Influence Maximization-Based Event Organization on Social Networks

Cheng-Te Li[(✉)]

National Cheng Kung University, Tainan, Taiwan
chengte@mail.ncku.edu.tw

Abstract. This paper is an ongoing work, and was presented as "Lightening Talk" in the DyNo workshop held at ECML/PKDD 2017. Online event-based social services allow users to organize social events by specifying the themes, and invite friends to participate social events. While the event information can be spread over the social network, it is expected that by certain communication between event hosts, users interested in the event themes can be as many as possible. In this work, by combining the ideas of team formation and influence maximization, we formulate a novel research problem, *Influential Team Formation* (ITF), to facilitate the organization of social events. Given a set L of required labels to describe the event topics, a social network, and the size k of the host team, ITF is to find a k-node set S that satisfying L and maximizing the *Influence-Cost Ratio* (i.e., the influence spread per communication cost between team members). Since ITF is proved to be NP-hard, we develop two greedy algorithms and one heuristic method to solve it. Extensive experiments conducted on Facebook and Google+ datasets exhibit the effectiveness and efficiency of the proposed methods. In addition, by employing the real event participation data in Meetup, we show that ITF with the proposed solutions is able to predict organizers of influential events.

Keywords: Event organization · Team formation
Influence maximization · Influential Team Formation
Social networks

1 Introduction

Team formation [6] in a social network is to find a set of experts such that not only a set of required labels is covered but also team members have lower communication cost with one another (i.e., well-connected in the underlying network). It is apparent that team formation can be applied to many real applications, such as searching for a group of employers to execute a project in a company, and composing an activity group for a cocktail party with particular themes. However, team formation techniques [1,4,7] are not applicable for organizing *Influential Events* in event-based social services (e.g. Meetup[1], Plancast[2], and

---

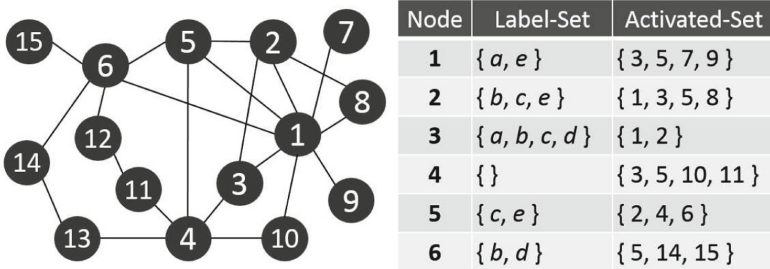[1] www.meetup.com.
[2] www.plancast.com.

Facebook Events). Here we consider influential event organization is to find a set of persons who are interested in the themes of an event, have better social interactions (i.e., lower communication cost), and can attract a large number of people to participate in the event. It is common and realistic to organize influential events. The real-world scenarios on the demand of influential events may include organizing technical conferences, fund raising for earthquake victims, and initiating anti-nuclear campaign. In such scenarios, people attempt to maximize the number of participants since more participants mean a success of the events. One may think *Social Influence Maximization* [5], which aims at finding a set of seeding users such that the number of influenced users can be maximized, seems to be a solution. However, influence maximization techniques [2,3,9] are not applicable for influential event organization because they consider neither the set of required labels, nor the communication between the selected seed nodes.

This work proposes a novel problem, *Influential Team Formation* (ITF), in a social network. Given a set of required labels and the size k of the team, the ITF problem is to find a set S of nodes such that (a) the query label set is covered by the discovered k-node set S, and (b) the *influence-cost ratio* of nodes in S is maximized. We propose the Influence-Cost Ratio (ICR) to quantify the influence spread of the selected k nodes per communication cost. ICR of a node set S is defined as $ICR(S) = \frac{\sigma(S)}{c(S)}$, where influence spread $\sigma(S)$ is the expected number of nodes activated by S while the communication cost $c(S)$ is the sum of all-pair shortest path lengths between nodes in S. A team can derive a higher ICR if the team members can lead to higher influence spread and are well-connected. The ITF problem is challenging since maximizing influence spread contradicts minimizing communication cost. Influence maximization tends to select well-separated nodes because their activated nodes can have less overlapping. But team formation prefers well-connected nodes since they can produce lower communication cost.

It is worthwhile to note that a team is a task-oriented group whose team members not only possess some skill labels to deal with the task, but also well collaborate with each other. Therefore, the team formation problem asks for a set of required skill labels as the input, and expects that the discovered team members are equipped with some of the required skill labels and have good communication among them. Since we aim at forming influential "teams", the selected team members (i.e., seeds) need to rely on a required set of skill labels and be well-connected to have good communication. In addition, "influential" teams also require the team members to be influential, i.e., team members should lead to higher influence spread in the social network. Consequently, the proposed ITF problem is a combination of team formation and influence maximization.

We create an example, as shown in Fig. 1, to exhibit the differences among team formation (TF), influence maximization (IM), and the proposed ITF. This example assumes the set of required labels is $\{a, b, c, d, e\}$ and $k = 3$. TF may select the set $S_{TF} = \{v_1, v_2, v_3\}$ since they cover more required labels and are well-connected. $ICR_{TF} = \frac{7}{3}$. IM will select the set $S_{IM} = \{v_1, v_4, v_6\}$ because

they can lead to highest influence spread. $ICR_{IM} = \frac{11}{5}$. ITF will find the set $S_{ITF} = \{v_1, v_5, v_6\}$ that leads to the highest $ICR_{ITF} = \frac{10}{3}$. It is because not only the union of the activated sets of v_1, v_5, and v_6 leads to the largest activated set (i.e., $\{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_9, v_{14}, v_{15}\}$), but also v_1, v_5, and v_6 are interconnected with a triangle structure in the network.

Node	Label-Set	Activated-Set
1	{ a, e }	{ 3, 5, 7, 9 }
2	{ b, c, e }	{ 1, 3, 5, 8 }
3	{ a, b, c, d }	{ 1, 2 }
4	{ }	{ 3, 5, 10, 11 }
5	{ c, e }	{ 2, 4, 6 }
6	{ b, d }	{ 5, 14, 15 }

Fig. 1. A toy example of a social network (left), and a table (right) that describes the set of required labels possessed by each node and the set of nodes activated by each node. Note that a subset of nodes is shown in the table. Nodes except for v_1 to v_6 do not contain any required label.

In this talk, we will present the ITF problem under the *Independent Cascade* (IC) model. In order to solve the ITF problem. We propose a greedy algorithm with quality guarantee. While the greedy solution is effective but very inefficient, we further develop two greedy methods: *ICR Greedy* (ICR-Greedy) and *Mixed Influence-Cost Greedy* (M-Greedy), and one heuristic method: *Similar Influence Search* (SimIS). ICR-Greedy iteratively selects nodes with highest marginal gain of ICR scores. M-Greedy combines the NewGreedy IM method [3] with the original TF algorithm [6] in an interweaving manner. SimIS integrates Group-PageRank [8] with a best-first search in the social network. To validate the proposed methods, we have simulation-based and prediction-based experiments. The simulation-based experiments conducted on two real social network datasets, Facebook and Google+, and the results show both M-Greedy and SimIS can generate the highest ICR scores with satisfying time efficiency. The prediction-based experiments are conducted using the real event participation data of the event-based social service Meetup. The goal is to validate whether ITF with the proposed solution can truly identify the organizers of influential events based on the required labels of the given event and the social network. The results exhibit satisfying accuracy.

References

1. Anagnostopoulos, A., Becchetti, L., Castillo, C., Gionis, A., Leonardi, S.: Online team formation in social networks. In: Proceedings of ACM International Conference on World Wide Web, WWW 2012, pp. 839–848 (2012)

2. Chen, W., Wang, C., Wang, Y.: Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2010, pp. 1029–1038 (2010)
3. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2009, pp. 199–208 (2009)
4. Kargar, M., An, A.: Discovering top-k teams of experts with/without a leader in social networks. In: Proceedings of ACM International Conference on Information and Knowledge Management, CIKM 2011, pp. 985–994 (2011)
5. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2003, pp. 137–146 (2003)
6. Lappas, T., Liu, K., Terzi, E.: Finding a team of experts in social networks. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2009, pp. 467–476 (2009)
7. Li, C.-T., Shan, M.-K., Lin, S.-D.: On team formation with expertise query in collaborative social networks. Knowl. Inf. Syst. **42**(2), 441–463 (2015)
8. Liu, Q., Xiang, B., Chen, E., Xiong, H., Tang, F., Yu, J.X.: Influence maximization over large-scale social networks: a bounded linear approach. In: Proceedings of ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, pp. 171–180 (2014)
9. Ohsaka, N., Akiba, T., Yoshida, Y., Kawarabayashi, K.-I.: Fast and accurate influence maximization on large networks with pruned Monte-Carlo simulations. In: Proceedings of AAAI Conference on Artificial Intelligence, AAAI 2014, pp. 138–144 (2014)