Research Statement

Yayati Gupta

December 8, 2017

1 Introduction

Networks are all-pervading. We are, both, comprised of them as well as, possessed by them. The brain network and protein-protein interaction networks reside well within us. While we are a part of several networks like friendship networks, collaboration networks, and sexual networks; yet we are surrounded by many networks such as road networks, World Wide Web, and computer networks. Such complex networks have been studied in detail. The study of the complex networks has today become more than just the study of nodes, edges, and their structural organization. A number of dynamic processes occur on top of these complex networks. Diffusion is one such process, the study of which has gained importance in the recent years. The edges in a complex network act as pathways for a number of things to diffuse over it. For example, the axons in a brain network help in the diffusion of electrical signals between neurons; the ties in a friendship network result in the diffusion of information, gossips, and rumors; the connections between computers lead to the transmission of files and viruses.

Today, the diffusion of information through online portals has become a question of great scientific interest. Online Social Networks (OSNs) like Facebook and Twitter provide a platform to fulfill people's penchant for information sharing, arguing and mudslinging. Used by approximately 1.4 billion people worldwide [5], Facebook's "Read, Like and Share" tradition has become a way of living. Understanding the diffusion of information on such networks can help us in diverse ways such as accelerating the spread of useful information, i.e., health-related advices or disaster management related announcements as well as for viral marketing of products and Internet memes[1]. It can also prevent the spread of malicious rumors and misinformation.

Understanding information diffusion is a multidisciplinary affair involving a variety of factors like the structure of these networks, the behavior, and psychology of the underlying individuals and the availability of the information diffusion data. Moreover, this research area is multifaceted where one can study a variety of questions based on dynamic networks, rumor spreading, viral marketing, link prediction, influence measures etc. During my Ph.D., I have identified a few important problems in this domain and have addressed them using a combined approach involving mathematics, sociology and computer science.

[1]Emails, tweets, images, audios, videos, jokes, texts; almost every kind of data can make its way over the Internet. Such digital artifacts spreading from person to person through the Internet are called Internet memes.

2 Dissertation Work

In this section, I briefly describe the key problems I addressed during my Ph.D.

2.1 Diffusion of Internet memes on OSNs

In the initial days of my Ph.D., I addressed the question of better modeling the diffusion of memes on OSNs. The current information diffusion models like linear threshold model and independent cascade model are challenged by the unavailability of parameters required to simulate them. These parameters correspond to the complex behavioral and psychological aspects of individuals and their relationships. For example, *"How many friends do it require to persuade a given individual?"*, or *"How frequently do two given people talk to each other?"*. Given an OSN having millions of nodes and billions of edges, such parameters are very hard to obtain. Hence, most of the research studies assume the values of these parameters to be equal for all the nodes or edges in the network to arrive at well-defined results. However, the correctness of this approach can be questioned. I addressed this issue in the first work of my dissertation.
The real world social networks have a peculiar structure which originates from a particular pattern of interactions between the underlying individuals. For example, people in one community usually do not talk to the people in other communities. This gives rise to community structure in these networks. Moreover, there are some highly influential people in these networks capable of easily reaching others in terms of diffusion of memes. On the other hand, the non-influential people in a network find it hard to make influential people adopt a meme. Such influential nodes occupy the structurally central position in the network, called the core. Remaining people constitute the periphery, the structure originated being called the core-periphery structure. Since the structure of OSNs originates from such particular patterns of communication between the individuals, it can be used to identify the parameters required to simulate the diffusion models. I used the well defined sociological principles like homophily and concepts from influence theory to identify these parameters.
I also proposed a complete framework to study the information diffusion on OSNs [3]. The framework comprises of two parts.

1. A generative model for OSNs.
2. A diffusion model based on the structure of the network.

We validated the framework with the help of diffusion data of the Higgs boson meme on Twitter as well as datasets of several other OSNs. We show that both the components of the above framework are required to achieve an information spreading pattern which mimics a real-world meme spreading pattern.

2.2 Diffusion of OSNs: The Fundamental Law of Social Networks

The second work of my thesis aimed at modeling the diffusion of an OSN in a network [2]. Most of the past research studies have focused on modeling the diffusion of products, innovations, and memes etc. but it is interesting that an OSN on which Internet memes diffuse, also diffuses from person to person through a network. This is because if one observes many of her friends using a

particular OSN, she[2] also tends to use that OSN. However, the diffusion of OSNs differs subtly from that of other entities. The two most important differentiating factors are

1. **adoption of only one OSN at a given time**: A person can adopt multiple Internet memes (or multiple ideas, innovations etc.) at a time, but he does not tend to use multiple OSNs[3] simultaneously.

2. **requirement of being adopted by a lot of friends**: One can adopt a product even if none of her friends use it, but she will adopt an OSN only when a significant fraction of her friends reside on it (since the major objective of an OSN is to connect with friends).

Now since every person wants to use only one OSN which is used by most of his friends, only one OSN tends to dominate the market at a time. This is evident from the history of OSNs as well. Orkut was the dominant OSN in India in 2008. But soon people started shifting to Facebook which led to the demise of Orkut in 2014. From that time, Facebook remained the dominating OSN. Google Plus was launched in 2011 but it could not attract the Facebook members leading to Facebook remaining the dominating OSN thereafter. The incident of people shifting from Orkut to Facebook but not from Facebook to Google Plus has been debated many times but no proper scientific theory has been put forward. We proposed a game theoretical model to explain the shifting behavior of people from one OSN to another. The model is based on the features of the OSNs and the attachment (a function of time) people have developed with these OSNs. We prove that the presence of at least one novel feature in a new OSN is a necessary (though not sufficient) condition for it to make its mark in the market. The theoretical study is backed up with experimental results conducted over the datasets of various OSNs as well as a broad survey on people's usage of these OSNs.

2.3 Virality of Internet Memes

While the first two works of my dissertation aimed at better modeling the diffusion of different kinds of entities on a social network, the third work addressed the quest of finding a method to achieve virality. We employed the understanding of diffusion of entities to induce artificial virality in an Internet meme (viral marketing). Today, Internet memes represent all kinds of information about products, ideas, innovations, startups' advertisements etc. Hence, OSNs have become a major channel to promote the sales of products and increase the visibility of firms, ideas, and innovations etc. However, one needs to utilize this channel strategically in order to achieve virality and make her product, idea or meme viral. One such strategy is selectively targeting the highly influential people (like celebrities, famous political figures etc.) in the network to adopt one's meme. Researchers have proposed a number of methods to identify these influential people. One widely accepted study has shown that the nodes lying in the core (structurally central location in the network) are the most influential in terms of meme diffusion [4]. However, targeting these core nodes with one's meme remains a challenge because of the reluctance of such nodes in adopting a meme. Their high social status makes them less susceptible to a meme's adoption. The third work of my dissertation experimentally discovers the presence of special kind of nodes in a network called pseudo-cores [1].

[2] I alternate between "he" and "she" to refer a non-specific gender in my statement.

[3] until and unless these OSNs have very different aims. For example, Facebook and LinkedIn have different aims. A person can use Facebook for connecting with friends and LinkedIn for maintaining professional contacts. But she mostly will not use Facebook and a similar OSN like MySpace parallelly. The statement has also been validated with the help of a survey conducted in our study.

These nodes are highly influential like the core nodes but easily susceptible unlike core nodes and hence are the most suitable candidates to be targeted for viral marketing.
Another aspect of this work addresses the issue of reachability to core/pseudo-core nodes. Directly approaching a core/pseudo-core node can be a challenging task as compared to reaching them through their friends. We utilize this strategy to intelligently pave a path from a given node to the core/pseudo-core nodes through a chain of friends. We propose two hill climbing algorithms to build such recommendation chains in the network. Hence, this work aims at increasing the practical realization of the ideas proposed for viral marketing.

3 Future Research

I believe that my thesis has laid foundation stones for many important problems holding strong relevance in the real world. However, all the studies, being the first of their kind, encompass wide scope for future research. I have contemplated on several related threads which can help one take these studies from foundation stones to standard theories and practical applications. I have described these related threads in the subsection "The Explorable Threads". In addition to these possible advancements, I am also interested in taking a slight diversion from my Ph.D. work and explore some novel and related problems. Two such diversions have been explained in the subsection "The Unexplored Threads".

3.1 The Explorable Threads

In this subsection, I highlight the possible advancements in my current research work.

3.1.1 Intertwining of Community Structure and Core-Periphery Structure

The first study of my dissertation has highlighted the importance of the structure of a social network in identifying the parameters associated with the information diffusion models. However, the complex intertwining of the community structure and core-periphery structure in these networks has yet not been completely understood. It is important to understand the coexistence of both of these structures in a network to completely determine the diffusion parameters. Today, an extremely large number of datasets are available online which can help us conduct a wide experimental study and observe this intertwining in a large number of networks. A better understanding of this coexistence will help us propose an improved theory and obtain stronger results for information diffusion.
Moreover, it will be interesting to extend this study beyond the boundary of social networks to other kinds of networks as well; like citation networks, road networks, and neural networks etc. One can then observe the difference in the intertwining of these structures in the social networks and another kind of networks.

3.1.2 From Global to Local

The third study proposed in my dissertation requires one to build a recommendation chain for which a person needs to identify his most influential friend in the network. Since we use coreness ("How core a node is?") as a measure of the influential power of a person, the problem boils down to determine the coreness of nodes in a network. However, this determination requires the global information of the network which is generally not available. If one could determine a way to find the

coreness of a node based on its local neighborhood, the practical relevance of the work will boost immensely. Such a study would be a great sell to the viral marketing researchers.

3.1.3 The Fundamental Law of Social Networks: Revisited

My dissertation work has proposed a model for shifting of users from one OSN to another based on several properties of OSNs and their usage. This study, I would say, is in its infant stage which when combined with the tools of machine learning, big data, statistics, and sociology, can help us answer one of the most compelling questions for social network scientists- "Will Facebook ever be overthrown? If yes, when and how?" Given there are so many OSNs entering and exiting the market, there is also an enormous amount of data to attack this problem. This particular thread though time-consuming, I believe will reap success in the long run.

3.2 The Unexplored Threads

In addition to the problems related to my Ph.D. work, I am also interested in exploring some new directions in this field which I describe next.

3.2.1 Information Diffusion on Multilayer Networks

Scientists have shifted from the study of independent social networks to multilayer networks in order to achieve a better understanding of the real-life scenario. Multilayer networks possess several layers where each layer represents a different network. The set of nodes in every layer is the same but the set of edges differ based on the relationship the particular layer represents. Multilayer networks have strong implications for the diffusion of memes on social networks. A meme diffusing on Facebook can be transmitted to Twitter by a person who is active on both these OSNs. This can be seen as the meme diffusing from one layer of the network (representing Facebook) to the other (representing Twitter). Understanding diffusion on multilayer networks is becoming a much-addressed research problem. In addition to better understanding the information diffusion on such networks, I am interested in the study of influential spreaders across various layers in these networks.

3.2.2 Surpassing the Core to Achieve Virality

A growing body of research, including my dissertation work, has focussed on the importance of core nodes in making a meme viral. Multiple discussions of this work with my colleagues and faculty members have yielded a question of a greater interest. Can periphery nodes collaborate to make an information viral without the involvement of core nodes? The question is of practical relevance because of the non-reachability of the core nodes. The problem can be addressed theoretically by designing an algorithm which generates a second core in the network by adding a minimum number of links between the periphery nodes.

4 Conclusion

I have addressed a number of related problems in the domain of complex networks during my Ph.D. The problems open up new directions for the exploration of diffusion in the online world using the tools of sociology, game theory, artificial intelligence, and mathematics. I see these problems as a

few important pieces of a big jigsaw puzzle solving which requires a continued prosperity in the research in these directions.

The study of complex networks is highly interdisciplinary and hence calls for contributions from a large and varied number of tools. Greater the number of tools, better the understanding. Hence, in my future research endeavors, I would like to extend my toolkit to include disciplines like big data, statistics and machine learning in order to better understand the domain of complex networks.

References

[1] Y. Gupta, D. Das, and S. Iyengar. Pseudo-cores: The terminus of an intelligent viral memes trajectory. In *Complex Networks VII*, pages 213–226. Springer, 2016.

[2] Y. Gupta, S. Iyengar, J. S. Saini, and N. Sridhar. Shifting behaviour of users: Towards understanding the fundamental law of social networks. *arXiv preprint arXiv:1507.07838*, 2015.

[3] Y. Gupta, A. Saxena, D. Das, and S. Iyengar. Modeling memetics using edge diversity. In *Complex Networks VII*, pages 187–198. Springer, 2016.

[4] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. Identification of influential spreaders in complex networks. *Nature physics*, 6(11):888–893, 2010.

[5] S. N. Statistics. Statistic brain (july 9, 2014).