OCTOPUS: An Online Topic-Aware Influence Analysis System for Social Networks

Ju Fan†, Jiarong Qiu†, Yuchen Li†, Qingfei Meng†, Dongxiang Zhang‡, Guoliang Li†, Kian-Lee Tan∗, Xiaoyong Du†

†Renmin University of China §SMU *NUS #UESTC +Tsinghua University
{fanj, jiarong_qiu, mengqingfei, duyong}@ruc.edu.cn; yuchenli@smu.edu.sg; tankl@comp.nus.edu.sg; zhangdo@uestc.edu.cn; liguoliang@tsinghua.edu.cn

Abstract—The wide adoption of social networks has brought a new demand on influence analysis. This paper presents OCTOPUS that offers social network users and analysts valuable insights through topic-aware social influence analysis services. OCTOPUS has the following novel features. First, OCTOPUS provides a user-friendly interface that allows users to employ simple and easy-to-use keywords to perform influence analysis. Second, OCTOPUS provides three powerful keyword-based topic-aware influence analysis tools: keyword-based influential user discovery, personalized influential keywords suggestion, and interactive influential paths exploration. These tools can not only discover influential users, but also provide insights on how the users influence the network. Third, OCTOPUS enables online influence analysis, which provides end-users with instant results. We have implemented and deployed OCTOPUS, and demonstrate its usability and efficiency on two social networks.

Keywords-Influence Analysis; Social Networks; Topic-Aware

I. INTRODUCTION

The prevalence of online social networks has prompted much attention on social influence analysis due to its potential value in many applications, such as viral marketing and rumor control. For example, an extensively-studied problem is influence maximization (IM) [3], [8], [4], [9], [7], [5] which aims to find a set of seed users that maximizes the number of users influenced by the seeds in a social network (influence spreads).

Recently, the wide adoption of social networks has brought a new demand for topic-aware influence analysis. Evidently, users in a social network may have various interests (which can be represented by topics), and the influence strength between two users (i.e., edge) is often topic-dependent. For example, a user may be influenced by her friend in some topics (e.g., sports), while remaining neutral/unaffected in others (e.g., politics). However, although some existing works [2], [1] have studied topic-aware influence analysis, they have some limitations. The first limitation is about usability. The existing approaches formalize the topics as probabilistic distributions (e.g., (0.7, 0.2, 0.1)) in a latent numerical space, which is very difficult, if not impossible, for end-users to interpret. The second limitation is about efficiency. In topic-aware influence analysis, each edge in the social network is topic dependent. This means that every query corresponds to a distinct topic distribution and results in a different graph. A straightforward solution that employs the traditional IM algorithms over the induced graph during query time is extremely expensive.

To address the limitations, we introduce an online topic-aware social influence analysis system OCTOPUS. Fed with a social network, OCTOPUS offers topic-aware influence analysis services with the following novel features.

First, OCTOPUS provides a user-friendly influence analysis interface (Figure 1): it allows users to employ simple and easy-to-use keywords to perform influence analysis and visualizes the results. The second feature is keyword-based influence analysis that consists of the following components. 1) OCTOPUS supports keyword-based influence maximization: given some keywords describing a topic, it finds the seed users with the maximum influence spread in that topic. 2) OCTOPUS supports personalized influential keywords suggestion that helps end-users explore how a target user influences the network, i.e., discovering the “selling points” of the user. 3) To provide illustrative interpretation, OCTOPUS visualizes the influential paths among users to facilitate interactive exploration. The third feature enables OCTOPUS to support online influence analysis, which gratifies the users with instant results. This is powered by our techniques proposed in [3], [6].

Equipped with these features, OCTOPUS could be useful in various applications. For example, social networks, such as Twitter, have become useful tools for political campaigns, where the tweets contain candidates’ political
standpoints. OCTOPUS can help publicity managers of the candidates in various aspects, including discovering who are the most influential candidates in certain standpoints (e.g., US-China relation), suggesting which standpoint of a candidate influences more people, and exploring the influential path from a candidate to the other. Another application is influence analysis in citation social networks. OCTOPUS can help end-users find influential researchers in an area (e.g., data mining), analyze the influential research contributions of a researcher, and visualize how a researcher influences the community. Last but not the least, OCTOPUS can also facilitate social media marketing, because businesses want to not only discover influential users for viral marketing, but also position their marketing strategy by identifying their influential product features (e.g., high-tech) to attract more users.

In this paper, we demonstrate OCTOPUS in two representative social networks. The first one is an open citation social network. We choose this network, because academic social data can be easily understood by the audience in ICDE 2018. The second network is a commercial social network. We deploy OCTOPUS to provide influence analysis on this network for advertising services, e.g., viral marketing.

To summarize, we make the following contributions. 1) We develop an online topic-aware social influence analysis system OCTOPUS for social network users. 2) OCTOPUS provides an easy-to-use keyword-based interface, and supports powerful online analysis tools. 3) We deploy OCTOPUS on an open academic social network, and a commercial network for system demonstration.

II. SYSTEM IMPLEMENTATION
A. System Overview
OCTOPUS provides a user-friendly keyword-based interface for end users to perform online topic-aware influence analysis. Figure 2 illustrates an architecture of OCTOPUS. Social Network Data. OCTOPUS considers not only the structure but also user-generated content (UGC) of a social network. Specifically, the data fed to OCTOPUS consists of 1) a social graph that models SN users and their relationships (e.g., friendship or follow) and 2) a set of social actions (UGC) from the users, such as reply/retweet in Twitter and citing actions in an academic social network. Topic-Aware Influence Modeling. Based on the social graph and action logs, OCTOPUS devises a topic-aware social influence model. It considers user-to-user influence strength (edge) is topic-dependent, as mentioned in Section I. To improve system usefulness and usability, OCTOPUS introduces keywords, which are extracted from action logs, and utilizes word topic modeling to provide keyword interpretations for topics. See Section II-B for details. User-Friendly Influence Analysis Interface. OCTOPUS provides a keyword-based interface to end-users. It allows users to employ simple and easy-to-use keywords to discover influential users in a topic, analyze personalized influential keywords, and explore influential paths in an online manner. Real-time Influence Analysis Engine. To enable online interaction with end-users, OCTOPUS devises real-time topic-aware influence analysis techniques. Section II-C presents topic-aware influence maximization and Section II-D discusses personalized influential keywords suggestion. We describe influential paths exploration in Section II-E. Note that this paper only provides high-level ideas of these techniques due to the space limit. More details on the algorithms and empirical evaluation can be found in our papers [3], [6].

B. Topic-Aware Influence Modeling
OCTOPUS utilizes the topic-aware independent cascade (IC) model [2]. It associates each graph edge $e = (u, v)$ with a set of activation probabilities $\langle pp_{u,v}^1, pp_{u,v}^2, \ldots, pp_{u,v}^Z \rangle$ over $Z$ topics, where $pp_{u,v}^z$ represents the likelihood that $u$ activates $v$ on topic $z$. It formalizes any item propagated in the social network (e.g., an ad or a product) as a topic distribution $\vec{\gamma} = \{\gamma^1, \gamma^2, \ldots, \gamma^Z\}$ in which $\gamma^z$ is the probability on topic $z$. Given a distribution $\vec{\gamma}$, the overall activation probability $pp_{u,v}$ over edge $e$ can be computed as $pp_{u,v} = \sum_z pp_{u,v}^z \cdot \gamma^z$. After computing $pp_{u,v}$, OCTOPUS utilizes the IC model for computing influence spreads.

However, topics are modeled by probabilistic distributions, which are impossible for end-users to interpret. OCTOPUS further introduces keyword distribution $p(w|z)$ over topic $z$, following convention of topic modeling studies, where $w$ is a keyword. Then, given a set of keywords $W$, we can derive a topic distribution $\vec{\gamma}$, which is captured by $W$, using the Bayesian formula (see [6]), and thus compute the influence spread of $W$. Note that both $pp_{u,v}^z$ and $p(w|z)$ can be derived from the action logs. For example, in an academic network, we extract distinct keywords from paper
titles and take them as $W$. Then, we regard a $v$’s paper citing a $u$’s paper as an item propagated from $u$ to $v$, and use the keywords in these papers to describe the item. Given a set of such items, we can jointly learn $pp_{u,v}$ and $p(w|z)$, e.g., using the Expectation-Maximization technique [2].

C. Keyword-Based Influence Maximization

To improve usability, OCTOPUS supports keyword-based influence maximization: given a set $W$ of keywords corresponding to a topic distribution $\gamma$, it finds the seed users with the maximum influence spread in that topic. For example, consider an academic citation network. End-users can find the set of influential users in data mining by simply typing in the keywords “data mining”. The challenge is the enormous number of potential queries, each of which corresponds to a topic distribution and results in a different probabilistic graph. A naïve solution is to compute $pp_{u,v}$ for each edge given the query and then employ the traditional IM algorithms. Obviously, this solution would be very expensive, and cannot be used for answering online keyword queries.

We develop efficient online algorithms that have a theoretical guarantee. We introduce a best-effort framework that estimates an upper bound of the influence spread for each user and then preferentially computes the exact influence spread for the users with larger upper bounds, so as to prune insignificant users. For effective bound estimation, we devise precomputation based, local graph based, and neighborhood based methods. Moreover, we devise a topic-sample-based algorithm that pre-computes seed sets for some offline-sampled topic distributions. Then, we use the samples to better estimate upper and lower bounds for pruning instead of directly answering the query, which also achieves theoretical guarantees. For algorithm details and performance evaluation, please refer to our paper [3].

D. Personalized Influential Keywords Suggestion

OCTOPUS helps explore how a target user in the social graph influences the network, i.e., discovering the “selling points” of the user. For example, a researcher may publish papers covering multiple topics, such as “clustering”, “rule mining”, “social network”, etc. It is desirable for the researcher to evaluate her most influential research contributions that affect more people in the academic community.

To support this useful feature in OCTOPUS, we introduce an algorithmic problem called personalized influential keywords suggestion in [6]: given a target user, it suggests a $k$-sized keyword set that maximizes the target user’s influence. Note that the suggested keywords are not from a predefined ontology. Instead, we examine all possible $k$-sized keyword sets, each of which corresponds to a topic distribution $\gamma$, as defined in our model (Section II-B), and find the one with the maximum influence spread. Our model can also make sure that the suggested keywords are consistent in topics. The challenge is the complexity of finding the optimal keyword set: we prove that it is not only a NP-hard problem but also NP-hard to be approximated within any constant ratio.

We introduce a sampling-based framework to estimate the influence spread of each $k$-sized keyword set by sampling edges in the social graph, and then select the one with the largest influence spread. Compared with traditional graph sampling methods, such as Monte Carlo sampling and Reverse Random set sampling [8], we devise a lazy propagation sampling algorithm that samples as few edges as possible to improve the efficiency. To achieve real-time influence spread computation, we introduce a novel index structure that maintains “influencers” of uniformly sampled users to avoid online sampling from scratch. We also devise effective pruning and delay materialization techniques for fast influence computation. We have shown that the methods not only have good approximation ratio, but also achieve real-time empirical performance, in our paper [6].

E. Influential Paths Visualization and Exploration

To provide illustrative interpretation of social influence, OCTOPUS visualizes the influential paths to users. For instance, a researcher may want to see how many people she has influenced, and, especially, how she influences them. To answer such questions, graph-based visualization that shows the direct/indirect influential paths could be helpful. The researcher can not only explore the paths to examine the “influences”, but also discover whether the influenced users form different “clusters”, which may represent different aspects of her research contributions. To support online influential paths visualization and exploration, OCTOPUS employs the maximum influence arborescence (MIA) [4]. The idea is to restrict all possible influence paths of user $u$ to a local tree structure rooted at $u$, where each path from $u$ to $v$ is the one with the largest activation probability among all $u$-to-$v$ paths. It further ignores all paths with the probabilities less than a threshold. We utilize d3js (https://d3js.org/) to visualize the paths and interact with the end-users.

III. Demo Scenarios

We demonstrate OCTOPUS on two social networks. The first one is an open citation network ACMCite (ACM-Citation-network-V8 from https://aminer.org/citation). We choose this network, because academic social data can be easily understood by the audience in ICDE 2018. Specifically, we process the raw data to construct a social graph with researchers and the citation relationships among the researchers. We take the papers as well as their citations as action logs. The second one is a commercial social network QQ from Tencent (http://www.qq.com/). Collaborating with Tencent, we deploy OCTOPUS to provide influence analysis for advertising services, e.g., viral marketing. The social graph consists of QQ users and their friendship. We focus on the users’ actions related to e-commerce products. For example, user $u$ posts an URL of iPhone X, and her friend $v$
Scenarios 1 - Keyword-Based Influential User Discovery.
Suppose that a user wants to analyze the influential researchers in data mining. Using OCTOPUS, the user can identify influential users in different "aspects". For example, "Rakesh Agrawal" and "Jiawei Han" are well known to be influential in data mining, frequent pattern mining, association rule mining, and frequent pattern mining. "Ian H. Witten" develops a popular data mining software WEKA. "John H. Holland" is a pioneer in genetic algorithms and adaptive systems. OCTOPUS can provide a set of keywords extracted from paper titles of the researcher, such as "network evolution", "small-world phenomenon", "link prediction". Obviously, these keywords are highly related to social networks, which capture the influential contributions of "Jure Leskovec". Moreover, OCTOPUS also provides an illustrative interpretation of keywords using a radar diagram. Suppose that a user selects the keywords "EM algorithm" as shown in Figure 1. Then, a radar diagram on the left-bottom of OCTOPUS interface shows the distribution over topics. For example, "EM algorithm" is very related to AI and machine learning, which also have connections to multimedia and HCI. OCTOPUS can allow an end-user to input keywords "game" and "viral marketing". OCTOPUS can suggest influential users on the topic "game" in the network, and the end-user can decide to push an ad to them. Moreover, OCTOPUS can also suggest influential keywords for a user, such as "Gum", "Strawberry" and "Xylitol".

Scenario 2 - Influential Keywords Suggestion.
Suppose that a user wants to discover the influential research contributions of "Jure Leskovec". OCTOPUS will provide a set of keywords extracted from paper titles of the researcher, for example, "network evolution", "small-world phenomenon", "link prediction". Obviously, these keywords are highly related to social networks, which capture the influential contributions of "Jure Leskovec". Moreover, OCTOPUS also provides an illustrative interpretation of keywords using a radar diagram. Suppose that a user selects the keywords "EM algorithm" as shown in Figure 1. A radar diagram on the left-bottom of OCTOPUS interface shows the distribution over topics. For example, "EM algorithm" is very related to AI and machine learning, which also have connections to multimedia and HCI. OCTOPUS can allow an end-user to input keywords "game" and "viral marketing". OCTOPUS can also suggest influential keywords for a user, such as "Gum", "Strawberry" and "Xylitol".

Scenario 3 - Interactive Influential Paths Exploration.
Suppose that a user wants to explore how "Michael Jordan" influences the community. OCTOPUS will visualize the influential paths (as shown in Figure 1, clipped due to the space limit), where the big yellow node is the queried user and the size of each node represents the effect of the user on influence. The user may find the influenced users roughly form some "clusters", which may represent different groups influenced by "Michael Jordan". OCTOPUS will highlight the paths through the node to help the user check more details. Similarly, OCTOPUS also supports the exploration of how a target user is influenced. For example, when typing in researcher "Archan Ganapathi", the user may find the influencers such as "Michael Jordan", "Michael Stonebraker", etc.

Demo Scenarios on the QQ network. In this network, we deploy OCTOPUS to provide influence analysis for advertising services, i.e., deciding which users in QQ should be pushed with an ad for "viral marketing". OCTOPUS can allow an end-user to input keywords "game" to find influential users on topic "game" in the network, and the end-user can decide to push an ad to them. Moreover, OCTOPUS can also suggest influential keywords for a user, such as "Gum", "Strawberry" and "Xylitol", which indicate the user is more influential for food-related products.

IV. CONCLUSION
In this paper, we introduced our OCTOPUS system to provide topic-aware social influence analysis. We presented the implementation of OCTOPUS and demonstrated its novel features on a user-friendly interface, powerful keyword-based influence analysis tools, and analysis efficiency.

Acknowledgement. This work was partly supported by NSF China (61602488, 61632016), and Tencent Social Ads Rhino-Bird Focused Research Grant. Guoliang Li was supported by the 973 program of China (2015CB358700), NSF China (61632016, 61472198, 61521002, 61661166012), and TAL education.

REFERENCES

1https://youtu.be/s5T6y1OCLMw