Soc. Netw. Anal. Min. (2015) 5:62
DOI 10.1007/s13278-015-0289-6

ORIGINAL ARTICLE

Probabilistic graphical models in modern social network analysis

Alireza Farasat[1,2] · Alexander Nikolaev[1] · Sargur N. Srihari[2] · Rachael Hageman Blair[3]

Received: 14 October 2014 / Revised: 14 August 2015 / Accepted: 23 August 2015 / Published online: 19 October 2015

Abstract The advent and availability of technology has brought us closer than ever through social networks. Consequently, there is a growing emphasis on mining social networks to extract information for knowledge and discovery. However, methods for social network analysis (SNA) have not kept pace with the data explosion. In this review, we describe directed and undirected probabilistic graphical models (PGMs), and highlight recent applications to social networks. PGMs represent a flexible class of models that can be adapted to address many of the current challenges in SNA. In this work, we motivate their use with simple and accessible examples to demonstrate the modeling and connect to theory. In addition, recent applications in modern SNA are highlighted, including the estimation and quantification of importance, propagation of influence, trust (and distrust), link and profile prediction, privacy protection, and news spread through microblogging. Applications are selected to demonstrate the flexibility and predictive capabilities of PGMs in SNA. Finally, we conclude with a discussion of challenges and opportunities for PGMs in social networks.

Keywords Probabilistic graphical modeling · Social network analysis · Bayesian networks · Markov networks · Exponential random graph models · Markov logic networks · Social influence · Network sampling

✉ Rachael Hageman Blair
hageman@buffalo.edu

[1] Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, USA

[2] Department of Computer Science and Engineering, University at Buffalo, Buffalo, USA

[3] Department of Biostatistics, University at Buffalo, Buffalo, USA

1 Introduction

Over 40 years ago, social scientist Allen Barton stated that "If our aim is to understand people's behavior rather than simply to record it, we want to know about primary groups, neighborhoods, organizations, social circles, and communities; about interaction, communication, role expectations, and social control." (Barton 1968 as reported in Freeman 2004). This sentiment is fundamental to the concept of *modularity*. The importance of structural relationships in defining communities and predicting future behaviors has long been recognized, and is not restricted to the social sciences (Freeman 2004).

Social network analysis (SNA) has a rich history that is based on the defining principle that links between actors are informative. The advent and availability of Internet technology has created an explosion in online social networks and a transformation in SNA. The analysis of today's social networks is a difficult *Big Data* problem, which requires the integration of statistics and computer science to leverage networks for knowledge mining and discovery (Manyika et al. 2011). Historically, scientists have had to rely on tractable records of social interactions and experiments (e.g., Milgram's small world experiment); now they have a luxury of accessing huge digital databases of relational social data. SNA relies on diverse data representations and relational information, which may include (among others), tracked relationships among actors, events, and other covariate information (Scott and Carrington 2011). Modeling social networks is especially challenging due to the heterogeneity of the populations represented, and

the broad spectrum of information represented in the data itself. Modern applications of SNA include, among others, the estimation of influence, privacy protection, trust (and distrust) microblogging, and web browsing.

In this review, we focus on probabilistic graphical models (PGMs), which have demonstrated promise in modeling social networks (Lauritzen 1996; Koller and Friedman 2009). PGMs represent a marriage between graph theory and probability theory that offers flexible modeling paradigms with good interpretably. The graphical representation consists of nodes connected by edges, which may be directed (Bayesian networks) or undirected (Markov networks). The relationship between nodes in a graph can be interpreted in terms of conditional independencies. These independencies can be read directly from the graph and enable a tractable decomposition of the joint distribution possible through the use of conditional probabilities. In this setting, the compact representation of a high-dimensional joint distribution of random variables $\{X_1, X_2, \ldots, X_p\}$ can be represented explicitly in a factorized form that has a graphical interpretation rooted in conditional independencies.

A powerful feature of the PGM modeling paradigms is the ability to perform probabilistic queries and reasoning on the graph at multiple levels (Koller and Friedman 2009). Queries of interest may include the estimation of probabilities (joint, conditional, or marginal), reasoning about variables in light of new evidence (causal, evidential, and inter-causal reasoning), and quantitative predictions through the use of the graph as a generator in simulations. Another attractive feature of PGMs is their inherent flexibility to model variables that follow different distributions, and the ability to bring in a priori information in to the learning process.

In this review, we outline the basic theory of PGMs, along with the parameter and structural learning. The topic of PGMs is extremely rich in content and theory. Several existing surveys on the topic of graphical models that are similar in spirit include (e.g., Goldenberg et al. 2010; Daud et al. 2010; Salter-Townshend et al. 2012; Srihari 2014). Our review differs from existing reviews in both style and content. One distinguishing feature is that we illustrate the different modeling paradigms using accessible and simple models. Simple examples facilitate a connection between theory and practice. Once this connection is established, we highlight more complex recent applications in SNA that differ from each other in both the nature of the data and objectives of the modeling. These applications reveal the inherent flexibility of PGMs to model a broad spectrum of data that target relevant open challenges and questions in SNA. We address both directed PGMs, known as Bayesian networks (BNs), and undirected PGMs, known as Markov networks, in Sects. 2 and 3, respectively. In Sect. 4, some of the current challenges are highlighted, comparisons between directed and undirected paradigms are made, and future directions and opportunities for PGM-based research in SNA are also highlighted.

2 Directed probabilistic graphical models

Bayesian networks (BNs) are a special class of PGMs that capture directed dependencies between variables, which may represent cause and effect relationships. The edges in a BN form a directed acyclic graphs (DAGs). The DAG architecture conveys a critical modeling assumption that there is no feedback via cycles in the graph. BNs obey the Markov assumption which states that each variable, X_i, is independent of its non-descendants (unconnected nodes), given its parents in G. Taken together, these assumptions enable the compact representation of the high-dimensional joint probability distribution of the variables in the model. Despite their flexibility, the use of directed graphs in SNA has been somewhat limited, although the applications that we highlight are diverse. We describe the basic principles of these directed PGMs and motivate them with applications in the literature, which showcase their utility in SNA.

Static Bayesian Networks Our major focus is static BNs, which utilize data from a single *snapshot* of a social community at a given time point. A DAG conveys precise information regarding the conditional independencies between modeled variables (nodes). For a set of random variables $\{X_1, X_2, \ldots, X_n\}$ is a network with the structure that encodes conditional independence relationships:

$$P(X_1, X_2, \ldots, X_n) = P(G) \prod_{i=1}^{n} P(X_i \mid \mathrm{pa}(X_i), \Theta_i), \qquad (1)$$

where $P(G)$ is the prior distribution over the graph G, $\mathrm{pa}(X_i)$ are the parent nodes of child X_i, and Θ_i denotes the parameters of the local probability distribution. The prior comes into play only when an expert cannot describe the graph and structural learning is required. Structures and relationships that are more likely (and less likely) can be embedded into $P(G)$ to influence searches through the posterior model space.

A simple and fully parameterized BN for a course at a University is shown in Fig. 1. This network can be viewed as a template model in which different sections of a course are taught by different professors ($\mathrm{Pr} \in \{p0, p1\}$). Similar templates can be used for various courses, e.g., Calculus, Introduction to Chemistry, etc. In this example, different teaching assistants (TA) vary in their teaching effectiveness ($\mathrm{TA}_q \in \{\mathrm{Poor}, \mathrm{Fair}, \mathrm{Good}\}$), and their own grade in the course influences their overall ability to convey the material in the class in well ($\mathrm{TA}_g \in \{A, B, C\}$). A student's grade ($\mathrm{Grade} \in \{A, B, C\}$) is caused by their intelligence

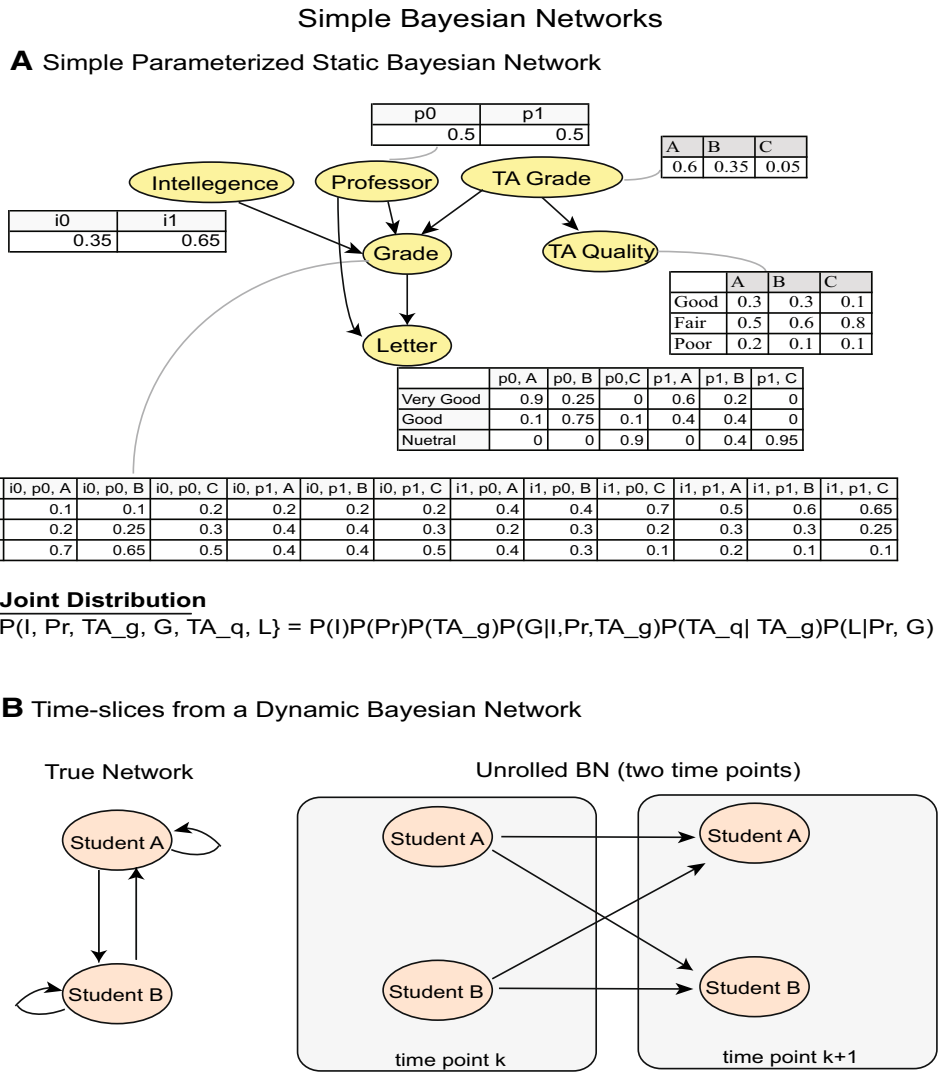


Fig. 1 **a** Simple example of a parameterized Bayesian Network of a University multi-section course. In this scenario, a student may enroll in a course taught by different `Professors`. A student's `Grade` is influenced by the student's `Intelligence`, the `Professor`, and the effectiveness of their TA as measured by the `TA grade` in the course. The `TA Quality` is a direct reflection of the TA's mastery of the material (their grade). The quality of a student's `Letter` of recommendation is dependent on the grade in the course, and also the professor. **b** Time slices from a dynamic Bayesian network of two students that interact in a study group. The true network is given to the left, which includes both feedback and a cycle (prohibited in static BNs). Relationships can *unrolled* in a DBN across discrete time points (*right*)

$\{i0, i1\}$, the professor of the course, and the grade of the TA. Finally, if a professor is asked to write a letter of recommendation for a particular student, for simplicity, this may be based on the performance of the student in the course and the professor (e.g., some are more prone to write positive letters). This scenario, although overly simplistic, may hold in large classroom settings where the teacher does not get to know the individual students well.

The Markov assumption is apparent through the conditional probability tables (CPDs) for each node, which depend only on the parents. The head nodes (top layer) have no parents, thus the CPD table is simply a marginal probability that will sum to one across the different states of the discrete variable. On the other hand, nodes with parents are conditional on the possible combinations of the parent states. It is evident that even for a small number of parents, and a small number of states for those parents, the CPD tables can grow quickly. In our example, $P(G \mid I, \mathrm{Pr}, \mathrm{TA}_g)$ has 12 possible states (or scenarios) that can occur that would influence grade-level.

Our toy example is an expert system that was written down according to knowledge about the modeling domain. Importantly, there are many scenarios that may be more or less realistic, which may include changing some edges or the addition of new variables. Nonetheless, it is often the case that a network structure can be accurately described by an expert. When a structure is prescribed for a BN, parameter learning is still required. This comes in the form of CPD tables for discrete distributions (Fig. 1). In our simple example, the probabilities for the CPD table may have been extracted from teaching evaluations, grades, or other means.

As demonstrated in our simple example, each child node is dependent on its parent nodes. The parameter learning can be viewed as a *local model* or *distribution* that involves only the child and the parents. These local models are the

building blocks of the graphical model and they make up the factors in the product for the joint distribution (Eq. 1). When the variables in the model are continuous, the specification of a local model requires distribution parameters. For example, if $\{X_1, X_2, X_3\}$ are Gaussian, and $X_1 \longrightarrow X_2 \longleftarrow X_3$, then a local model would be of the form $X_3 \sim N(\beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_3, \sigma^2)$. Therefore, in the continuous case, the local model can be viewed as a regression on the parents. Another popular local model is a conditional Gaussian Bayesian network (CG-BN), which gives rise to regressions in which a continuous child node may have parents that are both discrete or continuous (Lauritzen 1996). To enable factorization, CG-BNs prohibit the discrete children from having continuous parents.

Structural learning is required when the network is not known and has to be learned from the data. The objective function for maximization is the posterior probability of a graph, G, given by:

$$P(G \mid X) \propto P(X \mid G)P(G),$$

where $P(G)$ is the prior on the graph. The marginal likelihood, $P(X \mid G)$, requires complex integration over the parameters Θ:

$$P(X \mid G) \propto \int P(X \mid \theta, G)P(\theta \mid G)\mathrm{d}\theta,$$

which can be alleviated with the use of conjugate priors. In an effort to accelerate the learning process, and prevent from over-fitting, a *fan-in* assumption is typically adopted. This limits the number of parents that a node can have (e.g., a node can have no more than three parents). The graph prior $P(G)$ can be explicitly used to encourage certain relationships, and penalize against others (Mukherjee and Speed 2008; Hageman et al. 2011). Computation relies on the fact that each node in the network, together with the corresponding parents, represents a local model, which can be described by a regression. These individual regressions have priors on their parameters, $P(X \mid G)$, for example a normal-Wishart prior can be used for nodes that follow a normal distribution. In practice, the posterior in a graph is calculated as a product of the *local models*, which is valid representation under the Markov assumption. Possible local models are often *pre-computed* in an effort to ease the computational demand of the learning algorithms.

The structural learning problem concerns identifying a *global* network that assembles these local models in a optimal way. The process is a major challenge (NP-hard), as the number of possible networks is super-exponential with the number of nodes (Chickering et al. 2001). Structural learning methods rely on sampling-based approaches or a greedy optimization, e.g., hill climbing or simulated annealing (Heckerman 2008). Sampling-based approaches rely on Markov Chain Monte Carlo (MCMC) techniques that sample the posterior distribution by moving through model space according to a proposal distribution. The *proposal* represents the modification to the current graph, G_{curr}, which is then evaluated and potentially accepted, G_{new}, (kept in the sample) or rejected (not kept in the sample, another proposal is attempted) (Madigan et al. 1995). A widely used proposal for a new graph in the Markov chain is to either add, delete, or reverse a single edge (Fig. 2). This proposal is implemented within a Metropolis–Hastings framework. The acceptance criterion for a new graph is determined by:

$$\min\left\{1, \underbrace{\frac{P(G_{\text{new}} \mid X)}{P(G_{\text{curr}} \mid X)}}_{\text{BayesFactor}} \cdot \underbrace{\frac{Q(G_{\text{curr}} \mid G_{\text{new}})}{Q(G_{\text{new}} \mid G_{\text{curr}})}}_{\text{HastingsRatio}}\right\}. \quad (2)$$

The Bayes factor gives a measure of goodness of fit of the proposed graph relative to the current. The Hastings ratio is simply the neighborhood size of possible moves from the current and new graph, equivalently,

$\text{Neighborhood}(G_{\text{curr}})/\text{Neighborhood}(G_{\text{new}})$ (=5/6 in Fig. 2).

The directionality and causal structure of the inferred model make BN an attractive modeling paradigm for social networks that capture cause and effect relationships. Screen-based bayes net structure (SBNS) was developed as a search strategy for large-scale data, which relies on the adopted assumption of sparsity in the overall network structure (Goldenberg and Moore 2004). SBSN enforces the sparsity through a two stage process, which frames the structural learning problem as market basket analysis task. The algorithm relies on the theory of frequent sets and support, to first screen for local modules of nodes, and then connect them through a global structure search. The market basket framework lends itself to *transaction style* data, which is by nature large, sparse and binary. The learning problem is to identify an influence graph based on derived features of the binary transaction data. In this case, actors are assumed to be linked to each other indirectly through items or events. A simple example of individuals linked through a conference is shown in Fig. 3a. In this example, the conference attendance (transaction items) can be used to infer a network of social influence between individuals, which adds insights into the social hierarchy that are not apparent in classical interaction networks. The method was shown to be effective for modeling a variety of SNs, including citation networks, collaboration data, and movie appearance records (Berry and Linoff 1997).

Koelle et al. proposed applications of BNs to SNA for the prediction of novel links and pre-specified node features (e.g., leadership potential) (Koelle et al. 2006). The authors emphasize the advantage of BN to account for

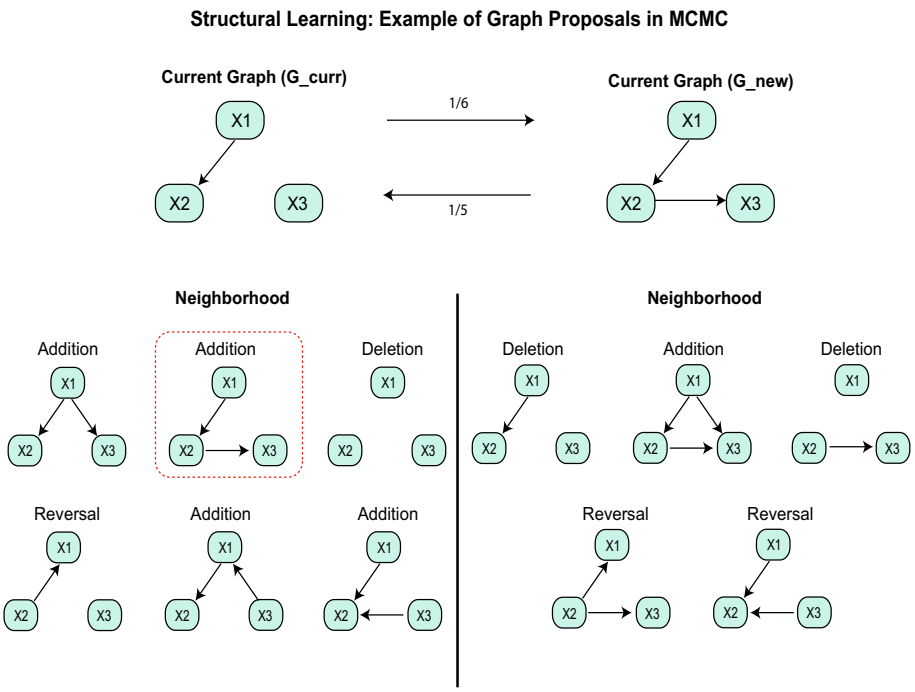


Fig. 2 A widely used proposal in an MCMC sampling for structural learning in a BN is the addition, reversal, or deletion of an edge from the current graph, G_{curr}, to form a new graph, G_{new}. The proposal move is selected at random from models in the *neighborhood* with a probability. The Metropolis–Hastings acceptance criteria are a function of the neighborhood size for the current and propose graphs and the overall fit of those graphs to the data as measured by the Bayes factor

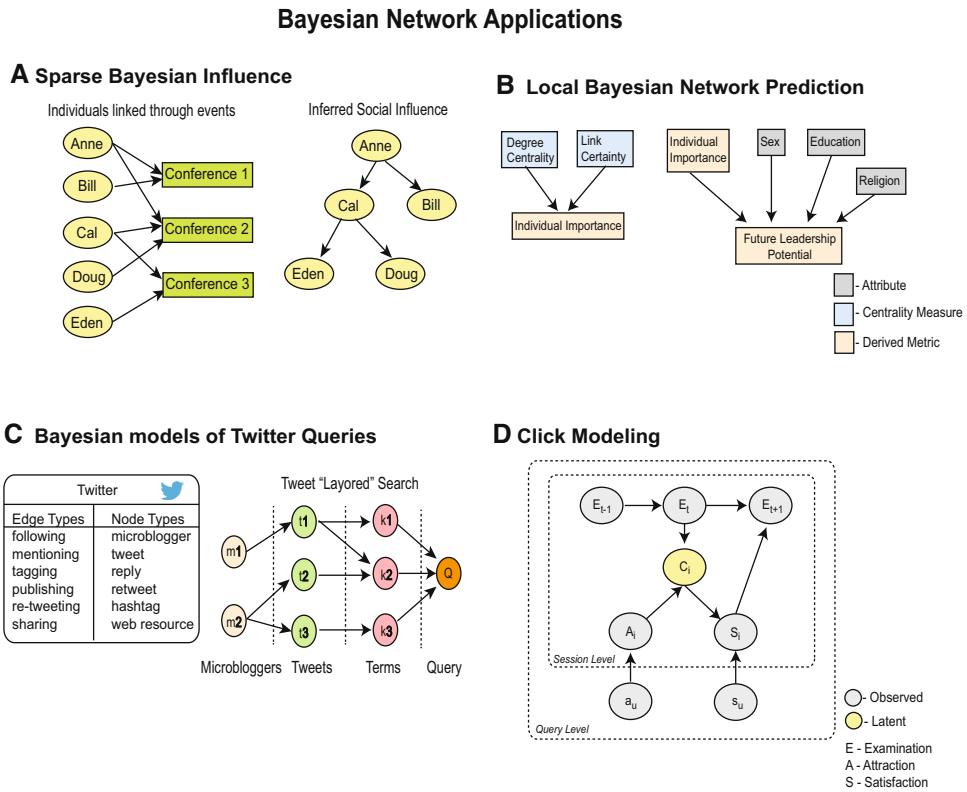


Fig. 3 Simplified schematics of select examples of Bayesian networks in social networks. **a** Inferring influence based on transaction style data that links actors to events. **b** DAGs can utilize network features such as attributes and centrality measures on the network itself to predict derived metrics, e.g., individual importance or leadership potential. **c** Twitter is a microblogging community, which can be queried using a retrieval model that is based on a Bayesian network. **d** An application of DBNs for click modeling in a browser. The temporal dimension is a click sequence that connects the examination of webpages with attraction and satisfaction

uncertainty, noise, and incompleteness in the network. For example, a topology-based network measures such as *degree centrality*, which is often used as a surrogate for *importance*, are subject to summarizations over incomplete and sometimes erroneous data. Comparatively, a BN affords more flexibility that enables measures such as *importance* to be estimated in a more data-dependent manner. Koelle et al. provide an example of combining topology-based network measures with covariate information (Fig. 3b). Directed inference of this type leverages small

local models, which can be naturally translated to regression or classification problems, depending on the child node (response variable). In this setting, the local BN can be evaluated at the node-level, ranked probability estimates can be used for predictive purposes, and the output serves as a surrogate for model fit on a given structure.

Privacy protection is a major concern amongst users in online social networks. Generally, people prefer that their personal information is shared in small circles of friends and family, and shielded from strangers. Despite this common desire, relatively simple BNs have been shown to be successful in the invasion of privacy though the inference of personal attributes, which have been shielded through privacy settings (He et al. 2006). The BNs operate under the often accurate assumption that friends in social circles are likely to share common attributes. In 2006, the recommendation by He et al. to improve privacy was to hide friend lists through privacy settings, and to request that friends hide their personal attributes. Practically speaking, setting the optimal privacy settings is complex, and can be a tedious and difficult for an average user (Lipford et al. 2008). In 2010, a privacy wizard template was proposed, which automates a persons privacy settings based on an implicit set of rules derived using Naive Bayes (the simplest BN) or decision tree methods (Fang and LeFevre 2010).

On the other side of the application spectrum, BNs are useful for recommending products and services, to users, taking into account their interests, needs and communications patterns. Belief propagation has been used to summarize belief about a product and propagate that belief through a BN (Ayday and Fekri 2010; Yang et al. 2013). Belief propagation is the process in which node marginal distributions (beliefs) are updated in light of *new evidence*. In the case of a BN, evidence (e.g., opinion or ratings) is absorbed and propagated through a computational object known as a junction tree, resulting in updated marginal distributions. Comparing the network marginals before and after evidence is entered and propagated conveys a system-wide effect of influence(s), and insights into how perception or ratings change when recommendations are passed through a network. Despite its simplicity, the BN approach has been shown to be competitive with the more classical collaborative filtering (CF)-based recommendation. Trust (and distrust) can be highly variable dynamic processes, which depends not only on distance from a recommender, but also, the characteristics of the network users (Wang and Vassileva 2003; Kuter and Golbeck 2007). Accounting for trust in recommendation systems is an open area of research

Microblogging networks represent another effective venue for rapidly disseminating information and influence throughout a community. Twitter is the most well-known microblogging network, in which posts (tweets) are short and time-sensitive with respect to the reference of current topics (Kwak et al. 2010). Users within microblogging networks of this type participate though the act of *following* and *being followed*, which gives rise naturally to directed associations (Java et al. 2007). With over 50 million tweets submitted daily, ranking and querying microblogs has become an important and active area of open research. Jabeur et al. proposed a retrieval model for tweet searches, which takes into account a number of factors, including hashtags, influence of the microbloggers, and the time (Jabeur et al. 2012a, b). A query relevance function was developed based on a BN that leverages the PageRank algorithm to estimate parameters, such as influence, in the model (Fig. 3c). The retrieval model was shown to outperform traditional methods for information retrieval on Twitter data from the TREC Tweets 2011 corpus (Ounis et al. 2011).

Dynamic Bayesian Networks The static BNs described depict a network at a single time point. This is most often an oversimplification of the true nature of the network, which is inherently dynamic. Modeling the dynamics of a network over the time-course can be achieved in the BN framework with additional modeling assumptions. Dynamic Bayesian networks (DBNs) provide compact representations for encoding structured probability distributions over arbitrarily long time-courses (Murphy 2002). State-space models, such as hidden Markov model (HMM) and Kalman filter models (KFMs), can be viewed as a special class of the more general DBN. Specifically, KFMs require unimodal linear Gaussian assumptions on the state-space variables. HMMs do not allow for factorizations within the state-space, but can be extended to hierarchical HMMs for this purpose. Overall, DBNs enable a more general representation of sequential or time-course data.

DBN modeling is achieved through the use of template models, which are instantiated, i.e., duplicated, over multiple time points. The relationships between the variables within a template are fixed, and represent the inherent dependencies between *ground variables* in the model. There are three types of edges in a DBN. Intra-time slice edges represent dependencies within a time slice. Persistent edges link the same variable in two time slices; for example, the velocity of a vehicle at a time slice is very dependent on the velocity of the vehicle in the previous time slice. Finally, inter-time slice edges connect different variables between time slices, for example, the velocity of a car may also be influenced by weather at the previous time slice.

The objective is to model a template variable over a discretized time-course, $X^0 \cdots X^T$, and represent $P(X^0 : X^T)$ as a function of the templates over the range of time points. Reducing the temporal problem to conditional template models makes the problem computationally tractable, but requires the specification of a fixed structure

across the entire time trajectory. In a DBN, the probability for a random variable X spanning the time-course can be given in factored form,

$$P(X^{0:T}) = P(X^0) \prod_{t=0}^{T-1} P(X^{t+1} \mid X^t),$$

where X^0 represents the initial state, and the conditional probability terms of the form $P(X^{t+1} \mid X^t)$ convey the conditional independence assumptions. The conditional representation of the likelihood is similar in spirit to the static BN representation, but conveys the conditional independence with respect to time. The Markov assumption enables this factorization, which has different, yet analogous meanings in static and dynamic BNs. In a DBN, the Markov assumption explains the memoryless property, i.e., that the current state depends on the previous and is conditionally independent of the past $(X^{t+1} \perp X^{0:t-1} \mid X^t)$. Comparatively, in static BNs, the Markov assumption only captures nodes' independence of their non-descendants, given the states of their parents.

Briefly, the learning paradigms are rather similar. Structural learning is typically achieved by the same scoring strategies, but with the added constraint that the structure must repeat over time (Friedman et al. 1998). Such a constraint alleviates the computational burden for search strategies. Additionally, the best initial structure can be searched for independently from the remainder of the time-course. The search is performed either through greedy hill climbing or sampling.

A major advantage of DBNs is that they can be enriched to accommodate more complex interactions that would violate DAG assumptions in a static BN. Figure 1b shows a simple example of a common situation where the true network has feedback in the form of self-loops and a cycle in the graph. This feedback is prohibited in a static BN, but can be captured in a DBN. In this scenario, two students form a study group, but also self-study, in an effort to improve learning outcomes. The *unrolled BN* captures these relationships for two time slices that contains persistent and inter-time slice edges. These relationships are preserved over a time series (e.g., a semester), thereby forming a *template model*.

Despite the fact that social networks are inherently dynamic, the applications of DBNs in SNA have been limited. Importantly, there have been many attempts to model social networks probabilistically over time, but not in the strict PGM context, which is the focus of this review; many of these advances are mentioned in the discussion. Chapelle et al. used DBNs to model web users' browsing history (Chapelle and Zhang 2009). The DBN extends the traditional and widely used cascade model for browsing behavior. The dynamic of click sequences for single click (Fig. 3d) takes into account the information at the query and session levels, differentiating perceived/actual attraction (a_u and A_i respectively) and perceived/actual satisfaction (s_u and S_i respectively) with links. At each click (time-step), the hidden binary variables for examination (E_i) and satisfaction (S_i) track the time progression to predict future clicks. The DBM approach was shown to outperform traditional methods, and highlighted the sensitivity of click modeling to measures of relevance and popularity at the query level.

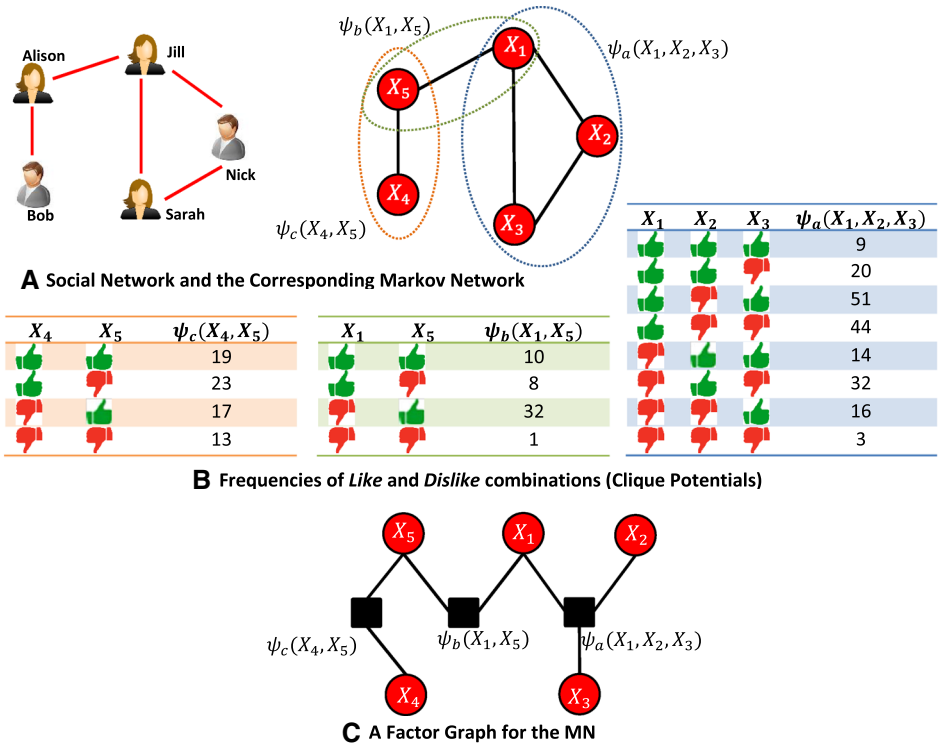
Meetings can be viewed as social events, in which valuable information is exchanged mainly through speech. Effectively processing, capturing, and organizing this information can be costly, but important in order to maximize the impact and information flow for participants. Dielman et al. cast the problem of meeting structuring as a DBN, which partitions meetings into sequences of actions or phases based on audio (Dielmann and Renals 2004). DBNs outperformed baseline HMMs in detecting *meeting actions* in a smart room, such as dialog, notes at the board, computer presentations, and presentations at the board.

Twitter and microblogs, in general, have become a major resource for the media to obtain *breaking news* or a the occurrence of a *critical event*. Recently, Sakaki et al. showed that tweet modeling via Kalman Filtering is effective for the prediction of earthquakes of a certain magnitude in Japan. Furthermore, they developed a reporting system *Torreter*, which is quicker than the existing government reporting system in warning registered individuals through email of an impending quake (Jansen et al. 2009).

3 Undirected probabilistic graphical models

Markov networks (MNs), also known as Markov random fields (MRFs), are PGMs with undirected edges. Similar to directed BNs, a MN is a representation of the joint probability distribution between random variables (represented by nodes), where the absence of an edge between two nodes implies conditional independence between the nodes, given the other nodes in the network. In this review, we restrict our focus to MNs, Markov logic networks (MLNs) and exponential random graph models (ERGMs), which can be viewed as generalizations of the random graphs (Frank and Strauss 1986), and are widely used in SNA (Newman et al. 2002). The basic formulation of these models and their utility in SNA will be highlighted.

Markov networks express the joint probability of given random variables by decomposing the network into smaller complete sub-graphs known as *cliques*, and use maximal cliques to capture the random variables' dependencies. A clique is a *maximal clique* if it cannot be extended to



X_4	X_5	$\psi_c(X_4, X_5)$
👍	👍	19
👍	👎	23
👎	👍	17
👎	👎	13

X_1	X_5	$\psi_b(X_1, X_5)$
👍	👍	10
👍	👎	8
👎	👍	32
👎	👎	1

X_1	X_2	X_3	$\psi_a(X_1, X_2, X_3)$
👍	👍	👍	9
👍	👍	👎	20
👍	👎	👍	51
👍	👎	👎	44
👎	👍	👍	14
👎	👍	👎	32
👎	👎	👍	16
👎	👎	👎	3

Fig. 4 **a** A social network and a corresponding MN model (**b**). The nodes of the MN are actors' decisions and the variable dependencies are defined based on the ties in the social network. The three tables show the frequencies of hypothetical (observed) *Like* and *Dislike* combinations. **c** A factor graph for the MN in **b**

include additional adjacent nodes. The clique representation enables a compact factorization of the probability density function (pdf). The joint pdf of n random variables, $X = \{X_1, \ldots, X_n\}$, with conditional (in)dependencies captured by a graph, can be expressed as:

$$P(X) = \frac{1}{Z} \prod_{C \in \Omega} \psi_C(X_C), \tag{3}$$

where C is a maximal clique in the set of maximal cliques Ω. Let X_C denote a subset of X comprised of the random variables that form clique C. Clique potential $\psi_C(X_C)$ is a function of these variables (e.g., the frequency of distinct realizations of the random variables forming the clique). A unique clique potential can be specified for each clique; the size of a clique can be one determinant of the corresponding clique potential; however, cliques of the same size may have different clique potentials. The clique potentials are positive functions that capture the dependence of the variables within the cliques (Koller and Friedman 2009). The normalizing constant, also known as the partition function, is given as:

$$Z = \sum_{X \in \chi} \prod_{C \in \Omega} \psi_C(X_C).$$

Each clique potential in a MN is specified by a *factor*, which can be viewed as a table of weights for each combination of values of variables in the potential. In some special cases of MNs such as log-linear models (Murphy 2012), clique potentials are represented by a set of functions, termed *features*, with associated weights (i.e., $\theta_C \phi_C(X_C) = \log(\psi_C(X_C))$, where $\phi_C(X_C)$ is a feature derived from the values of the variables in set X_C) and θ_C is the weight of $\phi_C(X_C)$ estimated at the parameter learning stage.

The *Hammersley–Clifford* theorem specifies the conditions under which a positive probability distribution can be represented as a MN. Specifically, the given representation (Eq. 3) implies conditional independencies between the maximal cliques and is, by definition, a Gibbs measure (Murphy 2012).

A simple example of the use of MNs to study the collective behavior in a social network is shown in Fig. 4. Suppose each member of a friendship network (Fig. 4a) is looking to make a decision about purchasing a given product, "liking" a post in an online forum, supporting a political party, participating in a school activity or choosing a family doctor. In this setting, the random variables in the nodes are binary and the edges indicate pairwise dependencies between them. Let $X_1, \ldots, X_5$ denote five random variables, each of which takes on the value of 1 or 0 to signal the member's attitude—*Like* or *Dislike*, respectively. Figure 4b depicts the MN with three cliques a, b and c, $\Omega = \{a, b, c\}$, with $X_a = \{X_1, X_2, X_3\}$, $X_b = \{X_1, X_5\}$ and $X_c = \{X_4, X_5\}$. The MN includes three clique potentials, $\psi_a(X_1, X_2, X_3), \psi_b(X_1, X_5)$ and $\psi_c(X_4, X_5)$, satisfying the requirements of the *Hammersley–Clifford*

theorem under the assumption that a member's decision can be affected by only their immediate friends and that it matters if those friends are also friends with each other. The joint probability function of $X_1, \ldots, X_5$ is expressed as:

$$P(X_1, \ldots, X_5) = \frac{1}{Z}\psi_a(X_1, X_2, X_3)\psi_b(X_1, X_5)\psi_c(X_4, X_5),$$

where $Z = \sum_{X_1, X_2, \ldots, X_5} \psi_a(X_1, X_2, X_3)\psi_b(X_1, X_5)\psi_c(X_4, X_5) = 3 \times 1 \times 13 + \cdots + 9 \times 10 \times 19$ (the summation is taken over all the 2^5 distinct realizations of the model's variables; the first explicitly written term above corresponds to the case where the variables are all zeros, and the last term corresponds to the case where the variables are all ones). To parameterize the MN and obtain a log-linear model, let $\theta_a\phi_a(X_1, X_2, X_3) = \log(\psi_a(X_1, X_2, X_3))$, $\theta_b\phi_b(X_1, X_5) = \log(\psi_b(X_1, X_5))$ and $\theta_c\phi_c(X_4, X_5) = \log(\psi_c(X_4, X_5))$. With $\Theta = \{\theta_a, \theta_b, \theta_c\}$, the joint probability function defined by the MN becomes:

$$P(X_1, \ldots, X_5) = \frac{1}{Z(\Theta)}\exp\{\theta_a\phi_a(X_1, X_2, X_3) + \theta_b\phi_b(X_1, X_5) + \theta_c\phi_c(X_4, X_5)\},$$

where $Z(\Theta)$ is the partition function. The set of the model parameters, Θ, would need to be estimated from any available data of known decisions made by the social network members. Figure 4c depicts a factor graph corresponding to the MN. Factor graphs are bipartite graphs used to specify the factorization of the probability distribution function, and also, to inform the computation of marginal probability distributions of MN variables (Murphy 2012).

MN specification problems, including parameters estimation and structure learning from data, can be quite challenging. The main difficulty in MN parameter estimation is that the maximum likelihood problem formulated with Eq. 3 has no analytical solution due to the complex expression of Z (Lee et al. 2006). The problem of finding the optimal structure of the MN using available data, similar to BNs, is even more challenging (Bromberg et al. 2009). Currently existing approaches to structure learning are either constraint-based or score-based (see Koller and Friedman 2009; Ding 2011; Schmidt et al. 2010 for more details).

MNs have found increased utility in SNA with the emergence of online social networks (OSNs) and digital social media (see Bonchi et al. 2011 for a review of key problems in SNA). The need to capture non-causal dependencies within and between data instances (e.g., profile information) and observed relationships (e.g., hyperlinks) in these applications is exacerbated by the presence of missing or hidden data in OSNs (Xiang and Neville 2013). A popular problem instance in this domain, that of user (missing) profile prediction, has been attacked using MNs (Taskar et al. 2002; Neville and Jensen 2007).

Along with the problem of predicting missing profiles, link prediction is among the most prominent problems in Big Data SNA. Multiple variations of MNs that have been used to estimate the probability that a (unobserved) link exists between nodes include Markov logic networks, relational Markov networks, relational Bayesian networks and relational dependency networks (Al Hasan and Zaki 2011; Chen et al. 2013; Tresp and Nickel 2013). Detection of community structures is another area of MN application (Newman 2006). Communities can be discovered through examination and subsetting (cutting) network relationships according to labels of interest, and through the use of weighted community detection algorithms. Social network clustering is especially challenging in a dynamic context, e.g. in mobile social networks (Humphreys 2007). Wan et al. employed undirected graphical models (i.e., conditional random fields) constructed from mobile user logs that include both communication records and user movement information (Wan et al. 2012).

Several generative models have been proposed, which are motivated by MNs, and explain the effects of selection and influence (e.g., see Aggarwal 2011). Modeling channeled spread of opinions and rumors, known more generally as diffusion modeling, is an active area of research in SNA (Bach et al. 2012). Several applications of diffusion models have been proposed for social networks including, but not limited to the spread of information (Cowan and Jonard 2004), viral marketing (Kempe et al. 2003), spread of diseases (Anderson and May 1979), the spread of cooperation (Santos et al. 2006). Given a social network, for each node, a corresponding random variable indicates the state of the node (e.g., product or technology adoption) and links in the network represent dependency (Wortman 2008).

Markov logic networks employ a probabilistic framework that integrates MNs with *first-order logic* such that the MN weights are positive for only a small subset of meaningful features viewed as templates (Richardson and Domingos 2006). Formally, let F_i denote a first-order logic formula, i.e., a logical expression comprising constants, variables, functions and predicates, and $w_i \in \mathcal{R}$ denote a scalar weight. An MLN is then defined as a set of pairs (F_i, w_i). From the MLN, the *ground Markov network*, $M_{L,C}$, is constructed (Richardson and Domingos 2006) with the probability distribution (Tresp and Nickel 2013),

$$P(X = x) = \frac{1}{Z}\exp\left(\sum_i w_i n_i(x)\right), \quad (4)$$

where $n_i(x)$ is the number of true groundings (e.g., true logic expressions based on observations) of F_i, i.e., such formulae that hold, in x.

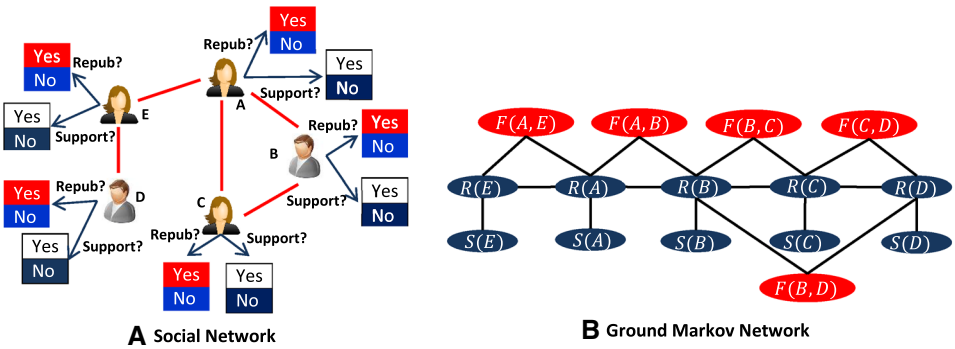


Fig. 5 An example of using MLN in the political science. **a** Depicts a social network of five senators with two attributes. The ground predicates (**b**) are denoted by 15 elliptical nodes. The red ones are captured by the links of the social networks and dark blue nodes indicate nodes' attributes. Two first-order logics, F_1 and F_2, determine the structure of the MLN. There exist five groundings of the F_1 (illustrated by the edges between the $R(x)$ and $S(x)'$ nodes) and 15 groundings of F_2 captured by the rest of the edges. Other examples of MLNs can be found in Tresp and Nickel (2013)

A simple example network of five senators is shown in Fig. 5a. In this setting, each senator supports one of two political parties (Democratic or Republican). Each senator in this network has two attributes (1) political affiliation, ($R(n)$ is 1 if senator n is a republican and 0 otherwise for $n \in \{A, B, C, D, E\}$), and (2) supporting a particular bill, ($S(n)$ is 1 if senator n supports the bill and 0 otherwise). Let $F(n, m)$, a binary symmetric function, denote the relationship between senators n and m ($n \neq m$). Suppose "Republicans do not support the bill" and "If two senators have a relationship and one is republican then so is the other" are two logical statements denoted by F_1 and F_2. The first-order logic format is given as follows:

$$F_1 : \forall n,\ R(n) \Rightarrow \neg S(n),$$
$$F_2 : \forall n, m\ F(n, m) \wedge R(n) \Rightarrow R(m).$$

The MLN is similar to the ground network (Fig. 5b). However, only the combinations of variables corresponding to logical statements, F_1 and F_2, are parameterized in the MLN (all the other weights are zero). Let $X = \{X_1, \ldots, X_{15}\}$ denote the set of all nodes in the $M_{L,C}$ where X_i indicate node i (e.g., X_1 is the node labeled $S(A)$ in Fig. 5b). In an MLN, clique potentials are defined similar to those in Markov networks. Now, one can use this MLN to find the probability that all senators in this example support the bill (i.e., $P(S(A) = 1, \ldots, S(E) = 1)$. To calculate the number of true groundings (n_1 and n_2), both F_1 and F_2 should be examined for all nodes in the observed network. More generally, this approach can be implemented to estimate missing profiles in social networks as well.

Many problems in statistical relational learning, such as link prediction (Domingos et al. 2008), social network modeling, collective classification, link-based clustering and object identification, can be formulated using instances of MLN (Richardson and Domingos 2006). Dierkes et al. used MLNs to investigate the influence of Mobile Social Networks on consumer decision-making behavior. With the call detail records represented by a weighted graph, MLNs were employed in conjunction with *logit* models as the learning technique based on lagged neighborhood variables. The resulting MLNs were used as predictive models for the analysis of the impact of *word of mouth* on churn (the decision to abandon a communication service provider) and purchase decisions (Dierkes et al. 2011).

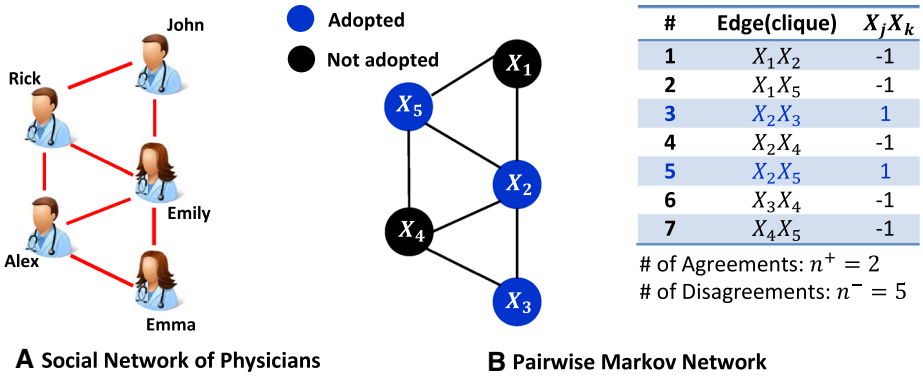
As mentioned above, link mining and link prediction problems can also be addressed using MLNs, since MLNs combine logic and probability reasoning in a single framework (Domingos et al. 2010). Furthermore, the ability of MLNs to represent complex rules by exploiting relational information makes them an appropriate alternative for collective classification (e.g., classification of publications in a citation network, or of hyperlinked webpages) (Crane and McDowell 2011).

The Ising model and its variations form a subclass of MN with foundations in theoretical physics. The Ising model is a discrete and pairwise MN, and is popular in applications in part due to its simplicity (Koller and Friedman 2009). The variables in the model, $X_1 \cdots X_p$, are assumed to be binary, and their joint probability is given as:

$$P(X, \Theta) = \exp\left(\sum_{(j,k)\in E} \theta_{jk} X_j X_k - \Phi(\Theta)\right) \quad \forall\, X \in \chi,$$

where $\chi \in \{-1, 1\}^p$, and $\Phi(\Theta)$ is the log of the partition function

$$\Phi(\Theta) = \log \sum_{x\in\chi} \left[\exp\left(\sum_{(j,k)\in E} \theta_{jk} x_j x_k\right)\right].$$



#	Edge(clique)	X_jX_k
1	X_1X_2	-1
2	X_1X_5	-1
3	X_2X_3	1
4	X_2X_4	-1
5	X_2X_5	1
6	X_3X_4	-1
7	X_4X_5	-1

Fig. 6 An example of implementing the Ising model to find the probability of adopting a new medication. **a** A sub-network of an physicians' advisory network with 5. **b** A pairwise Markov network is constructed where the cliques with size of at most 2 are involved

Special, efficient methods exist for learning the Ising Model parameters from data (Ravikumar et al. 2010). While the model has been originally found useful for understanding magnetism and phase transitions, its utility has later expanded to image processing, neural modeling, and studies of tipping points in economics and social domains (Afrasiabi et al. 2013).

In SNA, the Ising model can be employed to analyze factors such as network substructures and nodal features affecting the opinion formation process. A classical example within this is a study of *medical innovation* spread, namely the adoption of drug tetracycline by 125 physicians in four small cities in Illinois (Van den Bulte and Lilien 2001). A small subset of this network is illustrated in Fig. 6. The adoption status of node k is represented by $X_k \; \forall \; k = 1, \ldots, 5$, where $X_k = 1$ if adopted (blue nodes) and 0 otherwise (black nodes). Since the Ising model only captures pairwise dependencies between variables, the corresponding MN only considers cliques of size two (i.e., dyads); hence, one can concisely write the clique potentials in the form $\psi(X_k, X_j) = X_kX_j$. Let n^+ and n^- denote the number of agreements (i.e., cases with$X_kX_j = 1$ for some k and j) and the number of disagreements (i.e., cases with $X_kX_j = -1$ for some k and j), respectively. Assuming $\theta_{jk} = \theta_A$ if $X_kX_j = 1$ and $\theta_{jk} = \theta_D$ if $X_kX_j = -1$, $\Theta = \{\theta_A, \theta_D\}$, one can obtain $\sum_{(j,k)\in E} \theta_{jk}X_jX_k = \theta_An^+ - \theta_Dn^-$. Hence, the joint probability of all nodes' adoption status is:

$$P(X_1, \ldots, X_5, \Theta) = \exp(\theta_An^+ - \theta_Dn^- - \Phi(\Theta)),$$

where $\Phi(\Theta) = \log \sum_{x\in\chi}[\exp(\theta_An^+ - \theta_Dn^-)]$ is the partition function. For this small example, the table in Fig. 6b shows all clique potentials which are either -1 or 1 based on the network structure. The counts of agreements and disagreements are obtained next (e.g., $n^+ = 5$ and $n^- = 2$ in Fig. 6b). The partition function has 32 additive terms: each combination of X_k's leads to particular values of n^+ and n^- and all these values affect the value of $\Phi(\Theta)$.

In the model presented above, the counts of possible agreements and disagreements depend on the network structure, so the MN can be said to explore the impact of homophily on tetracycline adoption decisions. Note that the model's parameters, θ_A and θ_D, first need to be estimated from any given data (i.e., from a single observation of a network of (non)adopters); however, the approaches to such parameter estimation are beyond the scope of this paper.

Figure 7 depicts the entire physicians' advisory network from a data set prepared by Ron Burt from the 1966 data collected by Coleman et al. (1966) about the spread of medical innovation. The figure illustrates the physicians' network in two different time points and shows how physicians changed their opinions and adopted the new medication overtime. To find the probability of adoption, the Ising model can be modified by considering the impact of nodal attributes on the adoption.

Recently, the Ising Model has been used to examine social behaviors (Vega-Redondo 2007), including collective decision making, opinion formation and adoption of new technologies or products (Grabowski and Kosiński 2006; Krause et al. 2012). For example, Fellows et al. proposed a random model of the full network by modeling nodal attributes as random variates. They utilized the new model formulation to analyze a peer social network from the National Longitudinal Study of Adolescent Health (Fellows and Handcock 2012). Agliari et al. (2010) proposed a model to extract the underlying dynamics of social systems based on diffusive effects and people strategic choices to convince others. Through the adaptation of a cost function, based on the Ising model, for social interactions between individuals, they showed by numerical simulation that a steady-state is obtained through natural dynamics of social systems.

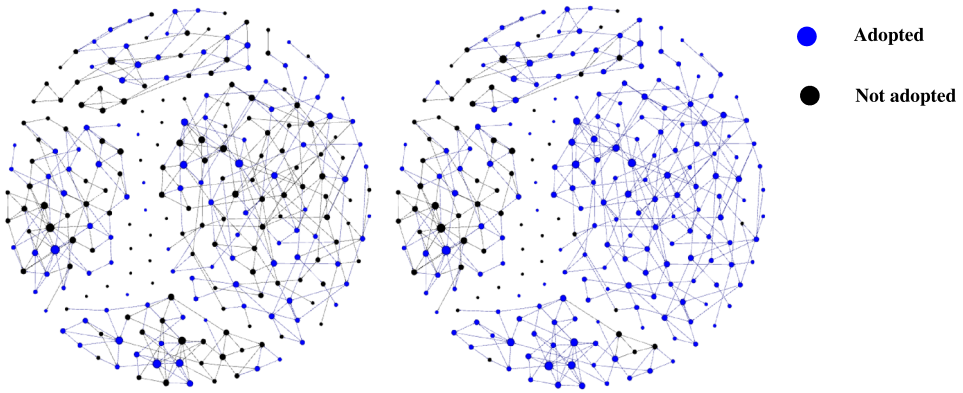


Fig. 7 The spread of new drug adoption through an advisory network of physicians: two snapshots at different time points, about 2 years apart (from *left* to *right*). The growth dynamics in the number of adopters can be analyzed with an Ising Model

Exponential random graph models (ERGMs) (Wasserman and Pattison 1996), also known as the p^*-class models, are among the most widely used network approaches to modeling social networks in recent years (Pattison and Wasserman 1999; Robins et al. 1999, 2007a). A social network of individuals is denoted by graph G_s with N nodes and M edges, $M \leq \binom{N}{2}$. The corresponding adjacency matrix of is denoted by $Y = [y_{ij}]_{N\times N}$, where y_{ij} is a random variable and defined as follows:

$$y_{ij} = \begin{cases} 1 & \text{if there exists a link between nodes } i \text{ and } j \ \forall i,j, i \neq j \\ 0 & \text{otherwise.} \end{cases}$$

Based on an ERGM, the probability of any observed network, y, is:

$$P(Y = y, \Theta) = \frac{1}{Z}\exp\left(\sum_{i=1}^{K} \theta_i f_i(y)\right), \tag{5}$$

where $f_i(y), i = 1, \ldots, K$, are called sufficient statistics (Morris et al. 2008; Lusher et al. 2012), or motifs based on configurations of the observed graph and $\Theta = \{\theta_1, \ldots, \theta_K\}$ is a K-vector of parameters (K is the number of different sufficient statistics used in the model). Network configurations used to compute sufficient statistics, including but not limited to network edge count (tie between two actors), as well as counts of 2-stars (two ties sharing an actor) and triads of various types, are related to communication patterns among actors in a social network (see Lusher et al. 2012 for more details about network configurations). The parameters of an ERGM describe the probabilities of a wide variety of possible configurations in social networks (Robins et al. 2001). Again, Z is called the normalization constant.

As an example, a social network of five individuals is assumed. Since the edges (ties) between nodes are considered as random variables, the given network is the most likely realization out of many possible networks. In this case, an ERGM constructs a probability distribution over all possible networks with five nodes. Figure 7 illustrates the social network (A) and the corresponding graph where edges, $y_{ij}\ \forall i,j = 1, \ldots, 5$ represent random variables along with five sufficient statistics ($f_i(y)\ i = 1, \ldots, 5$) including *edge*, *2-star*, *3-star*, *4-star and triangle* (B). The probability distribution of any possible network is obtained as follows:

$$P(Y = y, \Theta) = \frac{1}{Z}\exp(\theta_1 f_1(y) + \theta_2 f_2(y) + \theta_3 f_3(y) + \theta_4 f_4(y) + \theta_5 f_5(y)),$$

where y is any observed network with five nodes, $\Theta = \{\theta_1, \ldots, \theta_5\}$, the set of weights of sufficient statistics, are estimated through solving an optimization problem where the probability of the observed network is maximized. The exact computation of the normalization constant, Z, requires handling of many terms (all possible network realization must be considered and their corresponding sufficient statistics calculated). This challenge is conventionally handled using Markov Chain Monte Carlo (MCMC) sampling technique (Snijders et al. 2006).

Some of the first proposed models, e.g., random graphs and p_1 models (Frank and Strauss 1986), used Bernoulli and dyadic dependence structures, which are generally overly simplistic (Robins et al. 2007a). On the contrary, ERGMs are based on Markov dependence assumption (Frank and Strauss 1986) supposing that two possible ties are conditionally dependent when they share an actor (node). Moreover, Markov dependence assumption can be extended to attributed networks which assumes each node has a set of attributes influencing the node's possible incoming and outgoing ties (Robins et al. 2007a) (e.g.,

more experienced actors in an advisory network, more incoming ties). When nodal attributes are taken into account as random variables, ERGMs and MNs can be integrated to model the social network due to similarities that they share (see the Appendix and Fellows and Handcock 2012; Thiemichen et al. 2014; Lusher et al. 2012).

ERGMs have been widely employed to study the network and friendship formation (Song et al. 2014) and global network structural using local structure of the observed network (Uddin et al. 2013a). The observed network is considered as one realization from too many possible networks with similar important characteristics (Robins et al. 2007a). For example, Broekel and Hartog (2013) used ERGMs to identify factors determining the structure of inter-organizational networks based on the single observation. Schaefer and Simpkins (2014) used SNA to study the relation between weight status and friend selection and ERGMs to measure the effects of body mass index on friend selection.

Moreover, Goodreau et al. (2009) used ERGMs to examine the generative processes that give rise to widespread patterns in friendship networks. Cranmer and Desmarais used ERGMs to model co-sponsorship networks in the U.S. Congress and conflict networks in the international system. They determined that several previously unexplored network parameters are acceptable predictors of the U.S. House of Representatives legislative co-sponsorship network (Cranmer and Desmarais 2011).

The ERGMs have also been utilized in modeling the changing communication network structure and classifying networks based on the occurrence of their local features (Uddin et al. 2013a) and to identify micro-level structural properties of physician collaboration network on hospitalization cost and readmission rate (Uddin et al. 2013b). Finally, a ERGM-based model of clustering nodes considering their role in the network has been reported (Salter-Townshend and Murphy 2014).

4 Discussion

Mining social networks for knowledge and discovery has proven to be a very challenging and active research area. This review focussed on PGMs. The directed and undirected PGM paradigms were described and their applications to social networks were highlighted. An important consideration and major challenge is the issue of scalability, not only for PGMs, but for SNA, in general. Structural and parameter learning in high dimensions can be prohibitive. Moreover, for structural learning, both greedy- and sampling-based search strategies can get stuck at local minima, and many graphs may be likelihood equivalent. These numerical caveats can give rise to misleading networks, generating models, and subsequent predictions. In addition, ERGMs can exhibit *degeneracy*, which occurs when the generated networks show little resemblance to the generating model. Proposed modifications to the concept of *goodness of fit* have been proposed to safeguard against the problems of degeneracy (Goodreau 2007; Hunter et al. 2008).

In the majority of applications of PGMs (both directed and undirected) in SNA, the graphical structures are assumed to be either known or designed by human experts (i.e., captured directly by social networks), thereby the learning problem is limited to the parameter estimation. However, practically hand-constructed PGMs for SNA have many barriers: time taken to construct them varies from hours to months, experts can be costly or unavailable, the data may be huge and errors may lead to poor answers. On the other hand, structure learning is NP-hard with the hypothesis space being super-exponential ($2^{O(n^2)}$) networks.

Directed and undirected graphs share common interpretations in terms of conditional independences. Selection of a PGM modeling paradigm is not trivial and is driven by the data and ultimately what the user hopes to achieve with the model. When the relationship can be viewed in terms of cause and effect, BNs are more appropriate, and when the relationship is association, MNs are preferred. Inferences in both paradigms are met with challenges. The types of variables (continuous or discrete) have to be carefully considered. Modeling with a mixture of these variables is possible in the case of BNs under strict assumptions. However, the inference problem becomes more sensitive to sample size, as the parameters estimated for the local models are done so from a potentially reduced population, which can be severely subset by level factors of parent nodes. Another important learning task, outside of the scope of this review, is queries that involve the absorption of evidence (e.g., new data) in the network and propagation through the network. This process is known as belief propagation and it takes place on a factor graph (aka cluster graph). In the case of BNs, the factor graph is a factor tree (aka junction tree), and the propagation schemes give rise to exact inferences of marginal distributions (aka beliefs). On the other hand, in MNs the factor graph may have cycles, which does not ensure exact inference in terms of marginals, but has still been shown to be useful in practice, see Koller and Friedman (2009) for more details.

There are several opportunities to access open source data resources in order to develop and test methodologies for PGMs, and related areas. Max-Plank researchers have released OSN data used in publications, which includes crawled data from Flickr, YouTube, Wikipedia and

Facebook (Mislove et al. 2007; Cha et al. 2008, 2009; Viswanath et al. 2009). Several directed OSNs have been released in the Stanford Network Analysis Package (snap), e.g. from Epinions, Amazon, LiveJournal, Slashdot and Wikipedia voting (Stanford 2011). Recently, a Facebook dataset was released that exhibited convergence properties and was shown to be representative of the underlying population (Gjoka et al. 2010). Document classification datasets have also been released (Getoor 2012). A sample from the CiteSeer database contains 3312 publications from one of six classes, and 4732 links. The Cora dataset consists of 2708 publications classified into seven categories and the citation network has 5429 links. Each publication is described by a binary word vector which indicates the presence of certain words within a collection of 1433. WebKB consists of 877 scientific publications from five classes, contains 1601 links and includes binary word attributes similar to Cora. Terrorism databases are also publicly available (Division 1948; National Consortium for the Study of Terrorism and Responses to Terrorism 2015). The most extensive is the RAND Database of Worldwide Terrorism Incidents, which details terrorist attacks in nine distinct regions of the world across the time-span 1968–2009 (dates vary slightly depending on region) (Division 1948). Several well-known challenges may arise in the analysis and representation of terrorist network data, including incomplete information, latent variables influencing node dynamics, and fuzzy boundaries between terrorists, supporters of terrorists, and the innocent (Sparrow 1991; Krebs 2002). The DBLP computer science bibliography (http://dblp.uni-trier.de/db/) is a massive online database that contains bibliographic meta-data for over 2.6 million publications. There is also ample opportunity to enroll in various *data challenges*, which are often posed by corporations and operators of the networks themselves.

In this review, we surveyed directed and undirected PGMs, and highlighted their applications in modern social networks. Despite limitations that arise related to scalability and inference, it is our opinion that the utility of PGMs has been somewhat under-realized in the social network arena. It is indisputable that methods for understanding social networks have not kept pace with the data explosion. There are several relevant topics and opportunities in social networks, e.g., link predication, collective classification, modeling information diffusion, entity resolution, and viral marketing, where conditional independencies can be leveraged to improve performance. PGMs implicitly convey conditional independence and provide flexible modeling paradigms, which hold tremendous promise and untapped opportunity for SNA.

Acknowledgments A. N. is supported in part by a MURI grant (Number W911NF-09-1-0392) for Unified Research on Network-based Hard/Soft Information Fusion, issued by the US Army Research Office (ARO) under the program management of Dr. John Lavery, in part by the Academy of Finland Grant MineSocMed (Number 268078), and in part by the 2015 U.S. Air Force Summer Faculty Fellowship Program, sponsored by the Air Force Office of Scientific Research. R. H. B. is supported through NSF DMS 1312250.

Appendix

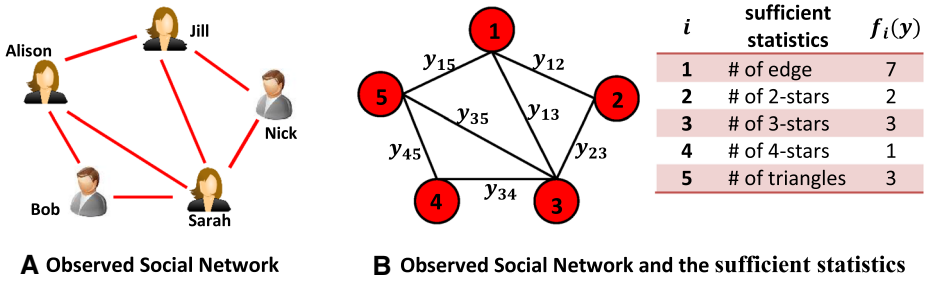
Similarity between MNs and ERGMs

While MNs and ERGMs have been developed in different scientific domains, they both specify exponential family distributions. MN models treat *social network nodes* as random variables, and hence, their utility is most obvious in modeling processes *on networks*; ERGMs, on the other hand, have been conceptualized to model *network formation*, where it is the *edge presence indicators* that are treated as random variables (these random variables are dependent if their corresponding edges share a node). But in fact, this application-related difference in *what to treat as random* is not fundamental. This Appendix works to more rigorously disclose the similarity between MNs and ERGMs by re-defining an ERGM as a PGM. We begin, however, by reviewing the branch of literature devoted exclusively to ERGMs.

Similar to MNs, a well-discussed problem of ERGMs for analyzing social networks is related to the challenge of parameters estimation (Robins et al. 2007b) due to the lack of enough observed data. Robins et al. (2007b) outline this and some other problems associated with ERGMs, e.g., degeneracy in model selection and bimodal distribution shapes (see also Handcock et al. 2003; Rinaldo et al. 2009; Snijders et al. 2006; Handcock et al. 2006).

The roots of ERGMs in the Principle of Maximum Entropy (Park and Newman 2004) and the Hammersley–Clifford theorem have been previously pointed out (Robins et al. 2001; Goldenberg et al. 2010). Here, we illustrate how MNs and ERGMs are similar in terms of the form and structure using most popular significant statistics in ERGMs; under the assumption of Markov dependence, for a given social network, one can build a corresponding Markov network via the following conversion: (1) each node in the Markov network will correspond to an edge in the social network [Fienberg called this construct a "usual graphical model" for ERGMs (Fienberg 2012)], (2) when two edges share a node in the social network, a link will be built between two corresponding nodes in the Markov network.

Corresponding to each possible edge in a social network, a node in an MN network is introduced; note the difference between the original social network and the MN network—they are not the same! Consider an ERGM with the



i	sufficient statistics	$f_i(y)$
1	# of edge	7
2	# of 2-stars	2
3	# of 3-stars	3
4	# of 4-stars	1
5	# of triangles	3

Fig. 8 **a** A social network with 5 nodes and **b** the corresponding realization network (*graph*) and sufficient statistics of the observed network

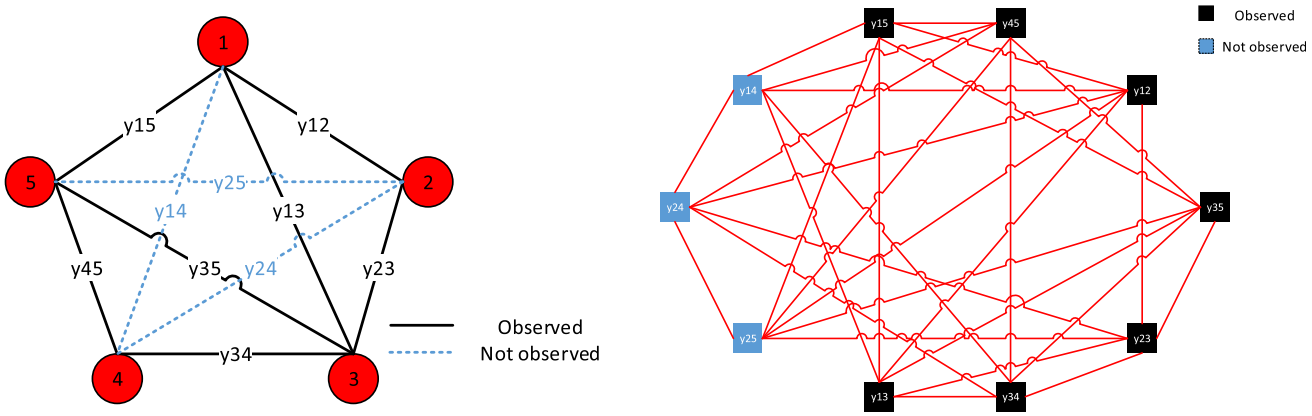


Fig. 9 A social network with five actors (*left*) and its corresponding Markov network (*right*)

significant statistics including the number of edges, $f_1(y)$, the number of k-stars, $f_i(y), i = 2, \ldots, N-1$ and the number of triangles, $f_N(y)$. In an MN, a maximum Entropy (maxent) model proposes the following form for the internal energy of the system, $E_c(x) = -\sum_i \alpha_{ci} g_{ci}$. Define, g_{ci} as i^{th} feature of clique $c \in \Omega$ and α_{ci} is its corresponding weight in G. Thus, $\psi_c(x) = \exp\{\beta_c \sum_{i=1}^{N} \alpha_{ci} g_{ci}\}$. Since there are too many parameters in the MN, they can be deducted by imposing homogeneity constraints similar to that of ERGMs (Robins et al. 2007a). Before imposing such constraints, these following facts are required.

It is straightforward to demonstrate that G encompasses cliques of size $\{3, \ldots, N-1\}$. In addition, all substructure in G_s can be redefined by features in G. Considering these points, we can rewrite the joint probability of all variables represented by the MN, $P(X)$, as follows:

$$P(X) = \frac{1}{Z(\alpha)} \prod_{c=1}^{C} \exp\left(\beta_c \sum_{i=1}^{N} \alpha_{ci} g_{ci}\right) = \frac{1}{Z(\alpha)} \exp\left(\sum_{c=1}^{C} \beta_c \sum_{i=1}^{N} \alpha_{ci} g_{ci}\right). \quad (6)$$

In (4), $Z(\alpha)$ is the partition function which is a function of parameters. The homogeneity assumption, here, means $\alpha_{ci} = \theta'_i \; \forall \; c = 1, \ldots, C$; then $P(X)$ is:

$$P(X) = \frac{1}{Z(\theta')} \exp\left(\sum_{i=1}^{N} \theta'_i \sum_{c=1}^{C} \beta_c g_{ci}\right). \quad (7)$$

In (5), let's $Z' = Z(\theta')$. In addition, we assume that $\sum_{c=1}^{C} \beta_c g_{ci}$ represented by f'_i, means that substructures i in all cliques c are added up by weight β_c. Finally, if we replace f'_i in (5):

$$P(X) = \frac{1}{Z'} \exp\left(\sum_{i=1}^{N} \theta'_i f'_i\right). \quad (8)$$

Comparing $P(Y = y)$ and (4) confirms that ERGMs and MNs are similar and under the following conditions they are identical:

1. $\theta_i = \theta'_i$,
2. $f_i = f'_i = \sum_{c=1}^{C} \beta_c g_{ci}$.

The following Numerical Example (the same example in the ERGM section) depicts similarities between ERGMs and MNs. The social network has five actors, $N = 5$ (Fig. 8). Considering Markov dependency assumption, there exists an unique corresponding Markov network shown in Fig. 9 with 10 nodes.There are 15 cliques (so-called factors) of size three or four,

$$\mathbf{\Phi} = \{\phi_1(y_{12}, y_{13}, y_{14}, y_{15}), \ldots, \phi_{15}(y_{24}, y_{45}, y_{25})\}.$$

As already mentioned, the joint probability function of all variables in each clique is proportional to the internal energy. For instance:

$$\phi_1(x) = \frac{1}{\lambda}\exp\{-\beta_1 E_c(y_{12}, y_{13}, y_{14}, y_{15})\},$$

where $E_1(x) = -\sum_i \alpha_{ci} g_{ci}$ and λ is the distribution parameter. This simple example shows that how ERGMs and MNs are the same in terms of the underlying concept and the expressed probability distribution.

References

Afrasiabi MH, Guérin R, Venkatesh S (2013) Opinion formation in Ising networks. In: Information theory and applications workshop (ITA), 2013, pp 1–10. IEEE

Aggarwal CC (2011) An introduction to social network data analytics. Springer, Berlin

Agliari E, Burioni R, Contucci P (2010) A diffusive strategic dynamics for social systems. J Stat Phys 139(3):478–491

Al Hasan M, Zaki MJ (2011) A survey of link prediction in social networks. In: Social network data analytics. Springer, Berlin, pp 243–275

Anderson RM, May RM et al (1979) Population biology of infectious diseases: Part i. Nature 280(5721):361–367

Ayday E, Fekri F (2010) A belief propagation based recommender system for online services. In: Proceedings of the fourth ACM conference on recommender systems, pp 217–220. ACM

Bach SH, Broecheler M, Getoor L, O'Leary DP (2012) Scaling MPE inference for constrained continuous Markov random fields with consensus optimization. In: NIPS, pp 2663–2671

Berry MJ, Linoff G (1997) Data mining techniques: for marketing, sales, and customer support. Wiley, New York

Bonchi F, Castillo C, Gionis A, Jaimes A (2011) Social network analysis and mining for business applications. ACM Trans Intell Syst Technol (TIST) 2(3):22

Broekel T, Hartog M (2013) Explaining the structure of inter-organizational networks using exponential random graph models. Ind Innov 20(3):277–295

Bromberg F, Margaritis D, Honavar V et al (2009) Efficient Markov network structure discovery using independence tests. J Artif Intell Res 35(2):449

Cha M, Mislove A, Adams B, Gummadi KP (2008) Characterizing social cascades in Flickr. In: Proceedings of the 1st workshop on online social networks (WOSN'08), Seattle, WA

Cha M, Mislove A, Gummadi KP (2009) A measurement-driven analysis of information propagation in the Flickr social network. In: Proceedings of the 18th annual World wide web conference (WWW'09), Madrid, Spain

Chapelle O, Zhang Y (2009) A dynamic Bayesian network click model for web search ranking. In: Proceedings of the 18th international conference on World wide web, pp 1–10. ACM

Chen H, Ku WS, Wang H, Tang L, Sun MT (2013) Linkprobe: probabilistic inference on large-scale social networks. In: 2013 IEEE 29th international conference on data engineering (ICDE), pp 290–301. IEEE

Chickering DM, Heckerman D, Meek C (2001) Large-sample learning of Bayesian networks in NP-hard. J Mach Learn Res 5(2004):1287–1330

Coleman JS, Katz E, Menzel H et al (1966) Medical innovation: a diffusion study. Bobbs-Merrill Company Indianapolis, New York

Cowan R, Jonard N (2004) Network structure and the diffusion of knowledge. J Econ Dyn Control 28(8):1557–1575

Crane R, McDowell LK (2011) Evaluating Markov logic networks for collective classification. In: Proceedings of the 9th MLG workshop at the 17th ACM SIGKDD conference on knowledge discovery and data mining

Cranmer SJ, Desmarais BA (2011) Inferential network analysis with exponential random graph models. Polit Anal 19(1):66–86

Daud A, Li J, Zhou L, Muhammad F (2010) Knowledge discovery through directed probabilistic topic models: a survey. Front Comput Sci China 4(2):280–301

Dielmann A, Renals S (2004) Dynamic Bayesian networks for meeting structuring. In: IEEE international conference on acoustics, speech, and signal processing, 2004. Proceedings (ICASSP'04), vol 5, p V-629. IEEE

Dierkes T, Bichler M, Krishnan R (2011) Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with Markov logic networks. Decis Support Syst 51(3):361–371

Ding S (2011) Learning undirected graphical models with structure penalty. arXiv:1104.5256

Division NSR (1948) Rand database of worldwide terrorism incidents. http://www.rand.org/nsrd/projects/terrorism-incidents.html

Domingos P, Kok S, Lowd D, Poon H, Richardson M, Singla P (2008) Markov logic. In: Probabilistic inductive logic programming. Springer, Berlin, pp 92–117

Domingos P, Lowd D, Kok S, Nath A, Poon H, Richardson M, Singla P (2010) Markov logic: a language and algorithms for link mining. In: Link mining: models, algorithms, and applications. Springer, New York, pp 135–161

Fang L, LeFevre K (2010) Privacy wizards for social networking sites. In: Proceedings of the 19th international conference on World wide web, pp 351–360. ACM

Fellows I, Handcock MS (2012) Exponential-family random network models (preprint). arXiv:1208.0121

Fienberg SE (2012) A brief history of statistical models for network analysis and open challenges. J Comput Graph Stat 21(4):825–839

Frank O, Strauss D (1986) Markov graphs. J Am Stat Assoc 81(395):832–842

Freeman L (2004) The development of social network analysis. Empirical Press, Vancouver

Friedman N, Murphy K, Russell S (1998) Learning the structure of dynamic probabilistic networks. In: Proceedings of the fourteenth conference on uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., San Mateo, pp 139–147

Getoor L (2012) Social network datasets. http://www.cs.umd.edu/sen/lbc-proj/LBC.html

Gjoka M, Kurant M, Butts CT, Markopoulou A (2010) Walking in Facebook: a case study of unbiased sampling of OSNs. In: INFOCOM, 2010 Proceedings IEEE, pp 1–9

Goldenberg A, Moore A (2004) Tractable learning of large Bayes net structures from sparse data. In: Proceedings of the twenty-first international conference on machine learning, p 44. ACM

Goldenberg A, Zheng AX, Fienberg SE, Airoldi EM (2010) A survey of statistical network models. Found Trends Mach Learn 2(2):129–233

Goodreau SM (2007) Advances in exponential random graph ($p*$) models applied to a large social network. Soc Netw 29(2):231–248

Goodreau SM, Kitts JA, Morris M (2009) Birds of a feather, or friend of a friend? using exponential random graph models to

investigate adolescent social networks*. Demography 46(1):103–125
Grabowski A, Kosiński R (2006) Ising-based model of opinion formation in a complex network of interpersonal interactions. Physica A: Stat Mech Appl 361(2):651–664
Hageman RS, Leduc MS, Korstanje R, Paigen B, Churchill GA (2011) A Bayesian framework for inference of the genotype–phenotype map for segregating populations. Genetics 187:1163–1170
Handcock MS, Robins G, Snijders TA, Moody J, Besag, J (2003) Assessing degeneracy in statistical models of social networks. Technical report, Working paper
Handcock M, Hunter D, Butts C, Goodreau S, Morris M (2006) Statnet: an r package for the statistical analysis and simulation of social networks. Manual. University of Washington
He J, Chu WW, Liu ZV (2006) Inferring privacy information from social networks. In: Intelligence and security informatics. Springer, Berlin, pp 154–165
Heckerman D (2008) A tutorial on learning with Bayesian networks. Springer, Berlin
Humphreys L (2007) Mobile social networks and social practice: a case study of dodgeball. J Comput Mediat Commun 13(1):341–360
Hunter DR, Goodreau SM, Handcock MS (2008) Goodness of fit of social network models. J Am Stat Assoc 103(481):
Jabeur LB, Tamine L, Boughanem M (2012a) Featured tweet search: modeling time and social influence for microblog retrieval. In: 2012 IEEE/WIC/ACM international conferences on Web intelligence and intelligent agent technology (WI-IAT), vol 1, pp 166–173. IEEE
Jabeur LB, Tamine L, Boughanem M (2012b) Uprising microblogs: a Bayesian network retrieval model for tweet search. In: Proceedings of the 27th annual ACM symposium on applied computing, pp 943–948. ACM
Jansen BJ, Zhang M, Sobel K, Chowdury A (2009) Twitter power: tweets as electronic word of mouth. J Am Soc Inf Sci Technol 60(11):2169–2188
Java A, Song X, Finin T, Tseng B (2007) Why we twitter: understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp 56–65. ACM
Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining, pp 137–146. ACM
Koelle D, Pfautz J, Farry M, Cox Z, Catto G, Campolongo J (2006) Applications of Bayesian belief networks in social network analysis. In: Proceedings of the 4th Bayesian modeling applications workshop, UAI conference
Koller D, Friedman N (2009) Probabilistic graphical models: principles and techniques. Massachusetts Institute of Technology, Cambridge
Krause SM, Böttcher P, Bornholdt S (2012) Mean-field-like behavior of the generalized voter-model-class kinetic Ising model. Phys Rev E 85(3):031126
Krebs VE (2002) Mapping networks of terrorist cells. Connections 24(3):43–52
Kuter U, Golbeck J (2007) Sunny: a new algorithm for trust inference in social networks using probabilistic confidence models. AAAI 7:1377–1382
Kwak H, Lee C, Park H, Moon S (2010) What is twitter, a social network or a news media? In: Proceedings of the 19th international conference on World wide web, pp 591–600. ACM
Lauritzen SL (1996) Graphical models. Oxford University Press, Oxford
Lee SI, Ganapathi V, Koller D (2006) Efficient structure learning of Markov networks using *l*_1-regularization. In: Advances in neural information processing systems, pp 817–824
Lipford HR, Besmer A, Watson J (2008) Understanding privacy settings in Facebook with an audience view. UPSEC 8:1–8
Lusher D, Koskinen J, Robins G (2012) Exponential random graph models for social networks: theory, methods, and applications. Cambridge University Press, Cambridge
Madigan D, York J, Allard D (1995) Bayesian graphical models for discrete data. Int Stat Rev 63(2):215–232
Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH (2011) Big data: the next frontier for innovation, competition, and productivity
Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B (2007) Measurement and analysis of online social networks. In: Proceedings of the 5th ACM/USENIX Internet measurement conference (IMC'07), San Diego, CA
Morris M, Handcock MS, Hunter DR (2008) Specification of exponential-family random graph models: terms and computational aspects. J Stat Softw 24(4):1548
Mukherjee S, Speed T (2008) Network inference using informative priors. PNAS 11158:14313–14318
Murphy KP (2002) Dynamic Bayesian networks: representation, inference and learning. PhD thesis, University of California
Murphy KP (2012) Machine learning: a probabilistic perspective. The MIT Press, Cambridge
National Consortium for the Study of Terrorism and Responses to Terrorism (START) (2015) University of Maryland. http://www.start.umd.edu/
Neville J, Jensen D (2007) Relational dependency networks. J Mach Learn Res 8:653–692
Newman ME (2006) Modularity and community structure in networks. Proc Natl Acad Sci 103(23):8577–8582
Newman ME, Watts DJ, Strogatz SH (2002) Random graph models of social networks. Proc Natl Acad Sci 99(suppl 1):2566–2572
Ounis I, Macdonald C, Lin J, Soboroff I (2011) Overview of the TREC-2011 microblog track. In: Proceedings of the 20th Text REtrieval Conference (TREC 2011)
Park J, Newman ME (2004) Statistical mechanics of networks. Phys Rev E 70(6):066117
Pattison P, Wasserman S (1999) Logit models and logistic regressions for social networks: II. Multivariate relations. Br J Math Stat Psychol 52(2):169–193
Ravikumar P, Wainwright MJ, Lafferty JD et al (2010) High-dimensional Ising model selection using l1-regularized logistic regression. Ann Stat 38(3):1287–1319
Richardson M, Domingos P (2006) Markov logic networks. Mach Learn 62(1–2):107–136
Rinaldo A, Fienberg SE, Zhou Y et al (2009) On the geometry of discrete exponential families with application to exponential random graph models. Electr J Stat 3:446–484
Robins G, Pattison P, Wasserman S (1999) Logit models and logistic regressions for social networks: III. Valued relations. Psychometrika 64(3):371–394
Robins G, Pattison P, Elliott P (2001) Network models for social influence processes. Psychometrika 66(2):161–189
Robins G, Pattison P, Kalish Y, Lusher D (2007a) An introduction to exponential random graph ($p*$) models for social networks. Soc Netw 29(2):173–191
Robins G, Snijders T, Wang P, Handcock M, Pattison P (2007b) Recent developments in exponential random graph ($p*$) models for social networks. Soc Netw 29(2):192–215
Salter-Townshend M, Murphy TB (2014) Role analysis in networks using mixtures of exponential random graph models. J Comput Grap Stat (just-accepted)

Salter-Townshend M, White A, Gollini I, Murphy TB (2012) Review of statistical network analysis: models, algorithms, and software. Stat Anal Data Min ASA Data Sci J 5(4):243–264

Santos FC, Pacheco JM, Lenaerts T (2006) Evolutionary dynamics of social dilemmas in structured heterogeneous populations. Proc Natl Acad Sci USA 103(9):3490–3494

Schaefer DR, Simpkins SD (2014) Using social network analysis to clarify the role of obesity in selection of adolescent friends. Am J Public Health 104(7):1223–1229

Schmidt MW, Murphy K, Fung G, Rosales R (2010) Structure learning in random fields for heart motion abnormality detection. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1–8

Scott J, Carrington PJ (2011) The SAGE handbook of social network analysis. SAGE Publications, London

Snijders TA, Pattison PE, Robins GL, Handcock MS (2006) New specifications for exponential random graph models. Sociol Methodol 36(1):99–153

Song X, Jiang S, Yan X, Chen H (2014) Collaborative friendship networks in online healthcare communities: an exponential random graph model analysis. In: Smart health, vol 8549. Springer, Switzerland, pp 75–87

Sparrow MK (1991) The application of network analysis to criminal intelligence: an assessment of the prospects. Soc Netw 13(3):251–274

Srihari S (2014) Probabilistic graphical models. In: Alhajj R, Rokne J (eds) Encyclopedia of social network analysis and mining. Springer, Berlin

Stanford (2011) Stanford network analysis package (snap). http://snap.stanford.edu

Taskar B, Abbeel P, Koller D (2002) Discriminative probabilistic models for relational data. In: Proceedings of the eighteenth conference on uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., USA, pp 485–492

Thiemichen S, Friel N, Caimo A, Kauermann G (2014) Bayesian exponential random graph models with nodal random effects (preprint). arXiv:1407.6895

Tresp V, Nickel M (2013) Relational models. In: Rokne J, Alhajj R (eds) Encyclopedia of social network analysis and mining. Springer, Heidelberg

Uddin S, Hamra J, Hossain L (2013a) Exploring communication networks to understand organizational crisis using exponential random graph models. Comput Math Organ Theory 19(1):25–41

Uddin S, Hossain L, Hamra J, Alam A (2013b) A study of physician collaborations through social network and exponential random graph. BMC Health Serv Res 13(1):234

Van den Bulte C, Lilien GL (2001) Medical innovation revisited: social contagion versus marketing effort1. Am J Sociol 106(5):1409–1435

Vega-Redondo F (2007) Complex social networks, vol 44. Cambridge University Press, Cambridge

Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in Facebook. In: Proceedings of the 2nd ACM SIGCOMM workshop on social networks (WOSN'09), Barcelona, Spain

Wan HY, Lin YF, Wu ZH, Huang HK (2012) Discovering typed communities in mobile social networks. J Comput Sci Technol 27(3):480–491

Wang Y, Vassileva J (2003) Bayesian network-based trust model. In: IEEE/WIC international conference on Web intelligence, 2003. WI 2003. Proceedings, pp 372–378. IEEE

Wasserman S, Pattison P (1996) Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p. Psychometrika 61(3):401–425

Wortman, J.: Viral marketing and the diffusion of trends on social networks (2008)

Xiang R, Neville J (2013) Collective inference for network data with copula latent Markov networks. In: Proceedings of the sixth ACM international conference on Web search and data mining, pp 647–656. ACM

Yang X, Guo Y, Liu Y (2013) Bayesian-inference-based recommendation in online social networks. IEEE Trans Parallel Distrib Syst 24(4):642–651