Theoretical Computer Science 429 (2012) 144–154

Contents lists available at SciVerse ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

On the approximability and hardness of minimum topic connected overlay and its special instances☆,☆☆

Jun Hosoda [c], Juraj Hromkovič [a,*], Taisuke Izumi [c], Hirotaka Ono [b], Monika Steinová [a], Koichi Wada [c]

[a] Department of Computer Science, ETH Zurich, Switzerland

[b] Department of Economic Engineering, Kyushu University, Japan

[c] Graduate School of Engineering, Nagoya Institute of Technology, Japan

ARTICLE INFO

Keywords:
Topic-connected overlay
Approximation algorithm
$\mathcal{APX}$
Hardness

ABSTRACT

In the context of designing a scalable overlay network to support decentralized topic-based pub/sub communication, the Minimum Topic-Connected Overlay problem (Min-TCO in short) has been investigated: given a set of t topics and a collection of n users together with the lists of topics they are interested in, the aim is to connect these users to a network by a minimum number of edges such that every graph induced by users interested in a common topic is connected. It is known that Min-TCO is $\mathcal{NP}$-hard and approximable within $O(\log t)$ in polynomial time.

In this paper, we further investigate the problem and some of its special instances. We give various hardness results for instances where the number of topics in which a user is interested in is bounded by a constant, and also for the instances where the number of users interested in a common topic is a constant. For the latter case, we present a first constant approximation algorithm. We also present some polynomial-time algorithms for very restricted instances of Min-TCO.

1. Introduction

Recently, the spreading of social networks and other services based on sharing content has allowed the development of many-to-many communication, often supported by these services. Publishers publish information through a logical channel that is consumed by subscribed users. This environment is often modeled by publish/subscribe (pub/sub) systems that can be classified into two categories. When 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, we speak about *content-based* pub/sub systems. Each channel in *topic-based* pub/sub systems is associated with a single topic and the messages are distributed to the users via channels by his/her topic selection. There are numerous implementations of pub/sub systems; for details see [1,4,6,7,21,22,24].

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). Many aspects of such a

☆ This research is partly supported by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research, 21500013, 21680001, 22650004, 22700010, 23104511, 23310104, Foundation for the Fusion of Science and Technology (FOST) and INAMORI FOUNDATION. The research is also partially funded by SNF grant 200021-132510/1.

☆☆ Some of the results of this paper were presented at MFCS 2011 and PODC 2011.

* Corresponding author.

E-mail addresses: juraj.hromkovic@inf.ethz.ch (J. Hromkovič), t-izumi@nitech.ac.jp (T. Izumi), hirotaka@en.kyushu-u.ac.jp (H. Ono), monika.steinova@inf.ethz.ch (M. Steinová), wada@nitech.ac.jp (K. Wada).

0304-3975/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2011.12.033

system can be studied (see [9,19]). Minimizing the diameter of the overlay network can minimize the overall time in which a message is distributed to all the subscribers. When minimizing the (average) degree of nodes in the network, the subscribers need to maintain a smaller number of connections. In this paper, we study the minimization of the overall number of connections in the system. A small number of connections may be necessary due to maintenance requirements or may be helpful since thus information aggregated into a single message can be broadcasted to the network and thus amortize the head count of otherwise small messages.

We study here the hardness of *Minimum Topic-Connected Overlay* (Min-TCO) which was studied in different scenarios in [2,9,16,17]. In Min-TCO, we are given a collection of users, a set of topics, and a user-interest assignment, and we want to connect users in an overlay network G such that all users interested in a common topic are connected and the overall number of edges in G is minimal. The hardness of the problem was studied in [9,2]. In [9], the inapproximability by a constant was proved and a logarithmic-factor approximation algorithm was presented. In [2], the lower bound on the approximability of Min-TCO was improved to $\Omega(\log(n))$, where n is the number of users.

Moreover, we focus here on special instances of Min-TCO. We study the case where, for each topic, there is a constant number of users interested in it. We also consider the case where the number of topics in which any user is interested is bounded by a constant. We believe that such restrictions on the instances have wide practical applications such as when a publisher has a limited number of slots for users or the user's application limits the number of topics that he/she can follow.

In the study of the general Min-TCO, we extend the method presented in [9] and design an approximation-preserving reduction from instances of the minimum hitting set problem to instances of Min-TCO. This reduction does not only prove a similar lower bound as in [2], but also shows that Min-TCO is $\mathcal{LOGAPX}$-complete and thus, concerning approximability, equivalent with such a famous problem as the minimum set cover. As our reduction is not blowing up the number of users interested in a common topic, the reduction is also an approximation-preserving reduction for the case where the number of users interested in a common topic is limited to a constant. Furthermore, we design a one-to-one reduction of these instances to special instances of the hitting set problem. As these special instances of the hitting set problem are constantly approximable, we immediately obtain the first approximation algorithm for our special instances. This, together with our approximation preserving reduction, shows that the restriction of Min-TCO to such special instances is $\mathcal{APX}$-complete. Finally, we present the first nontrivial parameterized algorithm for the instances of Min-TCO for which there is a small constant d such that for each topic the number of users interested in it is bounded by d. This algorithm is parameterized with respect to the size of the output.

For the case where the number of topics of Min-TCO is bounded from above by $(1 + \varepsilon(n))^{-1} \cdot \log\log n$, for $\varepsilon(n) \geq \frac{3/2 \log\log\log n}{\log\log n - 3/2 \log\log\log n}$ (n is the number of users), we present a polynomial-time algorithm that computes the optimal solution.

In the study of instances where the number of topics any user is interested in is restricted to a constant, we show that, if this number is at most 6, Min-TCO cannot be approximated within a factor of 694/693 in polynomial time, unless $\mathcal{P} = \mathcal{NP}$, even if any pair of two users is interested in at most three common topics.

The paper is organized as follows. Section 2 is devoted to the preliminaries and a summary of known results. The hardness, approximation results and the parameterized algorithm for instances of Min-TCO, where we limit the number of users interested in a common topic by a constant, are discussed in Section 3. This section also provides the discussion about $\mathcal{LOGAPX}$-completeness of the general Min-TCO. The results related to the instances of Min-TCO, where the number of topics that each user is interested in is constant, are presented in Section 4. Section 5 contains a polynomial-time algorithm that solves Min-TCO when the number of topics is small. The conclusion is provided in Section 6.

2. Preliminaries

In this section, we define basic notions used throughout the paper. We assume that the reader is familiar with notions of graph theory. Let $G = (V, E)$ be an undirected graph, where V is the set of vertices and E is the set of edges. Let $V(G)$ and $E(G)$ denote the set of vertices and the set of edges of G, respectively. We denote by $E[S]$ the set of edges of G in the subgraph induced by the vertices from $S \subseteq V$, i. e., $E[S] = \{\{u, v\} \in E \mid u, v \in S\}$. The graph induced by $S \subseteq V$ is denoted as $G[S] = (S, E[S])$. By $N[v]$ we denote the *closed neighborhood* of vertex v, i. e., $N[v] = \{u \in V \mid \{u, v\} \in E\} \cup \{v\}$. A graph G is called *connected*, if, for any $u_1, u_\ell \in V$, there exists a path $(u_1, u_2, \ldots, u_\ell)$ such that $\{u_i, u_{i+1}\} \in E$, for all $1 \leq i < \ell$. The set of all possible edges between vertices in S is denoted as $E_S = \{\{u, v\} \mid u, v \in S \wedge u \neq v\}$.

Let x be an instance of an optimization problem (in this paper, Min-TCO, Min-VC or Min-HS), then by $|x|$ we denote the size of this instance, i. e., the number of vertices and topics of an instance of Min-TCO and the number of elements and sets of an instance of Min-HS. For a set S, $|S|$ denotes the size of the set, i. e., the number of its elements.

The set of users or nodes of our network is denoted by $U = \{u_1, u_2, \ldots, u_n\}$. The topics are $T = \{t_1, t_2, \ldots, t_m\}$. Each user subscribes to several topics. This relation is expressed by the user interest function $\mathrm{INT} : U \to 2^T$. The set of all vertices of U interested in a topic t is denoted by U_t. For instance, if user $u \in U$ is interested in topics t_1, t_3 and t_4, then we have $\mathrm{INT}(u) = \{t_1, t_3, t_4\}$ and $u \in U_{t_1}, U_{t_3}, U_{t_4}$. For a given set of users U, a set of topics T, and an interest function INT, we say that a graph $G = (U, E)$ with $E \subseteq E_U$ is *t-topic-connected*, for $t \in T$, if the subgraph $G[U_t]$ is connected. We call the graph *topic-connected* if it is t-topic-connected for each topic $t \in T$. Note that the topic-connectedness property implies that a message published for topic t is transmitted to all users interested in this topic without using non-interested users as intermediate nodes.

The most general problem that we study in this paper is called *Minimum Topic Connected Overlay*:

Problem 1. Min-TCO is the following optimization problem:

Input: A set of users U, a set of topics T, and an user interest function INT : $U \to 2^T$.
Feasible solutions: Any set of edges $E \subseteq E_U$ such that the graph (U, E) is topic-connected.
Costs: Size of E.
Goal: Minimization.

In this paper we study also some of the special instances of the problem Min-TCO. We restrict the number of users that are interested in a common topic, i. e., the size of U_t, to a constant. We also study the instances where each user is interested in a constant number of topics. The definitions necessary for these special instances are summarized in the beginning of the corresponding section.

We refer here to the well-known *minimum hitting set problem* (Min-HS) and *minimum set cover problem* (Min-SC). In Min-HS, we are given a system of sets $\mathcal{S} = \{S_1, \ldots, S_m\}$ on n elements $X = \{x_1, \ldots, x_n\}$ (i. e., $S_j \subseteq X$). A feasible solution of this problem is a set $H \subseteq X$, such that $S_j \cap H \neq \emptyset$ for all j. Our goal is to minimize the size of H. The Min-SC is the dual problem to Min-HS. In this problem, we are given a system of sets $\mathcal{S} = \{S_1, \ldots, S_m\}$ on n elements $X = \{x_1, \ldots, x_n\}$, a feasible solution is a set $S \subseteq \mathcal{S}$ of sets such that for all i there exists j such that $x_i \in S_j \in S$ and the goal is the minimization of the size of S.

There are many modifications and subproblems of the hitting set problem that are intensively studied. In our paper, we refer to the d-HS problem – a restriction of Min-HS to instances where $|S_i| \leq d$ for all i.

The Min-HS and Min-SC is the same problem viewed from different perspectives [3], thus all the results concerning approximability of Min-SC carry over to Min-HS (but may differ by a constant factor). From the facts known about Min-SC it easily follows that Min-HS is $\mathcal{LOGAPX}$-complete [10] and d-HS is $\mathcal{APX}$-complete [20]. There is a well known d-approximation algorithm for d-HS [5]. Following from [12], it can be approximated with ratio $d - \frac{(d-1)\ln\ln n}{\ln n}$. Furthermore, d-HS is $\mathcal{NP}$-hard to approximate within a factor $(d - 1 - \varepsilon)$ due to [11] and it is not approximable within a factor better than d, unless the unique games conjecture fails [15].

We use the standard definitions from complexity theory (for details see [13]):

- For $\mathcal{NPO}$ problems in the class $\mathcal{PTAS}$, there exists an algorithm that, for arbitrary $\varepsilon > 0$, produces a solution in time polynomial in the input size (but possibly exponential in $1/\varepsilon$) that is within a factor $(1 + \varepsilon)$ from optimal.
- The $\mathcal{NPO}$ problems in the class $\mathcal{APX}$ are approximable by some constant-factor approximation algorithm in polynomial time.
- For $\mathcal{NPO}$ problems in the class $\mathcal{LOGAPX}$, there exists a polynomial-time logarithmic-factor approximation algorithm.

Thus

$$\mathcal{PTAS} \subseteq \mathcal{APX} \subseteq \mathcal{LOGAPX}.$$

Definition 1. Let A and B be two $\mathcal{NPO}$ minimization problems. Let I_A and I_B be the sets of the instances of A and B, respectively. Let $S_A(x)$ and $S_B(y)$ be the sets of the feasible solutions and let $cost_A(x)$ and $cost_B(y)$ be polynomially computable measures of the instances $x \in I_A$ and $y \in I_B$, respectively. We say that A is *AP-reducible* to B, if there exist functions f and g and a constant $\alpha > 0$ such that:

1. For any $x \in I_A$ and any $\varepsilon > 0$, $f(x, \varepsilon) \in I_B$.
2. For any $x \in I_A$, for any $\varepsilon > 0$, and any $y \in S_B(f(x, \varepsilon))$, $g(x, y, \varepsilon) \in S_A(x)$.
3. The functions f and g are computable in polynomial time with respect to the sizes of instances x and y, for any fixed ε.
4. The time complexity of computing f and g is nonincreasing with ε for all fixed instances of size $|x|$ and $|y|$.
5. For any $x \in I_A$, for any $\varepsilon > 0$, and for any $y \in S_B(f(x, \varepsilon))$

$$\frac{cost_B(y)}{\min\{cost_B(z) \mid z \in S_B(f(x, \varepsilon))\}} \leq 1 + \varepsilon \text{ implies}$$
$$\frac{cost_A(g(x, y, \varepsilon))}{\min\{cost_A(z) \mid z \in S_A(x)\}} \leq 1 + \alpha \cdot \varepsilon.$$

3. Results for Min-TCO when the number of users interested in a common topic is a constant

In this whole section, we denote by a triple (U, T, INT) an instance of Min-TCO. We focus here on the case where the number of users that share a topic t, i. e., $\max_{t \in T} |U_t|$, is bounded.

We present here a lower bound on the approximability, a constant approximation algorithm and an $\mathcal{APX}$-completeness proof for these restricted instances of Min-TCO.

3.1. Hardness results

It is easy to see that, if $\max_{t \in T} |U_t| \leq 2$, then Min-TCO can be solved in linear time, because two users sharing a topic t should be directly connected by an edge, which is the unique minimum solution.

Theorem 1. *If* $\max_{t\in T} |U_t| \leq 2$, *then* Min-TCO *can be solved in linear time.*

We extend the methods from [9] and design an AP-reduction from d-HS to Min-TCO, where $\max_{t\in T} |U_t| \leq d+1$.

Theorem 2. *For arbitrary* $d \geq 2$, *there exists an AP-reduction from* d-HS *to* Min-TCO, *where* $\max_{t\in T} |U_t| \leq d+1$.

Proof. Let $I_{\mathrm{HS}} = (X, \mathcal{S})$ be an instance of d-HS and let $\varepsilon > 0$ be arbitrary. We omit the subscript in the functions $cost_{d\text{-HS}}$ and $cost_{\text{Min-TCO}}$ as they are unambiguous. For the instance I_{HS}, we create an instance $I_{\mathrm{TCO}} = (U, T, \mathrm{INT})$ of Min-TCO with $\max_{t\in T} |U_t| \leq d+1$, with $|X|$ users that correspond to elements of X and $k = |X|^2 \cdot \lceil \frac{1+\varepsilon}{\varepsilon} \rceil$ new *special users* p_i, as follows (the function f in the definition of AP-reduction).

$$U = X \cup \{p_i \mid p_i \notin X \wedge 1 \leq i \leq k\},$$
$$T = \{t^i_{S_j} \mid S_j \in \mathcal{S} \wedge 1 \leq i \leq k\},$$
$$\mathrm{INT}(x) = \begin{cases} \{t^i_{S_j} \mid x \in S_j \wedge S_j \in \mathcal{S} \wedge 1 \leq i \leq k\} & \text{for } x \in X \\ \{t^i_{S_j} \mid S_j \in \mathcal{S}\} & \text{for } x = p_i. \end{cases}$$

Observe that the instance contains $k \cdot |\mathcal{S}|$ topics and its size is polynomial in the size of I_{HS}. The users interested in a topic $t^i_{S_j}$ ($S_j \in \mathcal{S}$) are exactly the elements that are members of set S_j in d-HS plus a special user p_i ($1 \leq i \leq k$). Let Sol_{TCO} be a feasible solution of Min-TCO on instance I_{TCO}. We partition the solution into levels. Level i is a set L_i of the edges of Sol_{TCO} that are incident with the special user p_i. In addition, we denote by L_0 the set of edges of Sol_{TCO} that are not incident with any special user. Therefore, $Sol_{\mathrm{TCO}} = \bigcup_{i=0}^{k} L_i$ and $L_i \cap L_j = \emptyset$ ($0 \leq i < j \leq k$).

We claim that, for any L_i ($1 \leq i \leq k$), the set of the non-special users incident with edges of L_i is a feasible solution of the instance I_{HS} of d-HS. This is true since, if a set $S_j \in \mathcal{S}$ is not hit, none of the edges $\{x, p_i\}$ ($x \in S_j$) is in L_i. But then the users interested in topic $t^i_{S_j}$ are not interconnected as user p_i is disconnected.

Let j be chosen such that L_j is the smallest of all sets L_i, for $1 \leq i \leq k$. We construct Sol_{HS} by picking all the non-special users that are incident to some edge from L_j (the function g in the definition of AP-reduction). Denote an optimal solution of d-HS and Min-TCO for I_{HS} and I_{TCO} by Opt_{HS} and Opt_{TCO}, respectively.

If we knew Opt_{HS}, we would be able to construct a feasible solution of Min-TCO on I_{TCO} as follows. First, we pick the edges $\{x, p_i\}$, $x \in Opt_{\mathrm{HS}}$, for all special users p_i, and include them in the solution. This way, for any topic $t \in \mathrm{INT}(p_i)$, we connect p_i to some element of X that is interested in t, too. To have a feasible solution, we could miss some edges between some elements of X. So, we pick all the edges between elements from X. The feasible solution of Min-TCO on I_{TCO} that we obtain has roughly cost

$$k \cdot cost(Opt_{\mathrm{HS}}) + |X|^2 \geq cost(Opt_{\mathrm{TCO}}).$$

On the other hand, if we replace all levels L_i ($1 \leq i \leq k$) by level L_j in Sol_{TCO}, we still have a feasible solution of Min-TCO on I_{TCO}, with cost possibly smaller. Thus

$$k \cdot cost(Sol_{\mathrm{HS}}) \leq cost(Sol_{\mathrm{TCO}}).$$

We use these two inequalities to bound the cost of Sol_{HS}:

$$k \cdot cost(Sol_{\mathrm{HS}}) \leq \frac{cost(Sol_{\mathrm{TCO}})}{cost(Opt_{\mathrm{TCO}})} \cdot \left(k \cdot cost(Opt_{\mathrm{HS}}) + |X|^2\right)$$

and thus

$$\frac{cost(Sol_{\mathrm{HS}})}{cost(Opt_{\mathrm{HS}})} \leq \frac{cost(Sol_{\mathrm{TCO}})}{cost(Opt_{\mathrm{TCO}})} \cdot \left(1 + \frac{|X|^2}{k}\right).$$

If $cost(Sol_{\mathrm{TCO}})/cost(Opt_{\mathrm{TCO}}) \leq 1+\varepsilon$ and $\alpha := 2$, then we have

$$\frac{cost(Sol_{\mathrm{HS}})}{cost(Opt_{\mathrm{HS}})} \leq (1+\varepsilon) \cdot \left(1 + \frac{|X|^2}{k}\right) \leq (1+\varepsilon) \cdot \left(1 + \frac{\varepsilon}{1+\varepsilon}\right) = 1 + 2\varepsilon.$$

It is easy to see that the five conditions of Definition 1 are satisfied and thus we have an AP-reduction. □

Corollary 1. *For any* $\delta > 0$ *and polynomial-time* α*-approximation algorithm of* Min-TCO *with* $\max_{t\in T} |U_t| \leq d+1$, *there exists a polynomial-time* $(\alpha+\delta)$*-approximation algorithm of* d-HS.

Proof. The approximation algorithm for d-HS would use Theorem 2 with $k := |X|^2 \cdot \lceil \frac{\alpha}{\delta} \rceil$. □

Our theorem also implies the following negative results on approximability. One of them holds if *unique games conjecture* is true. This conjecture is discussed, for example, in [23] and was introduced by Khot in [14].

Corollary 2. Min-TCO *with* $\max_{t\in T}|U_t| \leq d$

- *is* $\mathcal{NP}$*-hard to approximate within a factor of* $(d-2-\varepsilon)$*, for* $d \geq 4$ *and any* $\varepsilon > 0$.
- *admits no polynomial-time* $(d-1-\varepsilon)$*-approximation algorithm, for* $d \geq 3$ *and any* $\varepsilon > 0$*, unless the unique games conjecture fails.*

Proof. Otherwise, the reduction described in the proof of Theorem 2 would imply an approximation algorithm for d-HS with a ratio better than $d-1$ and d respectively. This would directly contradict theorems proven in [11,15]. □

The following corollary is an improvement of the already known results of [9] where an $O(\log|T|)$-approximation algorithm is presented, and of [2] where a lower bound of $\Omega(\log(n))$ on the approximability is shown. We close the gap by designing a reduction that can reduce *any* problem from class $\mathcal{LOGAPX}$ to Min-TCO preserving the approximation ratio up to a constant.

Corollary 3. Min-TCO *is* $\mathcal{LOGAPX}$*-complete.*

Proof. Min-TCO is in the class $\mathcal{LOGAPX}$ since it admits a logarithmic approximation algorithm as presented in [9]. Our reduction from the proof of Theorem 2 is independent of d and thus an AP-reduction from $\mathcal{LOGAPX}$-complete Min-HS to Min-TCO. □

3.2. A constant approximation algorithm

In this subsection, we present a reduction from Min-TCO with $\max_{t\in T}|U_t| \leq d$ to $O(d^2)$-HS thus showing that there exists a constant approximation algorithm for Min-TCO with $\max_{t\in T}|U_t| \leq d$ as d-HS is constantly approximable. Moreover, the constant approximation algorithm classifies this problem to be a member of the class $\mathcal{APX}$ and thus, since the $\mathcal{APX}$-hardness was proven in Section 3.1, we conclude that Min-TCO with $\max_{t\in T}|U_t| \leq d$ is $\mathcal{APX}$-complete.

Recall that a partition of vertices V in graph G is a pair (A, B), such that $A \subseteq V$, $B \subseteq V$, $A \cap B = \emptyset$, and $A \cup B = V$.

Definition 2. Let $V = \{v_1, \ldots, v_n\}$ be a set of vertices and for every partition (A_i, B_i) of V, let $E_i = \{\{u, v\} \mid u \in A_i \wedge v \in B_i\}$. Then we call the system $\mathcal{S} = \{E_1, \ldots, E_m\}$ of all sets of edges between vertices of all the partitions of V *a characteristic system* of edges on V. In other words, $\mathcal{S}$ contains all sets of edges that form a maximum bipartite graph on V.

In the following lemma, we show the basic properties of characteristic systems of edges.

Lemma 1. *Let* $\mathcal{S} = \{E_1, \ldots, E_m\}$ *be a characteristic system of edges on the set* V *of* n *vertices. Then*

1. $m = 2^{n-1} - 1$.
2. $|E_j| \leq \lfloor n/2 \rfloor \cdot \lceil n/2 \rceil$*, for all* j, $1 \leq j \leq m$.
3. *Any two sets* E_i *and* E_j *differ in at least* $n-1$ *elements* $(1 \leq i < j \leq m)$.
4. $H \subseteq E_V$ *is a hitting set of* $(E_V, \mathcal{S})$ *if and only if* (V, H) *is connected.*
5. *The set* $\mathcal{S}$ *is minimal with respect to set inclusion—no proper subset of* $\mathcal{S}$ *has property 4.*

Proof. Observe that the complementary graph (V, F_j) (with $F_j = E_V \setminus E_j$) contains two complete graphs—one on the vertices of A_j and other on the vertices of B_j, and it is a maximal graph (in the number of edges) that is not connected. We use this observation to prove the last two parts of our lemma.

Part 1: We count the different partitions (A_j, B_j) of the vertices V as each such partition determines a different set E_j of edges. There are 2^n ways how to distribute vertices from V into partitions. We have to subtract 2 possibilities for the cases where one of A_j or B_j is empty. Each of the other possibilities is counted twice – once when the vertices are present in A_j and once when they are present in B_j.

Part 2: Let the two sets of vertices A_j and B_j of a partition contain $k > 0$ and $n-k$ vertices. Then the size of E_j is $k \cdot (n-k)$. This function reaches its maximum for $k = n/2$ and thus we can conclude that, for all j, $1 \leq j \leq m$, we have $|E_j| \leq \lfloor n/2 \rfloor \cdot (n - \lfloor n/2 \rfloor) = \lfloor n/2 \rfloor \cdot \lceil n/2 \rceil$.

Part 3: Let us consider two different partitions (A_i, B_i) and (A_j, B_j) of the vertices V. The sets A_i and A_j must differ by at least one vertex. W.l.o.g., let the vertex $v \in A_i$ and $v \notin A_j$. Then, due to the transition of the vertex v from A_i to B_j, there are $|B_i|$ edges that are in E_i but cannot be in E_j, and there are $|A_i| - 1$ edges that are not in E_i, but are in E_j. Thus, the overall difference in the number of elements between the sets E_i and E_j is at least $|A_i| + |B_i| - 1 = n - 1$.

Part 4: First, we prove the if case. Suppose that H is a hitting set, but (V, H) is not connected. Since $\mathcal{S}$ contains complements of all maximal sets of edges that induce a disconnected graph, there exists j $(1 \leq j \leq m)$ such that $H \subseteq F_j$. But then, since E_j is complementary to F_j, it follows that $E_j \cap H = \emptyset$. Thus, H cannot be a hitting set as E_j is not hit.

For the only-if case, suppose that (V, H) is connected, but H is not a hitting set of $(E_V, \mathcal{S})$. Then there exists j such that E_j is not hit by H and thus $H \subseteq F_j$. Yet in such a case, by our assumption, (V, F_j) is not connected and thus (V, H) cannot be connected as well.

Part 5: Let $\mathcal{S}' = \mathcal{S} \setminus E_j$, and let $\left(E_V, \mathcal{S}'\right)$ be an instance of Min-HS. Then we claim that F_j is a hitting set of $(E_V, \mathcal{S}')$. First, observe that $F_j \neq \emptyset$ since E_j cannot contain all the edges. Moreover, for every $E_i \in \mathcal{S}'$, there exists $e \in E_i$ such that $e \notin E_j$. Then $e \in E_V \setminus E_j = F_j$ and thus F_j is a hitting set. However, by the definition of F_j, the graph (V, F_j) cannot be connected and thus, the if case of Part 4 does not hold. □

Now we are ready to present a simple one-to-one reduction of Min-TCO with $\max_{t\in T}|U_t| \leq d$ to $O(d^2)$-HS. The core concept is to construct a system of sets that has to be hit in $O(d^2)$-HS as a union over all the topics of the characteristic systems of edges on the vertices interested in the topic.

Theorem 3. *There exists a one-to-one reduction of instances of* Min-TCO *with* $\max_{t\in T}|U_t| \leq d$ *to instance of* $O(d^2)$-HS.

Proof. Let $I_{\mathrm{TCO}} = (U, T, \mathrm{INT})$ be an instance of Min-TCO with $\max_{t\in T}|U_t| \leq d$. For each topic $t \in T$ we define $\mathcal{S}_t$ to be the characteristic system of edges on vertices in U_t. Note that Lemma 1 holds for each $\mathcal{S}_t$ with $n := d$. We construct an $O(d^2)$-HS instance $I_{\mathrm{HS}} = (X, \mathcal{S})$ as follows:

$$X = \{\{u, v\} \mid u, v \in U \wedge u \neq v\}$$
$$\mathcal{S} = \bigcup_{t\in T} \mathcal{S}_t.$$

The system contains $\binom{|U|}{2}$ elements and at most $|T| \cdot \left(2^{d-1} - 1\right)$ sets in $\mathcal{S}$ and thus has a size polynomial in $|I_{\mathrm{TCO}}|$. Obviously, the construction of I_{HS} takes time polynomial in $|I_{\mathrm{TCO}}|$, too. We now show that a feasible solution of I_{TCO} corresponds to a feasible solution of I_{HS} and vice versa.

First, consider a feasible solution Sol_{HS} of I_{HS} and a topic $t \in T$. Due to our construction, the system $\mathcal{S}$ contains the characteristic system $\mathcal{S}_t$ on vertices U_t. Therefore, by Lemma 1 Part 4 and the fact that Sol_{HS} is a hitting set, we know that the graph induced by the edges in Sol_{HS} on vertices U_t is connected.

Now, consider a feasible solution Sol_{TCO} of I_{TCO}. By the following argument, we can easily see that Sol_{TCO} hits all the sets in $\mathcal{S}$. Let $P \in \mathcal{S}$ be a set that is not hit by Sol_{TCO}. Then there exists t such that $P \in \mathcal{S}_t$ and thus a set of the characteristic system was not hit and Sol_{TCO} is not a hitting set of $\mathcal{S}_t$. Yet in such a case, considering Lemma 1 Part 4, the subgraph induced on vertices U_t by edges from Sol_{TCO} cannot be connected and that is in contradiction with the definition of Min-TCO with $\max_{t\in T}|U_t| \leq d$. □

Theorem 4. *There exists a polynomial-time* $(\lfloor d/2\rfloor \cdot \lceil d/2\rceil)$*-approximation algorithm for* Min-TCO *with* $\max_{t\in T}|U_t| \leq d$.

Proof. We employ the reduction from Theorem 3 together with the well-known d-approximation algorithm for d-HS. Since the size of each set in $\mathcal{S}$ is at most $\lfloor d/2\rfloor \cdot \lceil d/2\rceil$ (Lemma 1 Part 2), by application of this approximation algorithm on $O(d^2)$-HS instance $(X, \mathcal{S})$ we obtain a $\lfloor d/2\rfloor \cdot \lceil d/2\rceil$ approximate solution of our Min-TCO instance with $\max_{t\in T}|U_t| \leq d$.

Note that our reduction is tight in the size of $\mathcal{S}$ as it is minimal (Lemma 1 Part 5), thus to achieve an improvement in the approximation algorithm, a different method has to be developed. □

Corollary 4. Min-TCO *with* $\max_{t\in T}|U_t| \leq 3$ *inherits the approximation hardness of* Min-VC.

Corollary 5. Min-TCO *with* $\max_{t\in T}|U_t| \leq d$ *is* $\mathcal{APX}$*-complete, for arbitrary* $d \geq 3$.

Proof. The $\mathcal{APX}$-hardness follows from the $\mathcal{APX}$-hardness of d-HS ([20]). Due to our reduction the problem belongs to the class $\mathcal{APX}$. □

3.3. Min-TCO *and parameterized complexity theory*

We shortly summarize the consequences of our reduction from Theorem 3 leading to a nontrivial parameterized algorithm for Min-TCO with $\max_{t\in T}|U_t|$ bounded by a constant. To express only the main factors in the time-complexity of algorithms if this section, we use O^*-notation which is the same as O-notation, except that polynomial factors are neglected.

Problem 2. Min-d-TCO(k) is the following parameterized problem:

Input: Instance of Min-TCO with $\max_{t\in T}|U_t| \leq d$ and a parameter k.
Goal: Decide whether there exists a feasible solution of the Min-TCO instance of size at most k.

Problem 3. d-HS(k) is the following parameterized problem:

Input: Instance of d-HS and a parameter k.
Goal: Decide whether there exists a feasible solution of the d-HS instance of size at most k.

Consider an instance of Min-TCO with $\max_{t\in T}|U_t| \leq d$ and n vertices. A straightforward parameterized algorithm for Min-d-TCO(k) has to consider all sets of at most k edges out of possible $\binom{n}{2}$ edges of the instance. This number of sets can be roughly estimated as

$$\sum_{i=0}^{k} \binom{n^2}{i} \leq k \cdot \binom{n^2}{k} \leq k \cdot n^{2k}.$$

Hence, the time complexity of such an approach is $O^*(n^{2k})$.

However, if we apply first our transformation from Theorem 3 and then the fixed-parameter algorithm from [18], we obtain a better algorithm for Min-d-TCO(k).

Theorem 5. Min-d-TCO(k) *with n users can be solved in time* $O(\alpha^k + n^2)$ *with*

$$\alpha = \frac{1}{2}\left(\left\lfloor \frac{n}{2} \right\rfloor \cdot \left\lceil \frac{n}{2} \right\rceil - 1\right) + \frac{1}{2}\left(\left\lfloor \frac{n}{2} \right\rfloor \cdot \left\lceil \frac{n}{2} \right\rceil - 1\right) \cdot \sqrt{1 + \frac{4}{(\lfloor n/2 \rfloor \cdot \lceil n/2 \rceil - 1)^2}}$$

which is approximately

$$\alpha = \left\lfloor \frac{n}{2} \right\rfloor \cdot \left\lceil \frac{n}{2} \right\rceil - 1 + O(n^{-2}).$$

Proof. Let (x, k) be an instance of Min-d-TCO(k). We apply our reduction from Theorem 3 to transform x to an instance y of $O(d^2)$-HS with dimension $d := \lfloor n/2 \rfloor \cdot \lceil n/2 \rceil$. Our transformation matches feasible solutions of both instances in a one-to-one manner, hence, (x, k) is a yes-instance of Min-d-TCO(k) if and only if (y, k) is a yes-instance of d-HS(k).

Hence, the fixed-parameter algorithm of Niedermeier and Rossmanith of [18] for d-HS(k) problem can be applied. The algorithm yields the claimed time complexity.

If we compute a rough estimation of the time complexity of this algorithm, we have

$$O(\alpha^k + n^2) = O^*(4^{-k} n^{2k}).$$

Hence, we can conclude that our algorithm is exponentially better than a trivial algorithm that exhaustively searches through all possibilities.

4. Hardness of Min-TCO when the number of connections of a user is a constant

It is natural to consider Min-TCO with bounded number of connections per user, i. e., to bound $\max_{u \in U} |\mathrm{INT}(u)|$, since for a fixed user the number of interesting topics is usually not too large. We show that, sadly, Min-TCO is $\mathcal{APX}$-hard even if $\max_{u \in U} |\mathrm{INT}(u)| \leq 6$. To show this, we design a reduction from *minimum vertex cover problem* (Min-VC) to Min-TCO. The minimum vertex cover problem is just a different name for d-HS with $d = 2$. For a better presentation, in this section, we refer to Min-VC instead of 2-HS.

Given is a graph $G = (V', E')$ and a positive integer k as an instance of Min-VC, where the goal is to decide whether the given graph has a solution of size at most k. We construct an instance of Min-TCO as follows. Let $V = V^{(1)} \cup V^{(2)}$ be the set of users, where $V^{(1)} = \{v^{(1)} \mid v \in V'\}$ and $V^{(2)} = \{v^{(2)} \mid v \in V'\}$. For each edge $e \in E'$, we prepare three topics, $t_e^{(0)}$, $t_e^{(1)}$ and $t_e^{(2)}$. The set of topics is the union of all these topics, i. e., $T = \bigcup_{e \in E'} \{t_e^{(0)}, t_e^{(1)}, t_e^{(2)}\}$. The user interest function INT is defined as

$$\mathrm{INT}(u^{(1)}) = \bigcup_{e \in E'[N[u]]} \{t_e^{(0)}, t_e^{(1)}\}$$

$$\mathrm{INT}(u^{(2)}) = \bigcup_{e \in E'[N[u]]} \{t_e^{(0)}, t_e^{(2)}\}.$$

The following lemma shows the relation between the solutions of the two problems.

Lemma 2. *The instance* (V, T, INT) *of* Min-TCO *defined as above has an optimal solution of cost* $k + 2|E'|$ *if and only if the instance* (V', E') *of* Min-VC *has an optimal solution of cost k.*
Moreover, any feasible solution H of (V, T, INT) *can be transformed into a feasible solution of* (V', E') *of cost at most* $|H| - 2|E'|$.

Proof. It is obvious that any feasible solution H of the instance of Min-TCO contains the edge $\{u^{(i)}, v^{(i)}\}$ $(i \in \{1, 2\})$, for every edge $e = \{u, v\}$, because only $u^{(1)}$ and $v^{(1)}$ (resp., $u^{(2)}$ and $v^{(2)}$) are interested in topic $t_e^{(1)}$ (resp., $t_e^{(2)}$).

Since each feasible solution H of (V, T, INT) must contain the edges $\{u^{(1)}, v^{(1)}\}$ and $\{u^{(2)}, v^{(2)}\}$, for every edge $e = \{u, v\} \in E'$, it is sufficient to consider only the topics $t_e^{(0)}$. The number of edges in H connecting a user from $V^{(1)}$ with a user from $V^{(2)}$ is at most $|H| - 2|E'|$.

For an edge $e = \{u, v\}$, the vertices that are interested in $t_e^{(0)}$ are $u^{(1)}$, $v^{(2)}$, $v^{(1)}$ and $u^{(2)}$. Since these four vertices have to be connected, H contains at least one edge of $\{u^{(1)}, u^{(2)}\}$, $\{v^{(1)}, v^{(2)}\}$, $\{u^{(1)}, v^{(2)}\}$ and $\{v^{(1)}, u^{(2)}\}$.

The optimal solution of (V, T, INT) contains at most two of these four edges, namely the edges $\{u^{(1)}, u^{(2)}\}$ and $\{v^{(1)}, v^{(2)}\}$. Observe that, for *each* edge f that is incident with vertex u in G, the edge $\{u^{(1)}, u^{(2)}\}$ connects the solution to be $t_f^{(0)}$-connected. The only topic that the other two edges connect is $t_{\{u,v\}}^{(0)}$ and thus they can be replaced by $\{u^{(1)}, u^{(2)}\}$ or $\{v^{(1)}, v^{(2)}\}$.

In any non-optimal solution, more than two of the four edges may be present and the replacement of edges $\{u^{(1)}, v^{(2)}\}$ and $\{v^{(1)}, u^{(2)}\}$ by $\{u^{(1)}, u^{(2)}\}$ and $\{v^{(1)}, v^{(2)}\}$, respectively, may lead to a decrease of the cost of the solution.

We assume that the solution of (V, T, INT) has been transformed so that it does not contain cross edges between $u^{(i)}$ and $v^{(3-i)}$ $(i \in \{1, 2\})$. The vertices that correspond to the edges between the two layers $V^{(1)}$ and $V^{(2)}$ form a feasible solution

of Min-VC. As discussed above, its size is at most $|H| - 2|E'|$ for a feasible solution H and exactly $|H| - 2|E'|$ for an optimal solution H. This proves one implication of the first claim and the second claim.

We now show that, if Min-VC has an optimal solution of size k, then the instance of Min-TCO has an optimal solution of size $k + 2|E'|$. From an optimal solution $W \subseteq V'$ of Min-VC, we construct the optimal solution of Min-TCO as $H = \{\{u^{(1)}, v^{(1)}\}, \{u^{(2)}, v^{(2)}\} \mid \{u, v\} \in E'\} \cup \{\{u^{(1)}, u^{(2)}\} \mid u \in W\}$. Clearly, the size of H is exactly $k + 2|E'|$. As W is the smallest set of vertices that covers all the edges of E', its corresponding edges of Min-TCO produce the minimal set of edges that connect every topic with superscript 0. Thus, H satisfies the connectivity requirement for every topic $t \in T$ and is optimal. □

We use the Min-VC on degree-bounded graphs, which is $\mathcal{APX}$-hard, to show lower bounds for our restricted Min-TCO. By the above reduction and the lemma, we prove the following theorem.

Theorem 6. Min-TCO *with* $\max_{v \in U} |\mathrm{INT}(v)| \leq 6$ *cannot be approximated within a factor of* 694/693 *in polynomial time, unless* $\mathcal{P} = \mathcal{NP}$*, even if* $|\mathrm{INT}(v) \cap \mathrm{INT}(u)| \leq 3$ *holds for every pair of different users* $u, v \in U$.

Proof. We prove the statement by contradiction. Suppose that there exists an approximation algorithm A for Min-TCO with the above stated restrictions that has the ratio $(1 + \delta)$.

Let $G = (V', E')$ be an instance of Min-VC and let G be cubic and regular (i. e., each vertex is incident with exactly three edges). We construct an instance I_{TCO} of Min-TCO as stated above and we apply our algorithm A to it to obtain a feasible solution Sol_{TCO}. From such a solution, by Lemma 2, we create a feasible solution of the original Min-VC instance Sol_{VC}. We denote by Opt_{TCO} and Opt_{VC} the optimal solutions of I_{TCO} and G, respectively.

Let d be a constant such that $d \cdot cost(Opt_{\mathrm{VC}}) = 3|V'|$. Since G is cubic and regular, $cost(Opt_{\mathrm{VC}}) \geq |E'|/3 = |V'|/2$ and thus $d \leq 6$.

Observe that, due to Lemma 2, $cost(Opt_{\mathrm{TCO}}) = cost(Opt_{\mathrm{VC}}) + 2|E'| = cost(Opt_{\mathrm{VC}}) + d \cdot cost(Opt_{\mathrm{VC}})$ and $cost(Sol_{\mathrm{TCO}}) \geq cost(Sol_{\mathrm{VC}}) + 2|E'| = cost(Sol_{\mathrm{VC}}) + d \cdot cost(Opt_{\mathrm{VC}})$. These two estimations give us the following bound

$$\frac{cost(Sol_{\mathrm{VC}}) + d \cdot cost(Opt_{\mathrm{VC}})}{cost(Opt_{\mathrm{VC}}) + d \cdot cost(Opt_{\mathrm{VC}})} \leq \frac{cost(Sol_{\mathrm{TCO}})}{cost(Opt_{\mathrm{TCO}})} \leq 1 + \delta.$$

The above inequality allows us to bound the ratio of our Min-VC solution Sol_{VC} and the optimal solution Opt_{VC}:

$$\frac{cost(Sol_{\mathrm{VC}})}{cost(Opt_{\mathrm{VC}})} \leq (1 + \delta) \cdot (d + 1) - d = 1 + \delta(d + 1) \leq 1 + 7\delta.$$

For $\delta := \frac{1}{693}$, we obtain a $\frac{100}{99}$-approximation algorithm for Min-VC on 3-regular graphs which is directly in contradiction with a theorem proven in [8].
□

Corollary 6. Min-TCO *with* $\max_{v \in U} |\mathrm{INT}(v)| \leq 6$ *is* $\mathcal{APX}$*-hard.*

Corollary 7. Min-TCO *with* $|\mathrm{INT}(v) \cap \mathrm{INT}(u)| \leq 3$*, for all users* $u, v \in U$*, is* $\mathcal{APX}$*-hard.*

This result is almost tight, the case when $|\mathrm{INT}(v) \cap \mathrm{INT}(u)| \leq 2$ is still open. The following theorem shows that Min-TCO with $|\mathrm{INT}(v) \cap \mathrm{INT}(u)| \leq 1$, for every pair of distinct users $u, v \in U$, can be solved in linear time.

Theorem 7. Min-TCO *can be solved in linear time, if* $|\mathrm{INT}(v) \cap \mathrm{INT}(u)| \leq 1$ *holds for every pair of users* $u, v \in U$, $u \neq v$.

Proof. We execute the following simple algorithm. First set the solution $E := \emptyset$. Then sequentially, for each topic t, choose its representative $v^* \in U$ ($t \in \mathrm{INT}(v^*)$) and add edges $\{\{v^*, u\} \mid u \in U_t \setminus \{v^*\}\}$ to the solution E. We show that, if $|\mathrm{INT}(v) \cap \mathrm{INT}(u)| \leq 1$, for all distinct $u, v \in U$, then the solution E is optimal.

Observe that, in our case, any edge in any feasible solution is present because of a unique topic. We cannot find an edge $e = \{u, v\}$ of the solution that belongs to the subgraphs for two different topics. (Otherwise $|\mathrm{INT}(v) \cap \mathrm{INT}(u)| > 1$ and our assumption would be wrong for the two endpoints of the edge e.) Thus, any solution consisting of spanning trees for every topic is feasible and optimal. Note that its size is $|T| \cdot (|U| - 1)$. □

Corollary 8. Min-TCO *with* $\max_{u \in U} |\mathrm{INT}(u)| \leq 2$ *can be solved in linear time.*

5. A polynomial-time algorithm for Min-TCO with bounded number of topics

In this section, we present a simple brute-force algorithm that achieves a polynomial running time when the number of topics is bounded by $|T| \leq \log\log|U| - \frac{3}{2}\log\log\log|U|$.

Theorem 8. *The optimal solution of* Min-TCO *can be computed in polynomial time if* $|T| \leq (1 + \varepsilon(|U|))^{-1} \cdot \log\log|U|$*, for a function*

$$\varepsilon(n) \geq \frac{3/2 \log\log\log n}{\log\log n - 3/2 \log\log\log n}.$$

Proof. Let (U, T, INT) be an instance of Min-TCO such that $|T| \leq (1 + \varepsilon(|U|))^{-1} \cdot \log\log|U|$. Moreover, $|T| > 2$, otherwise the problem is solvable in polynomial time. We shorten the notation by setting $t = |T|$ and $n = |U|$.

First observe that, if $u, v \in U$ and $\mathrm{INT}(u) \subseteq \mathrm{INT}(v)$, instead of solving instance (U, T, INT), we can solve Min-TCO on instance $(U \setminus \{u\}, T, \mathrm{INT})$ and add to such solution the direct edge $\{u, v\}$. Note that u has to be incident with at least one edge in any solution. Thus, the addition of the edge $\{u, v\}$ cannot increase the cost. Moreover, any other user that would be connected to u in some solution can be also connected to v. Thus, we can remove u, solve the smaller instance and then add u by a single edge. Such a solution is feasible and its size is unchanged. We say that user u is *dominated* by the user v if $\mathrm{INT}(u) \subseteq \mathrm{INT}(v)$.

Therefore, before applying our simple algorithm, we remove from the instance all the users that are dominated by some other user. We denote the set of remaining users (i. e., those with incomparable sets of interesting topics) by M. The largest system of incomparable sets on n elements is called a Sperner system and it is a well known fact that its size is at most $\binom{n}{\lfloor n/2 \rfloor}$. Since every user in M must have different set of interesting topics and these sets are all incomparable, we have

$$m = |M| \leq \binom{t}{\lfloor t/2 \rfloor} \leq \frac{2^t}{\sqrt{t}}.$$

(To verify the bound, consider t to be odd or even and use $\binom{2n}{n} \leq \frac{4^n}{\sqrt{3n+1}}$, $n \geq 1$.)

Our simple algorithm exhaustively searches over all the possible solutions on instance (M, T, INT) and then reconnects each of the removed users $U \setminus M$ by a single edge. The transformation to set M and the connection of the removed users is clearly polynomial. Thus, we only need to show that our exhaustive search is polynomial.

Observe that the size of the optimal solution is at most $t(m-1)$, as merged spanning trees, for all the topics, form a feasible solution. Our algorithm exhaustively searches over all possible solutions, i. e., it tries every possible set of i edges for $1 \leq i \leq t(m-1)$ and verifies the topic-connectivity requirements for such sets of edges. The verification of each set can be done in polynomial time. The number of sets it checks can be bounded as follows:

$$\begin{aligned}\sum_{i=1}^{t(m-1)} \binom{\binom{m}{2}}{i} &\leq \sum_{i=1}^{tm} \binom{m^2}{i} \\ &\leq tm \cdot \binom{m^2}{tm} \\ &\leq tm \cdot m^{tm} \\ &\leq m^{tm} \cdot O(\log^2 n).\end{aligned}$$

(Note that $tm \leq m^2/2$ and thus the binomial coefficient is maximal in tm. Otherwise the number of all possible choices of edges into a solution is polynomial in n.)

To check a polynomial number of sets, it is sufficient to bound the factor m^{tm} by a polynomial, i. e., by at most n^c for some $c > 0$. (In all our calculations, log stands for the binary logarithm, however any other logarithm can be used as the change will effect the exponent by a constant.) We consider two cases:

A: First assume that $t \leq \frac{\log\log n}{1+2\varepsilon(n)}$, then $m \leq 2^t \leq (\log n)^{(1+2\varepsilon(n))^{-1}}$.

We use the upper bounds on t and m to estimate the number of sets our exhaustive search has to check:

$$m^{tm} \leq (\log n)^{(1+2\varepsilon(n))^{-2} \cdot \log\log n \cdot (\log n)^{(1+2\varepsilon(n))^{-1}}} \leq n^c.$$

Then we take the logarithm of the inequality, leading to

$$(1 + 2\varepsilon(n))^{-2} \cdot \log\log^2 n \leq c \cdot (\log n)^{\frac{2\varepsilon(n)}{1+2\varepsilon(n)}}.$$

After another logarithm operation, we obtain the following inequality:

$$-2\log(1 + 2\varepsilon(n)) + 2\log\log\log n \leq \frac{2\varepsilon(n)}{1 + 2\varepsilon(n)} \cdot \log\log n + \log c.$$

We prove inequality (1) instead. In the end, we will see that the function $\varepsilon(n)$ is positive, except for the first few values. Thus, for large inputs, $2\log(1 + 2\varepsilon(n))$ is positive and thus the above inequalities will hold, too.

$$2\log\log\log n \leq \frac{2\varepsilon(n)}{1 + 2\varepsilon(n)} \cdot \log\log n. \tag{1}$$

We are now able to estimate the function $\varepsilon(n)$:

$$\varepsilon(n) \geq \frac{\log\log\log n}{\log\log n - 2\log\log\log n}. \tag{2}$$

Due to the following case, we use $\varepsilon(n) \geq \frac{3/2 \log\log\log n}{\log\log n - 3/2 \log\log\log n}$ that also satisfies (2) and is positive for $n \geq 16$.

B: To conclude the proof, assume that $\frac{\log\log n}{1+2\varepsilon(n)} < t \leq \frac{\log\log n}{1+\varepsilon(n)}$. Since we have both an upper and a lower bound on t, we can refine the estimation of m:

$$m \leq \frac{2^t}{\sqrt{t}} \leq (\log n)^{(1+\varepsilon(n))^{-1}} \cdot (1+2\varepsilon(n))^{1/2} \cdot (\log\log n)^{-1/2}.$$

We show that m^{tm} is polynomial in n similarly as in the previous case:

$$(\log n)^{(1+\varepsilon(n))^{-2} \cdot (\log\log n)^{1/2} \cdot (\log n)^{(1+\varepsilon(n))^{-1}} \cdot (1+2\varepsilon(n))^{1/2}} \leq m^{tm} \leq n^c.$$

Then we take the logarithm of the inequality, leading to

$$(1+\varepsilon(n))^{-2} \cdot (\log\log n)^{3/2} \cdot (1+2\varepsilon(n))^{1/2} \leq c \cdot (\log n)^{\frac{\varepsilon(n)}{1+\varepsilon(n)}}.$$

Assume that $(1+2\varepsilon(n))^{1/2} \leq 1+\varepsilon(n)$, except for the first few values, then it is sufficient to prove a simpler inequality:

$$(1+\varepsilon(n))^{-1} \cdot (\log\log n)^{3/2} \leq c \cdot (\log n)^{\frac{\varepsilon(n)}{1+\varepsilon(n)}}.$$

After another logarithm operation, we obtain the following inequality:

$$-\log(1+\varepsilon(n)) + 3/2 \cdot \log\log\log n \leq \frac{\varepsilon(n)}{1+\varepsilon(n)} \cdot \log\log n + \log c.$$

Again, assuming that $\log(1+\varepsilon(n)) > 0$ if n tends to infinity, to prove the above inequality, it is sufficient to show that

$$3/2 \log\log\log n \leq \frac{\varepsilon(n)}{1+\varepsilon(n)} \cdot \log\log n.$$

Thus we are able to bound the function $\varepsilon(n)$ as

$$\varepsilon(n) \geq \frac{3/2 \log\log\log n}{\log\log n - 3/2 \log\log\log n}.$$

Observe that $\varepsilon(n) > 0$ for $n \geq 16$, thus both assumptions that we made hold for $|U| \geq 16$ which concludes the proof. □

6. Conclusion

In this paper, we have closed the gap in the approximation hardness of Min-TCO by showing its $\mathcal{LOGAPX}$-completeness. We studied a subproblem of Min-TCO where the number of users interested in a common topic is bounded by a constant d. We showed that, if $d \leq 2$, the restricted Min-TCO is in $\mathcal{P}$ and, if $d \geq 3$, it is $\mathcal{APX}$-complete. The latter result, together with the constant approximation algorithm we presented, allows us to prove lower bounds on approximability of these special instances that match any lower bound known for any problem from the class $\mathcal{APX}$. Furthermore, we studied instances of Min-TCO where for a fixed user the number of interesting topics is bounded by a constant d. We presented a reduction that shows that such instances are $\mathcal{APX}$-hard for $d = 6$. In this reduction, any two users have at most three common topics, thus the reduction shows also that Min-TCO restricted in this way is $\mathcal{APX}$-hard. We also investigated Min-TCO with a bounded number of topics. Here we presented a polynomial-time algorithm for $|T| \leq (1+\varepsilon(|U|))^{-1} \cdot \log\log|U|$ and a function $\varepsilon(n) \geq \frac{3/2 \log\log\log n}{\log\log n - 3/2 \log\log\log n}$. The case where $|T| = \omega(\log\log|U|)$ and $|T| = o(|U|)$ remains to be a challenging open problem.

References

[1] Emmanuelle Anceaume, Maria Gradinariu, Ajoy Kumar Datta, Gwendal Simon, Antonino Virgillito, A semantic overlay for self- peer-to-peer publish/subscribe, in: Proc. of the 26th IEEE International Conference on Distributed Computing Systems (ICDCS 2006), 2006, p. 22.
[2] Dana Angluin, James Aspnes, Lev Reyzin, Inferring social networks from outbreaks, in: Proc. of the 21st International Conference on Algorithmic Learning Theory, ALT 2010, in: Lecture Notes in Computer Science, vol. 6331, Springer-Verlag, 2010, pp. 104–118.
[3] Giorgio Ausiello, Alessandro D'Atri, Marco Protasi, Structure preserving reductions among convex optimization problems, Journal of Computer and System Sciences 21 (1) (1980) 136–153.
[4] Roberto Baldoni, Roberto Beraldi, Vivien Quéma, Leonardo Querzoni, Sara Tucci Piergiovanni, Tera: topic-based event routing for peer-to-peer architectures, in: Proc. of the 2007 Inaugural International Conference on Distributed Event-Based Systems, DEBS 2007, in: ACM International Conference Proceeding Series, vol. 233, ACM, 2007, pp. 2–13.

[5] Reuven Bar-Yehuda, Shimon Even, A linear-time approximation algorithm for the weighted vertex cover problem, Journal of Algorithms 2 (2) (1981) 198–203.
[6] Antonio Carzaniga, Matthew J. Rutherford, Alexander L. Wolf, A routing scheme for content-based networking, in: Proc. of IEEE INFOCOM 2004, 2004, pp. 918–928.
[7] Raphaël Chand, Pascal Felber, Semantic peer-to-peer overlays for publish/subscribe networks, in: Euro-Par 2005 Parallel Processing, in: Lecture Notes in Computer Science, vol. 3648, Springer-Verlag, 2005, pp. 1194–1204.
[8] Miroslav Chlebík, Janka Chlebíková, Inapproximability results for bounded variants of optimization problems, in: Proc. of the 14th International Conference on Fundamentals of Computation Theory, FCT 2003, in: Lecture Notes in Computer Science, vol. 2751, Springer-Verlag, 2003, pp. 27–38.
[9] Gregory Chockler, Roie Melamed, Yoav Tock, Roman Vitenberg, Constructing scalable overlays for pub-sub with many topics, in: Proc. of the 26th Annual ACM Symposium on Principles of Distributed Computing, PODC 2007, ACM, 2007, pp. 109–118.
[10] Nadia Creignou, Sanjeev Khanna, Madhu Sudan, Complexity Classifications of Boolean Constraint Satisfaction Problems, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2001.
[11] Irit Dinur, Venkatesan Guruswami, Subhash Khot, Oded Regev, A new multilayered PCP and the hardness of hypergraph vertex cover, in: Proc. of the 35th Annual ACM Symposium on Theory of Computing, STOC 2003, ACM, New York, NY, USA, 2003, pp. 595–601.
[12] Eran Halperin, Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs, SIAM Journal on Computing 31 (5) (2002) 1608–1623.
[13] Juraj Hromkovič, Algorithmics for Hard Problems. Introduction to Combinatorial Optimization, Randomization, Approximation, and Heuristics, in: Texts in Theoretical Computer Science. An EATCS Series, Springer-Verlag, Berlin, 2003.
[14] Subhash Khot, On the power of unique 2-prover 1-round games, in: Proc. of the 34th Annual ACM Symposium on Theory of Computing, STOC 2002, ACM, New York, NY, USA, 2002, pp. 767–775.
[15] Subhash Khot, Oded Regev, Vertex cover might be hard to approximate to within 2-epsilon, Journal of Computer and System Sciences 74 (3) (2008) 335–349.
[16] Ephraim Korach, Michal Stern, The complete optimal stars-clustering-tree problem, Discrete Applied Mathematics 156 (4) (2008) 444–450.
[17] Ephraim Korach, Michal Stern, The clustering matroid and the optimal clustering tree, Mathematical Programming 98 (1–3) (2003) 385–414.
[18] Rolf Niedermeier, Peter Rossmanith, On efficient fixed-parameter algorithms for weighted vertex cover, Journal of Algorithms 47 (2) (2003) 63–77.
[19] Melih Onus, Andréa W. Richa, Minimum maximum degree publish-subscribe overlay network design, in: Proc. of IEEE INFOCOM 2009, IEEE, 2009, pp. 882–890.
[20] Christos H. Papadimitriou, Mihalis Yannakakis, Optimization, approximation, and complexity classes, Journal of Computer and System Sciences 43 (3) (1991) 425–440.
[21] Venugopalan Ramasubramanian, Ryan Peterson, Emin Gün Sirer, Corona: A high performance publish-subscribe system for the world wide web, in: Proc. of the 3rd Symposium on Networked Systems Design and Implementation, NSDI 2006. USENIX, 2006.
[22] Daniel Sandler, Alan Mislove, Ansley Post, Peter Druschel, Feedtree: sharing web micronews with peer-to-peer event notification, in: Proc. of the 4th International Workshop on Peer-to-Peer Systems, IPTPS 2005, in: Lecture Notes in Computer Science, vol. 3640, Springer-Verlag, 2005, pp. 141–151.
[23] David P. Williamson, David B. Shmoys, The Design of Approximation Algorithms, Cambridge University Press, 2011.
[24] Shelley Zhuang, Ben Y. Zhao, Anthony D. Joseph, Randy H. Katz, John Kubiatowicz, Bayeux: an architecture for scalable and fault-tolerant wide-area data dissemination, in: Network and Operating System Support for Digital Audio and Video, NOSSDAV 2001, ACM, 2001, pp. 11–20.