisure activities are included since these are elements of the scoring functions in some experiments, though the socializing mechanisms influence encounters between agents (friends or otherwise) at all activity types. Detailed analyses about the socializing behavior follow this summary.
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Score*</th>
<th>N alters at leisure (# alters)</th>
<th>N alters score*</th>
<th>Avg score-social score*</th>
<th>Time overlap with alters at leisure (alter-hrs)</th>
<th>Social score function and beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>175</td>
<td>0.0</td>
<td>0.0</td>
<td>175</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>184</td>
<td>0.2</td>
<td>0.9</td>
<td>183</td>
<td>2.2</td>
<td>10 N</td>
</tr>
<tr>
<td>3</td>
<td>187</td>
<td>0.2</td>
<td>1.3</td>
<td>186</td>
<td>2.1</td>
<td>10 N</td>
</tr>
<tr>
<td>4</td>
<td>175</td>
<td>0.1</td>
<td>0.0</td>
<td>175</td>
<td>1.8</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>175</td>
<td>3.2</td>
<td>0.0</td>
<td>175</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>204</td>
<td>3.5</td>
<td>13.5</td>
<td>190</td>
<td>1.8</td>
<td>10 N</td>
</tr>
<tr>
<td>7</td>
<td>175</td>
<td>0.0</td>
<td>0.0</td>
<td>175</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>185</td>
<td>0.1</td>
<td>0.9</td>
<td>173</td>
<td>2.6</td>
<td>10 T</td>
</tr>
<tr>
<td>102</td>
<td>187</td>
<td>0.3</td>
<td>1.7</td>
<td>185</td>
<td>2.2</td>
<td>10 N</td>
</tr>
<tr>
<td>104</td>
<td>176</td>
<td>0.3</td>
<td>0.0</td>
<td>176</td>
<td>1.8</td>
<td>0</td>
</tr>
<tr>
<td>1_501F2F</td>
<td>175</td>
<td>11.3</td>
<td>0.0</td>
<td>175</td>
<td>1.0</td>
<td>0</td>
</tr>
<tr>
<td>7_501F2F</td>
<td>175</td>
<td>9.8</td>
<td>0.0</td>
<td>175</td>
<td>1.1</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>201</td>
<td>0.2</td>
<td>2.4</td>
<td>167</td>
<td>3.5</td>
<td>24 T</td>
</tr>
<tr>
<td>32</td>
<td>201</td>
<td>0.2</td>
<td>2.4</td>
<td>167</td>
<td>3.5</td>
<td>24 T</td>
</tr>
<tr>
<td>42</td>
<td>223</td>
<td>0.2</td>
<td>2.5</td>
<td>187</td>
<td>3.7</td>
<td>24 T</td>
</tr>
<tr>
<td>51</td>
<td>190</td>
<td>0.0</td>
<td>0.0</td>
<td>190</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>52</td>
<td>237</td>
<td>0.2</td>
<td>2.4</td>
<td>198</td>
<td>4.5</td>
<td>24 T</td>
</tr>
<tr>
<td>72</td>
<td>209</td>
<td>0.2</td>
<td>2.4</td>
<td>174</td>
<td>3.5</td>
<td>24 T</td>
</tr>
<tr>
<td>4_4</td>
<td>175</td>
<td>0.2</td>
<td>0.0</td>
<td>175</td>
<td>1.8</td>
<td>0</td>
</tr>
<tr>
<td>13_4</td>
<td>220</td>
<td>0.4</td>
<td>6.0</td>
<td>190</td>
<td>2.8</td>
<td>24 T</td>
</tr>
<tr>
<td>23_4</td>
<td>189</td>
<td>0.4</td>
<td>2.1</td>
<td>187</td>
<td>2.1</td>
<td>10 N</td>
</tr>
</tbody>
</table>

*The total score consists of the standard score for the activity travel plus (in some experiments) a social score. This social score component is either based on the number of alters present ($\beta \ln(1+N)$, represented by $\beta N$) or the time overlap with alters present (face to face time: $\beta \ln(1+T)$, represented by $\beta T$).
7.4.1 Summary of the number of alters encountered at leisure activities

The average number of alters met at leisure is low if social ties are not established at face to face meetings. The random mechanisms of replanning do not align plans strongly, such that the incidence of alters crossing paths at activities is low. Using Experiments 1_501F2F and 7_501F2F as a measure of the maximum size of the social pool, with the initial locations and freely chosen locations, respectively, the reference random spatially embedded social network leads to roughly 2% of the agents which encounter one another face to face at leisure activities being previously tied socially. This is examined deeper in Figure 38 and Figure 39, which break down the number of agents co-present at activities who are alters and who are not known to each other.

The number of alters encountered at leisure activities is highest with a mechanism that permits social ties to be formed at leisure activities. Rewarding face to face meetings between agents has the second-strongest effect (3 versus 4). The models with denser social (102, fixed locations) or spatial (104, location choice) relationships also cause higher numbers of alters to overlap at activities: the former because each ego has more alters: the latter because of socially reinforced location choice sets that are spatially limited in scope (the agents have no choice but to encounter one another).

7.4.2 Summary of the duration of face-to-face leisure encounters

The duration of face-to-face contact at leisure is most strongly determined by whether a utility reward for face to face contact is awarded or not. Rewarded activities result in longer face-to-face contact. The next largest influence is the form of the scoring reward (Number versus Time; 2 versus 12) and the size of the scoring parameter (22 versus 12). The mechanism of adding social ties at the activities reduces the duration of social contacts (5, 6). The longest face to face encounter durations occur in models without penalties for being late to scheduled activities (42, 52) and with facilities open 24 hours (72).

7.4.3 Scoring the social interactions

The different scoring functions each dictate a different motivation for the agents in the experiments. The models without social interaction rewards consistently yield 175 points, regardless of the model mechanisms (1, 4, 5, 7, 4_4). The family of models allowing social rewards and time schedule overlaps (2, 12, 102) as well as secondary location choice (3, 23_4) provide approximately 185 points with the low parameter value and 200-220 points with the high value (22, 32, 72), depending on the boundary conditions enforced on the facilities in utility function. Eliminating the desired start and end times of the activities in an attempt to have the agents determine the best times for endogenously (socially) results in
utility of 223 (42), however note that the enforcement of late penalties is not possible without desired activity start times. Indeed, the highest utilities are attained in model 52 because there are no late penalties. The utility in Experiment 6 with social network evolution, secondary location choice, and socializing reward is 204, the highest of the models with standard boundary conditions.

The scores of experiments with different scoring functions cannot be compared because the agents are attempting to solve a different relaxation problem in each case. As an illustration, a very rough measure of the discrepancy between the mobility (travel behavior) result of a model with and without socializing utility rewards is obtained by subtracting the aggregate social score from the average total score to yield a mobility-only score (Table 22). The expectation is that the mobility score net the social reward would be lower than for the reference case, since the agents would be receiving compensation for less efficient travel decisions through the socializing reward. However the net mobility scores (standard MATSim scores) are generally higher in the experiments with social rewards. A reward for the total duration of the encounter (agent-hours) rather than the number of agents (compare 2 versus 12) makes a large difference of 12 utility points, for example. The socializing reward enables agents to travel at times, on routes, and to locations which are better on average, but which are not discovered by agents without the social motivation, and which do not have the same value to agents without social motivations.

7.4.4 Number of agents encountered by activity type

Figure 38 summarizes the average number of face-to-face encounters an ego can make at the various activity types (N other agents minus the ego): thus a measure of location choice (due to social reinforcement in the information exchange and/or utility, or else to the initial allocation) and time scheduling of the activities in the relaxed state of the models.
The initialization algorithm for the locations for the activities (Experiment 1) results in a fixed allocation of agents to facilities. When relaxed with schedule (and route) replanning, this spatial allocation results in the highest number of agents overlapping with a given ego in time/space at leisure activities (11.56 other agents) and the lowest at home (1.33). The work (3.38), shopping (4.61) and education (5.35) activities have intermediate numbers of agents overlapping in time. The absolute numbers are not meaningful since this is a 1% random sample of a real population. The relative sizes of the groups at the different activities is the relevant measure, here.

The fewest number of facilities are of the type, "education", which concentrates this activity in few places. Education is a rigidly scheduled activity which also increases the overlapping presence of agents. However only the younger portion of the population is engaged in it, so the size of the overlapping groups is reduced. The large number of work and home locations reduces the number of overlapping agents at these facilities, while again the rigid schedules would tend to increase it (as well as the fact that every agent will eventually return home, and most will participate in work, at some time during the day).
There are a similar number of leisure and shop locations (114 and 183), a small number relative to the number of work and home activities (Table 21). However the number of agents overlapping at leisure versus shopping is much larger in Experiment 1. It is necessary to find an explanation for this. There is no difference in the utility valuation of shopping and leisure activities in the utility function of Experiment 1 that could explain the difference (both have a desired duration of 2 hours and no fixed start times). Referring to Figure 37, it is evident that the difference in the number of overlapping agents is caused by the longer duration of leisure activities (each agent spends more time on average at leisure than at shopping), which affords more opportunities to overlap with other agents. The explanation of the longer activity durations as well as the larger numbers of agents encountered is found by looking at the opening times of the "shopping" facilities (8am-8pm), which are longer for leisure facilities (6am-12am). Leisure activities in the evening are especially popular (Figure 31).

The experiments with secondary location choice and the utility reward for face to face socializing at leisure activities (3, 6, 13_4, 23_4) have the largest influence on the number of agents encountered at leisure. Experiments 3 and 6 with the first type of location information exchange also strongly increase the number of agents co-present at shopping activities, whereas this increase is weak or not present in the second type of information exchange (13_4 and 23_4). The experiments with secondary location choice (and the first type of location information exchange) but without a utility reward (4 and 5) result in more agents overlapping at shopping activities but no increase in the number at leisure activities. Experiment 104 (identical to experiment 4 but with a spatially proximate social network) also has more agents overlapping at shopping than in the reference experiment.

The number of agents encountered at home and at school only varies in the experiments which do not have schedulable activities due to the lack of enforceable late-arrival penalties (42, 51, 52 and 24-hour facility open times (72).

### 7.4.5 Number of alters encountered by activity

Figure 39 shows the average size of groups of alters that encounter one another during the different activities, and Figure 40 examines the distribution of the group size without distinguishing by activity type. In these plots, the social network is an additional factor affecting the size of groups, above location choice and activity scheduling. The social networks are all initialized independently of the activity locations or the activity schedules. If a pair of agents have a social tie between them, and happen to participate in the same activity at the same time in the initial plans, this is a coincidence having to do with the fact that both the initial social network and the initial allocation of activity locations are spatially dependent. Depending on the socializing mechanisms used in each experiment, the face-to-face overlap becomes more or less strongly tied to the social network as the program iterates.
The group sizes are correspondingly much smaller than the number of all agents encountered. Large numbers of "zero"-sized groups, i.e. no alters met, reduces the average group size to tiny values, which is misleading due to the significant effect that larger group meetings have on social interactions (utility feedback, social network evolution, etc.).

Figure 39  Average number of an ego's alters co-present at the different activities

There is no reference case in Figure 39, since all the experiments shown have to do with one or another social mechanism. However several commonalities and contrasts can be summarized. The experiments with no secondary location choice and a face-to-face utility reward for socializing (2, 12, 22, 32, 42, 52, 72), as well as the experiments with secondary location choice but no utility reward (4) have very similar group sizes of alters in the different activities. The fewest alters are encountered at work (0.02 encounters per ego per work activity), then shopping (0.06-0.07), education (0.06-0.07), at leisure (0.13-0.16), and at home (0.16).

Experiment 3 with a social utility reward for leisure encounters has larger leisure groups (0.21) and Experiment 4_4 with no utility rewards for socializing but with the second type of location exchange, has larger shopping groups (0.11). Doubling the number of social ties (Experiment 102) doubles the group size of alters at the activities.
Spatially shrinking the social network (Experiment 104) greatly increases the relative incidence of spatially proximate alters, which are likely to reside at the same address and are encountered at "home" (= same building but not necessarily a household). The group sizes at shopping and leisure also increase, which is consistent with the finding (Table 20, Table 21) that the number of locations ultimately used in this Experiment 104 is smaller, concentrating activities spatially and socially into compartments along the lines of the fractured social network (Section 7.2.1).

Likewise, Experiments 13_4 and 23_4 with the second type of location information exchange, static social networks, and socializing utility, have a less efficient spread of spatial knowledge (Section 7.3.10) and concentrate activities in a smaller number of facilities than the experiments with the first type of location information exchange (in which information superfluous to an agent can still spread beyond to another agent). This results in larger group sizes at the shopping and leisure activities.

The group size of alters in Experiments 5 and 6 are large because of the rule for making new social ties, which allows each agent to make up to one new social tie per activity and iteration. In Experiment 6, the group size is larger at leisure activities, where the agents are rewarded for the number of alters they meet face to face.

The distributions of the number of alters encountered over all activities in a day's plan are summarized in Figure 40. The results can be grouped into 5 categories of similar qualitative outcomes. Experiments 2, 3, 4, 4_4, and 12 comprise the first group with the fewest friend encounters with the lowest variance. These models have the most highly constrained freedom to adapt plans, and/or the least powerful socializing feedback.

The second group of outcomes, Experiments 13_4, 23_4, and 102, have a slightly thicker tail with a higher incidence of encounters with larger groups of friends. The social information exchange mechanism in the experiments *_4 was already shown to have concentrated visits in certain locations more strongly than the other form of information exchange in which more superfluous information could flow through the social network (Table 21), thus increasing the size of groups at activities. Experiment 102 might be expected to have more alters encountering one another, despite only allowing rescheduling of plans, because there are twice as many alters per ego in this model.

The third group of results is the mapping of face-to-face encounters between all agents in the model in the reference cases: Experiments 1 and 7. These are not "alters" but time-space paths which cross one another at activities. The number of agents encountered indicates a maximum value for a model with no social reinforcement of location and scheduling. Obviously, the
number of agents crossing paths at activities is much higher than the number of alters crossing paths at activities in the other models.

Experiments 5 and 6 permit social network evolution by creating social ties between agents who cross paths at activities. The distribution of the number of alters encountering one another in a day is, as expected, between the minimum and maximum distributions of outcome groups 1 and 3.

Finally, Experiment 104 with a fixed social network with a shorter distance between alters results in a lot more friends encountering friends during the activities than the experiments in the first group (Experiment 4 can be directly compared to experiment 104), but fewer than in the evolving social networks that add social ties in face-to-face encounters.
Figure 40  The number of alters encountered at all activities

Experiments 2, 3, 4, 12, 4_4, 22, 32, 42, 52, 72
Experiments 1, 7 all face-to-face encounters (not alters)

Experiments 5 and 6

Experiments 13_4, 23_4, 102
7.4.6 Duration of face to face social interactions by activity type

The duration of face to face contact at leisure activities is the basis of some of the scoring rewards and is an interesting statistic to examine in detail, since the success of agents to coordinate their schedules is the crux of modelling the interaction between socializing and travel.

The total duration of face to face encounters at all activities is summarized in Figure 41, in units of agent-hours. The experiments are grouped by socializing and scoring mechanisms. Recall that the ideal activity durations (those awarding maximum marginal utility) are 12, 8, 6, and 2 hours for home, work, education, and leisure/shop, respectively. In general, the total duration of face-to-face encounters increases for all activity types together; there is correlation between alters in scheduling the leisure activities and the time allocation in other activities. The durations increase with more freedoms given to the agents in their time schedules and decrease in general with freedom of location choice, though this has a coupling with the scoring function. The durations also increase when utility rewards the duration directly, and with increasing utility parameter for face to face meetings at leisure activities. For a given socializing mechanism and utility reward for socializing, the social network density in geographic or social space does not seem to matter.

At the far right are the reference cases showing the duration of overlap of all agents at all activities at the initial locations and when secondary locations may be chosen freely with the standard utility function. The face-to-face overlaps are the shortest of all the experiments, and
because it is known that many agents are co-present (Figure 38), it indicates short duration of co-presence due to no mechanism beyond traffic congestion and other exogenous obligations (work and education schedules) aligning the schedules between agents. Freeing the location choice in Experiment 7 lengthens face-to-face contact, probably due to longer activity durations enabled by the more accessible locations (Figure 37).

At the left are the experiments with standard (social network) secondary location choice and no scoring feedback. These have the lowest numbers of agents co-present (Figure 38) as well as the shortest total face-to-face durations at leisure. The next group rewards face-to-face contact by the number of agents present (parameter value = 10) and also exhibits low face-to-face contact durations relative to other mechanisms. The mechanism of adding social ties during face-to-face encounters (5 and 6) reduces the total duration of face-to-face encounters relative to the other comparable experiments. Not allowing location choice at all (2, 102, 12) but rewarding either the number or duration of face to face contacts increases their duration over the results of the experiments with a comparable scoring function which allowed location choice. The largest total face-to-face durations occur in the experiments with total schedule flexibility (42, 52) and partial flexibility 72 (with 24-hour facilities) which have been shown to be of questionable legitimacy since the travel behavior is suspect.
The distributions of the face-to-face duration of encounters with alters is positive-skewed and can be qualitatively classified into 5 outcome groups (Figure 42):

- least skew, highly precise, low average duration (Experiment 5),
- medium skewed with low average (Experiments 4, 104, 4_4, and 6),
- higher average durations (2, 3, 12, 102, 23_4),
- high-skew, high-average duration (22, 32, 42, 72, 13_4),
- very high skew, high-average durations (52),

whereby only 22, 32, and 13_4 in the final two groups should be considered plausible models.

In general, a face to face utility reward promoting either more friends or more duration with friends does slightly lengthen the duration of social contacts (the second group above versus the third group). What stands out is the where Experiment 6 and 5 have been grouped. Though Experiment 6 has a utility reward for socializing face-to-face (like Experiment 3), its
face-to-face durations at leisure are short like the models 4, 104, and 4_4, which have no utility feedback. Likewise, the duration at leisure in Experiment 5 is in a class by itself: low average and very peaked. This points to the evolving social network mechanism as having the effect of shortening the durations of face-to-face contact relative to similar models without the evolving social networks (6 vs. 3, 5 vs. 4). Rewarding the time spent with friends rather than the number of friends met did not change the distribution of the face-to-face durations, but as noted earlier, this may be due to the parameters chosen for the utility function (2 vs. 12). Clearly, the higher socializing utility parameter leads to longer duration face-to-face contact (22 and 32 vs. 12, and 13_4 vs. 23_4). The different social information exchange algorithms did not make a difference (23_4 vs. 3, 4_4 vs 4), which is true also for the social network density in space (104 vs. 4) and social density (102 vs. 2).
Figure 42  Distribution of face to face duration with friends (friend-hrs) at leisure activities:

Experiment 5

Experiments 2, 3, 12, 102, and 23_4

Experiments 4, 104, 4_4, and 6

Experiments 22, 32, 42, 72, and 13_4
The utility reward for socializing is summarized in Figure 43 for those experiments valuating the duration of friend-encounters. The distribution of the scores ranges from Normal to negative-skewed, indicating that shorter encounters or smaller groups are just as likely or slightly more likely than longer durations or larger groups. Experiments 12 and 22 (32 is identical) differ by the parameter value of the utility function: 10.0 (Experiment 12) vs. 24.0 (Experiment 22), and the difference is not only a lower average, but a much tighter grouping and peaking of the scores in Experiment 12. Again, peaked distributions are taken to indicate homogeneity in the agent behavior. The higher utility reward for socializing enables the agents more flexibility in ways to realize higher net utility, for example by travelling at different times of day in the case of Experiments 12 and 22 (32). Experiment 13_4 shares a similar distribution of scores with Experiment 22 (32), though the peak is at slightly lower utility. Experiments 42, 52, and 72 should not be considered plausible on the basis of the travel behavior because of the lack of enforced activity start times, but the social face-to-face duration score seems consistent with the other experiments.
7.4.7 Face to face group size and time of day

The plots in Figure 44 illustrate the time-space overlap of agents who do not necessarily know each other. The profile in time of the number of other agents co-present with the ego at each activity exhibits some parallels to the time profile of the number of agents participating in activities (Figure 31), with the difference that the popularity of the facilities (Table 21) is convoluted into the measure presented here. Looking at Experiment 1, the number of other agents co-present with the ego (in 38335 activities undertaken) averages 5-7 during the daytime hours with slightly higher variance in the morning and afternoon (and slightly larger groups in the afternoon), and 1.2 overnight (agents living at the same address: not necessarily
households). Compare these values with Figure 38 and the distributions aggregated in time in Figure 40. The time profile varies only slightly across experiments.

The number of agents crossing paths at activities increases in the early morning and afternoon in Experiment 6 and Experiment 3 with socializing utility rewards for face-to-face contact, and the skew of the distribution is strongly positive at these times due to some very popular locations at those times of day (Figure 36 gives some idea of how the knowledge about facilities is distributed).

In Experiments 5 with an evolving social network but no reward, and 7 with free choice of secondary locations and no reward, there is a slight increase in the number of co-present agents in the morning and afternoon activities relative to Experiment 1, but this is hardly distinguishable from the base case of Experiment 1.

A socializing score in general (Experiments 2, 3, 6, 12, 22, 32, 102, 13_4, 23_4) tends to increase the number of hours in the day in which encounters take place, especially if the duration of social encounters is rewarded, and the effect is stronger with a higher reward: agents are still meeting in larger groups until 20:00-21:00 in these experiments, whereas they have stopped encountering each other in recognizably significant numbers in the experiments without a socializing reward. As seen earlier, a higher socializing reward is associated with higher variance in the group size and/or duration of contact. The agents are able to trade off the available socializing utility against a wider selection of locations and accept more scheduling compromises.

A higher social network density (Experiment 102) does not change the time profile of the number of agents crossing paths (compare Experiment 2), at least if the mechanism for improving plans does not include location choice. A spatially localized social network (104) with location choice replanning also has no discernable effect on the time profile (compare Experiment 4).

Relieving the constraint on facility opening times but enforcing late penalties spreads the agent overlap at activities into the evening (compare 72 to 2).

Relieving the constraints on the desired activity start times (42) and late penalty (51, 52) results in a broad choice of schedule for the agents and to them encountering one another in larger groups toward the beginning and end of the time period of the scenario, if there is a social score reward, and to a symmetric peak of group size with respect to time of day, peaking at midday, if there is none (51). This is an important finding because it complements the result of the travel behavior analysis in which these experiments do not yield usefully realistic behavior patterns. The experiments were an attempt to let work start times and durations, which are otherwise fixed exogenously, to emerge instead as a result of agent
interaction, as leisure and shopping schedules do in these models. The constraints in a MATSim configuration make this impossible. However, removing these constraints to different degrees, combined with the weak social network influence at the workplace (no reward for face-to-face working; and even if there was, the social networks used have a low number of work colleagues and/or customers that could interact to provide the utility) makes a model which is untenable. The activities cannot be scheduled. Stronger constraints are needed for those agents who might be free to schedule their work activities in order to arrive at sensible results. Certainly there should be heterogeneous utility in which a majority of the agents have an exogenous work schedule. For those which can schedule work time flexibly, a set of "work" relationships and corresponding utility reward for face-to-face participation at work would be necessary.
Figure 44  Number of other agents encountered at all activities versus time of day

Experiment 1

Experiment 2

Experiment 3

Experiment 4
Figure 44  Number of other agents encountered at all activities versus time of day (continued)
Figure 44  Number of other agents encountered at all activities versus time of day (continued)
Figure 44  Number of other agents encountered at all activities versus time of day (continued)
The distribution in time of encounters with alters (Figure 45) is a subsetting of the distributions in Figure 44, subject to filtering by the social network.
A formal quantitative measure of alters coordinating their schedules to meet face to face has not been defined and calculated. The number of alters meeting at each activity is small, which limits the statistical tools that can be used to describe each meeting (such as deviation from a mean arrival time of all alters at the activity). A qualitative comparison between the relationship between the time of day that alters overlap will therefore describe this coordination.

Figure 45  The number of alters encountered at all activities versus time of day
Figure 45  The number of alters encountered at all activities versus time of day (continued)

Experiment 6

Refer to the number of overlapping agents in a set of day plans (Figure 44) for comparison. The experiments with the standard, fixed, spatially embedded social network of average degree 12 (Experiments 2, 3, 4, 12, 22, 32, 42, 52, 72, 4_4, 13_3, 23_4) show very few social face-to-face encounters with alters (an average group size of 0.1). Those group meetings which do occur are outliers of between 1 and 5 alters, distributed equally throughout the day.

Experiment 102, with double the average number of alters but only allowing shifts in activity times, has a more frequent occurrence of slightly larger groups of alters encountering one another between 14:00 and 17:00.

Experiment 104 with a spatially localized social network and location choice as well as time shifts (but no utility reward for socializing) has even more alters meeting face to face in the afternoon, where ego encounters with 2 or more alters are at least 25% likely between the broader period of 11:00 and 18:00. This experiment with the spatially fractured social network combines group knowledge of proximate locations which are more likely to be frequented by friends with easier accessibility (proximity to locations). Though there is no socializing feedback via a utility reward, there is feedback through the location choice set that is reinforcing repeated meeting with the same alters.

The results of experiments 5 and 6 illustrate the instrumental mechanism of allowing the social networks to add ties based on face-to-face encounters (5 with no utility reward, 6 with utility reward); the social network growth mechanism essentially samples the encounter
profiles of Figure 44. As in Figure 44, the encounters with friends continue into the evening in Experiment 6 with social utility rewards, in contrast to Experiment 5. This face-to-face mechanism of adding social ties is behind the degree distribution, spatial distance between the alters, and the time overlap structures observed in these experiments. It is obviously the most direct (brute-force) method of simulating friend overlap at activities. But behaviorally it is not a good model with which to simulate joint activities because it is not governed by the agent motivations, i.e. the utility function. This separation is required by the MATSim framework between the long-term behavior exogenous to the traffic flow simulation and short-term plan relaxation carried out in the iterations (Section 4.4).

Figure 46 looks just at the number of alters encountered at the leisure activities, versus time of day. The distribution in time of the number of alters co-present at leisure activities reflects the distribution of friends at all activities in Figure 45. The time profile does not qualitatively change in any experiment. The average group size is larger at leisure than for all activities for Experiments 5 and 6. This is because the activities of other type with smaller groups are not included. As summarized in Figure 39, the face to face encounters are concentrated at the leisure activities.
Figure 46  Number of alters present at the same time and leisure activity versus the time of day of the activity

Experiment 2

Experiment 3

Experiment 4

Experiment 5
Figure 46  Number of alters present at the same time and leisure activity versus the time of day of the activity (continued)
Figure 46  Number of alters present at the same time and leisure activity versus the time of day of the activity (continued)
7.4.8 Distances travelled to participate in the different activities

The distance that all agents present at an activity have travelled from their home locations is summarized by activity in Table 23. The values can vary even if locations cannot change in some experiments, depending on which agents overlap at an activity in time. The locations for home, work, and education are fixed in all experiments in the initialization, and represent the distributions in a real population. Since agents cannot "visit" one another at each other's homes, the distance of all agents to the home location is zero. As is seen in the distances from each agent's home to each activity in its plan, the distance from each activity to the homes of all agents present also responds to the socializing mechanisms. When agents are free to choose their leisure and shopping locations, they choose to travel farther to each of these types of activities (Table 20), but they also encounter agents which have come from farther away (Table 23, Experiment 7). Experiments 4, 4_4, and 104 also do nothing more than this, though the choice of location is limited to what information percolates through the social network; here one sees that the agents in Experiment 104 with the spatially localized social networks come from locations closer to the activities than in Experiment 4 with a spatially larger social network. The similar experiment, Experiment 5, in which new social ties can be added at face to face meetings, actually contracts the radius to the homes of agents slightly (as seen Section 7.3.2). Reinforcing visits to the leisure locations where friends are likely to be found, the utility reward (Experiments 3, 6, 13_4, 23_4) combined with location choice, results in agents coming from much farther away to join not only leisure activities, but
shopping, as well. The new leisure location becomes a kind of anchor activity in the plan like home and work, and the new shopping location is likely more efficiently accessed in the plan.

Table 23 The average distance from home of all agents co-present at the different activities (km)

<table>
<thead>
<tr>
<th>Experiment</th>
<th>work</th>
<th>leisure</th>
<th>shop</th>
<th>home</th>
<th>education</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.7</td>
<td>6.9</td>
<td>5.9</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td>2</td>
<td>8.7</td>
<td>6.6</td>
<td>5.6</td>
<td>0.0</td>
<td>6.9</td>
</tr>
<tr>
<td>3</td>
<td>8.7</td>
<td>16.1</td>
<td>9.3</td>
<td>0.0</td>
<td>7.1</td>
</tr>
<tr>
<td>4</td>
<td>8.8</td>
<td>8.2</td>
<td>7.1</td>
<td>0.0</td>
<td>6.9</td>
</tr>
<tr>
<td>5</td>
<td>8.7</td>
<td>7.6</td>
<td>6.8</td>
<td>0.0</td>
<td>6.9</td>
</tr>
<tr>
<td>6</td>
<td>8.7</td>
<td>14.3</td>
<td>8.2</td>
<td>0.0</td>
<td>6.8</td>
</tr>
<tr>
<td>7</td>
<td>8.6</td>
<td>8.2</td>
<td>7.3</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td>12</td>
<td>8.7</td>
<td>6.7</td>
<td>5.6</td>
<td>0.0</td>
<td>6.8</td>
</tr>
<tr>
<td>102</td>
<td>8.7</td>
<td>6.6</td>
<td>5.7</td>
<td>0.0</td>
<td>6.9</td>
</tr>
<tr>
<td>104</td>
<td>8.8</td>
<td>6.9</td>
<td>5.9</td>
<td>0.0</td>
<td>6.6</td>
</tr>
<tr>
<td>1_501F2F</td>
<td>9.0</td>
<td>7.2</td>
<td>6.1</td>
<td>0.0</td>
<td>7.3</td>
</tr>
<tr>
<td>7_501F2F</td>
<td>8.8</td>
<td>7.1</td>
<td>6.3</td>
<td>0.0</td>
<td>7.1</td>
</tr>
<tr>
<td>22</td>
<td>8.7</td>
<td>6.9</td>
<td>5.8</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td>32</td>
<td>8.7</td>
<td>6.9</td>
<td>5.8</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td>42</td>
<td>8.6</td>
<td>7.0</td>
<td>5.9</td>
<td>0.0</td>
<td>6.8</td>
</tr>
<tr>
<td>51</td>
<td>8.7</td>
<td>7.0</td>
<td>6.1</td>
<td>0.0</td>
<td>7.1</td>
</tr>
<tr>
<td>52</td>
<td>8.7</td>
<td>7.1</td>
<td>6.1</td>
<td>0.0</td>
<td>7.1</td>
</tr>
<tr>
<td>72</td>
<td>8.7</td>
<td>6.7</td>
<td>5.6</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td>4_4</td>
<td>8.7</td>
<td>7.2</td>
<td>6.5</td>
<td>0.0</td>
<td>6.9</td>
</tr>
<tr>
<td>13_4</td>
<td>8.6</td>
<td>16.6</td>
<td>8.5</td>
<td>0.0</td>
<td>6.8</td>
</tr>
<tr>
<td>23_4</td>
<td>8.7</td>
<td>13.6</td>
<td>7.6</td>
<td>0.0</td>
<td>7.1</td>
</tr>
</tbody>
</table>

The values are significant to 3 digits after the decimal due to the large population but this significance does not shed light on the conclusions so the values are rounded, here.

7.4.9 Number of agents meeting: group size and distance travelled

The number of agents overlapping at an activity and the distance they travelled to the activity is summarized in Figure 47. This measure is a kind of inversion of the measure in Figure 34 (Distance from agents to their activities) and an indication of the attractiveness of locations
including the socializing opportunity that they afford. The plots show a combination of geography, location popularity (a combination of alter co-presence and social reward with social exchange of information), and accessibility. The results can be qualitatively grouped into 3 general forms:

- Experiments 1, 2, 7, 12, 102, 104, 22, 32, 42, 51, 52, 72, 4_4. These experiments have either no socializing utility and no location choice replanning, socializing utility with no location choice replanning, or location choice replanning with no socializing utility. They have a peak median group size of about 7 with an interquantile range from 3-12 agents, throughout a range of average travel distances between 3 and 18 km. The two social network forms (socially dense, 102; spatially contracted, 104) and the free individual choice of secondary locations (Experiment 7) do not affect the result apart from slightly reducing the spread of the size of the social pool (more seldom outliers in the boxplot).

- Experiments 3, 6, 13_4, 23_4. These experiments have a socializing utility reward and location choice replanning with two different information exchange mechanisms. They show a peak median of 50 or more other agents present per activity over a narrow range of longer distances of about 15-25km. At the peak, there are circa 100 other agents present in Experiment 3 at a distance of 18km and at 25km; 90 in Experiment 13_4 at 20km, and 75 in Experiment 6 at 18km.

- Experiments 4 and 5: These experiments have location choice replanning and no socializing reward. New social ties are made at all activities in Experiment 5. The mean number of other agents encountered in Experiment 4 peaks at 8 at a distance of 14km (interquartile range is circa 15) and in Experiment 5 the peak is 8 at a distance of 10km (interquartile range is circa 20).

In the first group, the agents have a broad range of choice (flat distribution) in which they can trade off a socializing pool with travel distance. This assumes homogeneous agents and no agent-specific socializing preferences. The experiments in the second group exhibit strong distance-dependent socializing pools that result from the scheduling and location choices with the socializing feedback. There are preferred distances to travel, which are longer than in either the first or third groups, and a coalescing of activities into large groups. Though these plots illustrate the entire population and not just the alters present, the effect of socializing within the social networks has effects that spread to the entire population and which change the numbers of agents co-present and the distance they travel to overlap. The experiments in the third group have a much weaker, yet clear peak of preferred, somewhat longer distances, and larger group sizes which are not influenced by utility feedback, but solely by information exchange and individual activity-travel time utility maximization.
Figure 47  Number of non-alter agents present at the same time and place versus the average distance the agents travelled to the activity

Experiments 1, 2, 7, 12, 102, 104, 22, 32, 42, 51, 52, 72, 4_4  
Experiments 3 (6, 13_4, 23_4 are similar)  
Experiments 4 and 5 (similar)

The subset of alters encountering one another versus distance to the activities (Figure 48) qualitatively elicits the willingness to travel to an encounter with a friend, with or without utility rewards. The number of encounters and the size of the groups is very small for those experiments with the reference static spatially embedded social network of degree 12. Only outlier occurrences of groups larger than 0 are observed (Experiments 2, 3, 4, 12, 22, 32, 42, 52, 72). This shows that the averages are not a useful guide to this behavior, since only a small proportion of the population is engaged in face to face meetings with friends.
In Experiments 5 and 6 however, agents meet with friends in groups of median size 2 or more for most distance classes up to 45km distant. The largest groups of 3-4 friends encounter one another at activities 3-15km (Experiment 5) or 6-30km (Experiment 6) from home.

The interpretation that the span of distances versus group size might define a range in which the travel distance is a worthwhile tradeoff for the encounter is not admissible however, because of the issue that these social ties are "awarded", or permitted to come into existence, independently of utility or any other semblance of travel budget or cost.

Likewise, these distributions show that the data may be problematic for inferentially establishing rules of thumb about behavior in the models regarding tradeoffs of distance-vs-willingess to socialize, since for each distance class, there is only a small number of non-zero group sizes relative to the total number of activities.
Figure 48  Number of alters present at the same time, activity, and place versus average distance the alters travelled to the activity

Experiment 2

Experiment 3

Experiment 4

Experiment 5
Figure 48  Number of alters present at the same time, activity, and place versus average distance the alters travelled to the activity (continued)
Figure 48  Number of alters present at the same time, activity, and place versus average distance the alters travelled to the activity (continued)

Experiment 42

Experiment 52

Experiment 72

Experiment 102
7.4.10 Interdependence of distance travelled and agent degree: location choice

The average distance to activities versus degree illustrates the relationship between a measure of the activity space of the agents and a measure of social connectivity. In the co-evolving models, at the aggregate level of the social network, there is strong correlation between the average distance from an agent's home location to its locations in its plan, and to the home locations of its alters. That is, the farther an agent travels to its activities, the farther away its alters live. In the other models, this relationship is fixed. The relationship between the distance to locations and agent degree is therefore similar to the plots in Figure 19 and no graphic is shown for this relationship.

Figure 49 investigates the effects of the information exchange mechanisms and utility rewards on the distance to activities. The information exchange and degree are related in the models (7.2.7). The distributions for Experiments 2, 22, 32, 42, 52, and 72 are the same; for 102 and 104 they are constant; and they emerge for the other experiments. In Experiments 2 (and equivalent), 102, and the Random12 experiment, there is a constant or slightly declining distance to all activities versus agent degree. In Experiment 104 this decrease is more marked. Either the distance to the locations in the agents' plans is independent of the number of alters they have, or the agents with the lower degrees visit locations that are farther away than the locations that agents with higher degree frequent. Thus in Experiment 104, "more friends ==
less travel," which is interesting, but meaningless since this is a random result from the initialization, which relies on similarities between the algorithm to allocate realistic locations of the anchor activities (home, work, education) from the Microcensus and the algorithm to generate social ties, which is strongly distance-dependent. The combined algorithms result in correlations giving the distribution in the figure.

In the cases where secondary location choice lets the distance to activities emerge, the distributions of distance to the facilities versus agent degree are also flat (no relationship) within model sensitivity (interquartile range of the data spread), though some statistically insignificant tendency for agents of higher degree to travel farther is arguable in Experiment 6.
Figure 49  Boxplots of the average distance between the home location of an ego and all of its activities versus the agent's degree

Experiment 2: Static degree 12 spatial social network (same as 12, 22, 32, 42, 52, 72)

Experiment 6: evolving with social utility reward

Experiment 102: Static degree 24 spatial social network

Experiment 5: evolving without social utility reward
Figure 49  Boxplots of the average distance between the home location of an ego and all of its activities versus the agent's degree (continued)

Nonspatial random social network of average degree 12

Experiment 104: Static degree 12 spatial social network spatially contracted

Experiment 3: Static degree 12 spatial social network

Experiment 4: Static degree 12 spatial social network
Figure 49  Boxplots of the average distance between the home location of an ego and all of its activities versus the agent's degree (continued)

Experiment 23_4: Static degree 12 spatial social network

Experiment 13_4: Static degree 12 spatial social network

7.5  Performance

The experiments of this scale (circa 10,000 agents, 5,000 facilities, transportation network of 5,000 links) requires 2GB of memory to run without a social network, and double that for the social network of degree 12 used here. The run times on a single 2.2GHZ processor are summarized for representative experiments in Table 24. The location choice with the first type of information exchange: exchange information with alters and then choose from one's own knowledge, takes 50%-90% longer than a MATSim run without a social network. The spatially localized social network in Experiment 104 ran faster, with acceleration in all steps of the iteration, presumably because there is less information passed across the social network and stored in the agents' knowledge, fewer exchanges, and fewer encounters. The experiments with a social network but no information exchange or location choice run in the same amount of time. This appears to be independent of the size of the social network (average degree 12 in Experiments 2 and 12 vs. 24 in Experiment 102). The experiment (7) with free choice of secondary locations runs faster than a standard run, presumably because there is less traffic congestion and fewer conflicts for slots that have to be resolved on the transportation infrastructure. The experiments with the second type of social exchange: choosing a location directly from an alter's knowledge, ran as fast as or only 12% slower than the reference case without the social interactions. The result for 500 iterations is still a run of 8-9 hours, however.
Table 24  Average run times of the iteration steps and total run time for 500 iterations for representative runs (seconds)

<table>
<thead>
<tr>
<th>Exper. replan</th>
<th>dump plans</th>
<th>mob-sim</th>
<th>time window</th>
<th>spatial</th>
<th>en-counte</th>
<th>info exchg</th>
<th>dissolve</th>
<th>links</th>
<th>net stats</th>
<th>iter 500 iters</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>18</td>
<td>4</td>
<td>34</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>55  7.6</td>
</tr>
<tr>
<td>4.4</td>
<td>12</td>
<td>4</td>
<td>42</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>57  7.9</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>4</td>
<td>48</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>58  8.0</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>4</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>58  8.1</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>4</td>
<td>49</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>58  8.1</td>
</tr>
<tr>
<td>102</td>
<td>6</td>
<td>4</td>
<td>49</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>58  8.1</td>
</tr>
<tr>
<td>23.4</td>
<td>15</td>
<td>4</td>
<td>47</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>65  9.1</td>
</tr>
<tr>
<td>104</td>
<td>8</td>
<td>4</td>
<td>43</td>
<td>4</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>67  9.3</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>5</td>
<td>50</td>
<td>4</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>83  11.5</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>5</td>
<td>49</td>
<td>5</td>
<td>0</td>
<td>19</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>87  12.0</td>
</tr>
<tr>
<td>3</td>
<td>19</td>
<td>5</td>
<td>62</td>
<td>5</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>100 13.9</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>5</td>
<td>57</td>
<td>6</td>
<td>0</td>
<td>20</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>101 14.0</td>
</tr>
</tbody>
</table>
8 Discussion of experimental results in the light of hypotheses and model verification

The introduction of social interactions in the traffic flow simulation model system is a way to provide the agents with activity choices that emerge collectively rather than being centrally directed. The descriptive analysis of Chapter 7 shows the broad scope and many dimensions of the social network phenomena caused by and expressed in activity-travel behavior and geography. The unfamiliarity of the territory makes it difficult to propose measures, and to discover anchoring trends, interdependencies and their strengths. This chapter attempts to summarize the observations from Chapter 7 and order them within the experimental framework, outlined in Chapter 6. These summarized observations are then used to support or refute the hypotheses about the behavior of the coupled model system, also in Chapter 6, concluding the verification (identifying the causes and effects) of the mechanisms used to demonstrate the social networks module. Chapter 9 assesses the module's usefulness in light of the verification conclusions.

8.1 Comparison of the experimental output

Some of the more important comparisons between the experiments are summarized here in the framework of Chapter 6.3 in order to systematically describe the effects of individual and combined algorithms for social interactions in the iterative directed relaxation environment of MATSim.

8.1.1 Baseline sensitivity: random seed effects

The baseline sensitivity of the MATSim model without social networks was investigated on the same road network, but with a lower capacity than the rest of the experiments. The variance of the results is very small, even in a model with relatively high levels of traffic congestion (Section 6.6.1, 7.1 and Table 17). Though the utility surface is constantly changing each iteration as a result of feedback from traffic loads on the transportation network, the iterative algorithm consistently finds the same aggregate temporal results (departure times, travel durations, activity durations) to within fractions of a percent. In models with social interactions, especially with higher road capacities and less congestion, deviations in these aggregate measures of more than fractions of a percent from a base case can be taken to be caused by the social interaction mechanism and not by randomness in the MATSim system.
Baseline social contact: mapping of face to face meetings in a model without a social network

Real social networks are hypothesized to have both a spatial and a non-spatial dynamic supporting their topology. This module permits generating any topologies within the range of totally spatially independent social ties to social ties which are totally determined by spatiotemporal proximity and activity patterns, such as face-to-face meetings. Variations between these extremes are also possible, depending on the definition of "social tie", such as mappings of agents having passed through the same facility in a given time frame (24 hrs, for example) with or without accounting for participation in similar activities.

Three bounding social networks are generated for the purposes of verifying those social networks which are allowed to evolve along with activity-travel behavior in the experiments, and the static spatially embedded social network used for the other experiments. The experiments establishing the bounding social networks are Random12 (nonspatial random social ties), 1_501F2F (mapping of all face-to-face meetings of agents at activities, with no social interactions and the initial activity locations), and 7_501F2F (mapping of all face-to-face meetings of agents at activities, with no social interactions and utility-maximizing secondary activity locations). The face-to-face interaction comprises agents engaged in the same activity at the same time in the same facility (Sections 7.2 and Table 18 and Table 19, as well as Sections 7.2.3, 7.4.1, and 7.4.2).

Random12, the non-spatially randomly generated social network, has an average distance between alters of 36km (independent of agent degree, Figure 19), a broad positive-skewed distribution of dyad distance (Figure 18), a symmetric Poisson-distributed degree (Figure 17), very short average geodesic path length, and extremely low clustering (Table 18). This social network has a large main component which excludes no agents and permits very quick but very homogeneous percolation of information. The degree distribution of the agents is independent of the activities (the number, timing, or location of activities in the plan) and the population density (Figure 24). But the number of social connections between population centers is clearly high. Population density affords the most potential for making connections (highest probability of finding alters) and the ties can be formed based on this consideration alone, because there is no weighting of the tie probability by distance. Thus a nonspatially generated social network has spatial characteristics that depend on an unbiased random sample of a specific geography.

The other extreme maps the potential agent interactions that could be made via face-to-face meetings with the reference configuration (Experiment 1) or the free secondary location choice configuration (Experiment 7) to construct a social network (Table 18). These social networks are completely determined by activity and geography embedding, the only
randomness coming from the initial facility allocations. These networks have average degree of 27 and 26, respectively, and also have large main components consisting of 98% of the agents, with the other components nearly entirely comprised of single agents who did not meet any others face-to-face in the day. The networks are highly clustered and have longer diameters and average geodesic path lengths than the random social network.

8.1.3 Time synchronization with a static social network and a utility reward for socializing

The experiments pertaining to time synchronization are Experiment 2 with utility reward $10 \times \ln(1 + N\text{friends})$, Experiment 12 with utility reward $10 \times \ln (1 + \text{Timefriends})$, and Experiment 22 with utility reward $24 \times \ln (1 + \text{Timefriends})$ (Sections 6.6.3, and Table 20 in Section 7.3).

While the aggregate travel statistics of all three of these experiments are not distinguishable from those in Experiment 1 (Table 20), relative to Experiment 1, these models cause distinctive peak spreading of travel volumes in the evening. The effect of valuating the total face-to-face time (Experiment 12) rather than the number of friends (Experiment 2) is to slightly widen the evening travel peak. A separate, additional peak of travel is generated in the late evening by the higher parameter in Experiment 22 (Figure 30). The travel profiles correspond to large displacements toward the evening of leisure activities (Figure 31): both a later peak (19:00 in Experiments 2 and 12 vs. 18:00 in Experiment 1) and more agents participating in leisure activities at this time. This peak occurs at 20:30 in Experiment 22 with a higher social utility parameter. The agents arrive home much later than without this utility reward. The slight lunchtime peak of shopping participation seen in Experiment 1 is eliminated. Given the same utility parameter, no difference in the aggregate time allocations of the agents can be detected between the utility reward for the number of alters making face-to-face contact (Experiment 2) versus the duration of face-to-face contact (Experiment 12).

Figure 37 gives insight into the duration of individual activities. As expected, the time-scheduling mechanism, combined with a socializing reward, increases the total duration that alters spend face-to-face at leisure activities. These experiments in fact yield the longest face-to-face durations with alters at leisure activities out of all the experiments (among models with sensible boundary conditions): 2.2 to 3.5 agent-hours (Figure 41, Figure 42). Given equal parameter values, the scoring function rewarding the number of alters met lengthens the duration of the leisure activity more than rewarding the duration of the encounters does (i.e. the duration of each agent encounter is less than 1 hour). Increasing the parameter value for the total duration of interactions at leisure (Experiment 12 versus 22) unexpectedly shortens the duration of leisure activities while lengthening the duration of shopping and education activities; activities where socializing is not rewarded in the utility (Figure 43).
explanation is the following: the higher utility parameter in Experiment 22 means that a high socializing utility for the leisure activities can be collected in a shorter leisure activity duration, after which the agent is able to realize even more plan utility by shifting its time allocation to shopping and education when the marginal returns to more time at leisure decrease again. The realized utility at leisure is much higher than in Experiment 12 (Figure 37) despite the shorter durations.

While the time scheduling mechanism has the largest effect on face-to-face duration, the average number of alters encountered face-to-face per ego is the lowest in these time-scheduling models compared to any other social feedback mechanism tested (Figure 39).

In all three time-coordination models, the socializing utility feedback has altered the time allocations at the different activities without changing the aggregate travel statistics over the day (Table 20). Leisure activities are essentially given a priority, which depends on the number of friends there. The resulting schedule shifts in response to the utility reward are specific to the geography and the topology of the social network, the desired schedule for the anchor (primary) activities, and the accessibility the agents have to the fixed activity locations. However, more precise statistical measures of time coordination, such as correlations of arrival times or deviations from a group-average arrival time are not useful since the average group size meeting face-to-face is very small in the models and in reality.

The qualitative evidence points to leisure time schedule coordination occurring by this mechanism, but the result is unpredictable because of the shifts caused in other activities. It is unexpectedly unclear how to relate the behavior to the mechanism or the model inputs. New measures are needed to understand the complex coupling between the social network and its geographical context with the timing of the activity-travel.

8.1.4 Time synchronization with a socially dense static social network and utility reward for socializing

The doubled density social network in Experiment 102 does not change the geographical embedding statistics in Table 19, but of course results in roughly twice the number of friends per unit area (Section 6.6.4). Doubling the number of agents in a social network (Experiment 102 vs. 2) was expected to cause a more noticeable schedule-coordination response in the model than Experiment 2, due to the higher number of friends giving more utility face-to-face at leisure activities.

Indeed, more utility was gained (187 points versus 184 in Experiment 2 and 175 without social network rewards). While no significant differences in aggregate travel behavior could be detected between either Experiment 102 or Experiment 2 and the base case (Table 20),
Experiment 2 did result in noticeable shifts of leisure activities to later times, with Experiment 102 shifting this yet again: the denser social network results in travel volume and activity profiles with an evening peak of travel volume which is higher and which occurs later than the peak in Experiment 2 (Figure 30, Figure 31). The number of non-alters encountered is about the same with the dense social network as with the standard social network (Figure 38).

The average leisure activity duration is however the longest of the experiments which value face-to-face meetings in utility, and the average activity duration of non-leisure activities is the lowest of all experiments (Figure 37). The cause is in the utility reward. Doubling the density of the social network provides roughly $\beta \ln(2)$ more utility for each leisure activity, as far as the number of agents per facility scales with number of friends (which is more or less the case: ). As in all the "duration-only" models, there is a large incentive to remain at a leisure activity as utility is accrued mutually by friends who arrive contemporaneously. The particular value of the utility parameter rewards meeting a friend higher than the time spent with the friend, since the time is generally less than 1 hour (i.e. $N$ friends $X$ $N$ hrs $< N$ friends). The number of alters co-present at leisure in Experiment 102 is marginally higher than Experiment 2 because this brings the most utility (Figure 39, Figure 40), and as a result, the total average face to face duration with friends (agent-hours) at leisure in Figure 41 is slightly higher than in Experiment 2.

The denser social network therefore causes a small but significantly different scheduling response in comparison to the standard social network, in the form of later and longer leisure activities, and slightly more average total time spent face-to-face with alters, who meet in slightly larger groups.

8.1.5 Baseline location choice: unconstrained secondary location choice of all available facilities

Experiment 7 with unconstrained secondary location choice provides the ideal distribution of secondary location facilities for the independently-acting agents, given the specific utility function and the fixed home-work-education locations. The expectations of experiment are summarized in Section 6.6.5.

Experiment 7 makes use of the two shopping facilities that were not frequented in the initialization, but also finds a solution in which the original 4989 facilities in the initialization (e.g. Experiment 1) are the ones used. This shows that the simple proximity- and trip-based algorithm used to allocate the initial activity locations results in a pool of locations in aggregate that is very similar to that which agents would choose were they choosing the secondary locations in order to maximize the plan's utility (Table 21, Figure 36).
However, the average activity-chain lengths and trip distances in this model are 7% longer than those in Experiment 1 (Table 20). Which agents choose which secondary locations is therefore different, dependent on the activity chain utility. Compare 1_501F2F with 7_501F2F in Table 22 to see that 11.3 agents meet face to face at activities in Experiment 1, versus 9.8 in Experiment 7. This is a redistribution of activities in space, since the travel time profiles in Figure 30 hardly differ.

The utility is the same in Experiment 1 and Experiment 7. The average trip duration is 2% longer for the correspondingly longer average trip distance in Experiment 7 (the speed is 58.8km/h vs. 55.7 km/h). Because time (duration) is scored, utility is lost despite the faster travel. Utility is gained in longer leisure durations relative to Experiment 1 (7.9hrs vs. 6.1hrs Figure 37), with little change to the durations of other activities. This does not have to do with any special leisure terms in the utility function (the function is exactly the same in the two experiments) but with a better accessibility of the agents to the leisure activities, which can be exploited to gain more utility from longer leisure activity durations. Late penalties are very similar in the two experiments.

The consistent indication is that the agents which can freely choose secondary locations travel farther in space on faster infrastructure and disperse themselves across the locations more. They gain no extra utility in doing so when the average utility is rounded to the nearest integer. In fact, they lose utility if the average is taken to the nearest tenth of a point. The fact that a more constrained model (route and time replanning) finds a higher final utility than a less constrained model (location, route, and time replanning) indicates a path-dependence or incomplete equilibrium in relaxing the model system which should be investigated further. There is nothing special algorithmically about the type of activity that is re-located. It is likely that allowing location choice replanning for each activity in turn would yield a similar result.

8.1.6 Secondary location choice with exchange of location knowledge on a static social network and standard utility

The social exchange mechanism "type 1" in Experiment 4 permits agents to exchange large amounts of information about secondary activity locations. The information is valued according to how many friends (geodesic distance = 1; i.e. nearest social neighbors) pass the information before it is used in location choice decisions. It was expected (Section 6.6.6) that Experiment 4 would distribute secondary locations in some manner between Experiment 1 and Experiment 7 (with no constraints on secondary location choice).

First of all, no difference in the travel volume or activity participation profiles in time can be detected between Experiments 4 and 7 (Figure 30, Figure 31). While time shifts were allowed
in all the experiments, noticeable differences in time shifts between them do not seem to be
effectuated by differences in the secondary location choice algorithm alone.

In Table 20, the average trip length in Experiments 4 is 9.1km, which is 0.2km shorter than
Experiment 7 and +0.4 more than Experiment 1: very small, but significant, differences,
showing that the result is indeed between that of Experiments 1 (the initialization) and 7
(relaxed secondary location choice). However the distance from each agent's home to its
activities is farther in Experiment 4 than in Experiment 7. In other words, when the secondary
location choice is mediated by the social exchange of knowledge, the plans have shorter
average trip distance between 2 locations in a plan, but longer distance from an agent's home
to these out-of-home activity locations. This suggests that the secondary locations are being
chosen closer to the farthest locations from home: moved farther from home and closer to the
other primary locations (work, education). More space is "consumed" with the social
information-passing mechanism for secondary location choice than with unconstrained
secondary location choice.

The utility score is the same in Experiments 1, 4, and 7. The speed is also high in Experiment
4, with short trip durations, like in Experiment 7. The distribution of the plan lengths, average
trip lengths, and distance to activities is very similar to Experiment 7. The social exchange
mechanism uses all the initial facilities (it cannot learn any additional ones; it could however
have forgotten an initial one, if proven not to have enough utility, but this was not the case).

The mean square error of the number of agent visits to each facility is much higher in
Experiment 4 than Experiment 7 for shop (22 vs. 12) and leisure (49 vs. 32) however,
indicating that the agents are more evenly distributed across locations in Experiment 7 and
more concentrated at certain locations in Experiment 4 (Table 21). A concentration was
expected (see Section 6.6.6) if the information about locations that is provided to agents is
influenced socially (either by weights on the information in Experiment 4, or by direct
querying of alters' knowledge in Experiment 4_4, see below).

Experiment 4 has one of the shortest shopping (leisure) durations of roughly 3.2hrs (5.5hrs)
(true for all the experiments with shopping/leisure location choice and no utility reward for
socializing, Experiments 4, 7, 4_4, 23_4, Table 22).

In the social information exchange mechanism in Experiment 4_4, the agents only choose one
new location from their friends' knowledge per iteration. The spread of knowledge is
hypothesized to be slower and less complete compared to Experiment 4 (i.e. less suitable
location at the end, as evidenced by lower utility scores) because only one new activity can be
learned by an agent per iteration, rather than a number of locations equal to the agent's degree.
While the aggregate travel statistics, time profiles of activity participation and travel volumes,
and utility are the same, there are some differences as a result of the different information exchange mechanism. First, the leisure activity durations in Experiment 4_4 are not extremely short like they are in Experiment 4, and are comparable to those of the other experiments. Another difference is that the number of non-friend agents co-present at activities in Experiment 4_4 is consistently lower than in Experiment 4 for all activity types, but especially for shopping. Experiment 4_4 has slightly larger groups of alters meeting face to face, though the difference between the experiments is small and no statistics are available to determine the significance of the difference. The variance of the number of agents choosing certain shopping locations is markedly higher in Experiment 4 than in Experiment 4_4, showing increased popularity effects for locations in Experiment 4.

The socially-mediated secondary location choice mechanisms illustrated in Experiments 4 and 4_4 do not change traffic volumes or the profile of activities throughout the day. But they do redistribute the agents among the (secondary) activity locations. This redistribution concentrates agents into certain locations relative to Experiment 7, with an unconstrained choice of secondary locations, or the reference, Experiment 1. Since there is no utility reward for face-to-face meetings in these experiments, this outcome is due only to the choice sets of alternative secondary locations (facilities) which vary among the agents because of the social weighting of the information about locations, and the social network topology allowing the information to spread. Allowing a lot of information to spread (Experiment 4) leads to more variation in the number of agents per location, interpreted to mean that some locations are markedly more popular than others simply because they are more likely to be mentioned socially, than if less information is allowed to spread (Experiment 4_4) or if information is uniformly accessible (Experiment 7).

### 8.1.7 Secondary location choice with a spatially contracted static social network and standard utility

Experiment 104 (Section 6.6.7) has a static, spatially contracted social network of degree 12, no utility reward for face to face socializing, and secondary location choice of the first type (exchange of location information with friends, choice of location from own knowledge). It is best compared with Experiment 4 which has the same socializing mechanisms but the standard social network that spreads relationships out more spatially.

The social network is more clustered and divided than either the "standard" social network used to initialized the other experiments, or a nonspatial random network. The clustering coefficient is 42x and 135x higher, respectively. It has 20 components of roughly equal size, the diameter is very large at 17 (vs. 10 for the standard initialization and 6 for a nonspatial random network), and the average geodesic path length is 5.4 vs. 4.0 (standard initialization social network) or 3.9 (for a nonspatial random network) (Table 18). The average distance to
alters is 3.5km vs. 17.6km (standard initialization) or 36.1km (nonspatial random) and the distance distribution is strongly peaked with an exponentially-decreasing right-hand tail (Table 19 and Figure 18). The geographical aggregation of the social network to a 3km x 3km grid shows a distinct lack of longer-distance social ties and higher incidence of shorter ties (Figure 26). The social network is spatially compact but topologically fractured, with strong cores and geodesically distant (tail-like) social chains.

Here, again, the experiment deals with new freedoms and incentives to change locations, and not schedules. Indeed, as in Experiment 4 and 7, there is no discernable difference in the traffic volumes with time (Figure 30), or the participation rate in activities with time (Figure 31). The alters spend the same total time together (Table 22, Figure 41).

Despite the small geographical scope of the social ties, the average distance to activities (and activity chain length) is the same as in Experiment 7 with the unconstrained secondary location choice, and slightly shorter than the comparable Experiment 4 with the standard social network initialization. The agents know about as many different facilities on average as they do in the comparable Experiment 4, circa 22 per agent (Figure 29), although as a population they have forgotten one shopping and one leisure activity relative to the reference case, Experiment 1.

Because the social network is fractured spatially into 20 components, the information flow and the distribution of agents amongst the facilities is different than in Experiment 4 and 7. The number of alters encountered at a leisure activity on average is 3x higher for Experiment 104 than the comparable Experiment 4. The denser social network steers more agents together in leisure activities. Figure 38 shows that the total number of agents present is higher in Experiment 4; however the number of these which are alters is higher in Experiment 104 (Figure 39). Figure 40 shows that this high average value is a result of a long, fat tail in a slower-decreasing exponential distribution of the number of alters encountered at leisure, meaning more frequent larger groups in Experiment 104, i.e. that alters effectively concentrate their activities in time-space to meet in large groups more frequently than in Experiment 4. The distance from the agents' homes to their place of leisure (shop) is 8.2km (7.1km) in Experiment 4 and only 6.9km (5.9km) in Experiment 104, again pointing to a very different distribution of agents to activities: in Experiment 104 more agents choose locations near to their homes.

The utility is one point higher than that of Experiment 4, which is significant according to the baseline sensitivity analysis of Table 17. A look at the time allocations shows why. First, the activity duration is comparable to that of Experiment 4 except for much longer working times (10 hrs vs. 7.5 hrs). Second, the trip travel durations are slightly shorter in Experiment 104
than in 4, due to slightly shorter distances. Thus the extra utility comes from a combination of the shorter travel times and longer working hours.

So, the agents find good secondary locations despite not having (geographically) far-flung social information nets. The average distance and the distribution of distances to the activities is indistinguishable from the other experiments without socializing utility scores, but the ratio of the average distance to all activities versus the average distance to friends is 2.0 for Experiment 104, while it is <0.5 for the other experiments. While it is clear in the latter case that two alters in the social network who live geographically distant from one another can exchange information about locations at opposite ends of the geographic region, providing a large and variable pool of potential locations to choose from, it was not expected that sufficient useful information about geography could propagate on geographically small social networks. Experiment 104 shows that even with extremely localized social networks, agents are able to communicate locations to one another and locate better secondary locations for their activities, at least as effectively as those in Experiments 4 or 7, and beyond the geographic boundaries of their ego networks. The agents' spatial knowledge overlaps sufficiently to pass on leisure and shopping activity locations that are useful to other agents. The ratio of activity radius to social radius may be a key parameter for an effective spatial information exchange, depending on the clustering, spatial distribution of alters, and the rules for information flow: in $k$ geodesic jumps through the social network, information might be expected to flow up to $k$ ego net radii, if all the alters live along a line. Successive jumps could quickly cover large territory, even if the ego nets are small. On the other hand, the initial social network with longer spatial social ties might not be any more effective at helping agents find their "best" secondary locations than a spatially contracted social network, because the occasional long-(geographic) distance social connections which may be helpful for communicating social goods like culture, language, values, and job opportunities, are less helpful for passing material goods or other services related to the physical world, like optimizing secondary location choice when travel time is at a premium in the utility function; a traveller in Zurich who is looking for a coffee shop near his work does not consult a friend in Honolulu about his neighborhood. Social ties much (geographically) closer to home will have more useful information.

Because of the tendency for agents to meet in larger groups with the spatially localized social network, it would be desirable to repeat 104 with scoring rewards for socializing (i.e. experiment 3 with a spatially more proximate social network) to see if the secondary location choice contracts to the geography of the social network, or remains cast over a broad geography, as in Experiments 104 and 3 (see 8.1.8).
8.1.8 Secondary location choice with exchange of location knowledge on a static social network and a utility reward for socializing

Experiment 3 combines Experiments 4 and Experiment 2 with knowledge that is localized socially through two kinds of socially-mediated exchange and utility feedback of $\beta \ln(1 + \text{NumFaceToFace})$ with $\beta = 10$ (Section 6.6.8).

The secondary locations are expected to be distributed in spatial concentrations greater than that in Experiment 4 to concentrate alters together; the number of friends encountered is expected to be higher; and the willingness to sacrifice utility on travel (congestion, distance) would be higher. This experiment results in high traffic volumes, and large shifts in activity time and location that were not realized in either Experiment 2 or Experiment 4.

In Table 20, the average trip length and activity chain length in Experiment 3 are 13.6km and 12.5km, respectively, which is 4.1km and 3.0km longer than Experiment 4. The secondary locations are not concentrated spatially (or linearly, along the activity chain) close to the anchor activities (work, education, home) of the agents, but a much longer distance away. The trip durations are also circa 7-8 minutes (80%) longer per trip, and travel speeds are extremely slow, (circa 14km/h slower).

The travel volume profiles with time show that with the utility reward for face to face meetings, the travel peaks are both higher and broader: more agents travel at once, and they do so for a longer duration, because they receive compensating utility at the social activity (Figure 30). The peak of participation in leisure activities shift to a later time (19:30 vs. 18:00) with a broad range of start- and end times and durations, and shopping is also distinctly later (19:00 vs. 17:30). (Figure 31).

In Table 21 there are one or two fewer unique leisure and shopping locations retained by the population in these experiments, showing a concentration or economization of activities in fewer locations than Experiment 4. The standard deviation of the number of activities (agent visits) at leisure facilities, is 143 for Experiment 3 (relative to 49 for Experiment 4). This shows that there is a high number of visits to a low number of facilities, rather than a more even spread of visits over facilities (Experiment 3 is similar to Experiment 6, as in Figure 36).

These experiments have the longest shopping durations and the longest leisure durations next to models which directly reward the time spent engaging alters (Figure 37).

A big difference in the number of non-alter agents encountered at leisure activities is seen in Figure 38. Indeed, this is the signature result of the mechanism in Experiment 3. By far the largest pools of potential social contacts at leisure activities are generated in this model (as
well as the similar model 23_4, below), some 10x as many as in other models (except Experiment 6, Section 8.1.10).

The time of day of socializing is also different for Experiment 3. The reward for face to face socializing directs agents (again, potential pools of social contacts rather than actual alters) to encounter one another in significantly larger groups in the afternoon, whereas this indication is not strong in other models, with exception of model 6 which is based on the same socializing mechanism (Figure 44). This is further evidence that the mechanism forces a coordination of activities in time.

While the distance from home of all agents co-present at the secondary activities is higher for Experiments 4 and 7 than in Experiment 1, it is much higher for Experiment 3 with socializing utility: 16.1km for leisure and 9.3km for shopping relative to 8.2km and 7.1km for Experiment 4 (Table 23). The large groups of 50 or more which cause such a high potential socializing pool to form in Experiment 3 arrive from a distance of circa 15-25km; without the socializing reward (Experiment 4), this pool is approximately 8 agents from a distance of 14 km (Figure 47). Not only does the socializing utility cause agents to travel farther and at more congested (popular) times, it causes, or enables, them to encounter more agents in bigger groups from farther away.

Experiment 23_4 repeats Experiment 3 with a different information exchange interaction: a location is chosen from an alter's knowledge rather than offered by each alter. The results can be compared directly to both Experiment 3 (to assess the different socializing mechanism) and Experiment 4_4 (to assess the effect of the scoring function, Section 8.1.6). The notable differences are summarized here. In Table 20, the average trip length and activity chain length in Experiment 23_4 are 13.6km and 12.5km, respectively, which is the same as Experiment 3 and 4.3km and 3.2km longer than Experiment 4_4. The standard deviation of the number of activities (agent visits) to leisure facilities, is 143 for Experiment 3 (relative to 49 for Experiment 4) and 122 for Experiment 23_4 (43 in Experiment 4_4). The distance from home of all agents co-present at the secondary activities is 16.1km (13.6km) for Experiments 3 (23_4) for leisure and 9.3km (7.6km) for shopping relative to 8.2km (7.2km) and 7.1km (6.5km) for Experiments 4 (4_4) (Table 23).

While the agents do not choose secondary locations quite as far from home in Experiment 23_4 as in Experiment 3, this same trend is seen between Experiment 4_4 and Experiment 4 and is therefore explained by the socializing interactions.

Qualitatively the same effects on location distance and the size of the pool of agents co-present at activities are observed in Experiment 3 and 23_4, pointing to the combination of the scoring function and location choice as being more important for learning about new
alternative locations and choosing them for plans than the particular social exchange mechanism.

It should be noted that the location choice mechanism with a face-to-face utility reward increases the number of agents encountered at activities, as does a spatially contracted social network. A major difference between the two is the much higher average distance to activities in the former, and the distance that co-present agents live from their activities.

The utility reward for face to face socializing is a very strong mechanism to encourage agent co-presence, in conjunction with location choice. Adaptation to the plans to increase co-presence occurs as much in space as in time, and both agents which are friends as well as those which are not friends in the social network have emergent coordination in their activities which overlap in time-space. The number of these agents who are actually alters is also highest in Experiment 3 (Figure 39) except for those models which permit new face-to-face ties to be added to the social networks (5 and 6).

The reward for meeting face to face as implemented is a time, location, and activity-specific attribute of the facility: credit in the utility function is given only when two agents meet all of these criteria. It serves as a variable for location choice as well as scheduling (and socializing). The "social interaction" utility term could also be defined more broadly to uncouple it from time or activity, as needed. Including only the facility and the activity might simulate a tendency for agents to perform a particular activity at a particular place. Considering only the facilities that alters visit in common, regardless of time-of-day or type of activity, is a more traditional facility attribute.

8.1.9 Secondary location choice with an evolving social network and no utility reward for socializing

This section compares Experiment 5 with the ability for agents to add a social tie each iteration at each activity, with Experiment 4 which has a static social network. Each experiment has an exchange of activity locations before replanning the activities with secondary location choice / scheduling /routing, and there is not utility reward for face to face socializing.

The emergent social network in Experiment 5 (Table 18) has a clustering coefficient 136 times higher than a nonspatial random graph of the same average degree and 38 times higher than the reference social network used in the initialization. Nearly all agents are in a single large component (97.8%) with a relatively long average path length 4.8 and diameter 12. The average dyad distance is 2/3 as long as in the initial social network, and distance to all alters and just over half as long. So this social network is spatially more compact than the initial
social network (these measures are still three times longer than the spatially compact social network in Experiment 104, however). The degree distribution has an exponential tail (Figure 17) characteristic of a growing social network. The link removal algorithm tends to reduce the tail thickness of this distribution.

The influence that the initial social network has on the final social network depends on how many social relationships are created each iteration (this equals the number that will be removed each iteration in order to keep the graph average degree constant). If the number of new social links is large each iteration through spatial encounters, and this number is removed again each iteration, the likelihood that social links remain from the initialization after 500 iterations is effectively zero. In practice, there are no surviving initial links after 50 iterations (earlier iterations were not written out so it was not recorded when the last initialized links were removed). In the final iteration, approximately 35% of the social ties were added during the iteration, and 65% were more than 1 iteration old (Figure 27). The average age of a social tie is 1.98 iterations (Table 18).

The distance to activities is roughly the same as in the reference case without a utility reward, Experiment 4 (7.5km vs. 7.6km). However the ratio of the distance to all activities versus the distance to all alters is 0.7, i.e. the spatial extent of activities is 70% the spatial extent of friends' homes. This ratio is 0.4 for Experiment 4 (Table 19), where the alters were spread more than twice as far away as the activities, and less than half that of Experiment 104, where alters were twice as close as activities. Experiment 5 thus finds a middle ground between the amount of area subtended by ego nets versus the area subtended by activity spaces. The secondary location choice mechanism, which weights the choice alternatives by the number of friends passing the information, consistently yields a radius of activities of ~7.5km regardless of the socializing algorithms and social networks used.

The distribution of the average distance to egos' alters for Experiment 5 is qualitatively of the same shape as that in the initialization (Figure 18). However the distance is distributed quite differently if analyzed with respect to degree. The more alters an ego has (higher the degree) in Experiment 5, the farther away the alters live, on average, from the ego. This is in contrast to the initialization, which has a strong distance-decay with degree (Figure 19, explained in Section 6.4.1).

The socializing mechanism drives agents with more activities to make more social ties (Figure 28), which is a hard-coded and strong determinant on the outcome of how the social network is related to activity-travel. This is an undesirable byproduct of models which cause agents to establish social ties upon meeting face-to-face: those agents with more activities will gain more alters. If one wants to study patterns of socializing as a function of how mobile (the
number of activities) an agent is, this algorithm will not be helpful because it mechanically enforces a certain relationship between degree and the number of activities.

The socializing results in the exchange of much new information about locations. Figure 29 shows that agents with more alters know about more distinct locations (compare to Experiment 4 with a flat distribution which is independent of degree).

No obvious differences in the travel volume or activity participation rates with time are ascertainable (Figure 30, Figure 31). The social network topology and/or whether the social network is evolving or static does not matter in the time profiles (Section 7.3.5). The utility function (and the model's boundary conditions) determines the time profiles of travel and activities. The average utility in each experiment is the same also, despite the differing social networks in Experiments 4 and 5.

There are also no significant differences between Experiment 5 and Experiment 4 in the aggregate distributions of distances to activities, distance as a function of agent degree, the number of unique facilities chosen, the variation in the number of agents choosing each facility, and the distance agents travel to activities versus the size of the group present.

The activity durations are longer in Experiment 5, especially for work, where Experiment 5 realizes the longest work durations of all plausible experiments (10.3 hrs vs. 7.5 hrs in Experiment 4, Figure 37). The ability to earn more utility for activities is afforded by faster travel to closer locations, resulting in lower disutility for travelling.

The number of alters at activities is of course very high for Experiment 5, since new social ties are created at activities. The number of alters encountered on average at leisure activities is 3.2 vs. only 0.1 in Experiment 4 (Table 22). The number of other agents in the potential social pool is similar to Experiment 4 for all activities, but the ability of agents to make new social ties from among this pool raises the number of agents which are alters at each activity (Figure 39, Figure 40). Raising the size of the pool of agents at activities is accomplished by a face-to-face utility reward.

The time overlap with alters is shorter at 1.5 agent-hrs vs. 1.8 agent-hrs (Figure 41, Figure 42). The agents overlap in time nearly the same way they do in Experiment 4, except that a higher proportion of them are alters who can socially interact.

Allowing the social network to evolve iteratively is an experiment which attempts to relax the social ties within the geography and the activity plans. The agents with more alters have a longer distance to their alters because the probability of meeting new friends is a function of how many activities one has, which is a very strong influence in the mechanism. No change in the time profiles of travel or activities relative to Experiment 4 was detected. The total utility
is the same as in Experiment 4, but the duration of activities is slightly longer because they are located closer to the agents' homes, and the trips have a slightly shorter duration. The average geographic ego net radius of the relaxed social network is longer than the distance from the agents' homes to the activity locations, but not as much longer as in the experiments with static social networks.

The effect of permitting the social network to evolve was less pronounced than expected in terms of travel behavior and location choice. Indeed in these initial investigations, it appears as though the topology of the social network is not important in the travel behavior if elements of the social network are not explicitly included in the utility function. The emergent social network is unlike the initializing social network or the non-spatial random social network. In particular, the spatial embedding resulting from this algorithm generates a similar average distance to friends, but distributed such that those agents with the most friends maintain them at longer distances.

8.1.10 Secondary location choice with an evolving social network and a utility reward for socializing

Experiment 6 compared to Experiment 3 provides a similar picture to the comparison between Experiment 5 and Experiment 4 (Section 8.1.9). The mechanism to add social ties capitalizes on the number of agents attracted to the same facilities through the social exchange of information employed in Experiment 3, allowing agents to make friends among the new social pools (Section 8.1.8). The social pools are made large in Experiment 3 by the utility reward for socializing. Establishing new ties among these pools builds a social network of extremely high clustering coefficients (Table 18), exponentially distributed distances to alters (Figure 18), and positive-skew degree distribution (Figure 17). The agent degree is a clear function of the number of activities the agent has in its plan (Figure 28).

Many of the characteristics of the time allocations, distances to facilities, and the number of agents encountered in Experiment 3 are retained, despite the social network adding and removing a large numbers of ties each iteration. As in Experiment 3, as a result of social network feedbacks, the agents prefer a handful of leisure facilities and gather there in large numbers (the capacities of the facilities were not enforced) (Figure 36). The time profile of traffic volume and participation in activities is similar between Experiment 3 and 6 (Figure 30 and Figure 31). The face to face socializing mechanism tends to make locations closer to the agent's homes and their average trip distance smaller (Table 20). The agents tend to travel in peak congestion and decrease their average speed, increase their disutility from travelling, and they compensate this with face to face utility rewards. Once again this illustrates the overriding importance of the utility function for determining the final travel behavior.
As seen in Experiment 5, while the evolving social network causes rapid and voluminous information exchange between the agents, the effect of this exchange on transportation behavior is secondary. The utility function has more bearing on the travel behavior. Adding contacts face-to-face enters this equation by increasing the utility at leisure activities, and the feedback slightly shortens the duration of this activity. More importantly, the tie-adding (and removal) mechanism generates a social network with geographic ego net radii roughly equal to the radius of daily activities, and distance to friends that increases with ego degree, with or without utility rewards for socializing.

The effect of the social network on transportation behavior is not large, compared to the effect of the utility function. The effect of the transportation behavior, the geography, and the link removal algorithm completely determine the social network, however.

8.1.11 Boundary conditions of the utility function and the facility opening times

One of the possible uses of a social interaction package in agent simulation is to allow the agents to schedule their activities endogenously, freeing the model system from exogenous boundary conditions to be able to observe natural relaxation of things like flexible work hours and the reactions to various road pricing schemes.

The effects of relaxing some of the constraints on the utility function and the facility opening times were assessed in conjunction with agent interactions and utility formulations that could substitute for these constraints. Experiments were made with time rescheduling and re-routing only (the traditional MATSim search space): socializing parameter value, minimum duration limit, late penalty, latest permissible start time, and facility opening times.

First, Experiment 22 (vs 12) increases the score parameter to test the sensitivity of the social interactions models to the parameter that rewards the total face-to-face time. The activity locations and the social networks are the same across the experiments. In the output, the emergent trip travel time is the same. Relative to Experiment 12, the leisure participation in Experiment 22 peaks later in the day. The travel volume has a distinct third volume peak after the work commute home that corresponds to this shift. The work and education activities are of significantly longer duration in Experiment 22 versus 12. The number of alters encountered at leisure and the duration of the encounter (agent-hrs) are both greater in Experiment 22 (Table 22, Figure 41). The effect of increasing this parameter is as expected: to make agents spend more time with one another at non-scheduled leisure activities. The mechanism succeeds in coordinating leisure behavior and even creates a new demand peak, even in conjunction with strong penalties for arriving late to scheduled activities (work, education) which tend to anchor the schedules to the exogenously determined desired participation times.
Experiment 32 is identical to Experiment 22 and is the test of the extent to which MATSim's utility function inadvertently causes activities which have no duration, but to which the agents must travel anyway. This is known to be possible but its frequency and significance have not been studied. Experiment 32 examines the possibility in the context of socializing, to ensure the validity (or at least plausibility) of the models presented here. It sets all "minimum activity durations" parameters to zero in order to immediately generate positive utility for durations longer than zero, and to avoid driving the model to schedule activities with zero duration (Section 6.6.11). There is no difference between the output of Experiments 22 and 32. Solutions with zero-duration activities is not a problem in the experiments examined here.

In Experiments 42, 51, and 52, the penalties for arriving late to activities aside from "education" are either 0 or are not enforceable due to there being no "latest possible" start time set. Experiment 42 is identical to Experiment 22, but with undefined latest starting times for activities. Experiment 52 is identical to Experiment 22, but with no late penalties assessed. In Experiment 51 there is no social network and no hope of controlling schedules via social interactions. It is a control against which the social interaction mechanisms can be measured. Despite having strict opening and closing times for the facilities, it does not yield usefully realistic behavior patterns. There are no morning and evening travel peaks, there is insufficient travel volume throughout the day, and impossible attempts by nearly all agents to travel at once during the final second of the day. Without the late penalty or any other indication of activity prioritization, there is effectively no way to evaluate one schedule versus another, as only the activity durations and avoiding traffic congestion matters. As a result, the iterative re-scheduling of appointments is not directed in any particular way: the start times of the first activities of the day are random, and the other activities follow according to the durations and travel times that bring the most utility.

The utility feedback for face to face socializing in Experiments 42 and 52 is only awarded for leisure activities. It succeeds in dividing leisure activities into slight morning and afternoon peaks, relative to Experiment 51, in which all leisure takes place in the afternoon (Figure 31). Some small effect of endogenous scheduling of leisure activities is therefore seen in the experiments with utility rewards for face to face meetings.

Including only leisure activities in the utility feedback is clearly insufficient for enforcing consistent endogenous scheduling of all activity types, but the experiments do show some effect on the agents' ability to coordinate the scheduling of leisure activities, despite very low average numbers of alters at leisure activities. The approach shows promise for endogenously regulating/coordinating face-to-face meetings of any activity type, like flexible work schedules. To replace exogenous schedules, the utility feedback and social networks would have to be more specifically tailored to the work activity, school activity, and home activity, meaning multiple kinds of social relationships and multiple kinds of reward for socializing,
depending on the context of the activity, before it would work to direct a plausible day schedule endogenously.

Experiment 72 is also identical to Experiment 22, but it does not enforce the facility opening times (they are available 24 hours/day). It is an experiment in the feasibility of endogenously determining facility opening times (see Section 7.3.4). Such a model might be desirable in determining latent demand for, say, evening shopping activities, etc. Comparing Experiment 72 with Experiment 22, and Experiment 52 with Experiment 22, it is evident that the main force driving the familiar two-peak travel volume pattern in time is the penalty for arriving late to a pre-determined scheduled work activity in the morning (leaving it when it has provided sufficient utility occurs in both models); i.e., the successful scheduling of agents in MATSim relies on avoiding the late penalty (which produces the morning peak) and accruing the maximum marginal utility for the activity duration (which produces the afternoon peak). The difference between Experiment 72 and Experiment 22 is the presence of the facility opening times. They clearly play a secondary role in shaping the schedules of the agents.

The ability for a well-specified MATSim experiment to produce plausible transportation behavior without exogenously imposed facility opening times means that it could be possible to let at least a portion of agents' activity start times and desired durations be determined endogenously based on social network mechanisms, without exogenously fixing them or the facility opening times. This would open a door to modelling flexible working hours, for example. But this work thus far does not suggest a clear method for how to do that in a way which can be calibrated to data. It makes sense to leave school hours exogenously fixed. Certainly the existing capabilities of MATSim to define agent-specific utility functions should be utilized to define behavioral rules in which a majority of the agents have an exogenous work or school schedule. For those which can schedule work time/place flexibly, a set of "work", "household", and "friend" relationships and corresponding utility reward for face-to-face participation at work, home, and other activities would be necessary.

8.2 Model verification: support and refutation of the hypotheses

8.2.1 Social ties

The effects on social network topology of embedding in activity travel and geography were expressed as hypotheses in Section 6.2.1.

1) a) The mapping of face to face meetings of agents in the relaxed state of a basic scenario (Experiment 1_501F2F) was expected to rarely exhibit agent overlaps, and thus many components and low average degree. The clustering coefficient was not predicted because it
is very sensitive to the agent overlap. In fact (Table 18 and Table 19), the average degree is high (= 27.2) relative to the standard initialization used for other experiments (degree = 12.0) and it is extremely positive-skewed (Figure 17), giving a much higher-than-Normal occurrence of agents who overlap with high numbers of other agents. The clustering is very high at 173 times that of a random network, and nearly all agents are in the main component of the social network. The graph has a rather long average geodesic path length and diameter, meaning that information flow would not be as fast as in a random graph. The agents overlapping at activities live 25% farther away from one another than the distance they live to the activities at which they meet.

The very high overlap (degree, clustering) comes as a surprise and is the result of the low number of leisure and shopping locations relative to the work and home locations, and the corresponding high number of individuals overlapping at these activities in time and space. Certainly "social networks", or mappings, of this type, would be helpful epidemiological tools.

Data pertaining to hypotheses b) the mapping of locations visited and activities pursued during the day (bipartite network of activity/locations and agents); and c) the mapping of locations visited during the day (bipartite mapping of locations and agents) was not analyzed yet.

2) The initial spatially embedded social network has a fat-tailed degree distribution and higher-than-random clustering because geography was used to introduce preference into choosing alters with which to forge social ties. The effect is stronger in the spatially compact social network of Experiment 104. Fat-tailed (positive-skewed) degree distributions emerged automatically when the probability of adding a social tie between agent $i$ and agent $j$ is a function of distance between the agents $i$ and $j$ which is not an inverse function of the population distribution with distance. This happens because the population is distributed in 2 dimensions: assuming an evenly distributed and unbounded population in space, the number of possible agents with which to social ties can be made at radius $r$ scales with $r^2$. If the probability of making a social tie scales at a lower (higher) rate than distance $d^{-2}$, the probability is higher (lower) of choosing an agent that is closer than distance $d$ than beyond distance $d$: a preferential choice resulting in a positive-skewed degree distribution and higher chance of two agents mutually knowing a third.

3) An equilibrium dynamic social network can be generated with the MATSim iterative directed system relaxation framework with a link removal rule that retains a constant average degree. The link removal algorithm used chooses links randomly for removal. This favors removing links from high-degree nodes, and shifts the degree distribution slightly to toward a Poisson distribution, away from a positive-skew distribution. The clustering coefficient
quickly tends toward a constant value in these evolving networks. Thus the generated graphs are deemed to be in equilibrium.

4) Including the mechanism which introduces friends of friends was not attempted, the focus during the research turning instead to travel behavior analysis rather than the generation of social networks.

8.2.2 Insights into travel behavior microsimulation

The hypotheses are stated in Section 6.2.2.

1) Evidence can be found of coordination of activity schedules: The evidence for coordinating activity schedules is qualitative but strong. The experiments which reward face-to-face socializing but do not allow location choice result in the most marked elongation of the duration of the activity where social interactions are rewarded. The experiments combining location choice with this reward result in the most marked coordination of both time and location of the activities and traffic volumes. The experiments which permit location choice as well as social network evolution (via face-to-face meetings) exhibit a similar pattern of schedule coordination, where the activity duration at which socializing is rewarded is shorter but more alters are present. Experiments without the utility reward for co-presence (i.e. which attempt to coordinate agent's co-presence only through shared alternative choice sets) do not exhibit schedule coordination.

2) A utility bonus for face to face socializing means that agents will trade off longer distance trips or longer travel times for the opportunity to meet with friends: This is exactly what the agents do; any utility reward compensates utility losses in the other dimensions of the agents' behavior, and the agent can sacrifice utility in these dimensions to gain in another. A reward for a particular activity can result in a plan which exhibits paradoxical "desire" to travel during a congested period, for example. This behavior, which leads to the familiar morning and afternoon travel demand peaks, is precipitated by "late arrival" penalties for scheduled activities in a standard MATSim configuration. The same effect is seen with social reinforcement for participation in leisure activities, which results in an emergent, endogenous scheduling of the leisure activity, using a utility reward instead of a penalty to enforce it.

3) Information exchange on a social network about secondary activity locations that reinforces choice sets results in secondary location choice that is better than a random allocation of locations (as measured by utility) but sub-optimal compared to free secondary location choice that is not socially-mediated: The experiments in which social networks mediate the secondary location choice result in the same or higher utility as the experiment in which the secondary locations could be chosen freely. The average distance to the secondary
locations chosen is not significantly different. However the number of agents choosing each facility does differ very slightly, as does the travel volume. The presumption is that the solution of secondary location choice is not unique, and either path-dependent with different combinations of location popularity and travel volumes giving similar utility, or that the system was incompletely relaxed in the experiments compared.

4) **Rewarding face to face social interactions in conjunction with free choice of secondary locations results in a different distribution of destination choice relative to the same secondary location choice algorithm without extra social utility rewards (exchange of location information alone):** The utility reward for face to face socializing has a strong influence on the distribution of the agents in the locations. The location choice mechanisms alone only cause slight differences. The combination of location choice and utility reward causes immensely popular locations to emerge. If no new friends are permitted to emerge from this social pool, only the number of agents present at the locations is high; if however, agents are permitted to choose new friends from this pool, both the number of friends as well as the number of agents is concentrated into certain activity locations. However, which locations are chosen, and their popularity, is not strongly affected by the friend-making mechanism.

5) **Rewarding face to face socializing at activities results in fewer facilities being chosen, and in these being more popular, than in a scenario without this reward:** Indeed some facilities are much more popular than others, but nearly the same number of facilities are chosen. It was expected that certain facilities would become abandoned, but this occurs only seldomly. Nearly all facilities are used by the agents because of their favorable accessibility to certain agents and because these agents may not be able to profit from socializing at other locations which are farther away, due to their 1) not having a social contact there, and/or 2) their not being able to schedule an activity of this type at a time when social overlap can occur, due to other schedule constraints.

6) **In models rewarding face-to-face co-presence, activities will be of shorter duration:** The activities in which the face-to-face co-presence is rewarded are shorter, whether the utility reward is related to the number of alters or the total duration of contact with the alters. However, a comparison of the strength of the two scoring mechanisms is not clear-cut because of the differing units, without calculating elasticities over parameter sweeps, which has not been done.

### 8.2.3 Social geography

This discussion refers to the social geography hypotheses in Section 6.2.3.
1) Social network effects on activity travel behavior can be distinguished from the effects of the geographic framework like population density, density of facilities, and accessibility measures: It is not straightforward to detect social network effects in travel behavior. Certainly the socializing mechanisms caused changes in the activity-travel behavior output. In the examples shown here, prior knowledge of the social dynamic was necessary in order to properly attribute the observed changes to the output which were caused by the socializing mechanisms. These can take the form of information exchanges, for example, which cannot be revealed in the travel behavior output of the simulation. The popularity of certain activity times and locations which was observed in some experiments was known to be a result of social utility feedback. If this was not known, one may also just as well attempt to explain it by some other attribute of the facility in an alternative model. Suggestions are made in Chapter 9 for how to use the agent model to support social interactions research in real-world observations.

2) In a social network which evolves with the activity-travel plan, or which mediates secondary location choices, the geographic extent of the ego networks is larger than the geographic extent of the locations used by the agent in the agent's final location choice: The evolving ego networks created by face-to-face meetings at activities do subtend more space (longer average distance between an ego and all of its alters) than the activities an agent participates in during a day (average distance between ego and its activities). This relationship holds in the two experiments with evolving social networks, though the ratio between the two distances differs, depending on whether face-to-face socializing is rewarded in the utility function (both distances are longer if there is a positive reward).

3) In a static social network, the geographic extent of the ego networks may be larger or smaller than the geographic extent of the locations used by the agent in the agent's final location choice: This is also true; the ego networks are spatially larger than the activity plans with the standard spatially-embedded social network, with or without the reward for face-to-face socializing. However this distance relationship is inverted in the experiment with the static social networks which are spatially very compact, but which mediate location information. These are smaller in geographic extent than the distance to the activities that the agents finally choose, when there is no feedback for socializing in the utility score. The main influence on the location choice is the utility function. The experiment with the spatially compact social network and location choice was not repeated with socializing score rewards.

4) In a social network which evolves with the activity-travel plan, an agent's spatial knowledge covers a much larger geographic area than the area subtended by the ego nets: The spatial analysis of all agents' knowledge was performed on the ensemble runs and then determined to be too time-consuming a calculation for the remaining experiments. The evidence from these experiments indicates that an ellipse encompassing 90% of the locations
in each agent's Knowledge structure covers the entire geographical scenario for most agents. The agents not socially connected to the main component with at least one social tie at any geodesic distance do not learn about new locations; the others retain information about the entire 100km x 100km region.

5) In a social network which evolves with the activity-travel plan, the link probability distribution vs. distance between agent home locations is exponentially decreasing: This hypothesis is true, whether there is a utility feedback for socializing or not. The average distance to alters of an individual ego increases with the ego's degree, however, which is the inverse of what is initialized in the "standard" initial social network used.

6) Social network effects can substitute for exogenously fixed variables and edogenize behavior in the model: social norms can replace exogenously established desired start time; desired duration; activity location. Activity locations become more popular when agents receive a utility reward for being co-present with alters. Removing the ability to punish agents for not upholding an exogenous schedule results in untenable models which cannot be scheduled, if no substitute mechanism is provided for valuating the appropriateness of start times. Utility rewards for co-presence with alters at leisure activities is effective at coordinating agent' schedules to meet at these activities. The indications are therefore, that, if structured carefully and comprehensively, with appropriate peer groups and reward structures in the utility, it is likely that activity start times could be endogenously generated by a social network mechanism ("joint planning"). Given the appropriate utility functions with decreasing returns to duration, but an endogenous "target" duration, the duration of these activities could also be successfully socially mediated. These mechanisms will work; calibrating them might be difficult and is an issue that has not been looked into in depth.

7) In a social network which evolves with the activity-travel plan, the agent's degree scales with the number of activities (will depend on the rules for making friendships face to face): The mechanism of adding a social tie with probability $p$ at each face-to-face meeting between agents at each activity explicitly couples the number of activities in an agent's plan to the number of social ties it can make. This mechanism is realistic in the sense that the social pool is larger the more out-of-home activities an agent undertakes. It is not realistic in that the MATSim activity plan only represents a single day's travel. Agents not travelling much on this day may have more activities on another day. Likewise, this evolution algorithm does not account for social relationships which do not result from face-to-face encounters during this specific day, such as long-term relationships and out-of-household family relationships (nor does it explicitly account for households).

8) In a social network which evolves with the activity-travel plan, the agent's degree scales with the size of the facilities visited (number of other agents at the facilities) it visits (related
to weighted accessibility measures of facilities): It is clear that the popularity of the facilities and the time slots increases if there is a utility reward for face-to-face encounters with alters. Not only alters, but all agents, are attracted to the facility and the time slot, whether or not new social ties can be made. If new social ties can be added at this socializing opportunity, one for each agent co-present, the degree of each agent scales proportionally to the number of agents at the facility. The link removal rule which maintains a constant average degree then tempers this growth somewhat. But high-degree nodes are still associated with high-popularity activities (== facility + activity type + time slot).
9 Conclusion: Emergence, usefulness, simplification

The thesis accomplished four things:

1) Modify the MATSim travel behavior simulation toolkit to include the capacity for users to program social interactions

2) Develop measures of socializing, geography, and activity travel that resolve their mutual influences on one another.

3) Demonstrate how various social networks, social exchange, and social scoring mechanisms influence and are influenced by the travel behavior simulation

4) Identify the most promising combinations for simulating the various dimensions of joint activity travel within MATSim.

The complicated analyses of the model output in the dimensions of activity patterns (scheduling, locations, and type of activities), travel behavior (timing, volumes, speeds), and social interactions (social network topology, evolution, information flows) forced setting modest goals for the initial evaluation of this modelling approach. Emphasis was placed on measures and tests which could feasibly show recognizable signatures of the influence of socializing on the activity-travel relaxation, and which could also offer insight into how to structure future simulations of joint travel. These were not tests of algorithms that could generate realistic travel behavior, but algorithms which could create controlled experiments in which expected signatures of their influence could be detected and described.

9.1 Highlights of what was observed

The statistical ensemble was a very small sample without social networks. It is the first published test of MATSim with different random seeds. It shows that the iterative targeted relaxation yields in repeatable results within very small error of 0.5%-2% for the most sensitive values like traffic speed. This is smaller than the variation of speed in observations. The model is reliably stable and it is assumed that variations in the output greater than this magnitude from the base case indicate the effect of a changed input or mechanism relative to the base case.

The module can instantiate a set of social relationships between agents from a number of standard network generation algorithms from the literature, from an existing social network that is read in, and/or supplement the algorithms with embedding in geography by using the
information from the agents' plans files. The memory required for the simulation increases proportional to the degree of the social network, which is very large compared to a road network object. In the examples this increase was up to 100%.

The social network generation algorithms produce the expected distributions if they are not geographically-embedded. But any geographical embedding (dependence of social ties on distance) will skew the distributions of the graph statistics and make the social network strongly dependent on the geography of the population (the shape of the perimeter and its characteristic length; or a variable population density).

The statistics calculated in the system are established broadly to cover as many dimensions of the output as thought likely to exhibit social network effects. Indeed, in some experiments the effect was only visible in a single output measure. But the analysis of the high dimensionality is tedious, despite the statistical indicators.

A range of emergent behaviors was detected in the output: locations and times of day becoming more popular, the size of groups meeting face to face increasing with utility rewards, shifts in travel volume profiles due to social influences, changes in activity durations. These behaviors are consistent with the driving mechanisms: similar mechanisms of similar magnitude resulted in similar changes to the output.

The activity-travel behavior is sensitive only to the utility function; changes in the alternative choice set caused by information exchanges on the social networks, different social network topologies, and different social network geographies could not move the model from the equilibrium that was achieved in models without social interactions. As expected, utility rewards for socializing result in sacrificing utility elsewhere (tradeoffs).

The tests of the utility function are also a first systematic record of this type of investigation in MATSim. In a model that does not reward face-to-face contact for work or school, but relies on exogenous schedules to force the agents' attendance, the penalties for being late are decisive in forming the rudimentary basis of the familiar two-phase demand peak in time. The facility opening times contribute a secondary influence on the peaks in the travel volume (and activity participation rate) with respect to time. These tests led to the insight that the penalties could override social network influences in models which might reward co-presence at work and would otherwise result in shifted demand peaks, but for reasons endogenous to the model. The penalties would have to be lifted in order to carry out this experiment, however this risks losing control of the time profile of the activities; most agents after all are subject to an exogenous work or school schedule. The recommendation is to enforce the late penalties only for those workers and students/pupils with fixed schedules but not for those with flexible schedules whose plans are to emerge with the social interactions.
9.2 Most valuable results

The work reinforced the warnings of Manski (1993) that recognizing the social network influence is not possible without knowing it is present and without knowing the form of the underlying social structures and the strengths of their influences. The popularity of certain activity times and locations which was observed to be a result of social utility feedback may also just as well be explained by some other attribute of the facility. Without prior knowledge of the social dynamic, a researcher cannot say for certain why certain agents behave in concert. In trying to explain such behaviors in real data, when information on socializing components of the behavior are not available, a researcher will typically need to resort to explanations using latent variables. However a simulation like this one which constructs similar behavior based on a social interaction mechanism can serve to support an argument that the real-life observations also occurred with a similar mechanism. Appropriate "networked actor" explanations of the observed behavior could then be attempted.

9.3 Limitations to the method

The statistical ensemble analyses was a very small sample. With a run time of 12 hours for the base MATSim model without the social networks, it is unlikely that large ensemble samples can be run of this model system. This is an unfortunate and insurmountable obstacle to understanding the complexity of the model system. A substitute is a sensitivity analysis to a handful of random seed parameters, but this is not very convincing to doubters (consumers) of the model output.

The limitations of the treatment of time in MATSim make it difficult to establish biographical and other social ties outside of the 24-hour period of the activity plans that affect activity travel behavior during the plan. The iterations to relax the activity-travel plans are not suited to simulating long-term relationships. The algorithm to add and remove social ties must be separate (uncoupled) from the utility function, since the utility cannot be taken to mean anything but the value of the day's plan within the day (or whatever time horizon is being planned, which will definitely always be a horizon in MATSim that is short relative to changes in activity rhythms, infrastructure changes, or changes in facility opening times, for example). The coupling would occur in microscopic representation of the utility of social ties: the long-term investment in a tie, the penalty for losing one, etc. Subsequently the risk and benefit of certain activity travel versus another to an agent's social well-being is not aptly captured. Nothing in the MATSim modelling environment can help in establishing these social ties except a household object which would at least give household ties automatically (it would be assumed that these exist cost-free). Establishing the long-term social links (i.e.
those which do not depend on the face-to-face contact in the representative daily travel pattern of the particular model) is an exercise external to MATSim.

The social network evolution as implemented to permit agents to meet face-to-face at activities was intended to retain part of the long-(geodesic) distance social relationships from the initialization, while augmenting them with a portion of relationships that are tied to the activity patterns. However, it has the undesirable effects of simply increasing the number of alters at these activities, forcing egos with many activities to have higher degree, and causing the average distance to alters to also scale with degree. Other alters from previous iterations are deleted on average within 3 iterations, and no links from the initialization remained after a very short time.

After the effort of adding and removing social links to create a social network with strange geographic characteristics, the travel behavior and the distribution of agents (whether friends or not) at facilities are very similar to the corresponding experiments with a static social network with properties more familiar in the literature. The extra time overhead for the evolution calculations make it more sensible to calculate these evolving social networks only once, if they are desired at all, and to use them again as static networks by reading them into the initialization. It is very likely that (statistically) nearly identical social networks could be generated by a bipartite network formulation with agent nodes and activity nodes (one node per activity and time bin, say hourly). One could then add random links in a second step if desired, up to the target average degree, for example with the distance-dependent probability for a social contact, as in the "standard" initialization. The link probabilities in the initialization could be modified with a homophilic index to fit some observed distributions (Watts et al., 2002). Another possible solution might be a modified link-removal algorithm which leaves a certain core of long-term relationships (e.g. from the initialization).

Enforcing the capacities of facilities was not attempted (Horni et al., 2008). This would not only temper the location choice results, it would also affect which agents could establish social ties in the experiments with social network evolution, and thus cap the degree distributions. This may be an alternative way to achieve an equilibrium social network while still using an algorithm which permits randomly adding social ties between agents meeting face-to-face.

In the MATSim version available for this work, agents could not visit homes of other agents for leisure or any other activities. This would be very important to include in more realistic studies.

New statistical and graphical representations would be desirable to more quickly analyze the output and to relate changes to the inputs.
9.4 Additional elements to improve and use in the future

True joint intentions were discussed in Chapter 5. The time- and space coordination of activities between alters is, in the end, indistinguishable from plans which were arranged between the agents by some interaction. But the process in MATSim is not a joint planning undertaking. There is no expression of intent and agreement to meet. One agent cannot be an element in another agent's day plan. Agents do not negotiate the changes to their plans (note: this element could be built into the framework of the social networks module as a "replanning" strategy). If one agent's randomly modified plan resulted in not meeting an alter at an activity, which in a previous iteration had rewarded both agents for meeting, both agents' utility would suffer. The plan which was successful earlier as a joint plan would now have a low score and would stand less of a chance of being chosen for execution the next iteration. This would risk both of them trying something completely different the next iteration. If the changes are small, near the end of the relaxation, the agents might meet each other again, but if not, the random independent mutations of plans could interrupt an emerging coordination pattern. Future implementations of "joint planning" should take care not to demote plans from "active" to "inactive" status unless other agents are not affected. However explicit joint planning, like agents negotiating an activity location or start time, is less important to represent in the model, since the results in this work show that this can emerge by the targeted iterative relaxation of MATSim.

Can calibrated geographic social networks be generated? A way to calibrate or to steer the emergent socializing behavior would be desirable. One method to influence the number of agents meeting face to face at an activity, while sticking to the utility maximization model approach, might be to introduce an equation for the number of alters present at an activity which has a "desired" target number, at which marginal utility is maximum, like the current utility function is specified with respect to activity duration. A distribution of the likelihood of meeting a certain number of agents at an activity in a given day, like the continuous curve generated by Frei and Axhausen (2008) in Section 2.2.4, may give such a target distribution. This kind of a model uses a structure similar to those models already in use in MATSim for activity duration. However implementing this type of optimization where agents must meet each other would rely on a functioning joint replanning and plan selection module as described above. It may also require definitions of specific types of relationships (work associate, in-household relative, out-of-household relative, etc.) in order to allocate the activity frequency to the right social links. A full-fledged Monte Carlo simulation based on Bayesian inference on joint distributions as in Butts (2000) could be used to generate validated fixed social networks with activity chains and locations (based on distances), provided an input dataset is detailed enough.
As the travel behavior is sensitive only to the utility function, in future models, attention should be paid to specifying the appropriate social network ties that influence the activity-travel, and the utility function associated with these social ties. Specific social roles such as accompaniment (travel leg only) versus joint activities should be defined, as well, as in Schlich and Axhausen (2004) (Section 2.2.4).

Finally, the code should be modularized with the goal being an out of the box implementation for any experimenter who does not want to modify existing Java classes, but wants to implement his own by extension of an API. This is a typical cycle of programming which is the reverse of ideal: the implementation is first tested, then the abstract classes or interfaces are extracted from it, and a new modular implementation replaces the original.

9.5 What is gained or lost by using social networks?

At the cost of roughly double the computing memory and 10%-50% more CPU time, and a lot of disk space for the iterations of social networks, the module provides the expected laboratory for social interaction experiments in a single-day simulation of activity travel. The biggest gain is the decentralized (emergent) ability to schedule activities in time and place.

9.5.1 Is realism important?

Reasonable assumptions have sufficed to produce credible qualitative signatures of expected sign (> , < 0) and order of magnitude in the model response. Two coupled models representing opportunity costs as transportation experts perceive them, and cascades of influence across a social network, are sufficient to illustrate the point that a range of plausible behaviors could emerge given different reasonable assumptions. Parameter fits and precise mappings of ego nets and individuals in the agent population to a real population are not necessary in order to learn from the coupled model system.

The focus of the analysis is not on the quantitative result but on how well the submodels produce behavior to reflect the intent of the researcher. The utility reward for socializing at leisure activities extends the duration (distance to) of leisure trips as intended and as observed. The distance-decay function for the probability of social ties did not result in strongly confining social relationships in space as expected; this may have something to do with the spatial density of the population, which was sparse due to the small scenario size required as a result of constraints on the calculating capacity of the computer. The models of socializing attempting to exchange spatial information did not exhibit particular heterogeneity of knowledge, at least as far as the locations visited by the agents. Agents were still able to find locations that in the aggregate mirror those that they would have chosen had they free choice.
However it is more likely that a model of socializing in which not regular face-to-face locations, but regular (or increasing) face-to-face encounters, regardless of location, would be a more realistic model. However, that agents with more friends knew about more locations is a direct result of this exchange algorithm, and another important corroboration of reality. The positive-skewed (exponential tailed) degree distributions with high clustering are small worlds that are realistic topologies for many documented human interactions. That these can result from purely geographic terms entering the tie probability function indicates, as other papers have also noted, a particular effect of geography on human contacts.

For purposes of increasing exposure of the model system to a planning audience and/or actual applications, increasing realism with fits to data is indispensable. Sections 9.1 and 9.3 discuss some of the challenges of using social network data, and Section 9.6 summarizes specifically what would need to be done for a calibrated simulation with a social network.

9.5.2 Do they improve understanding, computational speed, or results of travel-demand simulations?

The social network mechanism does provide emergent activity and travel coordination in time and space, which is not possible with a centrally-governing mechanism.

9.5.3 For what kinds of travel research questions could they be most useful/ignored?

Social networks are not necessary for determinations of regularly-scheduled or highly inelastic appointments like most people's work and education obligations. They are useful for emergent scheduling of flexible activities (location, time, duration) and they will be useful for mode choice. They proved less useful for the communication of information in the experiments described here, but this could change if different information (or influence, virus, etc.) were the focus of investigation on the social network.

9.6 What, if any, kinds of relationships should be explicitly accounted for in transportation planning?

Models with proxy variables for social interaction, e.g. groupings of agents based on behaviors or personal characteristics (Schlich and Axhausen, 2004), were not compared with the social network models. If these categorical variables are available, they are (because of their lack of specificity) probably more accurate than a social network. However they will not provide the same kind of emergent coordination from microscopic interdependencies. The overhead for running a social network model is not incredibly high, and the gains of explicit social interactions are potentially quite enriching to the result. The mechanism of face-to-face
utility rewards certainly has enough power to coordinate activity schedules and locations in a reasonable runtime overhead. Employing a social interaction to regulate flexible schedules between agents based on the mutual utility of joint participation will work.

But the specification of the social network will always be the difficult task. The generation of the social network must be specified in the context of the activity travel that is to be governed socially. The experimenter must answer the questions:

- Which decisions are the agents making (activity type, location, start time, duration, mode)?
- Which social relationships are important for the decisions (boss/employee, teacher/student, peers)?
- And which consistent utility function will be used in conjunction with social activity and activity-travel (co-presence, duration of co-presence, decreasing returns, minimum durations)?

The social network, utility function, and the replanning relaxation must be structured interdependently and to the purpose of the activities being modelled, if the model is to have meaning. This means overcoming the issues of long- and short-term utility rewards, the conditions for initiating and dissolving social contacts, social priorities, etc.

One of the most immediate and relevant social structures that could be incorporated with some credibility is the household. Adamowicz et al. (2005) provide a review paper of different models of household decision making for tasks such as the allocation of travel and financial resources, time obligations, etc., and rule and/or utility-based models that have been proposed to represent them in simulation.

### 9.7 Validation

For anyone to want to use (apply) a model, it has to bear resemblance to reality (Bankes, 1993). When MATSim is coupled with social networks, the same validation of activity duration and start time, distance travelled, travel speed or duration, and traffic volumes need to be performed and calibrated that need to be done in regular MATSim runs. This is an ongoing challenge for every complex multi-agent simulation package.

However, the social networks output provides additional information about the duration of face-to-face contact, group size, mutual travel to the activity, group composition (ego net, statistical cluster, joint activity travel), social network topology, and geographic embedding of the social network, which would have to be compared to reality. Datasets are lacking and it is unlikely that they will be gathered to the detail they are available in the model.
It is realistic to expect that a global social network which influences travel behavior will never be sampled to sufficient geodesic distance and in sufficiently high numbers of individuals and travel-behavior situations to be a useful validating reference. Frei et al. (2008) report a cost of 110 CHF per useful response in their survey of trusted alters. Carrasco (2008b) reports interview durations of over 3 hours per respondent. The results of the snowball experiment of Kowald et al. (2009) are not in, but the researchers encourage participation with a cash incentive of 20 CHF per mailed questionnaire (whether it is returned or not), over and above a considerable recruitment and interview cost. These survey expenses of time and money mean that the samples will be small and have few distinctive structures, such that they could have been drawn from many possible network topologies. Models estimated on these samples to provide extrapolations of large social networks will likely have low explanatory power. Depending on the underlying structures which have been overlooked in the survey, the resulting extrapolated social networks could be dramatically incorrect in any number of measures (Borgatti et al., 2004; Paez et al., 2008).

Most likely, researchers wishing to justify social network models with data will have to take individual parameters separately from samples of social networks and travel behavior datasets and use them in a composite reference in the different dimensions of the problem.

A wiser use of the multi-agent-based social travel modelling system is to use consensus assumptions appropriate for the problem at hand, and to concentrate on questions of how the social and activity-travel system react to the assumptions. Attempts to validate the social networks beyond coarse statistical ranges provided by the literature on social networks is not advised. Indeed, the MATSim system is itself experimental and only validated in a handful of measures in selected contexts. If validation of a particular integrated social network model is desired, statistical inquiries should be performed at aggregate levels similar to the data that is being collected.
10 References


Schlich, R. (2004) 12 weeks of leisure travel survey, *Travel Survey Metadata Series*, ETH Zürich, Zürich,


---

A 1 Curriculum Vitae

**Jeremy Keith Hackney**

Phone: +41 1 633 33 25  
http://www.ivt.ethz.ch/people/jhackney/index_EN  
E-Mail: hackney@ivt.baug.ethz.ch  

**ETH-Hönggerberg**  
Inst.f.Verkehrsplanung/Transportsysteme  
8093 Zürich  
SWITZERLAND

**Education**

**PhD, expected 1/2010**  
Inst. for Transport Planning and Systems (IVT), Swiss Fed. Inst. of Technology (ETH), Zurich, Switzerland.  
Dissertation: "Agent modelling of social networks and travel behavior."  
Advisers: Kay W. Axhausen (ETH Zurich), Eric J. Miller (U. Toronto)

**Paul Seydel Fulbright Fellow, 9/1998-9/1999**  
Interdisciplinary Center for General Ecology (IKAÖ), University of Berne, Berne, Switzerland.  
Project: "The Use of Lightweight Electric Vehicles and Effects on Mobility."

**S.M. Civil Engineering (Technology and Policy), 1997**  
Massachusetts Institute of Technology, Cambridge, U.S.A  

**Non-Degree Language Studies 1994**  
Università per Stranieri, Perugia, IT.

**B.S. Engineering Physics and Magna Cum Laude Physics, 1993**  
University of Colorado, Boulder, U.S.A.

---

**Research, Teaching, and Work Experience**

**Scientific Researcher**, 5/2003-present, IVT, ETH Zurich, Switzerland.


Associate Scientist, 2002-2003, Environmental and Societal Impacts Group (Value of weather information for the electric power industry), National Center for Atmospheric Research, Boulder, U.S.A.

Associate Scientist, 2000-2002, Climate and Global Dynamics Division (Computer simulation of longwave absorption), National Center for Atmospheric Research, Boulder, U.S.A.

Associate Scientist, 1998, High Altitude Observatory (Computer simulation of global atmospheric tides), National Center for Atmospheric Research, Boulder, U.S.A.

T.A. Alternative Fuel Vehicle Technology and Policy (Graduate Level), 1997, Technology Policy Program, Massachusetts Institute of Technology.


T.A. The History of Technology Seminar, 1992-1993 (3 Semesters), Herbst Program for Humanities in Engineering, University of Colorado at Boulder.

Recent Projects

Travel impacts of social networks and social networking tools (Volkswagen Foundation), 2008-present
Co-P.I.: Iterative (snowball) sampling of social networks and mobility patterns in Zurich.

Travel behavior in a dynamic spatial and social context (Swiss National Fund), 2006-2008
P.I.: Agent-based simulation approach to time use in social and geographic contexts.

Traffic Monitoring in Canton Zurich (Canton Zurich), 2004-2005
P.I.: Mapping GPS speed measurements and timetable data to a road network model and extrapolating to non-measured regions using regression with spatial correlation terms.

European Dataset for Long-Distance Travel (ETIS BASE) (EU), 2003-2005
Assistant: Assembly of Pan-European dataset of transportation supply (sea, air, road, rail) and LOGIT modelling of transportation mode and destination choices at high spatial resolution.

Peer Review Publications


Hackney, J., M. Kowald, and A. Frei (to appear) Applications of Social Networks in Transportation Science (chapter is in German), in Stegbauer, C. and R. Häußling, (eds.) Handbuch Netzwerkforschung, xxx-xxx, VS-Verlag, Wiesbaden.


**Peer-Review Conference papers and Presentations**


**Other Relevant Working Papers and Project Reports**


Hackney, J.K., M. Bernard and K.W. Axhausen (2006) Predicting link speeds with floating car data, presentation at the Urban Data Committee, Transportation Research Board Annual Meeting, January 2006, presentations of the Transportation Planning and Systems Group, 163, IVT.


**Invited Presentations**


**Awards and Funded Proposals**

**Swiss National Fund**, 2005 "Travel behaviour in a dynamic spatial and social context: Modelling the interdependence of social network interactions and spatial choices", CHF 90,288, 2 years.


**Conferences Organized**


**Professional Affiliations**

International Association for Travel Behavior Research Student Member

World Conference on Transportation Research Special Interest Group 11: Transport and the Environment

Transportation Research Board Climate Change Subcommittee

Transportation Research Board Time Use and Activity Patterns Subcommittee

**Personal**

U.S.A. citizen, born 1971