On the Approximation of Special Instances of Minimum Topic-Connected Overlay*

Monika Steinová

Department of Computer Science, ETH Zurich, Switzerland, monika.steinova@inf.ethz.ch

Abstract. The design of a scalable overlay network to support decentralized topic-based publish/subscribe communication is nowadays a problem of great importance. We investigate here one such design problem called Minimum Topic-Connected Overlay. Given a collection of users together with the lists of topics they are interested in, connect these users to a network by a minimum number of edges such that every graph induced by users interested in one common topic is connected. It is known that this problem is APX-hard and approximable by a logarithmic factor. We focus here on hardness properties of some special instances. We study the problem where, for each topic, there are at most three users interested in it. Surprisingly, we show that even with such strong restriction, the problem stays NP-hard and it inherits the approximation hardness of the well-known vertex cover problem.

1 Introduction

Nowadays, many internet applications support many-to-many communication based on sharing content: publishers publish information through a logical channel that is consumed by subscribed users. The situation is often modeled by publish/subscribe (pub/sub) systems. These systems can be classified into two categories. In content-based pub/sub systems the channels are associated with a collection of attributes and the messages are delivered to a subscriber only if their attributes match user defined constraints. Each channel in topic-based pub/sub systems is associated with a single topic. The messages are here distributed via channels to the user by his/her topic selection. There are numerous implementations of pub/sub systems, for details see [1–4, 9–11].

In our paper, we focus on topic-based peer-to-peer pub/sub systems. In such a system, subscribers interested in a particular topic have to be connected without the use of intermediate agents (such as servers). There are many aspects studied here (see [5, 8]). One may need a small diameter of the overlay network to minimize the total time in which a message is distributed to all subscribers. If an (average) degree of nodes in the network is minimized, the subscribers need to maintain (in average) a smaller number of connections. We regard here the minimization of the overall connections in the system, i.e., the minimum number

* The research was partially funded by SNF grant 200021-132510/1.
of edges the overlay network needs to have to satisfy the demands of subscribers. A small number of edges may decrease the maintenance requirements as the network needs only a small number of connections. All messages sent via one edge of the network may be aggregated to a single message and thus amortize the head count of otherwise small messages.

We study here the hardness of Minimum Topic-Connected Overlay (Min-TCO) that was introduced by Chocler et al. in [5]: Given a collection of vertices (subscribers), a set of topics and a vertex interest assignment, we want to connect vertices in an overlay network \( G \) such that all vertices interested in a common topic are connected and the overall number of edges in \( G \) is minimal. The problem is APX-hard and approximable by a logarithmic factor ([5]). We focus here on special cases of Min-TCO, namely on instances where, for each topic, there are at most three users interested in it. We show that, surprisingly, the problem stays NP-hard and inherits the approximation hardness of the famous minimum vertex cover problem.

2 Preliminaries

In this section, we define basic notions used throughout the paper. We use standard graph theory terminology. Let \( G = (V,E) \) be a graph and \( T \subseteq V \), then by \( G[T] \) we denote the graph induced by vertices \( T \) on \( G \).

The problem we study here is a subclass of Min-TCO and it is defined as follows.

**Problem 1.** Min-\( d \)-TCO is the following optimization problem:

**Input:** Set of vertices \( V \), a set of topics \( T \) and a vertex interest assignment \( F : V \rightarrow 2^{|T|} \), such that there are at most \( d \) vertices interested in a topic \( t \), i.e., \( |\{v \in V \mid t \in F(v)\}| \leq d \).

**Feasible solutions:** Set of edges \( E \subseteq V \times V \) such that, for all \( t \in T \), the graph \( G[\{v \in V \mid t \in F(v)\}] \) is connected.

**Costs:** Size of \( E \).

**Goal:** Minimization.

Let \((V,T,F)\) be an instance of Min-\( d \)-TCO, let \( S \) be some feasible solution and let \( t \subseteq T \). We define an edge interest assignment of \( S \) as \( F_S(s) = F(u) \cap F(v) \) for any edge \( s = \{u,v\} \in S \). For both vertex and edge interest assignments we define their “inverse interest assignments” \( F' \) and \( F'_S \) as follows: \( F'(t) = \{v \in V \mid F(v) \cap t \neq \emptyset\} \) and \( F'_S(t) = \{e \in S \mid F_S(e) \cap t \neq \emptyset\} \).

Recall that a vertex cover of a graph \( G = (W,E) \) is a set \( C \subseteq W \) of vertices such that every edge of \( G \) is incident to at least one vertex from \( C \). The minimum vertex cover problem (VC) is the problem of finding a vertex cover of minimum cardinality in a given graph. There is a well known 2-approximation algorithm for VC, it is NP-hard to approximate it within a factor 1.3606 ([6]) and VC is not approximable within a factor better than 2 unless the unique games conjecture fails ([7]).
3 Hardness results for Min-3-TCO

**Theorem 1.** If there exists a polynomial time $\alpha$-approximation algorithm solving VC, then there is a polynomial time $\alpha$-approximation algorithm solving Min-3-TCO.

**Proof.** Let $I_{3TCO} = (V, T, F)$ be an instance of Min-3-TCO. W.l.o.g. assume that $|F'(\{t\})| = 3$ for all $t \in T$. (Otherwise the two vertices interested in a topic $t$ have to be connected in any solution. Therefore they can be contracted to a single vertex and then the problem can be solved on the smaller instance.)

Observe that the three edges between three vertices interested in a common topic form a triangle. In order to guarantee that these three vertices are connected in a solution, we have to pick at least two out of these three triangle edges.

We employ this fact in our reduction to VC. We reduce our instance $I_{3TCO}$ to an instance $I_{VC} = (W, E)$ of VC as follows. The vertices in $W$ represent all possible edges in Min-3-TCO, i.e., $W = \{w_{xy} \mid x, y \in V \land F(x) \cap F(y) \neq \emptyset\}$. Two such vertices are connected by an edge if and only if the two corresponding edges of $I_{3TCO}$ belong to the same triangle, i.e., $E = \{\{w_{xy}, w_{xz}\} \mid F(x) \cap F(y) \cap F(z) \neq \emptyset\}$.

A feasible solution of Min-3-TCO on $I_{3TCO}$ is a feasible solution of VC on $I_{VC}$ and vice versa. Moreover, observe that this reduction is one-to-one, i.e., edges in the feasible solution of $I_{3TCO}$ correspond to vertices in the feasible solution of $I_{VC}$ and vice versa. Therefore we can conclude that the reduction preserves the approximation ratio.

**Corollary 1.** There is a polynomial-time 2-approximation algorithm solving Min-3-TCO.

**Theorem 2.** If there exists a polynomial time $\alpha$-approximation algorithm for Min-3-TCO, then there is a polynomial time $(\alpha + \varepsilon)$-approximation algorithm for VC, for any $\varepsilon > 0$.

**Proof.** Let $I_{VC} = (W, E)$ be an instance of VC and let $A$ be an $\alpha$-approximation algorithm for Min-3-TCO. For any $\varepsilon > 0$, we transform $I_{VC}$ to an instance of Min-3-TCO, we apply algorithm $A$ on it, from the approximate solution of Min-3-TCO we create a solution of $I_{VC}$ and we show that such solution cannot be too expensive with respect to the optimal solution of $I_{VC}$.

For the instance $I_{VC}$, we create an instance $I_{3TCO} = (V, T, F)$ of Min-3-TCO with $|W| + k$ vertices ($k$ will be fixed later) as follows.

\[
V = W \cup \{p_i \mid p_i \notin W \land 1 \leq i \leq k\},
\]

\[
T = \{t_{uv}^i \mid \{u, v\} \in E \land 1 \leq i \leq k\},
\]

\[
F(w) = \begin{cases} 
\{t_{wx}^i \mid \{w, x\} \in E \land 1 \leq i \leq k\} & \text{for } w \in W \\
\{t_{xy}^i \mid \{x, y\} \in E\} & \text{for } w = p_i
\end{cases}
\]
Observe that the instance contains $k|E|$ topics. The vertices interested in a topic $t_{i}^{e}$ ($e \in E$) are exactly the two incident vertices of $e$ in VC and a special vertex $p_{i}$ ($1 \leq i \leq k$). A small example when path of length 3 is reduced is shown in Figure 1.

![Fig. 1. Reduction of a simple path to an instance of Min-3-TCO. Gray vertices are special vertices, dashed lines denote the possible edges in a solution of Min-3-TCO.](image)

We apply algorithm $A$ on instance $I_{3TCO}$ and denote its solution by $Sol_{3TCO}$. We divide the solution to levels. Level $i$ is a set $S_{i}$ of edges of $Sol_{3TCO}$ that are incident with special vertex $p_{i}$. In addition, we denote by $S_{0}$ the set of edges of $Sol_{3TCO}$ that are not incident with any special vertex $p_{i}$. Therefore $Sol_{3TCO} = \bigcup_{i=0}^{k} S_{i}$ and $S_{i} \cap S_{j} = \emptyset$ ($0 \leq i < j \leq k$).

We claim that, for any $S_{i}$ ($1 \leq i \leq k$), the set of non-special vertices incident with edges of $S_{i}$ is a feasible solution of VC. This is true since, if an edge $e = \{u,w\} \in E$ is not covered, none of the two edges $\{u,p_{i}\}$ and $\{w,p_{i}\}$ is in $S_{i}$. But then the vertices interested in topic $t_{uw}^{e}$ are not connected together as always at least 2 edges are necessary.

Let $j$ be chosen such that $S_{j}$ is the smallest of all sets $S_{i}$, for $1 \leq i \leq k$. We construct $Sol_{VC}$ by picking all non-special vertices that are incident to some edge from $S_{j}$. Denote the optimal solution of VC and Min-3-TCO for $I_{VC}$ and $I_{3TCO}$ by $Opt_{VC}$ and $Opt_{3TCO}$, respectively. We use the fact that $k \cdot |Opt_{VC}| + |E| \geq |Opt_{3TCO}|$ and $k \cdot |Sol_{VC}| \leq |Sol_{3TCO}|$ and we bound the size of $Sol_{VC}$:

\[
k \cdot |Sol_{VC}| \leq |Sol_{3TCO}|
\leq \alpha \cdot |Opt_{3TCO}|
\leq \alpha \cdot (k \cdot |Opt_{VC}| + |E|)
\leq \alpha \cdot k \cdot |Opt_{VC}| + \alpha \cdot |E|
\]
and thus

\[ |\text{Sol}_{VC}| \leq \alpha \cdot |\text{Opt}_{VC}| + \frac{\alpha \cdot |E|}{k}. \]

Since \( \alpha \) is constant and \( |E| \) is fixed for the instance, we pick \( k \) such that \( k \geq \frac{\alpha \cdot |E|}{\varepsilon} \). Note that \( k \) is polynomial in the size of the input and thus the size of the instance \( I_{3TCO} \) is polynomial in the size of \( I_{VC} \). We can bound the size of \( \text{Sol}_{VC} \) both as:

\[ |\text{Sol}_{VC}| \leq \alpha \cdot |\text{Opt}_{VC}| + \varepsilon \quad \text{and} \quad |\text{Sol}_{VC}| \leq (\alpha + \varepsilon) \cdot |\text{Opt}_{VC}|. \]

**Corollary 2.** There is no polynomial time \((1.3606 - \varepsilon)\)-approximation algorithm for Min-3-TCO (for any \( \varepsilon > 0 \)).

**Corollary 3.** If the unique game conjecture holds, there is no polynomial time \((2 - \varepsilon)\)-approximation algorithm for Min-3-TCO (for any \( \varepsilon > 0 \)).

**Proof.** Otherwise, the reduction described in Theorem 2 would imply an approximation algorithm for VC with ratio better than 2. This would contradict [7].

**Acknowledgment**

The author would like to thank Hans-Joachim Böckenhauer and Richard Královič for helping to improve the presentation of this paper.

**References**


