A FRAMEWORK FOR INFORMATION DIFFUSION OVER SOCIAL NETWORKS RESEARCH
Outlining Options and Challenges

Juan Yao and Markus Helfert
School of Computing, Dublin City University, Dublin, Ireland

Keywords: Information Diffusion, Social Networks, Social Influence.

Abstract: Information diffusion is a phenomenon in which new ideas or behaviours spread contagiously through social networks in the style of an epidemic. Recently, researchers have contributed a plethora of studies, approaches and theoretical contributions related to various aspects of the diffusion phenomenon. There are many options and approaches. However there are only rare research articles consolidating and reviewing the various options. In this paper, we aim to contribute an overview of the most prominent approaches related to the studies of the diffusion phenomenon. We present a framework and research overview for this area. Our framework can assist researchers and practitioners to identify suitable solutions and understand the challenges in the information diffusion over social networks research.

1 INTRODUCTION

Information diffusion within social networks has known an increased interest in recent years. It is a ubiquitous process in human social networks. Recently, the emergence of mobile, email and online social networking have gradually transformed communication among people. Social interaction through these platforms leaves extensive digital traces by its very nature. The availability of such rich data has lead to a number of empirical and theoretical studies on information diffusion process at a more precise, quantitative level. These studies have shed light on several important principles of such phenomenon. The objective of this paper is to review various recent studies related to information diffusion process in social networks, summarize their findings and possible extensions, and present directions for future research.

2 MODELS & FORMAL DESCRIPTION

While the outcomes of the diffusion processes are easily visible, their inner workings have remained elusive. Models of diffusion process aim to provide a clear picture of how such dynamic process unfolds by defining the way information flows between individuals. Here we consider a collection of probabilistic and game-theoretic diffusion models; summarize some of the challenges faced when modelling the diffusion phenomenon and common approaches to address these challenges.

2.1 The Dynamics of Adoption Process

One of the most common approaches of modelling the adoption process is to explicitly represent the step-by-step dynamics of adoption. Typically, it assumes the dynamic process unfolds in discrete time unit with each individual following certain rule when making adopting decision. A set of individuals are chosen to be initial active set which corresponds to the early adopters of the behaviour. Starting with the initial active set the process then unfolds as follows: at each time step, individuals who were active at previous time step remain active; an individual will be activated according to certain rules such as a threshold number of his neighbours have been activated. The dynamic process continues until no more activation is possible.

In the economics literature, diffusion models have been studied from a game-theoretic perspective (Morris, 2000). This approach builds on work investigating how a new technology A might be spread through a social network of individuals who

Yao J. and Helfert M. (2010).
A FRAMEWORK FOR INFORMATION DIFFUSION OVER SOCIAL NETWORKS RESEARCH - Outlining Options and Challenges.
In *Proceedings of the 5th International Conference on Software and Data Technologies*, pages 491-494
DOI: 10.5220/0002926204910494

are currently users of technology of B. The diffusion process is modelled as the dynamic of a coordination game played on the social network, in which the adoption of a common strategy between players has a higher payoff. In particular, in every time step, each player in a social network has two available choices A and B. Each player receives a positive payoff for each of his neighbours that has the same choice as he does, in addition to an intrinsic benefit that he derives from his choice.

2.2 Modelling Social Influence

An underlying premise many diffusion models build on is that people are influenced by their neighbours in the social network when making adopting decision. Social influence determines to a large extent what we adopt and when we adopt it. One simple way to capture this effect is to assign each individual in the network a threshold value (Granovetter, 1978). The threshold value indicates the personal tendency of an individual to adopt the behaviour when his neighbours do. In addition to the number of adopted friends, how those friends are connected to one another could also have an impact on individual's propensity of adopting. It is argued that (Burt, 2005) if two actors related to the same individual are also related to each other, they have greater power over that individual than if they were unrelated. A primitive approach (Katona, Zubcsek and Sarvary, 2009) to incorporate this idea is to postulate the adoption likelihood of individual increases as a function of the density of relationships among their adopted neighbours.

Another way to encode neighbour influence is by using infection rate, inspired by the epidemic models. In the Independent Cascade Model (Goldenberg, Libai and Mullen, 2001), every time an individual contacts with an active neighbour, he has a constant chance of getting activated. Obviously, a constant infection rate seems not accurate enough. Kempe, Kleinberg and Tardos (2003) refined the Independent Cascade Model to interpret the idea that an individual's receptiveness to influence depends on the past history of interactions with his neighbours. Contrary to a constant rate, an individual's probability of being activated is a function of the set of neighbours have already tried and failed to influence him.

It's commonly accepted that social influence affects adopting decision. However, it is not the only factors that drive information diffusion. Van den Bulte and Stremersch (2004) argue that S-shaped diffusion curves can also result from heterogeneity in the intrinsic tendency to adopt. In the context of new product diffusion, individuals are different with respect to subject matter expertise, strength of opinion, personality traits, media exposure or perceived adopting costs. Efforts have been made to capture this effect. For instance, Hartline, Mirrokni and Sundararajan (2008) model a buyer's decision to buy an item is influenced by the set of other buyers that own the item and the price at which the item is offered.

2.3 Modelling Diffusion of Competing Technologies

Most models discussed above typically focus on the diffusion of one behaviour or technology. What is often ignored is that several different behaviours may coexist in a system at the same time and possibly compete with each other. Some may have a better chance to survive and spread than others. In fact, this scenario frequently arises in the real world. In consumer market, producers of consumer technologies often must introduce a new product into a market where a competitor will offer a comparable product, which makes them vie for sales with competing word-of-mouth cascades. Thus it is important to examine the diffusion of competing technologies in social networks.

As mentioned earlier, such phenomena can be modelled as a coordination game played on the edges of the social network with multiple equilibria. An influential paper by Morris (Morris, 2000) provided a set of elegant game-theoretic characterizations for when these qualitatively different types of equilibria arise in terms of the underlying network topology and the quality of technology A relative to technology B. In recent work, Immorlica *et al.* (2007) incorporates compatibility between technologies and discuss how this effects the diffusion. Their results show that in some cases, for one technology to survive the introduction of another, the cost of adopting both technologies must be balanced with a narrow, intermediate range.

3 EMPIRICAL STUDIES & FINDINGS

Leskovec, Adamic and Huberman's (2006) argues that while above models address the question of how influence spreads in a network, they are based on assumed rather than measured influence effects. The wide variety of rules theoretical diffusion models posed on individuals' behaviour, even if plausible, are often lacking empirical support. Furthermore, most of the diffusion models take a single snapshot of the evolving network and then build upon this static network topology. As such, it becomes unclear how accurately existing models render real-world diffusion phenomena. Cointet and Roth (2007) suggest that future investigations of the diffusion mechanisms should begin with adequate empirical protocols, then propose adapted modelling frameworks. Most of the recent empirical studies on information diffusion focus on addressing the following fundamental questions:

- What are the characteristics of social influence?
- What are the patterns of information cascade?
- How does network structure affect diffusion process?

3.1 Characteristics of Social Influence

If we view the diffusion process as a cascade of social influence, a natural starting point is to understand the local mechanism of the influence. Social influence can be described as the actions of an individual can induce his or her friends to behave in a similar way. Backstrom *et al.* (Backstrom *et al.*, 2006) investigated the membership problem in online communities and measured how propensity of individuals to join a community depends on friends already within the community. Specifically, they measured the joining probability as a function of the number of friends already in the community. The adoption curve exhibits kind of a diminishing returns pattern in which it continues increasing, but more and more slowly, even for large numbers of friends.

This kind of diminishing returns pattern has also been observed in many other studies. For example, Christakis and Fowler (Christakis and Fowler, 2007) studied the spread of obesity in large social network over 32 years and found out that people were most likely to become obese when friends became obese. However, Kleinberg (Kleinberg, 2007) argued that the dependence of adopting probability on the number of friends adopted expressed in this way reflects an aggregate property of the full population, and does not imply anything about an particular individual's response to their friends' behaviour.

Anagnostopoulos *et al.* (Anagnostopoulos *et al.*, 2008) argues that while these studies have established the existence of correlation between user actions and social affiliations, they do not address the source of the correlation. There are factors such as homophily (Mcpherson, Lovin and Cook, 2001) or unobserved confounding variables that can induce statistical correlation between the actions of friends in a social network. Homophily is the tendency for people to choose relationships with people who are similar to them, and hence perform similar actions. Recently, Anagnostopoulos *et al.* (Anagnostopoulos *et al.*, 2008) proposed a statistical test to distinguish social influence from correlation using time series data of user actions. Crandall *et al.* (Crandall *et al.*, 2008) developed techniques for identifying and modelling the interactions between social influence and social selection process and found an elaborate interplay between the two factors.

3.2 Patterns of Information Cascade

Although above studies have shed light on the mechanisms of social influence, the overall patterns by which the influence spreads through social networks have been a mystery. Several recent studies have been conducted to illustrate the existence of cascade and observe patterns of cascading behaviour. In particular, Leskovec et al. (Leskovec, Singh and Kleinberg, 2006) consider information cascades in a recommendation network. According to their observation, the distribution of cascade sizes can be approximated by a heavy-tailed distribution. Generally cascades are shallow but occasional large bursts also occur. Another recent study on cascading behaviour in large blog network (Leskovec *et al.*, 2007) found that blog posts do not have a bursty behaviour; they only have a weekly periodicity.

Liben-Nowell and Kleinberg (Liben-Nowell and Kleinberg, 2008) traced the information cascade process on a global scale by using methods to reconstruct the propagation of massively circulated Internet chain letters. Contrary to predictions that large-scale information spreads widely and reaches many people in very few steps, their results show that the progress of these chain letters proceeds in a narrow but very deep tree-like pattern, continuing for several hundred steps.

3.3 Role of Network Structure

Researchers have long emphasized the important role played by the network structure in determining properties of information diffusion. However, the way such dynamic process is affected by network structure is still poorly understood. How widely does information spread? Does it spread only in local region? Does it spread quickly on a dense network? Several studies on real-world network data have been conducted to address questions like these.

Granovetter (Granovetter, 1973)'s weak ties hypothesis states that weak ties typically act as connectors between different communities or circles of friendship. Using mobile call records, Onnela *et al.* (Onnela *et al.*, 2007) have observed a coupling between interaction strengths and the network's local structure, confirming the weak tie hypothesis. Specifically, they found that weak ties appear to be crucial for maintaining the network's structural integrity, but strong ties play an important role in maintaining local communities. In addition, they investigated how the dynamics of different tie strengths influence the spread of information in the network. They show that the coupling between tie strength and network structure significantly slows the diffusion process, resulting in dynamic trapping of information in communities and find that both weak and strong ties have a relatively insignificant role as conduits for information.

4 CONCLUSIONS

In this paper we summarize three major challenges faced when modelling diffusion phenomenon. We review recent studies that measure diffusion phenomenon empirically. We believe our framework can assist researchers and practitioners to understand the challenges in studying information diffusion over social networks and identify suitable solutions to address those challenges.

REFERENCES

Anagnostopoulos, A., Kumar, R., and Mahdian, M., 2008. Influence and correlation in social networks. In *KDD '08*, pages 7-15, New York, NY, USA. ACM.

Burt, R. S., 2005. *Brokerage and Closure: An Introduction to Social Capital*. Oxford University Press, USA.

Backstrom, L., Huttenlocher, D., Kleinberg, J., and Lan, X., 2006. Group formation in large social networks: membership, growth, and evolution. In *KDD '06*, pages 44-54, New York, NY, USA. ACM.

Cointet, J. P. and Roth, C., 2007. How realistic should knowledge diffusion models be? *Journal of Artificial Societies and Social Simulation*, 10(3).

Christakis, N. A. and Fowler, J. H., 2007. The spread of obesity in a large social network over 32 years. *N Engl J Med*, 357(4):370-379.

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., and Suri, S., 2008. Feedback effects between similarity and social influence in online communities. In *KDD '08*, pages 160-168, New York, NY, USA. ACM.

Granovetter, M. S., 1973. The strength of weak ties. *The American Journal of Sociology*, 78(6):1360-1380.

Granovetter, M., 1978. Threshold models of collective behavior. *The American Journal of Sociology*, 83(6):1420-1443.

Goldenberg, J., Libai, B., and Mullen, E., 2001. Talk of the network: a complex systems look at the underlying process of word-of-mouth. Marketing Letters, 12(3):211-223.

Hartline, J., Mirrokni, V., and Sundararajan, M., 2008. Optimal marketing strategies over social networks. In *WWW '08*, pages 189-198, New York, NY, USA. ACM Press.

Immorlica, N., Kleinberg, J., Mahdian, M., and Wexler, T., 2007. The role of compatibility in the diffusion of tecfhnologies through social networks. In Proc. 8th EC, pages 75-83.

Kempe, D., Kleinberg, J., and Tardos, Ã., 2003. Maximizing the spread of influence through a social network. In *KDD '03*, New York, NY, USA. ACM.

Kleinberg, J., 2007. *Cascading Behavior in Networks: Algorithmic and Economic Issues*. Cambridge University Press.

Katona, Z., Zubcsek, P. P., and Sarvary, M., 2009. Network effects and personal influence: Diffusion of an online social network. Working paper.

Leskovec, J., Adamic, L. A., and Huberman, B. A., 2006. The dynamics of viral marketing. In *EC '06: Proceedings of the 7th ACM conference on Electronic commerce*, pages 228-237, New York, NY, USA. ACM.

Leskovec, J., Singh, A., and Kleinberg, J., 2006. Patterns of influence in a recommendation network. pages 380-389.

Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., and Hurst, M., 2007. Cascading behavior in large blog graphs.

Liben-Nowell, D. and Kleinberg, J., 2008. Tracing information flow on a global scale using internet chain-letter data. *Proceedings of the National Academy of Sciences*, 105(12):4633-4638.

Morris, S., 2000. Contagion. *Review of Economic Studies*, 67:57-68.

Mcpherson, M., Lovin, L. S., and Cook, J. M., 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415-444.

Onnela, J. P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.-L., 2007. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332-7336.

Van den Bulte, C., Stremersch, S., 2004. Social Contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Science*, 23(4):530-544.