Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 223 (2023) 370–372

Procedia Computer Science

www.elsevier.com/locate/procedia

XII Latin-American Algorithms, Graphs and Optimization Symposium (LAGOS 2023)

Fair subgraph selection for contagion containment

(Brief Announcement)

Esther M. Arkin[a], Rezaul A. Chowdhury[a], Mayank Goswami[b,*], Jason Huang[a], Joseph S. B. Mitchell[a], Valentin Polishchuk[c], Rakesh Ravindran[d]

[a]*Stony Brook University, 100 Nicolls Rd, Stony Brook 11794, USA*
[b]*Queens College CUNY, 65-30 Kissena Blvd., Flushing 11367, USA*
[c]*Linköping University, Linköping 581 83, Sweden*
[d]*CUNY Graduate Center, 365 5th Ave, New York 10016, USA*

Abstract

We present a new class of problems where the goal is to select a "fair" subgraph H of a given graph $G = (V, E)$, such that H decomposes into many small components. A subgraph $H \subset G$ is (P, d) fair if every vertex $v \in P$ has the same degree d in H, where $P \subset V$ and $d > 0$ are input parameters.

These problems arise when the goal is to allow individuals to equally participate in activities in such a way that the connected components within an interaction graph, which models potential interactions among people, are of the smallest possible size, so that the spread of the contagion, and the difficulty of contact tracing in case of infection, is minimized. Within a preference graph that models the set of preferred choices for each individual when selecting among available options of where to conduct any particular type of activity (e.g., which gym to attend), we seek to compute the fair subgraph of assignments of individuals to these options, so that the number of people in each connected component ("interaction community") of the resulting subgraph is minimized, and everyone is given the same number of options for every activity.

We show that the fair subgraph selection problem is NP-hard, even for very restricted versions. We then formulate the problem as an integer program, and give a polynomial time computable lower bound on the optimal solution.

1. Introduction

We explore to what extent interactions can be limited largely within social groups, which we call *interaction communities*, while affording a significant degree of normal activities that come with everyday life,

* Corresponding author. Tel.: +1 718-997-3566 ; fax: +1 718-997-3566.
E-mail address: mayank.goswami@qc.cuny.edu

1877-0509
10.1016/j.procs.2023.08.250

such as going to stores, restaurants, gyms, salons, etc. We propose a model in which it is possible for people to go about their activities, while bounding the size of the resulting interaction communities within which a contagion might spread. This proposed approach is meant to be an optimized compromise between the extreme of a complete "lock down" (in which people are unable to perform many/most activities) and a completely open, unrestricted "business as usual", with no limits on freedom of choice among activities.

Of great importance to us is the fact that *all users be treated fairly*, and that the above optimization does not end up favoring some subset of the people. To this end, we demand that the chosen subgraph be "fair", in that each user be given the same freedom to perform the same number of activities as anyone else.

2. Problem: Fair Subgraph Selection for Contagion Containment

We are given a set, P, of people residing within a locality and a set, $A = \{A_1, ..., A_m\}$, of *activities*, with each activity A_i being a set of establishments at which the ith activity can be conducted. All the user preferences are modeled by the following Preference Graph.

Definition 1. *(Preference Graph) A preference graph $G = (V, E)$ is such that $V = P \cup (\bigcup_{i=1}^{m} A_i)$ and $E \subseteq P \times \bigcup_{i=1}^{m} A_i$. P is the vertex set corresponding to people, A_i the set of locations for activity i. E satisfies the condition that $\forall 1 \le i \le m, \forall p \in P, \exists a_i \in A_i$ such that $(p, a_i) \in E$. That is, every person p has at least one preferred establishment a_i for activity i.*

In the above, we assume w.l.o.g that the preference list is homogeneous. That is, every person has some preference to all activities in the neighborhood. We remark that heterogenous lists can be handled by adding virtual nodes. As mentioned, we want to select fair subgraphs, formally defined as follows.

Definition 2 (Fair Subgraph)**.** *Given a graph $G = (V, E)$, disjoint subsets $P, Q \subset V$, and a positive integer $d > 0$, a subgraph $H \subset G$ is set to be (P, Q, d)-fair if for any $v \in P$, v is adjacent to exactly d vertices from Q in H.*

We are now in a position to state our problem.

Definition 3 (FSS-MinMaxCS : Fair Subgraph Selection - Minimize Maximum Component Size)**.** *The **input** to the FSSCC problem is a preference graph $G = (P \cup (\bigcup_{i=1}^{m} A_i), E)$, and an integer $d > 0$. The **output** is a subgraph $H \subset G$ such that H is (P, A_i, d)-fair for every $1 \le i \le m$, and subject to this constraint, the size of $L \cap P$, where L is the largest connected component in H, is as small as possible.*

The fairness condition above means that every person has exactly d options available for every activity. In fact, one could let d depend on the type of activity i, by specifying a sequence of d_is $(1 \le i \le m)$ instead of d. Note that the objective function $L \cap P$ is the number of the people in the largest component of H, since we do not care about activity nodes.

We remark that one may also consider other objective function, such as maximizing the number of components of H, which is equivalent to minimizing the *average* size of a component of H.

In the remainder, we will restrict our attention to the case when $d = 1$, i.e., every user must be assigned *exactly one option for every activity*. As we will see, this version already poses many interesting challenges.

Related Work: In the Independent Cascade (IC) model of epidemics, an infected node infects each of its neighbors independently with probability p. Several works have studied this model for large classes of graphs such as preferential attachment model [1, 2] and the configuration model [3], observing that the size of the outbreak is related to the *largest component* in the random graph resulting when each edge of the graph is kept independently with probability p. As remarked in the recent work [4] these works only consider tree-like graphs that have few or no short cycles; the authors in [4] give an algorithm that predicts the probability and final size of an outbreak by accessing only $O(1)$ nodes in a general graph. Note that this entire line of work *assumes the interaction graph is given*, and then predicts sizes of outbreaks. In contrast, our problem asks to find a matching of users to activities that *results in a small interaction graph to begin with*, on which one can then study any contagion model. Our problem is also related to the critical node deletion problem [5, 6, 7].

3. Theoretical Results

Result 1: Hardness. We prove that not only is the general version NP-hard, but so is a simple special case.

Theorem 4. *MinMaxCS is NP-hard even when there are only two activities, i.e., when $m = 2$. In fact, MinMaxCS is NP-hard for $m = 2$ even when every person has exactly 2 preferred choices for each of the 2 activities.*

Result 2: IP formulation. We show that the following IP captures our problem.

$$
\begin{aligned}
&\min \quad S \\
&\text{s.t.} \\
&S \geq \sum_{j \in P \setminus i} u_{ij} + 1, \quad \sum_{j \in N(i) \cap A_k} y_{ij} = 1, && \forall i \in P, \forall A_k \in A, \\
&u_{ij} \geq y_{ik} + u_{kj} - 1, && \forall i, j \in V, i \neq j, \forall k \in N(i), k \neq j, \\
&u_{ij} \geq y_{ij}, && \forall (i, j) \in E, \\
&u_{ij} = u_{ji},\ y_{ij} = y_{ji}, \quad y_{ij} \in \{0, 1\},\ u_{ij} \in \{0, 1\}, \forall i, j \in V
\end{aligned}
$$

Here: a_i is 1 if i^{th} type of activity is open for business, 0 otherwise. y_{ij} (resp. u_{ij}) is 1 if i is connected to j via an edge (resp. a path), 0 otherwise. Finally, $N(i)$ denotes the set of neighbors of node i.

Result 3: A Lower Bound for the 2-Activity Problem. We provide a lower bound on the optimal solution of the two-activity version ($m = 2$), which we will call MinMaxCS2. Note that $OPT \geq |P|/\min(|A_1|, |A_2|)$. A better lower bound can be derived from a related, but simpler problem of *minimizing the most crowded activity*, termed *MinMaxCrowded*2- For the two activity problem, match people to activities in order to minimize the number of people assigned to the most populated establishment.

Theorem 5. *$OPT(MinMaxCS2) \geq OPT(MinMaxCrowded2)$. Furthermore, $OPT(MinMaxCrowded2)$ can be computed in time polynomial in the input size of the preference graph.*

Acknowledgements Mayank Goswami is supported by National Science Foundation CCF-1910873. Joseph Mitchell and Esther Arkin are partially supported by NSF (CCF-2007275). Valentin Polishchuk is supported by the Swedish Transport Administration and the Swedish Research Council.

References

[1] B. Bollobás, O. Riordan, Robustness and vulnerability of scale-free random graphs, Internet Mathematics 1 (1) (2004) 1–35.
[2] O. Riordan, The small giant component in scale-free random graphs, Combinatorics, Probability and Computing 14 (5-6) (2005) 897–938.
[3] M. Molloy, B. Reed, A critical point for random graphs with a given degree sequence, Random Structures & Algorithms 6 (2-3) (1995) 161–180.
[4] Y. Alimohammadi, C. Borgs, A. Saberi, Algorithms using local graph features to predict epidemics, in: Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), SIAM, 2022, pp. 3430–3451.
[5] M. Yannakakis, Node-and edge-deletion NP-complete problems, in: Proceedings of the tenth annual ACM Symposium on Theory of Computing (STOC), 1978, pp. 253–264.
[6] S. Shen, J. C. Smith, R. Goli, Exact interdiction models and algorithms for disconnecting networks via node deletions, Discrete Optimization 9 (3) (2012) 172–188.
[7] M. Lalou, M. A. Tahraoui, H. Kheddouci, The critical node detection problem in networks: A survey, Computer Science Review 28 (2018) 92–117.