Chapter 4

A SURVEY OF MODELS AND ALGORITHMS FOR SOCIAL INFLUENCE ANALYSIS

Jimeng Sun
IBM TJ Watson Research Center, USA
jimeng@us.ibm.com

Jie Tang
Tsinghua University, China
jietang@tsinghua.edu.cn

Abstract Social influence is the behavioral change of a person because of the perceived relationship with other people, organizations and society in general. Social influence has been a widely accepted phenomenon in social networks for decades. Many applications have been built based around the implicit notation of social influence between people, such as marketing, advertisement and recommendations. With the exponential growth of online social network services such as Facebook and Twitter, social influence can for the first time be measured over a large population. In this chapter, we survey the research on social influence analysis with a focus on the computational aspects. First, we present statistical measurements related to social influence. Second, we describe the literature on social similarity and influences. Third, we present the research on social influence maximization which has many practical applications including marketing and advertisement.

Keywords: Social network analysis, Social influence analysis, Network centrality, Influence maximization

Social influence refers to the behavioral change of individuals affected by others in a network. Social influence is an intuitive and well-accepted phenomenon in social networks[22]. The strength of social influence depends on many factors such as the strength of relationships between people in the networks, the network distance between users, temporal effects, characteristics of networks and individuals in the network. In this chapter, we focus on computational aspect of social influence analysis and describe the measures and algorithms related to it. More specifically, we aim at qualitatively or quantitatively measuring the influence levels of nodes and edges in the network.

First we present standard measures and concepts of social networks and their connection to social influence measures such as centrality, closeness and betweenness. These measures are fundamental concepts about social network analysis, and are also deeply related to the importance or influence of nodes or edges in the networks.

Second, we present the qualitative and quantitative social influence analysis and applications, which has been well studied in sociology research. Much of the work focuses on differentiating social correlation and social influence. Many qualitative models and tests have been proposed to explain social phenomena in social networks. However, most studies are limited to smaller scale data sets and macro-level observation, partly because of the lack of high-quality longitudinal data on social networks.

Finally, we survey influence maximization techniques, which go beyond simple statistic measures such as centrality. We also present applications of influence maximization. These include methods for predicting customer behavior and online advertising through viral marketing.

1. Influence Related Statistics

A social network is modeled as a graph $G = \{V, E\}$, where V is the set of nodes, and E is the set of edges. As is the convention, the links correspond to actors (people) and the links correspond to social relationships. At the local level, social influence is a directional effect from node A to node B, and is related to the edge strength from A to B. On a global level, some nodes can have intrinsically higher influence than others due to network structure. These global measures are often associated with nodes in the network rather than edges. We next present the concepts and measures at a local and global level respectively.

1.1 Edge Measures

Edge measures relate the influence-based concepts and measures on a pair of nodes. This explains the simple influence-related processes and interactions between individual nodes.

Tie strength. According to Granovetter's seminal work [32], the tie strength between two nodes depends on the overlap of their neighborhoods. In particular, the more common neighbors that a pair of nodes A and B may have, the stronger the tie between them. If the overlap of neighborhoods between A and

B is large, we consider A and B to have a strong tie. Otherwise, they are considered to have a weak tie. We formally define the strength $S(A, B)$ in terms of their Jaccard coefficient.

$$S(A, B) = \frac{|n_A \cap n_B|}{|n_A \cup n_B|}$$

Here, n_A and n_B indicate the neighborhoods of A and B, respectively. Sometimes, the tie strength is defined under a different name called embeddedness. The embeddedness of an edge is high if two nodes incident on the edge have a high overlap of neighborhoods. When two individuals are connected by an embedded edge, it makes it easier for them to trust one another, because it is easier to find out dishonest behavior through mutual friends [33]. On the other end, when embeddedness is zero, two end nodes have no mutual friends. Therefore, it is riskier for them to trust each other because there are no mutual friends for behavioral verification.

A corollary from this tie strength is the hypothesis of *triadic closure*. This relates to the nature of the ties between sets of three actors A, B, and C. If strong ties connect A to B and A to C, then B and C are likely to be connected by a strong tie as well. Conversely, if A-B and A-C are weak ties, B and C are less likely to have a strong tie. Triadic closure is measured by the clustering coefficient of the network[35, 67]. The clustering coefficient of a node A is defined as the probability that two randomly selected friends of A are friends with each other. In other words, it is the fraction of pairs of friends of A that are linked to one another. This is naturally related to the problem of triangle counting in a network. Let n_Δ be the number of triangles in the network and $|E|$ be the number of edges. The clustering coefficient is formally defined as follows:

$$C = \frac{6n_\Delta}{|E|}$$

The naive way of counting the number of triangles n_Δ is expensive. An interesting connection between n_Δ and the eigenvalues of the network was discovered by Tsourakakis [66]. This work shows that n_Δ is approximately equal to the third-moment of the eigenvalues (or $\sum \lambda_i^3$, where λ_i is the ith eigenvalue). Given the skewed distribution of eigenvalues, the triangle counts can be approximated by computing the third moment of only a small number of the top eigenvalues. This also provides an efficient way for computing the clustering coefficient.

Weak ties. When the overlap of the neighborhoods of A and B is small, the connection A-B is considered to be a weak tie. When there is no overlap, the connection A-B is a local bridge [32]. In the extreme case, the removal of A-B may result in the disconnection of the connected component containing A and

B. In such a case, the connection A-B may be considered a global bridge. It may be argued that in real networks, global bridges occur rarely as compared to local bridges. However, the effect of local and global bridges is quite similar.

Edge Betweenness. Another important measure is the edge betweenness, which measures the total amount of flow across the edge. Here, we assume that the information flow between A and B are evenly distributed on the shortest paths between A and B. Freeman [27, 28] first articulated the concept of betweenness in the context of sociology. One application of edge betweenness is that of graph partitioning. The idea is to gradually remove edges of high betweenness scores to turn the network into a hierarchy of disconnected components. These disconnected components will be the clusters of nodes in the network. More detailed studies on clustering methods are presented in the work by Girvan and Newman [29].

1.2 Node Measures

Node-based centrality is defined in order to measure the importance of a node in the network. Centrality has attracted a lot of attention as a tool for studying social networks [28, 9]. A node with high centrality score is usually considered more highly influential than other nodes in the network. Many centrality measures have been proposed based on the precise definition of influence. The main principle to categorize the centrality measures is the type of random walk computation involved. In particular, the centrality measures can be grouped into two categories: radial and medial measures [9]. Radial measures assess random walks that start or end from a given node. On the other hand, medial measures assess random walks that pass through a given node. The radial measures are further separated into volume measures and length measures based on the type of random walks. Volume measures fix the length of walks and find the volume (or number) of the walks limited by the length. Length measures fix the volume of the target nodes and find the length of walks to reach the target volume. Next we introduce some popular centrality measures based on these categories.

Degree. The first group of the centrality measures is that of the radial and volume-based measures. The simplest and most popular measure in this category is that of degree centrality. Let A be the adjacency matrix of a network, and $deg(i)$ be the degree of node i. The degree centrality c_i^{DEG} of node i is defined to be the degree of the node:

$$c_i^{DEG} = deg(i).$$

One way of interpreting the degree centrality is that it counts the number of paths of length 1 that starts from a node. A natural generalization from this perspective is the $K - path$ centrality which is the number of paths of length at most k that start from a node.

Another class of measures are based on the diffusion behavior in the network. The Katz centrality [40] counts the number of walks starting from a node, while penalizing longer walks. More formally, the Katz centrality c_i^{KATZ} of node i is defined as follows:

$$c_i^{KATZ} = e_i^T(\sum_{j=1}^{\infty}(\beta A)^j)\mathbf{1}$$

Here, e_i is a column vector whose ith element is 1, and all other elements are 0. The value of β is a positive penalty constant between 0 and 1.

A slight variation of the Katz measure is the Bonacich centrality [8] which allows for negative values of β. The Bonacich centrality c_i^{BON} of node i is defined as follows:

$$c_i^{BON} = e_i^T(\frac{1}{\beta}\sum_{j=1}^{\infty}(\beta A)^j)\mathbf{1}$$

Here the negative weight allows to subtract the even-numbered walks from the odd-numbered walks which have an interpretation in exchange networks [9]. The Katz and the Bonacich centralities are special cases of the Hubbell centrality [37]. The Hubbell centrality c_i^{HUB} of node i is defined to be

$$c_i^{HUB} = e_i^T(\sum_{j=0}^{\infty}X^j)\mathbf{y}$$

Here, X is a matrix and y is a vector. It can be shown that $X = \beta A$ and $y = \beta A\mathbf{1}$ lead to the Katz centrality, and $X = \beta A$ and $y = A\mathbf{1}$ lead to the Bonacich centrality. The eigenvector centrality [7], the principal eigenvector of the matrix A, is related to the Katz centrality: the eigenvector centrality is the limit of the Katz centrality as β approaches $\frac{1}{\lambda}$ from below [9].

Closeness. The second group of the centrality measures is that of the radial and length based measures. Unlike the volume based measures, the length based measures count the length of the walks. The most popular centrality measure in this group is the Freeman's closeness centrality [28]. It measures the centrality by computing the average of the shortest distances to all other nodes. Then, the closeness centrality c_i^{CLO} of node i is defined as follows:

$$c_i^{CLO} = e_i^T S\mathbf{1}.$$

Here S be the matrix whose (i, j)th element contains the length of the shortest path from node i to j and $\mathbf{1}$ is the all one vector.

Node Betweenness. As is the case for edges of high betweenness, nodes of high betweenness occupy critical positions in the network structure, and are therefore able to play critical roles. This is often enabled by a large amount of flow, which is carried by nodes which occupy a position at the interface of tightly-knit groups. Such nodes are considered to have high betweenness. The concept of betweenness is related to nodes that span *structural holes* in a social network. We will discuss more on this point slightly later.

Another popular group of the centrality measures is that of *medial* measures. It is called ‘medial’ since all the walks *passing through* a node are considered. The most well-known centrality in this group is the Freeman’s betweenness centrality [27]. It measures how much a given node lies in the shortest paths of other nodes. The betweenness centrality c_i^{BET} of node i is defined as follows:

$$c_i^{BET} = \sum_{j,k} \frac{b_{jik}}{b_{jk}}$$

Here b_{jk} is the number of shortest paths from node j to k, and b_{jik} be the number of shortest paths from node j to k that pass through node i.

The naive algorithm for computing the betweenness involves all-pair shortest paths. This requires $\Theta(n^3)$ time and $\Theta(n^2)$ storage. Brandes [10] designed a faster algorithm with the use of n single-source-shortest-path algorithms. This requires $O(n+m)$ space and runs in $O(nm)$ and $O(nm+n^2\log n)$ time, where n is the number of nodes and m is the number of edges.

Newman [52] proposed an alternative betweenness centrality measure based on random walks on the graph. The main idea is that instead of considering shortest paths, it considers all possible walks and computes the betweenness from these different walks. Then, the Newman’s betweenness centrality c_i^{NBE} of node i is defined as follows:

$$c_i^{NBE} = \sum_{j\neq i\neq k} R_{jk}^{(i)}.$$

Here $R^{(i)}$ be the matrix whose (j,k)th element $R_{jk}^{(i)}$ contains the probability of a random walk from j to k, which contains i as an intermediate node.

Structural holes. In a network, we call a node a structural hole if it is connected to multiple local bridges. A canonical example is that a person’s success within a company or organization often depends on their access to local bridges [12]. By removing such a person, an “empty space” will occur in the network. This is referred to as a *structural hole*. The person who serves as a structural hole can interconnect information originating from multiple noninteracting parties. Therefore, this person is structurally important to the connectivity of diverse regions of the network. Another interesting point is that the interests of the actor representing a structural hole and of the organization may not be aligned. For the organization, accelerating the information flow between groups could be beneficial, which requires building of bridges. However, this building of bridges would come at the expense of structural hole's latent power of regulating information flow at the boundaries of these groups.

2. Social Similarity and Influence

A central problem for social influence is to understand the interplay between similarity and social ties [20]. A lot of research has tried to identify influence and correlation in social networks from many different aspects: social similarity and influence [2, 20]; marketing through social influence [21, 55], influence maximization [41]; social influence model and practice through conformity, compliance and obedience [18, 24], and social influence in virtual worlds [23, 5].

2.1 Homophily

Homophily [43] is one of the most fundamental characteristics of social networks. This suggests that an actor in the social network tends to be similar to their connected neighbors or "friends". This is a natural result, because the friends or neighbors of a given actor in the social network are not a random sample of the underlying population. The neighbors of a given actor in the social network are often similar to that actor along many different dimensions including racial and ethnic dimensions, age, their occupations, and their interests and beliefs. McPherson et al. [48] provide an extensive review of research in the long and rich history on homophily. Singla et al. [60] has conducted a large-scale experiment of homophily on real social networks, which includes data from user interactions in the MSN Messenger network and a subset of Microsoft Web search data collected in the summer of 2006. They observe that the similarities between friends is significantly larger than a random pairwise sample, especially in attributes such as age, location and query category. This experiment confirms the existence of homophily at a global scale in large online social networks.

The phenomenon of homophily can originate from many different mechanisms:

- *Social influence:* This indicates that people tend to follow the behaviors of their friends. The social influence effect leads people to adopt behaviors exhibited by their neighbors.

- *Selection:* This indicates that people tend to create relationships with other people who are already similar to them;

- *Confounding variables:* Other unknown variables exist, which may cause friends to behave similarly with one another.

These three factors are often intertwined in real social networks, and the overall effect is to provide a strong support for the homophily phenomenon. Intuitively, the effects of selection and social influence lead to different applications in mining social network data. In particular, recommendation systems are based on the selection/social similarity, while viral marketing [21, 55] is based on social influence. To model these different factors, several models have been proposed [36, 20].

Generative models for selection and influence. Holme and Newman [36] proposed a generative model to balance the effects of selection and influence. The idea is to initially place the M edges of the network uniformly at random between vertex pairs, and also assign opinions to vertices uniformly at random. With this initialization, an influence- and selection-based dynamic is simulated. Each step of the simulation either moves an edge to lie between two individuals whose opinions agree (selection process), or we change the opinion of an individual to agree with one of their neighbors (influence process). The results of their simulation confirmed that the selection tend to generate a large number of small clusters, while social influence will generate large coherent clusters. Thus, this interesting model suggests that these two factors both support clusters in the network, though the nature of such clusters is quite different.

Every vertex in the Holme-Newman model [36] at a given time can only have one opinion. This may be an oversimplification of real social networks. To address this limitation, Crandall et al. [20] introduced multi-dimensional opinion vectors to better model complex social networks. In particular, they assumed that there is a set of m possible activities in the social network. Each node v at time t has an m-dimensional vector $v(t)$, where the ith coordinate of $v(t)$ represents the extent to which person v is engaging in activity i. They use cosine similarity to compute the similarity between two people. Similar to the Holme-Newman model, Crandall et al. also propose a more comprehensive generative model which samples a person's activities based on their own history, their neighbors' history, and a background distribution. Crandall's model is arguably more powerful, but also requires more parameters. Therefore more data is required in order to learn the parameters. Finally, they applied their model and conducted a predictive modeling study on wikipedia and live journal datasets. The benefit of the proposed similarity model are still inconclusive.

Quantifying influence and selection. Subsequent to the work in [20], Scripps et al. [58] proposed the formal computational definitions of selection and influence. We formally define selection and influence as follows:

$$Selection = \frac{p(a_{ij}^{t} = 1 | a_{ij}^{t-1} = 0, \langle \mathbf{x}_i^{t-1}, \mathbf{x}_j^{t-1} \rangle > \epsilon)}{p(a_{ij}^{t} = 1 | a_{ij}^{t-1} = 0)}$$

Here, the denominator is the conditional probability that an unlinked pair will become linked and the numerator is the same probability for unlinked pairs whose similarity exceeds the threshold ϵ. Values greater than one indicate the presence of selection.

$$Influence = \frac{p(\langle \mathbf{x}_i^{t}, \mathbf{x}_j^{tT} \rangle > \langle \mathbf{x}_i^{t-1}, \mathbf{x}_j^{t-1} \rangle | a_{ij}^{t-1} = 0, a_{ij}^{t} = 1)}{p(\langle \mathbf{x}_i^{t}, \mathbf{x}_j^{t} \rangle > \langle \mathbf{x}_i^{t-1}, \mathbf{x}_j^{t-1} \rangle | a_{ij}^{t-1} = 0)}$$

Here, the numerator is the conditional probability that similarity increases from time $t-1$ to t between two nodes that became linked at time t and the denominator is the probability that the similarity increases from time $t-1$ to t between two nodes that were not linked at time $t-1$. As with selection, values greater than one indicate the presence of influence.

Based on this definition, Scripps et al. [58] present a matrix alignment framework by incorporating the temporal information to learn the weight of different attributes for establishing relationships between users. This can be done by optimizing (minimizing) the following objective function:

$$\min_W \sum_{t=1}^{T} \|A^t - X^{t-1} W X^{(t-1)\top}\|_F^2 \tag{4.1}$$

where the diagonal elements of W correspond to the vector of weights of attributes and $\|\cdot\|_F$ denotes the Frobenius norm. Solving the objective function (Eq. 4.1) is equivalent to the problem of finding the weights of different attributes associated with users. A distortion distance function is used to measure the degree of influence and selection.

The above method can be used to analyze influence and selection. However, it does not differentiate the influence from different angles (topics). Several theories in sociology [32, 42] show that the effect of the social influence from different angles (topics) may be different. This can be easily understood by observing different social phenomenon for different angles. For example, colleagues have strong influence on one another's work, whereas friends have strong influence on one another's daily life. Thus, there are several challenging problems in terms of differentiating the social influences from different angles (topics). A number of key questions arise in this context. (a) How to quantify the strength of those social influences? (b) How to construct a model and estimate the model parameters for real large networks?

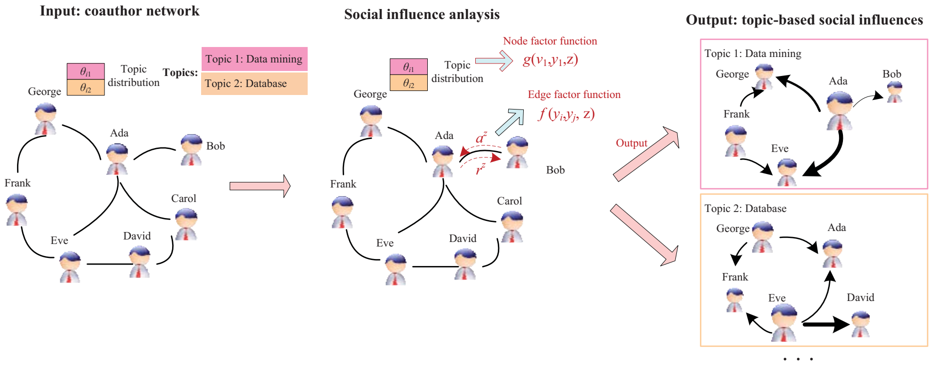

Figure 4.1. Topic-level Social influence Analysis on Co-author Network.

The motivation can be further explained using Example Figure 4.1. The left figure illustrates the input, which is a co-author network of 7 researchers, and the topic distribution of each researcher. For example, George has the same probability (.5) on both topics, “data mining” and “database”; The right figure shows the output of our social influence analysis: two social influence graphs, one for each topic, where the arrows indicate the direction and strength. We see that Ada is the key person on “data mining”, while Eve is the key person on “databases”. Thus, the goal is to effectively and efficiently obtain the social influence graphs for real and large networks.

To address this problem, Tang et al. [63] propose a Topical Factor Graph (TFG) model to formalize the topic-level social influence analysis into a unified graphical model, and present Topical Affinity Propagation (TAP) for model learning. In particular, the goal of the model is to simultaneously capture the user topical distributions (or user interests), similarity between users, and network structure. Figure 4.2 shows the graphical structure of the proposed model. The TFG model has a set of observed variables $\{v_i\}_{i=1}^N$ and a set of hidden vectors $\{\mathbf{y}_i\}_{i=1}^N$, which correspond to the N nodes in the input network.

The hidden vector $\mathbf{y}_i \in \{1, \ldots, N\}^T$ models the topic-level influence from other nodes to node v_i. Each element y_i^z takes the value from the set $\{1, \ldots, N\}$, and represents the node that has the highest probability to influence node v_i on topic z.

For example, Figure 4.2 shows a simple example of an TFG. The observed data consists of four nodes $\{v_1, \ldots, v_4\}$, which have corresponding hidden vectors $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_4\}$. The edges between the hidden nodes indicate the four social relationships in the original network (or edges in the original network).

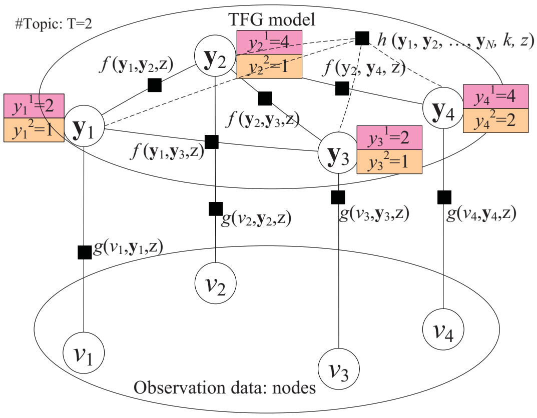

Figure 4.2. Graphical representation of the topical factor graph model. $\{v_1 \ldots v_4\}$ are observable nodes in the social network; $\{y1 \ldots y_4\}$ are hidden vectors defined on all nodes, with each element representing which node has the highest probability to influence the corresponding node; $g(.)$ represents a feature function defined on a node, $f(.)$ represents a feature function defined on an edge; and $h(.)$ represents a global feature function defined for each node, i.e. $k \in \{1 \ldots N\}$.

Three types of feature functions are defined in order to capture the network information: node feature function $g(v_i, \mathbf{y}_i, z)$, edge feature function $f(\mathbf{y}_i, \mathbf{y}_j, z)$, and global feature function $h(\mathbf{y}_1, \ldots, \mathbf{y}_N, k, z)$.

The node feature function g describes the local information on nodes (e.g., attributes associated with users or topical distribution of users). The edge feature function f describes the correlation between users via the edge on the graph model, and the global feature function captures constraints defined on the network. Based on the formulation, an objective function is defined by maximizing the likelihood of the observation.

$$P(\mathbf{v}, \mathbf{Y}) = \frac{1}{Z} \prod_{k=1}^{N} \prod_{z=1}^{T} h(\mathbf{y}_1, \ldots, \mathbf{y}_N, k, z)$$

$$\prod_{i=1}^{N} \prod_{z=1}^{T} g(v_i, \mathbf{y}_i, z) \prod_{e_{kl} \in E} \prod_{z=1}^{T} f(\mathbf{y}_k, \mathbf{y}_l, z) \tag{4.2}$$

Here, Z is a normalizing factor; $\mathbf{v} = [v_1, \ldots, v_N]$ and $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$ corresponds to all observed and hidden variables, respectively. The feature function f, g, and h can be defined in multiple different ways. For example, in the work described in [63], f is defined with binary values in order to capture the existence of the edge between two users; the node feature function g is defined according to the similarity of two users on a topic; and the global feature function h is defined as a constraint.

Based on this formulation, the task of social influence is cast as that of identifying the node which has the highest probability to influence another node on a specific topic along with the edge. This is the same as that of maximizing the likelihood function $P(\mathbf{v}, \mathbf{Y})$.

2.2 Existential Test for Social Influence

Anagnostopoulos et al. [2] try to differentiate social influence from homophily or confounding variables by proposing the shuffle test and edge reversal test. The idea of shuffle test is that if social influence does not play a role, even though an agent's probability of activation could depend on her friends, the timing of such an activation should be independent of the timing of other agents. Therefore, the data distribution and characteristics will not change even if the exact time of occurrence is shuffled around. The idea of edge-reversal test is that other forms of social correlation (than social influence) are only based on the fact that two friends often share common characteristics or are affected by the same external variables and are independent of which of these two individuals has named the other as a friend. Thus, reversing the edges will not change our estimate of the social correlation significantly. On the other hand, social influence spreads in the direction specified by the edges of the graph, and hence reversing the edges should intuitively change the estimate of the correlation. Anagnostopoulos and et al. [2] test their models using tagging data from Flickr and validate social influence as a source of correlation between the actions of individuals with social ties.

The proposed tests in [2] assume a static network, which is true in many real social networks. LaFond and Neville [25] propose a different randomization test with the use of a relational autoregression model. More specifically, they propose to model the social network as a time-evolving graph $G_t = (V, E_t)$ where V is the set of all nodes, E_t is the set of all edges at time t. Besides G_t, the nodes have some attribute at time t denoted by X_t. The main idea is that selection and social influence can be differentiated through the autocorrelation between X_t and G_t . On the one hand, the selection process can be represented as a causal relationship from X_{t-1} to G_t, which means the node attributes at time $t-1$, i.e., X_{t-1}, determines the social network at G_t. On the other hand, the social influence can be represented as the causal relation from G_{t-1} to X_t, which means the social network at time t, i.e., G_t, determines the node attributes at time t, i.e., X_t .

Aral et al. [3] propose a diffusion model for differentiating selection and social influence. In particular, their intuition is that although the diffusion patterns created by peer influence-driven contagions and homophilous diffusion are similar, the effects are likely to result in significantly different dynamics. Influence-driven contagions are self-reinforcing and display rapid, exponential, and less predictable diffusion as they evolve, whereas selection-driven diffusion processes are governed by the distributions of characteristics over nodes. In [3], they develop a matched sample estimation framework to distinguish influence and homophily effects in dynamic networks.

Social influence in Healthcare. Christakis and Fowler studied the effect of social influence on health related issues including alcohol consumption [56], obesity [16], smoking [17], trouble sleep [49], loneliness [13], happiness [26]. In these studies, they use longitudinal data covering roughly 12,000 people and correlate health status and social network structure over a 32-year period. They found that clusters of nodes with similar health status in the network. In another word, people tend to be more similar in health status to their friends than in a random graph. The main focus of all these studies is to explain why homophily of health status is present. The analysis in Christakis and Fowler argues that, even accounting for effects of selection and confounding variables, there is significant evidence for social influence as well. The evidence suggests that health status can be influenced by the health status of the neighbors. For example, their obesity study [16] suggests that obesity may exhibit some amount of "contagion" in the social network. Although people do not necessarily catch it as the way one catches the flu, it can spread through the underlying social network via the mechanism of social influence. Similar observations of their study on alcohol consumption[56] discover that clusters of drinkers and abstainers were present in the network at all time points, and the clusters extended to 3 degrees of separation through the social network. These clusters were not only due to selective formation of social ties among drinkers but also seem to reflect social influence. Changes in the alcohol consumption behavior of a person's social network had a statistically significant effect on that person's subsequent alcohol consumption behavior. The behaviors of immediate neighbors and co-workers were not significantly associated with a person's drinking behavior, but the behavior of relatives and friends was.

2.3 Influence and Actions

Influence is usually reflected in changes in social action patterns (user behavior) in the social network. Recent work [31, 68] has studied the problem of learning the influence degree from historical user actions, while some other work [58, 64] investigates how social actions evolve in the context of the network, and how such actions are affected by social influence factors. Before introducing these methods, we will first define the time-varying attribute augmented networks with user actions:

DEFINITION 4.1 ***Time-varying attribute-action augmented network**: The time-varying attribute-action augmented network is denoted as $G^t = (V^t, E^t, X^t, Y^t)$, wheren V^t is the set of users and E^t is the set of links between users at time t, X^t represents the attribute matrix of all users in the network at time t and Y^t represents the set of actions of all users at time t.*

For all actions, they define a set of action tuples as $\mathbf{Y} = (v, y, t)$, where $v \in V^t, t \in 1, \cdots, T$, and $y \in Y^t$.

Learning influence probabilities Goyal et al. [31] study the problem of learning the influence degrees (called probabilities) from a historic log of user actions. They present the concept of user influential probability and action influential probability. The assumption is that if user v_i performs an action y at time t and later ($t' > t$) his friend v_j also perform the action, then there is an influence from v_i on v_j. The goal of learning influence probabilities [31] is find a (static of dynamic) model to best capture the user influence and action influence information in the network. They give a general user influential probability and action influential probability definitions as follows:

- User Influence Probability

$$infl(v_i) = \frac{|\{y|\exists v, \Delta t : prop(a, v_i, v_j, \Delta t) \wedge 0 \leq \Delta t|}{Y_{v_i}}$$

- Action Influence Probability

$$infl(y) = \frac{|\{v_i|\exists v_j, \Delta t : prop(a, v_j, v_i, \Delta t) \wedge 0 \leq \Delta t\}|}{\text{number of users performing } y}$$

where $\Delta t = t_j - t_i$ represents the difference between the time when user v_j performing the action and the time when user v_i performing the action, given $e_{ij} = 1$; $prop(a, v_i, v_j, \Delta t)$ represents the action propagation score.

Goyal et al. [31] propose three methods in order to approximate the action propagation $prop(a, v_i, v_j, \Delta t)$: 1) static model (based on Bernoulli distribution, Jaccard Index, and Partial Credits), 2) Continuous Time (CT) Model, and 3) Discrete Time (DT) Model. The model can be learned with a two-stage algorithm. Finally, the learned influence probabilities have been applied to action prediction and the experiments show that the Continuous Time (CT) model can achieve a better performance than other models on the Flickr social network with the action of “joining a group”.

Social action tracking The main advantage of methods proposed in [31] is that the model is scalable and it is effective for a large social network. One limitation is that it ignores the correlation between user actions, and it also does not consider the attributes associated with each user node.

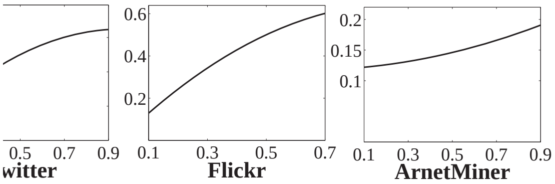

Figure 4.3. Social influence. The x-axis stands for the percentage of one's friends who perform an action at $t - 1$ and the y-axis represents the likelihood that the user also performs the action at t.

To address this problem, Tan et al. [62] propose the social action tracking problem. This problem discusses how to simultaneously model the social network structure, user attributes and user actions over time. They perform an analysis on three real social networks: Twitter[1], Flickr[2], and Arnetminer[3]. On Twitter, the action is defined as whether a user discusses the topic "Haiti Earthquake" on his microblogs (tweets). On Flickr, the action is defined as whether a user adds a photo to his favorite list. In the case of Arnetminer, the action is defined as whether a researcher publishes a paper on a specific conference (or journal). The analysis includes three aspects: (1) social influence; (2) time-dependency of user actions; (3) action correlation between users. Figure 4.3 [62] shows the effect of social influence. We see that with the percentage of one's friends performing an action increasing, the likelihood that the user also performs the action is increased. For example, when the percentage of one's friends discussing "Haiti Earthquake" on their tweets increases the likelihood that the user herself posts tweets about "Haiti Earthquake" is also increased significantly. Figure 4.4 illustrates how a user's action is dependent on his historic behaviors. It can be seen that a strong time-dependency exists for users' actions. For instance, on Twitter, averagely users who posted tweets about "Haiti Earthquake" will have a much higher probability ($+20$ to 40%) to post tweets on this topic than those who never discussed this topic on their blogs. Figure 4.5 shows the correlation between users' actions at the same timestamp. An interesting phenomenon is that friends may perform an action at the same time. For example, on Twitter, two friends have a higher probability ($+19.6\%$) to discuss the "Haiti Earthquake" than two users randomly chosen from the network.

In order to model and track social influence and user actions, Tan et al. [62] propose a Noise-Tolerant Time-varying Factor Graph Model (NTT-FGM), which is based on three intuitions:

[1] http://www.twitter.com, a microblogging system.

[2] http://www.flickr.com, a photo sharing system.

[3] http://arnetminer.org, an academic search system.

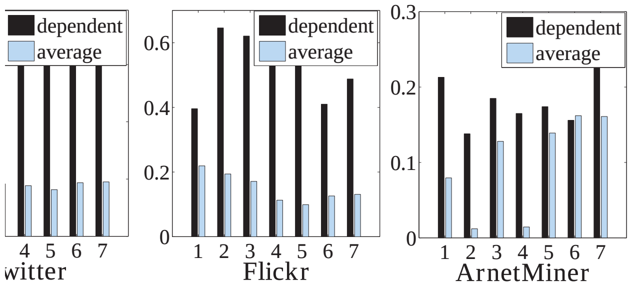

Figure 4.4. Time-dependency of user actions. The x-axis stands for different timestamps. "dependent" denotes the likelihood that a user performs an action which was previously performed by herself; "average" represents the likelihood that a user performs the action.

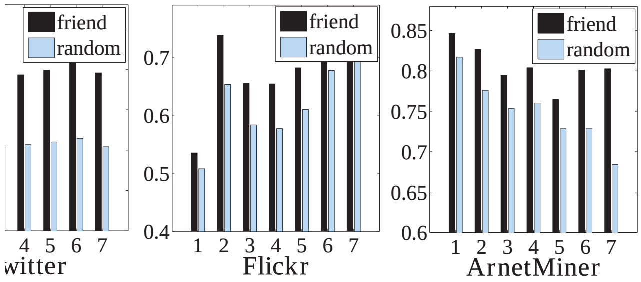

Figure 4.5. Action correlation. The x-axis stands for different time windows. "friend" denotes the likelihood that two friends perform an action together; "random" represents the likelihood that two random users perform the action together.

1 User actions at time t are influenced by their friends' historic actions (time $< t$).

2 User actions at time t are usually dependent on their previous actions.

3 User actions at a same time t have a (strong) correlation.

Moreover, the discrete variable y_i^t only models the user action at a coarse level, but cannot describe the intention of the user to perform an action. Directly modeling the social action Y would inevitably introduce noise into the model. A continuous variable for modeling the *action bias* is favorable. Thus, the concept of latent action state is presented:

DEFINITION 4.2 ***Latent action state****: For each user action y_i^t, we define a (continuous) latent state $z_i^t \in [0, 1]$, which corresponds to a combination of the observed action y_i and a possible bias, to describe the actual intensity of the intention of the user to perform the action.*

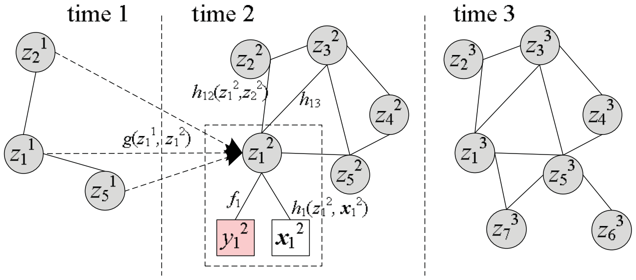

Figure 4.6. Graphical representation of the NTT-FGM model. Each circle stands for a user's latent action state z_i^t at time t in the network, which is used to characterize the intention degree of the user to perform the action; the latent state is associated with the action y_i^t, a vector of attributes $\mathbf{x}_i^t$, and depends on friends' historic actions $\mathbf{z}_{\sim v_i}^{t-1}$ and correlates with friends' actions $\mathbf{z}_{\sim v_i}^t$ at time t; $g(.)$ denotes a factor function to represent the friends' influence on a user's action; $h_i(.)$ represents a factor defined on user v_i's attributes; and $h_{ij}(.)$ represents a factor to capture the correlation between users' actions.

Figure 4.6 shows the graphical structure of the NTT-FGM model. An action of user v_i at time t, i.e., y_i^t is modeled by using a (continuous) latent action state z_i^t, which is dependent on friends' historic actions $\mathbf{z}_{\sim v_i}^{t-1}$ (where $\sim v_i$ represents friends of user v_i in the network), users' action correlation $\mathbf{z}_{\sim v_i}^t$, and users' attributes $\mathbf{x}_i^t$. Specifically, in the NTT-FGM model, each discrete action is mapped into the latent state space and the action bias is modeled using a factor function. For example, for $y_i^t = 1$, a small value of its corresponding z_i^t suggests that a user v_i has a low intention to perform the action, thus a large action bias $|y_i^t - z_i^t|$. Next, influence between users is modeled using the latent states based on the same assumption: latent states of user actions at time t are conditionally independent of all the previous states given the latent states at time $t-1$. Finally, the correlation between actions is also modeled in the latent state space. A Markov random field is defined to model the dependency (correlation) among the continuous latent states.

Thus, given a series of attribute augmented networks $\mathbf{G} = \{G^t = (V^t, E^t, X^t, Y^t)\}$, $t \in \{1, \cdots, T\}$ and $V = V^1 \cup V^2 \cup \ldots \cup V^T, |V| = N$, the joint distribution over the actions $\mathbf{Y}$ given $\mathbf{G}$ can be written as follows:

$$p(\mathbf{Y}|\mathbf{G}) = \prod_{t=1}^{T} \prod_{i=1}^{N} f(y_i^t|z_i^t) f(z_i^t|\mathbf{z}_{\sim v_i}^{t-1}) f(z_i^t|\mathbf{z}_{\sim v_i}^t, \mathbf{x}_i^t) \tag{4.3}$$

where notation $\sim v_i$ represents neighbors of v_i in the social network. The joint probability has three types of factor functions:

- Action bias factor: $f(y_i^t|z_i^t)$ represents the posterior probability of user v_i's action y_i at time t given the continuous latent state z_i^t;

- Influence factor: $f(z_i^t|\mathbf{z}_{\sim v_i}^{t-1})$ reflects friends' influence on user v_i's action at time t;

- Correlation factor: $f(z_i^t|\mathbf{z}_{\sim v_i}^t, \mathbf{x}_i^t)$ denotes the correlation between users' action at time t.

The three factors can be instantiated in different ways, reflecting the prior knowledge for different applications. Finally, in the work [62], all the three factor function are defined by quadratic functions due to two reasons: the quadratic function is integrable and it offer the possibility to design an exact solution to solve the objective function (joint probability). Finally, the model is learned using an EM-style algorithm and for scale up to large-scale data sets, a distributed learning algorithm has been designed based on the MPI (Message Passing Interface).

Mixture model for user actions Manavoglu et al. [47] propose a mixture-model based approach for learning individualized behavior (action) models for Web users where a behavior model is a probabilistic model describing which actions the user will perform in the future.

They first build a global behavior model for the entire population and then personalize this global model for the existing users by assigning each user individual component weights for the mixture model, and then use these individual weights to group the users into behavior model clusters. Finally they show that the clusters generated in this manner are interpretable and able to represent dominant behavior patterns.

They claim that they are able to eliminate one of the biggest problems of personalization, which is the lack of sufficient information about each individual. This is achieved by starting with a global model and optimizing the weights for each individual with respect to the amount of data available for him or her.

Specifically, for each action in a user session, the history $H(U)$ is defined by the ordered sequence of actions, which have been observed so far. Their behavior model for individual U is a model, that predicts the next action A^{next} given the history $H(U)$. Therefore the problem is to infer this model, $P(A^{next}|H(U), Data)$, for each individual given the training data. For example, the Markov model is often used for such problems.

In the first stage, they use a global mixture model to capture the ordered sequence of actions for an individual U as follows:

$$P(A^{next}|H(U), Data) = \sum_{k=1}^{N_c} \alpha_k P(A^{next}|H(U), Data, k) \tag{4.4}$$

where $\sum_{k=1}^{N_c} \alpha_k = 1$, α_k is the prior probability of cluster k, and $P(A^{next}|H(U), Data, k)$ is the distribution for the k-th component. For the global model, the different α_k take on the same values across all the users.

There are two ways to model the cluster-specific distributions: a first order Markov model and the maximum entropy model. Regarding the Markov model, we can use the following way to model the k-th cluster distribution

$$P(A^{next}|H(U), Data, k) \propto \theta_{0,k} \prod_{h=1}^{|H(U)|} \theta_{h \to (h+1),k} \qquad (4.5)$$

where $\theta_{0,k}$ is the probability of observing $H(U)_0$ as the first action in the history, and $\theta_{h \to (h+1),k}$ is the probability of observing a transition from action number h to action number $h+1$ in the history.

In the second stage, we personalize the mixture model by using individual cluster probabilities, $\alpha_{U,k}$'s, for each user as follows:

$$P_U(A^{next}|H(U), Data) = \sum_{k=1}^{N_c} \alpha_{U,k} P(A^{next}|H(U), Data, k) \qquad (4.6)$$

where $\sum_{k=1}^{N_c} \alpha_{U,k} = 1$. The component distribution, $P(A^{next}|H(U), Data, k)$, is the same as in global mixture model: either maximum entropy or Markov model, which is fixed across all users.

An EM-style algorithm is utilized to estimate the model parameters.

2.4 Influence and Interaction

Besides the attribute and user actions, influence can be also reflected by the interactions between users. Typically, online communities contain ancillary interaction information about users. For example, a Facebook user has a Wall page, where her friends can post messages. Based on the messages posted on the Wall, one can infer which friends are close and which are acquaintances only. Similarly, one can use follower and following members on Twitter to infer the strength of a relationship.

Xiang et al [68] propose a latent variable model to infer relationship strength based on profile similarity and interaction activity, with the goal of automatically distinguishing strong relation- ships from weak ones. The model attempts to represent the intrinsic causality of social interactions via statistical dependencies. It distinguishes interaction activity from user profile data, and integrates two types of information by considering the relationship strength to

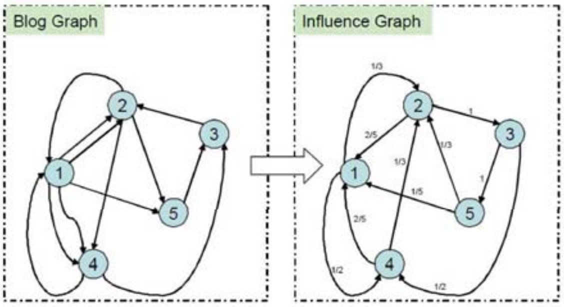

Figure 4.7. From blog graph to influence graph

be the hidden effect of user profile similarities, as well as the hidden cause of the interactions between users.

The input to the problem can be considered an attribute-augmented network $G = (V, E, X)$ with interaction information $\mathbf{m}_{ij} \subset M$ between users, where $\mathbf{m}_{ij}$ is a set of different interactions between users v_i and v_j. The model also uses continuous latent variable z, but for each link rather than action. The latent variable can be further treated as the strength of the social influence.

There are some methods aiming to model social influence using a link analysis method. The basic idea is similar to the concept of random walks. Java et al. [39] employ such a method to model the influence in online social networks.

Figure 4.7 shows the conversion of a blog network into an influence graph. A link from u to v indicates that u is influenced by v. The edges in the influence graph are the reverse of the blog graph to indicate this influence. Multiple edges indicate stronger influence and are weighted higher. In the influence graph, the direction of edges is opposite as the blog graph. And the influence weight can be calculated

$$W_{u,v} = \frac{C_{u,v}}{d_v}$$

Based on the influence graph, they proposed several typical applications, such as spam detection and node selection. The classical PageRank and HITS algorithms can be also employed here.

2.4.1 Influence and Friendship Drift.

Sarkar et al. [57] study the problem of friendships drifting over time. They explore two aspects of social network modeling by the use of a latent space model. First, they generalize a static model of relationships into a dynamic model that accounts for friendships drifting over time. Second, they show how to make it tractable to learn such models from data, even as the number of entities n gets large. The generalized model associates each entity with a point in p-dimensional Euclidean latent space. The points can move as time progresses but large moves in latent space are improbable. Observed links between entities are more likely if the entities are close in latent space. They show how to make such a model tractable (subquadratic in the number of entities) by the use of the following characteristics (a) appropriate kernel functions for similarity in latent space; (b) the use of low dimensional KD-trees; (c) a new efficient dynamic adaptation of multidimensional scaling for a first pass of approximate projection of entities into latent space; and (d) an efficient conjugate gradient update rule for non-linear local optimization in which amortized time per entity during an update is O($logn$). They use both synthetic and real data on up to 11,000 entities which indicate near-linear scaling in computation time and improved performance over four alternative approaches. We also illustrate the system operating on twelve years of NIPS co-authorship data.

2.4.2 Influence and Autocorrelation. Autocorrelation refers to correlation between values of the same variable (e.g., action or attribute) associated with linked nodes (users) [51]. More formally, autocorrelation in social networks, and in particular for influence analysis, can be defined with respect to a set of linked users $e_{ij} = 1, e_{ij} \in E$ and an attribute matrix X associated with these uses, as the correlation between the values of X on these instance pairs.

Neville et al. provide an overview of research on autocorrelation in a number of fields with an emphasis on implications for relational learning, and outline a number of challenges and opportunities for model learning and inference [51]. Social phenomena such as social influence, diffusion processes, and the principle of homophily give rise to autocorrelated observations as well, through their influence on social interactions that govern the data generation process.

Another related topic is referred to as collective behavior in social networks. Essentially, collective behavior modeling is to understand the behavior correlation in the social network. For this purpose, much work has been done. For example, Tang and Liu [65] aim to predict collective behaviors in social media. In particular, they try to answer the question: given information about some individuals, how can we infer the behavior of unobserved individuals in the same network?

They attempt to utilize the behavior correlation presented in a social network to predict the collective behavior in social media. The input of their problem is the same as Definition 4.1. They propose a framework called *SocDim* [64], which is composed of two steps, which are those of social dimension extraction and discriminative learning respectively. In the instantiation of the framework *SocDim*, modularity maximization is adopted to extract social dimensions. There are several concerns about the scalability of *SocDim*:

- The social dimensions extracted according to modularity maximization are dense.

- The modularity maximization requires the computation of the top eigenvectors of a modularity matrix which will become a daunting task when the network scales to millions of node.
- Networks in social media tend to evolve which entails efficient update of the model for collective behavior prediction.

2.4.3 Influence and Grouping Behavior. Grouping behavior, e.g., user's participation behavior into a forum, is an important action in the social network. The point of influence and grouping behavior is to study how different factors influence the dynamics of grouping behaviors.

Shi et al. investigated the user participation behavior in diverse online forums [59]. In that paper, they are mainly focused on three central questions:

1 What are the factors in online forums that potentially influence people's behavior in joining communities and what is the corresponding impact?

2 What are the relationships between these factors, i.e. which ones are more effective in predicting the user joining behavior, and which ones carry supplementary information?

3 What are the similarities and differences of user grouping behavior in forums of different types (such as news forums versus technology forums)?

In order to answer the first question, they analyze four features that can usually be obtained from a forum dataset:

1 *Friends of Reply Relationship.* Use this feature to describe how users are influenced by the numbers of neighbors with whom they have ever had any reply relationship.

2 *Community Sizes.* Use community size as the measurement to quantify the 'popularity' of information.

3 *Average Ratings of Top Posts.* Aside from the popularity of information, we are also interested in how the authority or interestingness of information impacts user behavior.

4 *Similarities of Users.* This is the only feature with dependency: if two users are 'similar' in a certain way, what is the correlation of the sets of communities they join?

Their first discovery is that, despite the relative randomness, the diffusion curve of influence from users of reply relationships has very similar diffusion patterns. However, the reasons that people are linked together are very different. They also investigate the influence of the features associated with communities, which include the size of communities and the authority or the interestingness of the information in the communities. They find that their corresponding information diffusion curves show some strong regularities of user joining behavior as well, and these curves are very different from those of reply relationships. Furthermore, we analyze the effects of similarity of users on the communities they join, and find two users who communicate more frequently or have more common friends are more likely to be in the same set of communities.

In order to answer the second question, we construct a bipartite graph, whose two sets of nodes are users and communities, to encompass all the features and their relationships in this problem. Based on the bipartite graph, we build a bipartite Markov Random Field (BiMRF) model to quantitatively evaluate how much each feature affects the grouping behavior in online forums, as well as their relationships with each other. BiMRF is a Markov random graph with edges and two-stars as its configuration, and incorporates the node-level features we have described as in a social selection model. The most significant advantage of using the BiMRF model is that it can explicitly incorporate the dependency between different users' joining behavior, i.e., how a user's joining behavior is affected by her friends' joining behavior. The results of this quantitative analysis shows that different features have different effectiveness of prediction in news forums versus technology forums.

Backstrom et al. [4] also explore a large corpus of thriving online communities. These groups vary widely in size, moderation and privacy, and cover an equally diverse set of subject matter. They present a number of descriptive statistics of these groups. Using metadata from groups, members, and individual messages, they identify users who post and are replied-to frequently by multiple group members. They classify these high-engagement users based on the longevity of their engagements. Their results show that users who will go on to become long-lived, highly-engaged user experience significantly better treatment than other users from the moment they join the group, well before there is an opportunity for them to develop a long-standing relationship with members of the group. They also present a simple model explaining long-term heavy engagement as a combination of user-dependent and group dependent factors. Using this model as an analytical tool, they show that properties of the user alone are sufficient to explain 95% of all memberships, but introducing a small amount of group-specific information dramatically improves our ability to model users belonging to multiple groups.

3. Influence Maximization in Viral Marketing

Social influence analysis has various real-world applications. Influence maximization in viral marketing is an example of such an important application. In this section, we will introduce the problem of influence maximization and review recent research progress. We will also introduce relevant work on representative user and expert discovery.

3.1 Influence Maximization

The problem of influence maximization can be traced back to the research on "word-of-mouth" and "viral marketing" [6, 11, 21, 38, 46, 55]. The problem of often motivated by the determination of potential customers for marketing purposes. The goal is to minimize marketing cost and more generally to maximize profit. For example, a company may wish to market a new product through the natural "word of mouth" effect arising from the interactions in a social network. The goal is to get a small number of influential users to adopt the product, and subsequently trigger a large cascade of further adoptions. In order to achieve this goal, we need a measure to quantify the intrinsic characteristics of the user (e.g., the expected profit from the user) and the user network value (e.g., the expected profit from users that may be influenced by the user). Previously, the problem has mainly been studied in marketing decision or business management. Domingos and Richardson [21] formulated this problem as a ranking problem using a Markov random field model. They further present an efficient algorithm to learn the model [55]. However, the method models the marketing decision process in a "black box". How users influence each other once a set of users have been marketed (selected), how they will influence their neighbors and how the diffusion process will continue are problems which are still not fully solved. Kempe et al. [41] took the first step to formally define the process in two diffusion models and theoretically proved that the optimization problem of selecting the most influential nodes in the two models is NP-hard. They have developed an algorithm to solve the models with approximation guarantees. The efficiency and scalability of the algorithm has been further improved in recent years [14, 15]. We will skip the work in marketing or business and focus on the formulation of the problem and model learning.

3.1.1 Diffusion Influence Model. There are quite a few classical models of this problem. Here, we review some of them. For ease in explanation, we associate each user with a status: active or inactive. Then, the status of the chosen set of users to market (also referred to as "seed nodes") is viewed as active, while the other users are viewed as inactive. The problem of influence maximization is studied with the use of this status-based dynamic. Initially all users are considered inactive. Then, the chosen users are activated, who may further influence their friends (neighbor nodes) to be active as well. The simplest model is to quantify the influence of each node with some heuristics. Some examples are as follows:

1) High-degree heuristic. It chooses the seed nodes according to their degree d_v. The strategy is very simple but also natural because the nodes with more neighbors would arguably tend to impose more influence upon its direct neighbors. This consideration of high-degree nodes is also known in the sociology literature as “degree centrality”.

2) Low-distance Heuristic. Another commonly used influence measure in sociology is distance centrality, which considers the nodes with the shortest paths to other nodes as be seed nodes. This strategy is based on the intuition that individuals are more likely to be influenced by those who are closely related to them [26].

3) Degree discount heuristic. The general idea of this approach is that if u has been selected as a seed, then when considering selecting v as a new seed based on its degree, we should not count the edge $\overrightarrow{vu}$ towards its degree. This is referred to as SingleDiscount. More specifically, for a node v with d_v neighbors of which t_v are selected as seeds already, we should discount v's degree by $2t_v + (d_v - t_v)\, t_v p$.

4) Linear threshold model. In this family of models, whether a given node v will be active can be based on an arbitrary monotone function of the set of neighbors of v that are already active. We associate a monotone threshold function f_v which maps subsets of v's neighbors to real numbers in $[0, 1]$. Then, each node v is given a threshold θ_v, and v will turn active in step t if$f_v(S) > \theta_v$, where S is the set of neighbors of v that are active in step $t - 1$.

Specifically, in [41] the threshold function $f_v(S)$ is instantiated as $f_v\,(S) = \sum_{u \in S} b_{v.u}$ where $b_{v.u}$ can be seen as a fixed weight, subject to the following constraint:

$$\sum_{u neighbors of v} b_{v,u} \leq 1$$

5) General cascade model. We first define an incremental function $p_v(u, S) \in [0, 1]$ as the success probability of user u activating user v, i.e., user u tries to activate v and finally succeeds, where S is those of v's neighbors that have already attempted but failed to make v active. A special version of this model used in [41] is called Independent Cascade Model in which $p_v(u, S)$ is a constant, meaning that whether v is to be active does not depend on the order v's neighbors try to activate it. And a special case of Independent Cascade Model is weighted cascade model, where each edge from node u to v is assigned probability $1/d_v$ of activating v.

One challenging problem in the diffusion influence model is the evaluation of its effectiveness and efficiency. From the theoretical perspective, Kempe et al. [41] prove that the optimization of their two proposed models, i.e., linear threshold model and general cascade model is NP-hard. Their proposed approximation algorithms can also theoretically guarantee that the influence spread is within $(1 - 1/e)$ of the optimal influence spread. From an empirical perspective, Kempe et al. [41] show that their proposed models can outperform the traditional heuristics in terms of the maximization of social influences. Recent research mainly focuses on the improvement of the efficiency of the algorithm. For example, Leskovec et al. [44] present an optimization strategy referred to as "Cost-Effective Lazy Forward" or "CELF", which could accelerate the procedure by up to 700 times with no worse effectiveness. Chen et al. [14] further improve the efficiency by employing a new heuristics and in [15] they extend the algorithm to handle large-scale data sets. Another problem is the evaluation of the effectiveness of the models for influence maximization. Some recent work has been proposed in [21] and [55], though these methods are designed only for small data sets. It is still a challenging problem to extend these methods to large data sets.

3.1.2 Learning to Predict Customers. Viral Marketing aims to increase brand awareness and marketer revenue with the help of social networks and social influence. Direct marketing is an important application, which attempts to market only to a select set of potentially profitable customers. Previously, the problem was mainly addressed by constructing a model that predicts a customer's response from their past buying behavior and any available demographic information [45]. When applied successfully, this approach can significantly increase profits [53]. One limitation of the approach is that it treats each customer independently of other customers in terms of their actions. In reality, a person's decision to buy a product is often influenced by their friends and acquaintances. It is not desirable to ignore such a networking influence, because it can lead to severely suboptimal decisions.

We will first introduce a model that tries to combine the network value with customer intrinsic value [21]. Here, the intrinsic value represents attributes (e.g., customer behavior history) that are directly associated with a customer. Such attributes might affect the likelihood of the customer to buy the product, while the network value represents the social network (e.g., customers' friends), which may influence the customer's buying decision.

The basic idea here is to formalize the social network as Markov random fields, where each customer's probability of buying is modeled as a function of both the intrinsic desirability of the product for the customer and the influence of other customers. Formally, the input can be defined as: consider a social network $G = (V, E)$, with n potential customers and their relationships recorded in E, and let $\mathbf{x}_i$ indicate the attributes associated with each customer v_i. We assign a boolean variable y_i to each customer that takes the value 1 if the customer v_i buys the product being marketed, and 0 otherwise. Further let NB^i be the neighbors of v_i in the social network and z_i be a variable representing the marketing action that is taken for customer v_i. z_i can boolean variable, with $z_i = 1$ if the customer is selected to market (e.g., be offered a free product), and $z_i = 0$ otherwise. Alternatively, z_i could be a continuous variable indicating a discount offered to the customer. Given this, we can define the marketing process for customer v_i in a Markov random field as follows:

$$\begin{aligned} P(y_i|\mathbf{y}_{NB^i}, \mathbf{x}_i, \mathbf{z}) &= \textstyle\sum_{C(NB^i)} P(y_i, \mathbf{y}_{NB^i}|\mathbf{x}_i, \mathbf{z}) \\ &= \textstyle\sum_{C(NB^i)} P(y_i|\mathbf{y}_{NB^i}, \mathbf{x}_i, \mathbf{z})P(\mathbf{y}_{NB^i}|\mathbf{X}, \mathbf{z}) \end{aligned} \tag{4.7}$$

where $C(NB^i)$ is the set of all possible configuration of the neighbors of v_i; and $\mathbf{X}$ represent attributes of all customers. To estimate $P(\mathbf{y}_{NB^i}|\mathbf{X}, \mathbf{z})$, Domingos and Richardson [21] employ the maximum entropy estimation to approximate the probability based on the independent assumption, i.e.,

$$P(\mathbf{y}_{NB^i}|\mathbf{X}, \mathbf{z}) = \prod_{v_j \in NB^i} P(y_j|\mathbf{X}, \mathbf{z}) \tag{4.8}$$

The marketing action z is modeled as a Boolean variable. The cost of marketing to a customer is further considered in the Markov model. Let r_0 be the revenue from selling the product to the customer if no marketing action is performed, and r_1 be the revenue if marketing is performed. The cost can be considered as offering a discount to the marketed customer. Thus the expected lift in profit from marketing to customer v_i in isolation (without influence) can be defined as follows:

$$ELP_i{}^1(Y, z) = r_1 P(y_i = 1|Y, f_i{}^1(M)) - r_0 P(y_i = 1|Y, f_i{}^0(z)) - c \tag{4.9}$$

where $f_i{}^1(z_i)$ be the result of setting z_i to 1 and leaving the rest of $\mathbf{z}$ unchanged, and similarly for $f_i^0(z_i)$.

Thus the global lift in profit for a particular choice $\mathbf{z}$:

$$ELP_i{}^1(Y, \mathbf{z}) = \sum_{i=1}^{n} [r_i P(X_i = 1|Y, \mathbf{z}) - r_0 P(X_i = 1|Y, z_0) - c_i]$$

A customer's total value is the global lift in profit from marketing to him

$$ELP(Y, f_i{}^1(z_i)) - ELP(Y, f_i{}^0(z_i))$$

and his network value is the difference between his total and intrinsic values. The model can be also adjusted to a continuous version with no qualitative difference.

In marketing context, the goal for modeling the value of a customer is to find the **z** that can maximize the lift in profit. Richardson and Domingos propose several approximation algorithms in [21] to solve this problem. And they make further contribution to this field in their later paper [55] by showing an tractable way to calculate the optimal M by directly solving the equation:

$$r\Delta_i(Y)\frac{d\Delta P_i(z,Y)}{dz} = \frac{dc(z)}{dz}$$

where z denotes the market action. The network effect $\Delta_i\left(Y\right) = \sum\limits_{j=1}^{n} w_{ji}\Delta_j\left(Y\right)$ is the total increase in probability of purchasing in the network (including y_i) that results from a unit change in $P_0(y_i)$ when w_{ji} indicates how much v_j can influence v_i. And the $\Delta P_i\left(z,Y\right) = \beta_i\left[P_0\left(X_i = 1|Y, M_i = z\right) - P_0\left(X_i = 1|Y, M_i = 0\right)\right]$ denotes the immediate change in customer v_i's probability of purchasing when he is marketed to with marketing action z.

Some other work aims to find the optimal marketing strategy by directly maximizing the revenue rather than social influence. Work [34] makes some investigation in this field. The basic idea is as follows: since a customer who owns a product can have an impact on potential buyers, it is important to decide the sequence of marketing, as well as the price to offer the buyers. Thus a simple marketing strategy, called influence-and-exploit strategy, is introduced, which basically consists of an influence step and an exploit step. In the influence step, the seller starts by giving some products for free to some specially chosen customers (hopefully the most influential ones), and in the exploit step, the seller try to sell products to the remaining customers with a fixed optimal price. Hartline et al. [34] also prove that the influence-and-exploit method works as a reasonable approximation of the NP-hard problem of finding the optimal marketing strategy.

3.1.3 Maximizing the Spread of Influence. Kempe et al. [41] propose the linear threshold model and the independent cascade model. The optimal solution to either model is NP-hard. The solution here is to use a submodular function to approximate the influence function. Submodular functions have a number of very nice tractability properties in terms of the design of approximation algorithms. One important property that is used in the approach is as follows. Given a function f that is submodular, taking only non-negative values, then we have

$$f(S \cup \{v\}) \geq f(S)$$

for all elements v and sets S. Thus, the problem can be transformed into finding a k-element set S for which $f(S)$ is maximized. The problem, can be solved using a greedy hill-climbing algorithm which approximates the optimal solution within a factor of $(1 - 1/e)$. The following theorem formally defines the problem.

THEOREM 4.3 *[19, 50] For a non-negative, monotone submodular function f, let S be a set of size k obtained by selecting elements one at a time, each time choosing an element that provides the largest marginal increase in the function value. Let $S^\star$ be a set that maximizes the value of f over all k-element sets. Then $f(S) \geq (1 - 1/e) \cdot f(S^\star)$; in other words, S provides a $(1 - 1/e)$-approximation.*

The model can be further extended to assume that each node v has an associated non-negative weight w_v, which can be used to capture the human prior knowledge to the task at hand, e.g., how important it is that v be activated in the final outcome.

To adapt the model to a more realistic scenario, we may have a number of m of different marketing actions M_i available, each of which may affect some subset of nodes by increasing their probabilities of becoming active; however, different nodes may respond to marketing actions in different ways. Thus a more general model can considered [41]. More specifically, we can introduce investment t_i for each marketing action M_i. Thus the goal is to reach a maximum profit lift while the total investments do not exceed the budget. A marketing strategy is then an m-dimensional vector **t** of investments. The probability that node v will become active is determined by the strategy and denoted by $h_v(\mathbf{t})$. By assuming that the function is non-decreasing and satisfies the "diminishing returns" property for all $t \geq t'$ and $a \geq 0$:

$$h_v(t + a) - h_v(t) \leq h_v(t' + a) - h_v(t') \tag{4.10}$$

Satisfying the above inequality corresponds to an interesting marketing intuition: the marketing action would be more effective when the targeted individual is less "marketing-saturated" at that point. Finally, the objective of the model is to maximize the expected size of the final active set. Given an initial set A and let the expected size of the final active set is $\sigma(A)$, then the expected revenue of the marketing strategy **t** can be defined as:

$$g(\mathbf{t}) = \sum_{A \subset V} \sigma(A) \cdot \prod_{v \in A} h_v(t) \cdot \prod_{u \in V - A} (1 - h_u(\mathbf{t})) \tag{4.11}$$

We note that if A be the active set of nodes, then the inactive set of nodes is denoted by $V - A$. We can use the submodular property in order to optimize the function corresponding to the revenue of the marketing strategy. We can design a greedy hill-climbing algorithm, which can still guarantee an approximation within a constant factor. A proof of this result may be found in [41].

3.2 Other Applications

3.2.1 Online Advertising. Social influence analysis techniques can also be leveraged for online advertising. For example, the work in [54] proposes methods for identifying brand-specific audiences without utilizing the user private information. The proposed method takes advantage of the notion of "seed nodes", which can specifically indicate the users (or browsers) who exhibit brand affinity. Yet another term "brand proximity" is a distance measure between candidate nodes and the seed nodes. For each browser b_i we use $\overrightarrow{\phi}_{b_i} = [\phi^1{}_{b_i}, \phi^2{}_{b_i}, ..., \phi^P{}_{b_i}]$ to denote the effect of the P proximity measures. Then we can discover the best audiences for marketing by ranking the candidate nodes b_i with respect to $\overrightarrow{\phi}_{b_i}$ based on some monotonic function $score(b_i) = f_i(\overrightarrow{\phi}_{b_i} \cdot \overrightarrow{I}_q)$. The selection vector $\overrightarrow{I}_q = [0, ..., 1, ..., 0]$ holds a 1 in the q-th row. The proximity measures P can be chosen from a pool. Finally, the authors show that the quasi-social network extracted from the data corresponds well with a real social network. This means that the modeled "friends" on the virtual network accurately reflect the relationships between friends or relatives in the real world.

Another tractable approach for viral marketing is through frequent pattern mining, which is studied by Goyal et al. in [30].Their research focuses on the actions performed by the users, under the assumption that users can see their friends' actions. The authors formally define leaders in a social network, and introduce an efficient algorithm aiming at discovering the leaders. The basic formation of the problem is that actions take place in different time steps, and the actions which come up later could be influenced by the earlier taken actions. This is called the propagation of influence. The notion of leaders corresponds to people who can influence a sufficient number of people in the network with their actions for a long enough period of time. Aside from the normal leaders, there are other kinds of users who only influence a smaller subset of people. These users are called tribe leaders. The algorithm for finding leaders in a social network makes use of action logs, which sorts actions chronologically.

3.2.2 Influential Blog Discovery. In the web 2.0 era, people spend a significant amount of time on user-generated content web sites, a classic example of which are blog sites. People form an online social network by visiting other users' blog posts. Some of the blog users bring in new information, ideas, and opinions, and disseminate them down to the masses. This influences the opinions and decisions of others by word of mouth. This set of users are called opinion leaders.

In order to tackle this problem, we can first define the following properties for each blogger:

- **R**ecognition: An influential blog post is recognized by many people. This generally means that there are a lot of inlinks to the article.
- **A**ctivity generation: Blogs often have comments associated with them. A large number of comments indicates that the contents of the article encourages discussion. This indicates that the blog is influential.
- **Novelty**: Normally a novel blog is one that with less outgoing links.
- **E**loquence: Longer articles posted on blog sites tend to be more eloquent, and can thus be more influential.

The work in [1] presents a model which takes advantages of the above four properties to describe the influence flow in the influence-graph consisting of all the blogger pages. Basically, the influence flow probability is defined as follows:

$$InfluenceFlow(p) = w_{in} \sum_{m=1}^{|\iota|} I(p_m) - w_{out} \sum_{n=1}^{|\theta|} I(p_n) \tag{4.12}$$

w_{in} and w_{out} is the weight to describe the contribution of incoming and outgoing links. Finally, the influence of a blog is defined as:

$$I(p) = w(\lambda) \times (w_{com}\gamma_p + InfluenceFlow(p)) \tag{4.13}$$

where w_{com} denotes the weight that can be used to regulate the contribution of the number of comments (γ_p) towards the influence of the blog post p.

In another work [61], Song et al. associate a hidden node v_e to each node v to represent the source of the novel information in blog v. More specifically, let $Out(v)$ denote the set of blogs that v links to. The information novelty contribution of entry v_e is then calculated as:

$$Nov\left(v_e|Out(v_e)\right) = \min_{O_e \in Out(v_e)} Nov(v_e|O_e) \tag{4.14}$$

The information novelty provided by the hidden node of blog v is measured as the average of the novelty scores of the entries it contains.

$$Nov\left(v|Out(v)\right) = \frac{\sum\limits_{v_e \in V} \left(Nov\left(v_e|Out(v_e)\right)\right)}{card\left(Set\left(v_e\right)\right)} \tag{4.15}$$

where $card(\cdot)$ denotes total number of entries of interest in blog v. Then, the problem can be formulated as solving the InfluenceRank IR.

$$IR^T(I-(1-\beta)\alpha W-(1-\beta)\alpha a\cdot e^T) = (1-\beta)(1-\alpha)e^T+\beta.Nov^T \quad (4.16)$$

with $IR^T \cdot e = 1$.

As the InfluenceRank can be fitted in a random walk framework, α is the probability that the random walk follows a link. β reflects how significant the novelty is to the opinion leaders we expect to detect. e is the n-vector of all ones and a is the vector with components a_i = 1 if i-th row of W corresponds to a dangling node, and 0, otherwise, where W is the normalized adjacent matrix.

4. Conclusion

Social influence analysis aims at qualitatively and quantitatively measuring the influence of one person on others. As social networking becomes more prevalent in the activities of millions of people on a day-to-day basis, both research study and practical applications on social influence will continue to grow. Furthermore, the size of the networks on which the underlying applications need to be used also continues to grow over time. Therefore, effective and efficient social influence methods are in high demand.

In this chapter, we focus on the computational aspects of social influence analysis and describe different methods and algorithms for calculating social influence related measures. First, we cover the basic statistical measure of networks such as centrality, closeness and betweenness; second, we present the social influence and selection models. These covers the fundamental concepts on influence; third, we present the influence maximization and its application for viral marketing.

In the future, an important and challenging research area is to develop efficient, effective and quantifiable social influence mechanisms to enable various applications in social networks and social media. This area lies in the intersection of computer science, sociology, and physics. In particular, scalable and parallel data mining algorithms, and scalable database and web technology have been changing how sociologists approach this problem. Instead of building conceptual models and conducting small scale simulations and user studies, more and more people now rely on large-scale data mining algorithms to analyze social network data. This provides more realistic results for large-scale applications. This chapter provides an introduction of the problem space in social influence analysis. The area is still in its infancy, and we anticipate that more techniques will be developed for this problem in the future.

References

[1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. In *Proceedings of the 1st ACM International Conference on Web Search and Data Mining (WSDM'08)*, pages 207–217, 2008.

[2] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'08)*, pages 7–15, 2008.

[3] S. Aral, L. Muchnika, and A. Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):2154421549, 2009.

[4] L. Backstrom, R. Kumar, C. Marlow, J. Novak, and A. Tomkins. Preferential behavior in online groups. In *Proceedings of the international conference on Web search and web data mining (WSDM'08)*, pages 117–128, 2008.

[5] E. Bakshy, B. Karrer, and L. A. Adamic. Social influence and the diffusion of user-created content. In *EC '09: Proceedings of the tenth ACM conference on Electronic commerce*, pages 325–334, New York, NY, USA, 2009. ACM.

[6] F. M. Bass. A new product growth for model consumer durables. *Management Science*, 15(5):215–227, 1969.

[7] P. Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2:113–120, 1972.

[8] P. Bonacich. Power and centrality: a family of measures. *American Journal of Sociology*, 92:1170–1182, 1987.

[9] S. P. Borgatti and M. G. Everett. A graph-theoretic perspective on centrality. *Social Networks*, 2006.

[10] U. Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 2001.

[11] J. J. Brown and P. H. Reingen. Social ties and word-of-mouth referral behavior. *The Journal of Consumer Research*, 14(3):350–362, 1987.

[12] R. S. Burt. *Structural holes: The social structure of competition*. Harvard University Press, Cambridge, MA, 1992.

[13] J. T. Cacioppo, J. H. Fowler, and N. A. Christakis. Alone in the Crowd: The Structure and Spread of Loneliness in a Large Social Network. *SSRN eLibrary*, 2008.

[14] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD International*

Conference on Knowledge Discovery and Data Mining (SIGKDD'09), pages 199–207, 2009.

[15] W. Chen, Y. Wang, and S. Yang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'10)*, pages 807–816, 2010.

[16] N. A. Christakis and J. H. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357, 2007.

[17] N. A. Christakis and J. H. Fowler. The collective dynamics of smoking in a large social network. *N Engl J Med*, 358(21):2249–2258, May 2008.

[18] R. B. Cialdini and N. J. Goldstein. Social influence: compliance and conformity. *Annu Rev Psychol*, 55:591–621, 2004.

[19] G. Cornuejols, M. L. Fisher, and G. L. Nemhauser. Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. *Management Science*, 23(8):789–810, 1977.

[20] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'08)*, pages 160–168, 2008.

[21] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'01)*, pages 57–66, 2001.

[22] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[23] P. W. Eastwick and W. L. Gardner. Is it a game? evidence for social influence in the virtual world. *Social Influence*, 4(1):18–32, 2009.

[24] S. M. Elias and A. R. Pratkanis. Teaching social influence: Demonstrations and exercises from the discipline of social psychology. *Social Influence*, 1(2):147–162, 2006.

[25] T. L. Fond and J. Neville. Randomization tests for distinguishing social influence and homophily effects. In *Proceeding of the 19th international conference on World Wide Web (WWW'10)*, 2010.

[26] J. H. Fowler and N. A. Christakis. Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study. *British Medical Journal, Vol. 3, January 2009*, 2008.

[27] L. C. Freeman. A set of measure of centrality based on betweenness. *Sociometry*, 40:35–41, 1977.

[28] L. C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1:215239, 1979.

[29] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821–7826, June 2002.

[30] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Discovering leaders from community actions. In *Proceeding of the 17th ACM conference on Information and knowledge management (CIKM'08)*, pages 499–508, 2008.

[31] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *Proceedings of the 3st ACM International Conference on Web Search and Data Mining (WSDM'10)*, pages 207–217, 2010.

[32] M. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[33] M. Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510, 1985.

[34] J. Hartline, V. Mirrokni, and M. Sundararajan. Optimal marketing strategies over social networks. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 189–198, New York, NY, USA, 2008. ACM.

[35] P. W. Holland and S. Leinhardt. Transitivity in structural models of small groups. *Small Group Research*, 2:107124, 1971.

[36] P. Holme and M. E. J. Newman. Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review*, 74(056108), 2006.

[37] C. Hubbell. An input-output approach to clique identification. *Sociometry*, 28:377–399, 1965.

[38] G. J., L. B., and M. E. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12:211–223(13), August 2001.

[39] A. Java, P. Kolari, T. Finin, and T. Oates. Modeling the spread of influence on the blogosphere. In *Proceeding of the 15th international conference on World Wide Web (WWW'06)*, 2006.

[40] L. Katz. A new index derived from sociometric data analysis. *Psychometrika*, 18:39–43, 1953.

[41] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'03)*, pages 137–146, 2003.

[42] D. Krackhardt. *The Strength of Strong ties: the importance of philos in networks and organization in Book of Nitin Nohria and Robert G. Eccles*

(Ed.), Networks and Organizations. Cambridge, Harvard Business School Press, Hershey, USA, 1992.

[43] P. Lazarsfeld and R. K. Merton. Friendship as a social process: A substantive and methodological analysis. *Freedom and Control in Modern Society*, page 1866, 1954.

[44] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'07)*, pages 420–429, 2007.

[45] C. X. Ling and C. Li. Data mining for direct marketing: Problems and solutions. In *KDD'98*, pages 73–79, 1998.

[46] V. Mahajan, E. Muller, and F. M. Bass. New product diffusion models in marketing: A review and directions for research. *The Journal of Marketing*, 54(1):1–26, 1990.

[47] E. Manavoglu, D. Pavlov, and C. L. Giles. Probabilistic user behavior models. In *Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03)*, page 203, Washington, DC, USA, 2003. IEEE Computer Society.

[48] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.

[49] S. C. Mednick, N. A. Christakis, and J. H. Fowler. The spread of sleep loss influences drug use in adolescent social networks. *PLoS ONE*, 5(3):e9775, 03 2010.

[50] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14(1):265–294, 1978.

[51] J. Neville, O. Simsek, and D. Jensen. Autocorrelation and relational learning: challenges and opportunities. In *Proceedings of the ICML-04 Workshop on Statistical Relational Learning*, 2004.

[52] M. E. J. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 2005.

[53] G. Piatetsky-Shapiro and B. M. Masand. Estimating campaign benefits and modeling lift. In *KDD'99*, pages 185–193, 1999.

[54] F. Provost, B. Dalessandro, R. Hook, X. Zhang, and A. Murray. Audience selection for on-line brand advertising: privacy-friendly social network targeting. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 707–716, New York, NY, USA, 2009. ACM.

[55] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'02)*, pages 61–70, 2002.

[56] J. N. Rosenquist, J. Murabito, J. H. Fowler, and N. A. Christakis. The Spread of Alcohol Consumption Behavior in a Large Social Network. *Annals of Internal Medicine*, 152(7):426–433, 2010.

[57] P. Sarkar and A. W. Moore. Dynamic social network analysis using latent space models. *SIGKDD Explor. Newsl.*, 7(2):31–40, 2005.

[58] J. Scripps, P.-N. Tan, and A.-H. Esfahanian. Measuring the effects of preprocessing decisions and network forces in dynamic network analysis. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'09)*, pages 747–756, 2009.

[59] X. Shi, J. Zhu, R. Cai, and L. Zhang. User grouping behavior in online forums. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining(KDD'09)*, pages 777–786, New York, NY, USA, 2009. ACM.

[60] P. Singla and M. Richardson. Yes, there is a correlation: - from social networks to personal behavior on the web. In *Proceeding of the 17th international conference on World Wide Web (WWW'08)*, pages 655–664, 2008.

[61] X. Song, Y. Chi, K. Hino, and B. L. Tseng. Identifying opinion leaders in the blogosphere. In *Proceedings of the 15th ACM international conference on Information and knowledge management (CIKM'06)*, pages 971–974, 2007.

[62] C. Tan, J. Tang, J. Sun, Q. Lin, and F. Wang. Social action tracking via noise tolerant time-varying factor graphs. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'10)*, pages 807–816, 2010.

[63] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'09)*, pages 807–816, 2009.

[64] L. Tang and H. Liu. Relational learning via latent social dimensions. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'09)*, pages 817–826, 2009.

[65] L. Tang and H. Liu. Scalable learning of collective behavior based on sparse social dimensions. In *Proceeding of the 18th ACM conference on Information and knowledge management(CIKM'09)*, pages 1107–1116, New York, NY, USA, 2009. ACM.

[66] C. E. Tsourakakis. Fast counting of triangles in large real networks without counting: Algorithms and laws. In *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008)*, pages 608–617, 2008.

[67] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, pages 440–442, Jun 1998.

[68] R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *Proceeding of the 19th international conference on World Wide Web (WWW'10)*, pages 981–990, 2010.