Maximizing the Spread of Influence through a Social Network

A Literature Review on Online Social Networks

Yang Sun

2634****

SOUTHEAST UNIVERSITY - MONASH UNIVERSITY JOINT GRADUATE SCHOOL

8 May 2015

Maximizing the Spread of Influence through a Social Network

Abstract

In a digital age, social networks help people communicate one another in their daily life. Maximizing the spread of influence among people through a social network is thus important. In this article, I examine four papers from some journals on this topic during the recent ten years, and try to explain what a social network is, and briefly introduce some background information. In order to maximize the influence, what set of individuals we should target, how we can measure the spread of influence, in particular, what methods can be used to solve this problem. I also have evaluation and conclusion of the current states of social networks related search nowadays.

Keywords: social networks, maximize influence, set of individuals

Introduction

A social network is the graph of relationships and interactions within a group of individuals - play a fundamental role as a medium for the spread of information, ideas, and influence among its members [1], like Twitter, Renren Network, Wechat. In the past, people tended to give a call or send text to contact with others, it is quite inconvenient and high cost. With the development of social networks, the ways of traditional communication are replaced by them. Either handling official business or chatting with friends has been transferred on the social networks. For example, Twitter, a microblogging service less than three years old, commands more than 41 million users as of July 2009 and is growing fast [2]. People use Wechat to chat with friends, use Gmail to handle official business, use YouTube to watch videos. We can know that what big power social networks are.

Under this circumstance, in order to make people's life more convenient and efficient, maximizing the spread of influence among people is quite important. Here there is a question, what set of individuals should we target? Using online social networks, you can influence others and also leverage the power of others' influence [3], so we should focus on the right relationships and the right people, social networks provide us tremendous leveraging power - one can create word of mouth. Word of mouth is the passing of information from person to person by oral communication, which could be as simple as telling someone the time of day [4]. In this perspective, data mining is very important. And another question is how we can measure the influence. We can take a social network as a directed graph, and every user that using this social network is a node that is a part of the directed graph. If a user starts to use the social network, the corresponding node becomes active; if a user does not use the social network, the

corresponding node is just inactive. At the same time, the active node can affect its neighbor nodes. Using such a method, we can know easily how much the influence is according to different social networks. The most important question is that what methods we can use to simulate the process of maximizing the influence. There are two models. One is the Linear Threshold Model, and the other is the Independent Cascade Model. I will introduce them in the Body of the Review section.

Scope and Method

In this review report, basically I illustrate what a social network is. In order to maximize the spread of influence through a social network, I focus on the following three questions:

1. What set of individuals should we target?
2. How can we measure the influence?
3. What methods can be used to solve this problem?

However, I focus more on the methods. What methods people use to deal with the spread of influence?

The four papers are all given by my adviser, and the main topic is online social networks. My adviser gives me ten papers, I scan all of the ten papers. At last, I choose four papers that are best related to the three questions.

Body of the Review

According to the introduction, we all know that what a social network is, what a social network's function is, how much a social network affect people' daily life in a digital age. Now we can research that how to maximize the spread of influence through a social network.

What set of individuals should we target?

There are thousands of social networks on the Internet for us to select and use. Different people will select different social networks according to their own interests and characters. For example, college students generally use Renren network, why? Because students can upload the photos that they take, share videos that they like, follow the official account that is full of college information. Business men mainly use Google Gmail to handel their business deals, why? Because Google Gmail provides service for people all over the world, with the development of globalization, business men often need to communicate with partners abroad, Google Gmail is a good medium for them.

Basically Goldenberg et al. [4] conclude that a social network has 4 levels:

Randomly distributed networks: Exponential random graph models of social networks became state-of-the-art methods of social network analysis in the 1980s. This framework has the capacity to represent social-structural effects commonly observed in many human social networks, including general degree-based structural effects commonly observed in many human social networks as well as reciprocity and transitivity, and at the node-level, homophily and attribute-based activity and popularity effects, as derived from explicit hypotheses about dependencies among network ties. Parameters are given in terms of the prevalence of small subgraph configurations in the network and can be interpreted as describing the combinations of local social processes from which a given network emerges. These probability models for networks on a given set of actors allow generalization beyond the restrictive dyadic independence assumption of micro-networks, allowing models to be built from theoretical structural foundations of social behavior.

Scale-free networks: A scale-free network is a network whose degree distribution follows a power law, at least asymptotically. In network theory a scale-free ideal network is a random network with a degree distribution that unravels the size distribution of social groups. Specific characteristics of scale-free networks vary with the theories and analytical tools used to create them, however, in general, scale-free networks have some common characteristics. One notable characteristic in a scale-free network is the relative commonness of vertices with a degree that greatly exceeds the average. The highest-degree nodes are often called "hubs", and may serve specific purposes in their networks, although this depends greatly on the social context. Another general characteristic of scale-free networks is the clustering coefficient distribution, which decreases as the node degree increases. This distribution also follows a power law. The Barabási model of network evolution shown above is an example of a scale-free network.

Large-scale networks: Large-scale network is a term somewhat synonymous with "macro-level" as used, primarily, in social and behavioral sciences, in economics. Originally, the term was used extensively in the computer sciences (see large-scale network mapping).

Complex networks: Most larger social networks display features of social complexity, which involves substantial non-trivial features of network topology, with patterns of complex connections between elements that are neither purely regular nor purely random (see, complexity science, dynamical system and chaos theory), as do biological, and technological networks. Such complex network features include a heavy tail in the degree distribution, a high clustering coefficient, assortativity or disassortativity among vertices, community structure, and hierarchical structure. In the case of agency-directed networks these features also include reciprocity, triad significance profile (TSP, see network motif), and other features. In contrast, many of the mathematical models of networks that have been studied in the past, such as lattices and random graphs, do not show these features.

In a word, this problem's main idea is relationship classification. The links in a social network often reflect social relationships among users. People try to classify the relationships of different social network links based on a small subset of known social relationships and certain social network property. By taking different relationships into account, we further investigate the impact of relationship classification on information propagation problem. After leveraging the social relationships, people exploit these relationships to maximize the information propagation.

How should we measure the influence?

In order to solve this problem, we should firstly set up a social network constructor. An online social network is often composed of users, links, and groups. As in all online social networks, to participate fully in an online social network, a user (often a human being) must register with the site. The user profile collected by the site contains the volunteered information about the users, which could be bogus sometimes. After a user registers in a site, the user can create links to other users in the same social network. Here users form links for various reasons: the users can be a real world acquaintances, or business contacts; they can share some common interests; or they are interested in each other's contents. For a user u, the set of users with whom u has links are called the contacts of the user u.

In a word, a social network is represented as a directed graph. Each user is considered as a node. Each node can be either active (use a social network) or inactive. By the "word-of-mouth" effects, each node's tendency to become active increases monotonically as more of its neighbors become active. Here node can switch to active from inactive, but does not switch in the other direction.

For example, a social network $G = (V, E)$, a labeling $\ell()$ for a small subset $K \subset E$ of edges, and the set of all possible labels C that can be used. Here a labeling $\ell(e)$ of an edge $e = (u, v)$ denotes that the relation between u and v is of type $\ell(e)$. Our objective is to assign each link e in E a label from C such that the accuracy is maximized. Here we define the accuracy as the ratio of the number of links which are labeled correctly to the total number of unlabeled links.

What methods can be used to solve this problem?

This is the most important question to be addressed. There are two models to address this question. One is the Linear Threshold Model and the other is the Independent Cascade Model.

Linear Threshold Model

In this model, a node is influenced by each neighbor according to a weight. Each node has a threshold which is chosen uniformly at random from the interval [0, 1]. A node becomes active if weight is more than threshold.

Independent Cascade Model

In this model, if node *A* succeeds in step *t*, then node *B* becomes active in step *t*+1. Whether or not node *A* succeeds, it cannot make any further attempts to activate node *B* in subsequent rounds (it means independent).The process runs until no more activation is possible.

The Linear Threshold and Independent Cascade Models are two of the most basic and widely-studied diffusion models, but of course many extensions can be considered. There is another way to set up, proposing a general framework that simultaneously includes both of these models as special cases. For the sake of concreteness in the introduction, we will discuss our results in terms of these two models in particular.

Interpretation and Conclusion

In this paper, in order to maximize the spread of influence through a social network, I propose three important questions. As for the first question - what set of individuals should we target? Different social networks have different targeting set of individuals. We need to use efficient and effective methods for link classification in online social networks and for maximizing the influence in such networks. As for the second question - how can we measure the spread of influence? We should set up a model and can take a social network as a directed graph and take a user as a node. We can measure the spread of influence by using such a model. As for the final question -what methods can we use to solve this problem? I introduce two models. One is the Linear Threshold Model, and the other is the Independent Cascade Model. These two models can simulate the process of maximizing the spread of influence through a social network.

References

[1] Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 137-146.

[2] Haewoon K., Changhyun L., Hosung P., & Sue M. (2010). What is Twitter, a Social Network or a News Media? *International World Wide Web Conference Committee*, ACM, 26-30.

[3] Tang, S., Yuan, J., Mao, X., Li, M., Chen, W., & Dai, G. (2011). Relationship classification in large scale online social networks and its impact on information propagation. *Proceedings of IEEE Conference on Computer Communications* (*INFOCOM 2011*), 2291-2299.

[4] Goldenberg, J., Libai, B., & Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3), 211-223.