Visual-VM: A Social Network Visualization Tool for Viral Marketing

Cheng Long, Raymond Chi-Wing Wong
The Hong Kong University of Science and Technology
{clong, raywong}@cse.ust.hk

Abstract—**The paper presents Visual-VM, a social network visualization tool, where the main focus is to provide utilities for viral marketing (e.g., *influence maximization*). Besides, Visual-VM utilizes the location information of each user (which could be estimated from the user's profile) for social network visualization, which is not used in existing social network visualization tools. Visual-VM also supports common utilities for social network exploration.**

Keywords-**Social network visualization, viral marketing**

I. INTRODUCTION

Viral marketing [1-4] is an advertising strategy which utilizes the "word-of-mouth" effect among the friends in social networks. Specifically, instead of covering *massive* users directly as traditional advertising methods do, viral marketing targets a *limited* number of *initial* users (e.g., by providing *incentives*) and utilizes their social relationships, such as friends, families and co-workers, to further spread the awareness of the product among individuals. Each individual who gets the awareness of the product is said to be *influenced*. The number of all influenced individuals corresponds to the *influence* incurred by the initial users.

The propagation process of viral marketing within a social network can be described as follows. At the beginning, the advertiser targets a set of initial users, e.g., by providing some incentives, which we call *seeds*. Then, the seeds initiate the diffusion process of the product information in the social network. Many models studying how the above diffusion process works have been proposed. Among them, the *Independent Cascade Model* (IC model) [5] and the *Linear Threshold Model* (LT model) [6] are the two widely-used models.

As mentioned above, in a viral marketing campaign, a company (advertiser) first targets a limited number of seeds and then these seeds would initiate the diffusion process of the product information in the social network. Thus, the most critical issue for viral marketing is to decide which users should be targeted as seeds at the beginning.

To answer this question, existing studies on viral marketing mainly consider the following two scenarios.

- The budget of how many seeds could be targeted is given, e.g., k, and the goal is to maximize the influence resulted from the diffusion process initiated by the seeds. The seed selection problem in this scenario is called **influence maximization** [1, 2].
- The influence requirement has been specified, e.g., at least J users should be influenced, and the goal is to minimize the number of seeds to be used. The seed selection problem in this scenario is called **seed minimization** [3, 4].

On the other hand, *social network visualization* [7-9], which refers to the techniques and implementations for visualizing a social network, have attracted much attention. For example, in [9], the authors designed a tool called Vizster which depicts the social network in a way such that users can view, explore, navigate and analyze the social network easily. However, all these social network visualization tools do not provide the utilities for facilitating viral marketing campaigns.

In this paper, we develop a social network visualization tool called *Visual-VM*, which not only provides common utilities for visualizing, exploring and analyzing the social network, but also provides utilities for viral marketing campaigns, e.g., influence maximization. Besides, Visual-VM provides the utility for simulating the two widely-used diffusion models, the IC model and the LT model. To the best of our knowledge, Visual-VM is the first social network visualization tool that provides utilities for viral marketing.

The social network visualization method employed by Visual-VM is based on an intuitive idea that one can simply depict the social network on a world map based on the users' location information which is available in the users' profiles. In the literature of social network visualization, how to design the layout of the users when depicting the social network is a critical issue. Existing social network visualization methods usually use some topological structure such as tree for this purpose. Instead, Visual-VM utilizes the spatial location of the users which has two advantages. First, it is simple and fast. Second, the community structures within the social network are usually maintained when depicting the social network since people who live near to one another usually form a community and these people would also appear near to one another in the visualized social network.

In the following, we provide some background knowledge in Section II, introduce the design and implementation details of Visual-VM in Section III, and demonstrate Visual-VM in Section IV. We conclude our paper in Section V.

II. BACKGROUND

A. Social Network Visualization

In the literature, people usually present social networks as graphs, where each user in the social network is represented by

a node and each relation between two users is represented by an edge between the nodes corresponding to these two users. Thus, to visualize a social network, they depict a graph which represents the social network [7-9]. One critical issue is how to design the layout of the nodes in the graph such that the graph could be viewed in a clear and meaningful way. Several methods have been proposed for this purpose, e.g., random layout, force-based layout and tree layout. The random layout, though simple, suffers from several drawbacks, e.g., the graph depicted in this way could be messed up and the community structures within the social network are probably destroyed. The force-based layout simulates a physical system where the nodes are regarded as *repelling* objects and the edges are regarded as *springs*. The repelling force of the nodes pushes the nodes apart while the edges hold the nodes together, which results in an equilibrium of the layout of the nodes. The tree layout corresponds to a breadth-first-search from a node in the graph. All these methods do not utilize the *spatial* information embedded by the users for social network visualization as Visual-VM does.

B. Viral Marketing

Motivated by the fact that social network plays a fundamental role in spreading ideas and innovations, Domingoes and Richardson [10, 11] proposed to use social networks for marketing use, which is referred to as viral marketing. Mainly two application scenarios of viral marketing, namely influence maximization and seed minimization, have been extensively studied.

Influence maximization [1, 2]. The *influence maximization* problem is a discrete optimization problem [1]: *Given a social network $G(V, E)$ and an integer k, the problem is to find k seeds such that the incurred influence is maximized.* Kempe et al. [1] proved that the influence maximization problem is NP-hard for both the IC model and the LT model.

The influence maximization problem has been extended into the setting with multiple products instead of a single product. In [2], Datta et al. proposed to promote multiple products via marketing in the same social network, where each user is assumed to have a *seed constraint* which specifies the maximum times of being selected as seeds for the products. Given the quota of the number of the seeds for each product, the problem is to select a set of seeds for each product such that the seed constraint of each user is satisfied and the overall influence incurred by the seeds (for all products) is maximized. Note that this problem based on multiple products is more general than that based on single product. Thus, in the following, we focus on the influence maximization problem with multiple products. Two algorithms, called *Greedy* and *Fair-Greedy*, are designed for the influence maximization problem in [2]. Greedy is a simple greedy algorithm, which repeatedly selects a seed that incurs the largest *gain* in terms of influence. Fair-Greedy is a variation of Greedy, which tries to balance the influence corresponding to each product when selecting the seeds for the products. We implemented both Greedy and Fair-Greedy in Visual-VM as the influence maximization utility.
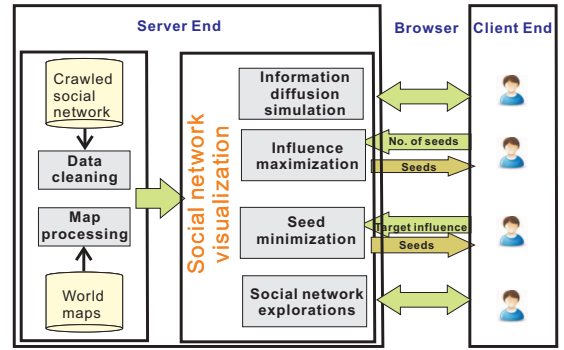

Fig. 1. The architecture of Visual-VM

Seed minimization [3, 4]. The *seed minimization* was first proposed in our previous work [3, 4], where we consider the IC model and the LT model as the underlying diffusion models. Specifically, the problem is defined as: *given a positive number J, the problem is to select a set of seeds such that the influence incurred by these seeds is at least J and the size of the seed set is minimized.* In [3], we proved that the seed minimization problem is NP-hard and thus we developed an approximate algorithm called *SM-Greedy* which provides a certain degree of error guarantee.

III. DESIGN AND IMPLEMENTATIONS

A. Architecture

We present the architecture of Visual-VM in Figure 1. Visual-VM is a web application which has the server end and the client end. The core of Visual-VM is at the the server end, which involves two stages. The first one is the data preprocessing stage. At this stage, it performs a *data cleaning* process on the crawled social network and estimates the coordinates of each user. These coordinates are used for depicting the social network. Besides, it performs a *map processing* task for preparing the map data which is for social network visualization use. The second one is the serving stage, which includes five utilities: *social network visualization*, *information diffusion simulation*, *influence maximization*, *seed minimization* and *social network exploration*. The social network visualization utility is the basis for each of the other utilities. The implementation details of each of these utilities will be introduced in Section III-B.

B. Implementations

Data Cleaning. The major data used for developing Visual-VM is the social network data which includes a set of users each of which has its profile and a set of relations among the users. We crawled our social network data from Twitter, which includes 6,499 users and 884,676 relations among the users. The profile of each user includes 25 types of information, e.g., his/her name, age, living city and followers. Since we want to depict each user on a world map, which relies on the locations (coordinates) of the users, we need to collect/estimate the coordinates of each user first. To this end, we utilize the living city information of each user, since it is possible to estimate the location of each user based on the city he/she

lives in. We have three steps. First, we perform a data cleaning task on the city information of each user since multiple users might refer to the same city with different names (e.g., both "LA" and "Los Angeles" corresponds the same city). Second, we crawl the coordinates of the cities from the website [12]. Third, we assign the users with the coordinates such that the locations of all users living in the same city together form a circle in the area corresponding to that city on the map.

Map Processing. We have two types of maps, namely the geographical world map and the political world map. As will be introduced later, the social network is depicted by layers, which provide different levels of details of the social network. Thus, we prepare the corresponding map for each layer, each has a distinct resolution. We used an image compression program called PNG optimizer [13] for this processing step.

Social Network Visualization. In Visual-VM, we have a main frame which displays the world map. On the world map, we depict the social network as follows. For each user in the social network, we plot a solid dot on the world map with its location set to be that of the user. For each relation between two users in the social network, we plot a solid line between the two solid dots corresponding to these two users.

Since a social network is usually in a large scale, we depict the social network with 4 layers. The most detailed layer called the *individual layer* corresponds to the original social network itself. The next layer which is less detailed is called the *city layer*, where the social network corresponds to the set of cities and the set of relations between cities (there exists a relation between two cities if there exists a relation between two users come from each of the two cities). Similarly, we have the *province layer* and the *country layer*. We implemented a zoom-in bar for switching among these layers.

In the context of Visual-VM, each relation between two users has its weight, which indicates the strength of this relation. Thus, when visualizing the social network, we provide the option such that one can specify a threshold and only those edges with weights at least the threshold will be displayed and used. This functionality was implemented in the "resolution index" at the bottom of the main frame.

Besides, we provide a small overview map, which shows the rectangle region of the world map (and the social network depicted on it) that enclose the current area displayed in the main frame of Visual-VM. Within the overview map, we have a box with a red boundary which indicates the exact region displayed in the main frame. One can easily navigate the world map by moving this box on the overview map.

In addition, we have a console at the left-side of the main frame, where the query results (e.g., the seeds returned by the influence maximization utility) could be displayed and the advertiser can do some operation on these query results. This console could be hidden/shown by the advertisers easily.

Information Diffusion Simulation. Given a set of seeds, this utility simulates the information diffusion process initiated by these seeds visually. We capture the information diffusion process by a *traversal* from the seeds on the social network.
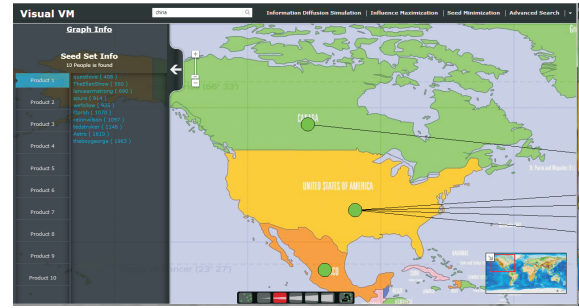

Fig. 2.   The main interface of Visual-VM

Same as the existing studies on viral marketing [1-4], we use *Monte-Carlo sampling* for estimating the influence incurred in the information diffusion process. Besides, we visualize the information diffusion process in Visual-VM by dynamically highlighting those users when they are first influenced and also those edges through which the information diffuses.

Influence Maximization. This utility receives the information of a set of products and the quota of seeds for each product. Advertisers can specify which influence maximization algorithm (Greedy or Fair-Greedy in [2]) to be used and also which diffusion models (the IC model or the LT model) to be adopted. The utility then finds the seeds by executing the influence maximization algorithm chosen. After that, it returns the seeds found and presents them in the console.

Seed Minimization. This utility receives the information of the targeted number of users to be influenced (i.e., influence requirement) as input and returns as few seeds as possible such that the influence incurred by these seeds exceeds the requirement. We plan to use the SM-Greedy algorithm in [3] for implementing this utility. Again, advertisers can choose the diffusion model for the viral marketing.

Social Network Exploration. This component mainly includes keyword search, neighborhood query and advanced search. With the keyword search utility, we can search any elements that match or partially match the keywords and the results will be shown in the console at the left side. With the neighborhood query utility, one can conveniently acquire the neighbors of a given user, which are also shown in the console. With the advanced search utility, one can specify the relation among different searching criteria, e.g., OR and AND, for searching.

IV. USER INTERFACES & DEMONSTRATION

A. User Interfaces

The main interface of Visual-VM is shown in Figure 2. We briefly describe it as follows. The central part corresponds to the main frame where the social network is depicted. The top menu bar includes the main utilities of Visual-VM, namely information diffusion simulation, influence maximization, seed minimization and advanced search. The left-side area is the console where some users (seeds) are shown. The zoom-in bar next to the console controls the visualization layer of the social network. The resolution index is shown at the bottom which controls which edges to be displayed. The overview

map is at the right-bottom area, within which there is a box with a red boundary.

The interfaces of some Visual-VM's utilities will be covered in Section IV-B.

B. Demonstration

Due to the page limit, we demonstrate the following three utilities of Visual-VM only.

Information Diffusion Simulation. We have two steps. First, we need to specify a set of seeds which will initiate the information diffusion process, and choose the underlying diffusion model. To select a set of seeds, we can go to the individual layer which presents the users exactly and click those users in order to select them as seeds. Figure 3 shows the interface for selecting seeds, where each dot represents a user. When clicking a dot (selecting the corresponding user as a seed), the dot's color changes to blue. After we finish selecting the seeds, we click the "finish" button on the top-right corner and the diffusion process will be initiated. Second, the diffusion process could be visualized in a way such that when a user influences another user, the line between the two dots corresponding to these two users is highlighted dynamically. Besides, we can trace the diffusion process iteration by iteration in the console. For example, the console in Figure 3 shows 4 iterations and the users can trace the details of each iteration there.
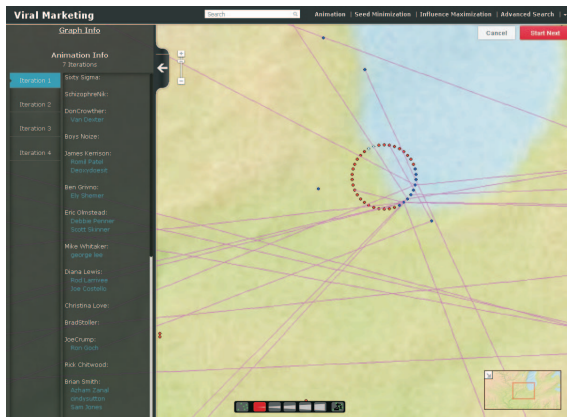

Fig. 3. Demo of information diffusion

Influence Maximization. We have two steps. First, we need to specify the information of the quota of seeds for each product. To do this, we click the influence maximization menu and a dialog box as shown in Figure 4 appears. In this dialog box, we input the quota of seeds for each product, the graph representation method, the influence maximization algorithm and the diffusion model. Second, Visual-VM executes the selected influence maximization algorithm based on the input information and returns a set of seeds for each product, which are presented in the console as shown in Figure 5. Then, the advertiser could explore the set of seeds for each product. For example, he/she can click one of the returned seeds and the profile of this seed would be displayed at the center of the main frame as shown in Figure 5.

Social Network Exploration. One can perform the keyword search, neighborhood query and advanced search. For keyword
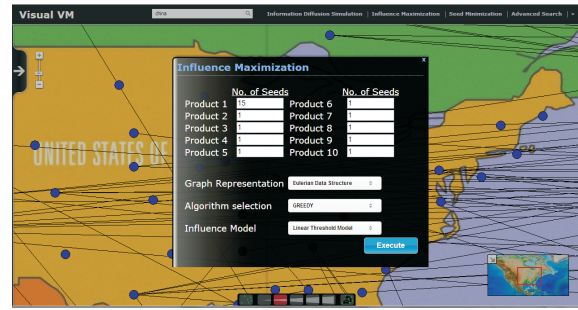

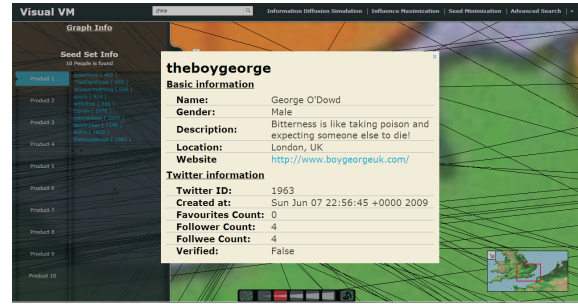Fig. 4. Demo of influence maximization (1)

Fig. 5. Demo of influence maximization (2)

search, we can type several keywords in the textbox located in the menu bar and search the users that match or partially match the keywords. For the neighborhood query, we can right-click a specific user and his neighbors would be shown in the console. For the advanced search, we click the advanced search menu and specify sophisticated searching criteria for searching. Again, all the results will be shown in the console.

V. CONCLUSION

In this paper, we presents a social network visualization tool called Visual-VM for viral marketing use. Visual-VM also supports the common utilities for social network exploration.

REFERENCES

[1] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *SIGKDD*, 2003.

[2] S. Datta, A. Majumder, and N. Shrivastava, "Viral marketing for multiple products," in *ICDM*, 2010.

[3] C. Long and R. C. W. Wong, "Minimizing seed set for viral marketing," in *ICDM*. IEEE, 2011, pp. 427–436.

[4] C. Long and R. C.-W. Wong, "Viral marketing for dedicated customers," *Information Systems*, vol. 46, pp. 1–23, 2014.

[5] M. Granovetter, "Threshold models of collective behavior," *The American Journal of Sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.

[6] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Marketing Letters*, vol. 12, no. 3, pp. 211–223, 2001.

[7] M. Sköld, *Social network visualization*. Thesis, 2008.

[8] F. B. ViAcgas and J. Donath, "Social network visualization: Can we go beyond the graph," in *Workshop on Social Networks*, vol. 4, 2004.

[9] J. Heer and D. Boyd, "Vizster: Visualizing online social networks," in *IEEE Symposium on Information Visualization*. IEEE, 2005, pp. 32–39.

[10] P. Domingos and M. Richardson, "Mining the network value of customers," in *KDD*, 2001.

[11] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *SIGKDD*, 2002.

[12] "itouchmap," http://itouchmap.com/latlong.html, 2011, [Online].

[13] "Pngoptimizer," http://psydk.org/PngOptimizer.php, 2011, [Online].