Social Network Analysis and Mining, 5(1), 62:1-62:18,2015
Springer.

Noname manuscript No.
(will be inserted by the editor)

Probabilistic Graphical Models in Modern Social Network Analysis

Alireza Farasat · Alexander Nikolaev · Sargur N. Srihari · Rachael Hageman Blair

Received: date / Accepted: date

Rachael Hageman Blair
Department of Biostatistics
State University of New York and Buffalo
E-mail: hageman@buffalo.edu

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 describe recent applications to social networks. Modern SNA is flooded with challenges that arise from the inherent size, scope, and heterogeneity of both the data and underlying population. As a flexible modeling paradigm, PGMs can be adapted to address some SNA challenges. Such challenges are common themes in Big Data applications, but must be carefully considered for reliable inference and modeling. For this reason, we begin with a thorough description of data collection and sampling methods, which are often necessary in social networks, and underlie any downstream modeling efforts. PGMs in SNA have been used to tackle current and relevant challenges, including the estimation and quantification of importance, propagation of influence, trust (and distrust), link and profile prediction, privacy protection, and news spread through micro-blogging. We highlight these applications, and others, to showcase 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

1 Introduction

Over forty 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 [48].

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 [99]. SNA 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. However, this gain in information comes at a price; many of the statistical tools for analyzing such databases break due to the enormity of social networks and complex interdependencies within the data. False discovery rates are not easily controlled, which makes the identification of meaningful signals and relationships difficult [42]. Moreover, sampling networks is typically required, which can propagate selection bias through and downstream inference procedures.

SNA relies on diverse data representations and relational information, which may include (among others), tracked relationships among actors, events, and other covariate information [130]. 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. In this review, we focus on Probabilistic Graphical Models (PGMs), a flexible modeling paradigm, which has been shown to be an effective approach to modeling social networks [81,91]. Modern applications, including the estimation of influence, privacy protection, trust (and distrust) microblogging, and web-browsing, are presented to highlight the flexibility and utility of PGMs in addressing current and relevant problems in modern SNA.

PGMs provide a compact representation of a high-dimensional joint probability distribution of variables, by utilizing conditional independencies in the network of these variables; such a network, with local (in)dependency specifications, is called a model. PGM modeling is rooted in probabilistic reasoning, querying and also can also be used for generative purposes (sampling) [81]. In this review, we outline the basic theory, and model parameter and structural learning, but emphasize practical application and implementation of these

models to solve modern problems in SNA. We describe some of the unique statistical challenges that arise in using PGMs in SNA. The challenges are not isolated to PGMs. Rather, they propagate from the very foundation of the model - the data, through the local statistical models of the links and nodes, and finally to the graphical model. This review is organized from the *bottom-up*: from data sampling, to directed and undirected graphical models.

This paper is structured as follows. Section 2 provides an overview on data collection methods for SNA, reviews the challenges that arise in network sampling, and cites some network data repositories. In Section 3, directed probabilistic graphical models, static and dynamic, are discussed accompanied by application examples in SNA. Section 4 turns to undirected graphical model types and their applications. Section 5 concludes the paper and outlines future directions and challenges for PGM-based research in SNA.

2 Data collection and sampling

Data collection from social networks is a fundamental challenge that inherently affects downstream analysis through sampling bias [11, 19]. The reproducibility and generalization of any statistical analysis performed depends critically on the sample population, and how representative they are of the true population. In traditional observational and clinical studies, randomization and large sample size are important aspects of experimental design [28]. The object of a study may be driven by attributes such as the presence of a disease, or a covariate such as profession, age, preferences, etc. In contrast, SNA focuses primarily on the relations among actors, not the actors themselves and their individual attributes. For this reason, the population is not usually comprised of actors sampled independently; rather, the sampling scheme is driven by ties among the actors.

Snowball sampling begins with an actor, or a set of actors, and moves through the network by sampling ties [13]. Snowball methods are useful for identifying modules within a population, e.g., leaders, sub-cultures, and communities. The inability to include isolated actors that are directly tied in, but may be informative to the analysis, is a major limitation. Other disadvantages include the overestimation of connectivity, and the sensitivity of the sample to the initialization setting(s) of the snowball(s). Improvements on snowball sampling have been proposed to address some of these limitations [8, 44, 66, 133].

An alternative approach is to target actors in an *ego-centric* manner. There are two main sampling designs, with and without alter peer connections [63]. In this setting, a set of focal actors is selected, and their first-level ties are identified. In ego-centric networks with alter connections, those first-level ties are examined to determine connections between them. Ego-centric network without alter connections simply rely on focal actors and first-level ties; with

this approach, the extrapolation and generalization to the whole network is not possible.

Online Social Networks (OSNs) present unique challenges due to their massive size and the nature of the heterogenous attributes. A number of factors complicate the data collection process. Individuals can customize personal privacy setting, limiting crawlers from obtaining information and ultimately creating a missing data problem for the analyst. The diversity and dynamic nature of the data itself makes pages difficult to parse for collection purposes. Furthermore, the sampling is critical for tractable inference and analyses of large-scale OSNs. In most OSNs, we are faced with *hidden populations*, i.e., with unknown population size or the underlying distributions of the variables (edges or actors). In these cases, access to the network is facilitated through neighbors only. Crawling (through neighbors), either by random walks or graph transversals, is one of the most widely-used network exploration technique for OSNs.

- ***Random Walk:*** Metropolis-Hastings algorithm is a widely-used Markov Chain Monte Carlo (MCMC) method for sampling social networks [26]. The random walk starts at a random (or targeted) node and proceeds iteratively, moving between nodes i to j according to a transition probability. As $n \longrightarrow \infty$, the sampling distribution approaches the stationary distribution of actor characteristics, as if each sampled individual was uniformly drawn from the underlying population. In practice, the heuristic diagnostics are performed to assess convergence; the success of the methods can also depend on the starting point of the chain. Even with multiple chains, mixing can be slow and the chain can get stuck in regions of the graph. Note that these features are common to applications of MCMC methods, and not restricted to OSNs [52].

- ***Graph transversals:*** Several graph transversal methods have been applied to OSNs. These techniques differ only slightly in the order in which they systematically visit nodes in the network. Breadth-first search (BFS) and snowball sampling visit the graph through neighbor nodes [57]. Depth-first search (DFS) explores the graph from the seed node through the children nodes, and backtracks at dead-ends.

Factors such as sample size, as well as seed and algorithm choice can introduce bias into the statistical analysis of a network. Several authors have performed detailed investigations of the efficiency and bias associated with sampling algorithms using different OSNs [18,92]. Breadth First Search (BFS) is the most widely used method for OSN sampling and has been shown to be biased toward high-degree nodes [87,160]. Variants of the M-H algorithm have been proposed: Metropolized Random Walk with Backtracking (MRWB), M-H Random Walk (MHRW), Re-Weighted Random Walk (RWRW) and Unbiased Sampling to Reduce Self Loops (USRS), which aim to reduce or correct sample

bias [54, 125, 139, 152].

Publicly Available Data:
Several data resources have been created to house a wealth of diverse social network data. These resources are usually open source, requiring, at a minimum, a user agreement. Leveraging these resources is ideal for the development and testing of methodologies related to SNA. Max-Plank researchers have released OSN data used in publications, which includes crawled data from Flickr, YouTube, Wikipedia and Facebook [20, 21, 101, 149]. Several directed OSNs have been released in the Stanford network analysis package (snap), e.g. from Epinions, Amazon, LiveJournal, Slashdot and Wikipedia voting [138]. Recently, a Facebook dataset collected with MHRW was released, which exhibited convergence properties and was shown to be representative of the underlying population [54]. However, MHRW and UNI data sets contain only link information, thereby prohibiting attribute based analyses.

Document classification datasets have also been released [53]. A sample from the CiteSeer database contains 3, 312 publications from one of six classes, and 4, 732 links. The Cora dataset consists of 2, 708 publications classified into seven categories and the citation network has 5, 429 links. Each publication is described by a binary word vector which indicates the presence of certain words within a collection of 1, 433. WebKB consists of 877 scientific publications from five classes, contains 1, 601 links and includes binary word attributes similar to Cora. Terrorism databases are also publicly available [38, 141, 142]. 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) [38]. 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 [85, 136].

An alternative option to access data is to enroll in *data challenges*, which are often posed by corporations and operators of the networks themselves. For example, the Nokia mobile data challenge data was released in 2012 [90]. The data follows 200 users throughout the course of a year, and includes: usage (full call and message log), status (GPS readings, operation mode, environment (accelerometer samples, wi-fi access points, bluetooth devices), personal (full contact list, calendar), and user profile. Formal requests are required to use of this data, ensuring use for search and development, and prohibiting commercial use. Twitter has just posed TREC 2013, a collection of 240 million tweets (statutes) collected over a two month period [100]. This is the third year of TREC releases. The use of this data requires registration for a competition that centers around a competition. The 2013 competition centers around real-time ad-hoc search tasks.

3 Directed Probabilistic Graph Models

Bayesian Networks (BNs) are a special class of PGMs that capture directed dependencies between variables, which may represent cause-and-effect relationships. We describe two different branches of BNs, static and dynamic, which may be used to model social networks at a single time point or across a series of time point respectively. Both rely on the Markov assumptions, which enables the compact representation of the high-dimensional joint probability distribution of the variables in the model. Arguably, the use of directed graphs in SNA has been somewhat limited, although the applications themselves 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 utilize data from a single *snapshot* of a social community at a given time-point, described by Directed Acyclic Graph (DAGs). A DAG conveys precise information regarding the conditional independencies between modeled variables (nodes). The resulting graph, G, can be translated directly into a factored representation of the joint distribution [67, 91]. BNs obey the Markov condition which states that each variable, X_i, is independent of its non-descendants (unconnected nodes), given its parents in G. Under these assumptions, a BN for a set of variables $\{X_1, X_2, \ldots X_n\}$ is a network with the structure that encodes conditional independence relationships:

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

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.

Depending on the data and modeling objectives, BN learning may require up to two layers of inference: structural and parameter learning. Identifying the DAG that best explains the data is an NP-hard problem [27]. Structural inference can be conducted by sampling the posterior distribution to obtain an ensemble of feasible graphs, or through the implementation of a greedy hill-climbing algorithm, to identify a single graph structure that best approximates the Maximum a Posteriori (MAP) probabilities [68]. In many applications of SNA, the structure is often assumed, at least to some degree. In this case, the statistical inference problem is local parameter inference conditional on the assumed structure of the network.

The directionality and causal structure of the inferred model makes BN an attractive modeling paradigm for social networks that captures and conveys cause and effect relationships in a problem setting. Such examples, may manifest in decision making (influence). 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 [55]. Sparsity in BN is a popular assumption that can safeguard against over-fitting [68]. SPSN enforces the sparsity through a two stage process, which frames the structural learning problem as Market Basket Analysis task [12]. 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. In this case, actors are assumed to be linked to each other indirectly through items or events (Figure 1A). The learning problem is to identify an influence graph based on derived features of the binary transaction data. The method was shown to be effective for modeling a variety of SNs, including citation networks, collaboration data, and movie appearance records [12].

Koelle *et al.* proposed applications of BNs to SNA for the prediction of novel links and pre-specified node features (e.g., leadership potential) [80]. The authors emphasize the advantage of BN to account for 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*, is 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 (Figure 1B). 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 [65]. Generally, people prefer that their personal information is shared in small circles of friends and family, and shielded from strangers [24]. 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 [65]. 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 [96]. 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 [43].

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 [9,159]. Belief propagation is the process in which node marginal distributions (beliefs) are updated in light of *new evidence* [82]. 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 Filterting (CF)-based recommendation [158]. 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 [88,153]. 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 [89]. Users within microblogging networks of this type participate though the act of *following* and *being followed*, which gives rise naturally to directed associations [75]. With over 50 million tweets submittted daily, ranking and querying micrblogs has become an important and active area of open research [25, 97, 105, 110, 114]. 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 [72,73]. A query relevance function was developed based on a BN that leverages the PageRank algorithm to estimate parameters, such as influence, in the model (Figure 1C). The retrieval model was shown to outperform traditional methods for information retrieval on Twitter data from the TREC Tweets 2011 corpus [111].

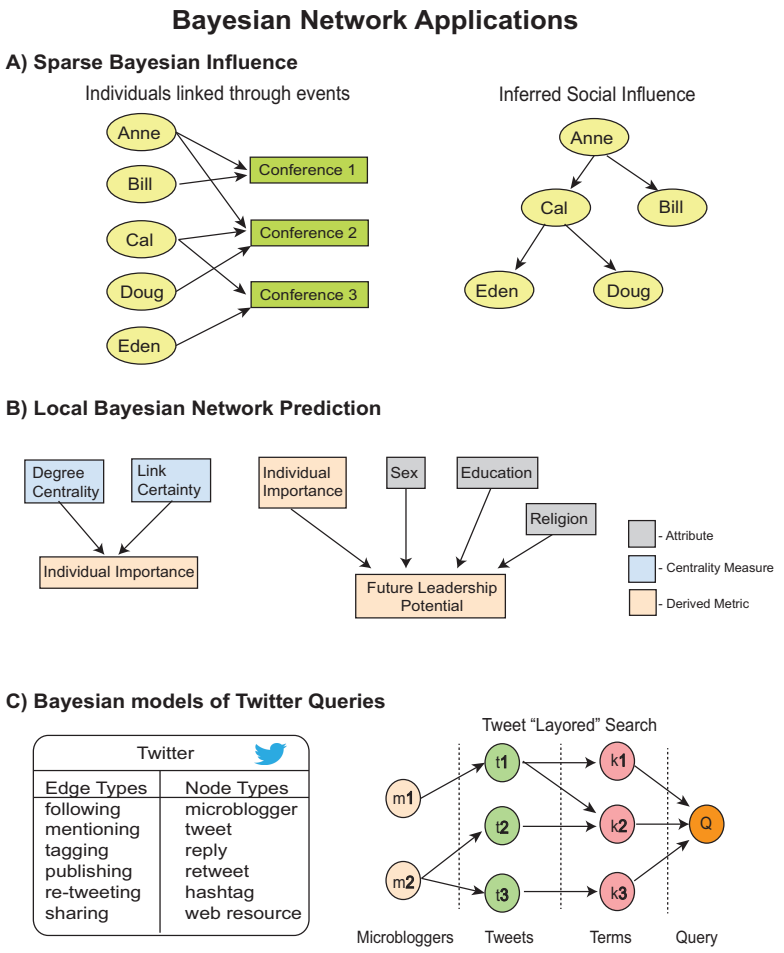

Fig. 1 Simplified schematics of select examples of Bayesian Networks in social networks. (**A**) Inferring sparse Bayesian influence based on transaction style data, which links actors to events. (**B**) Local models can be used to assess predict local metrics, such as individual importance or leadership potential, from attributes and centrality measures on the network itself. (**C**) Twitter is a microblogging community, which can be queried using a retrieval model described by a Bayesian Network.

Thus far, the BNs discussed summarize information at a single time-point. This represents an oversimplification of the true nature of the networks described, which are inherently dynamic [137]. In the described SN applications, the dynamic aspects are simplified by extracting data from a snapshot (or series of snapshots) of the SN across a time-period. The discretization, e.g., coarse or fine, can bias the results of the analysis. Discretization can give rise to many of the issues related to data collection discussed in Section 2. Modeling the dynamics of a network over the time-course can be achieved in the

BN framework with additional modeling assumptions.

Dynamic Bayesian Networks Dynamic Bayesian Networks (DBNs) provide compact representations for encoding structured probability distributions over arbitrarily long time courses [103]. 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. 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. The objective is to model a template variable over a discretized time course, $X^0 \dots X^T$, 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,

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

where X^0 represents the initial state, and the conditional probability terms of the form $P\left(X^{t+1} \mid X^t\right)$ 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 memorylessness property, i.e., that the current state depends on the previous and is conditionally independent of the past $\left(X^{t+1} \perp X^{0:t-1} \mid X^t\right)$. Comparatively, in static BNs, the Markov assumption only captures nodes' independence of their non-descendants, given the states of their parents.

Both DBNs and static BNs represent joint distributions of random variables. Similar to static BNs, DBNs also may require up to two layers of inference, structural and parameter learning. 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 [49]. 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. Several options exist for parameter learning, including junction trees, belief prorogation, and EM algorithm [33,78,132].

Despite the fact that social networks are typically 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 discussed in Section 5. Chapelle *et al.* used DBNs to model web users' browsing history [22]. The DBN extends the traditional and widely-used cascade model for browsing behavior to a more general model [77]. The dynamic studied here is that of click sequences, which is illustrated in Figure 3 for a single click (one time instance). The model 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.

DBNs and HMMs are very popular in the area of speech recognition [115, 162]. Meetings are social events, in which valuable information is exchanged mainly through speech. Effectively processing, capturing, and organizing this information can be costly, but is critical 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 [35]. Data including speaker order, location detected from microphone array, talk rate, pitch, and overall energy (enthusiasm). DBNs outperformed baseline HMMs in detecting *meeting actions* in a smart room, such as dialogue, 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.* modeled Twitter activity using KFMs in an effort to identify event and event location [124]. Each Twitter user is assumed to represent a sensor that monitors tweet features such as keywords, locations of tweets, their length and content. Support Vector Machines (SVM) are first used for event classification, followed by a Kalman filter to identify the location and the path itself. Location information of the quake is estimated through parameter learning at each time-point. Through tweet modeling, the authors were able to predict 96% of Japan's earthquakes of a certain magnitude. 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 [74]. This important and highly cited work can be generalized in this paradigm to model and predict other events.

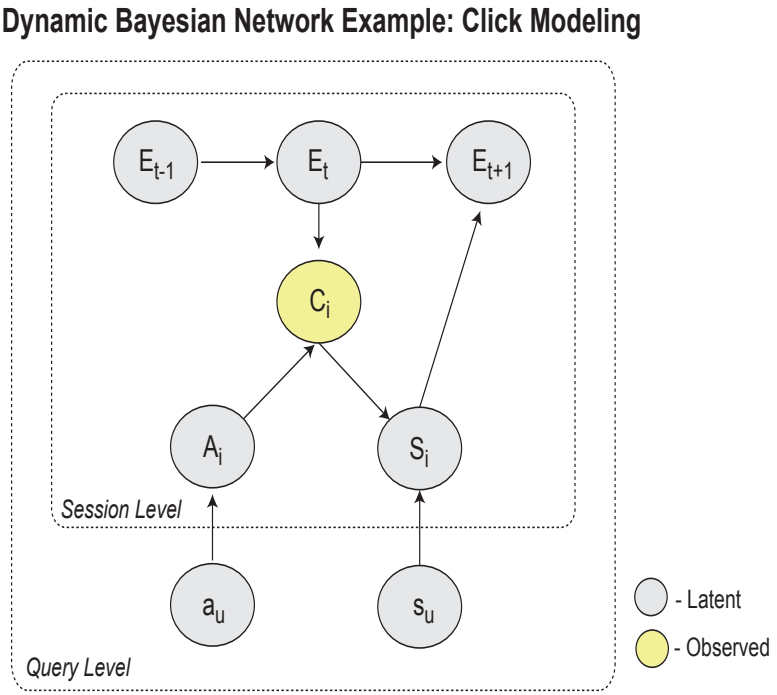

Fig. 2 An example of a time instance in a DBN used for click modeling in a browser. The temporal dimension is click sequence, which can be progressed through binary latent variables depicting satisfaction (S_i) and examination (E_i). Attraction (A_i) and satisfaction (S_i) are modeled at the session level, as well as the query level (a_u and s_u), which is assumed to be time invariant.

4 Undirected Probabilistic Graph Models

Markov Networks (MNs), also known as Markov Random Fields (MRFs), are PGMs with undirected edges. Similar to directed BNs, a MN graph is a representation of the joint distribution between variables (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 [47], and are widely used in SNA [109]. The basic formulation of these models and their utility in SNA will be highlighted.

Markov Networks can be decomposed into smaller complete sub-graphs known as *cliques*. A clique is a *maximal clique* if it cannot be extended to

included addition adjacent nodes. Clique representation enables a compact factorization of the probability density function (pdf). Specifically, the pdf captured by a graph G can be represented in the form:

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

where C is a maximal clique in the set of maximal cliques Ω, and $\psi_C(x_C)$ is the clique potential. The clique potentials are positive functions that capture the variable dependence within the cliques [82]. 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 [104], clique potentials are represented by a set of functions, termed *features*, with associated weights (i.e., $\phi_C(X_C) = \log(\psi_C(X_C))$, where $\phi_C(X_C)$ is a feature derived from the values of the variables in set X_C).

The *Hammersley-Clifford* theorem specifies the conditions under which a positive probability distribution can be represented as a MN. Specifically, the given representation (Equation 1) implies conditional independencies between the maximal cliques and is, by definition, a Gibbs measure [61].

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 Equation 1 has no analytical solution due to the complex expression of Z [93]. The problem of finding the optimal structure of G [76] using available data, similar to BNs, is even more challenging [16]. Currently existing approaches to structure learning are either constraint-based or score-based (see [37, 81, 106, 123, 129, 161] for more details).

MNs found use in SNA with the emergence of online social networks (OSNs) and digital social media (see [14] 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 [156]. A popular problem instance in this domain, that of user (missing) profile prediction, has been attacked using MNs [107, 117, 140].

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 [5, 23, 143, 145].

Detection of community substructures is another area of MN application [41, 108]. Social network clustering is especially challenging in a dynamic context, e.g. in Mobile Social Networks [70]. 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 [151]. 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. Relational Markov Networks can be used for labeling relationships in a social network with given content and link structure [150].

Several generative models have been proposed, which are motivated by MNs, and explain the effects of selection and influence (e.g., see [2]). Modeling channeled spread of opinions and rumors, known more generally as diffusion modeling, is an active area of research in SNA [10, 94, 119]. Several applications of diffusion models have been proposed for social networks including, but not limited to the spread of information [30], viral marketing [77], spread of diseases [7], the spread of cooperation [127]. 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 [155].

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 [117]. 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 [117] with the probability distribution [145],

$$P(X = x) = \frac{1}{Z} \exp\left(\sum_i w_i n_i(x)\right), \tag{2}$$

where $n_i(x)$ is the number of true groundings (e.g., logic expressions) of F_i, i.e., such formulae that hold, in x. Figure 3 gives an example of a ground MLN represented as a pairwise MN (left) for two individuals [104].

Many problems in statistical relational learning, such as link prediction [39], social network modeling, collective classification, link-based clustering and object identification, can be formulated using instances of MLN [117]. Dierkes *et al.* used MLNs to investigate the influence of Mobile Social Networks on

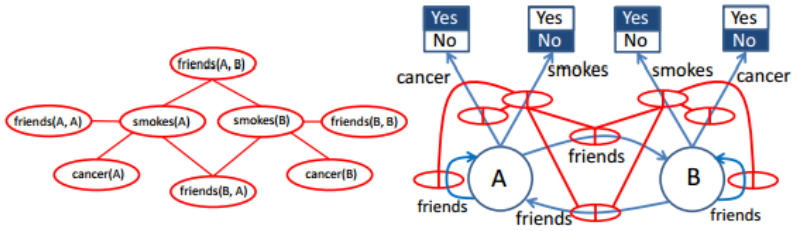

Fig. 3 An example of MLN with two entities (individuals) A and B, the unary relations "smokes" and "cancer" and the binary relation "friend". The ground predicates are denoted by eight elliptical nodes. Two formulas, F_1 ("someone who smokes has cancer") and F_2 ("friends either both smoke or both do not smoke") are captured. There exist two groundings of the F_1 (illustrated by the edges between the "smokes" and "cancer" nodes) and four groundings of F_2 captured by the rest of the edges [145].

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 [36].

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 [40,131]. 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) [31,34].

The Ising model and its variations form a subclass of MN with foundations in theoretical physics [6]. The Ising model is a discrete and pairwise MN, and is popular in applications in part due to its simplicity [82]. The variables in the model, $X_1 \dots X_p$, 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 $\chi \in \{0,1\}^p$, and $\Phi(\Theta)$ is the log of the partition function

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

Special, efficient methods exist for learning the Ising Model parameters from data [116]. 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 [1].

In SNA, the Ising model can be employed to analyze factors such as network sub-structures and nodal features affecting the opinion formation process. A classical example within this are is a study of *medical innovation* spread, namely the adoption of drug tetracycline by 125 physicians in four small cities in Illinois [17]. Figure 4 depicts the physicians' advisory network from a data set prepared by Ron Burt from the 1966 data collected by Coleman, Katz and Menzel [29] 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.

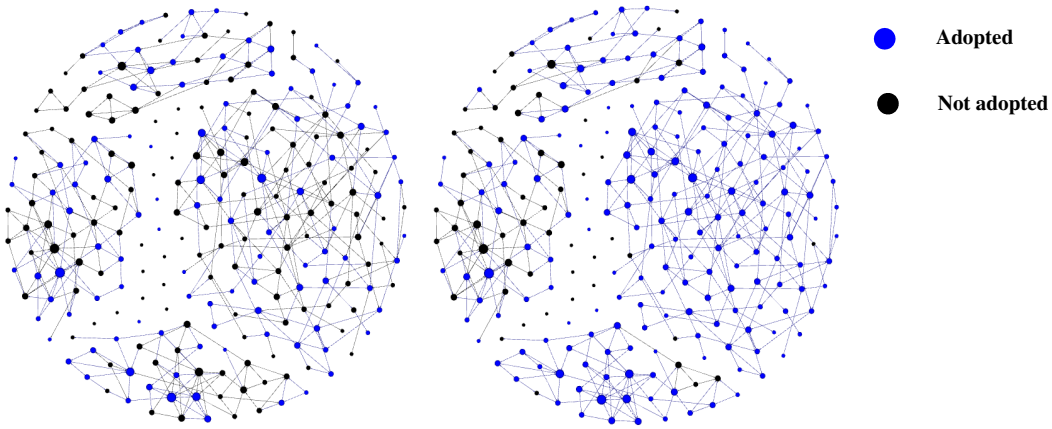

Fig. 4 The spread of new drug adoption through an advisory network physicians: two snapshots at different time points, about two years apart (from left to right). The growth dynamics in the number of adopters can be analyzed with an Ising Model.

Recently, the Ising Model has been used to examine social behaviors [148], including collective decision making, opinion formation and adoption of new technologies or products [50, 60, 84]. 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 [45]. Agliari *et al.* proposed a model to extract the underlying dynamics of social systems based on diffusive effects and people strategic choices to convince others [3]. 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.

Exponential Random Graph Models (ERGMs) [154], also known as the p^*-class models, are among the most widely-used network approaches to modeling social networks in recent years [47,113,120,121,134]. 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 an observed network, x, is:

$$P(Y = y, \Theta) = \frac{1}{Z} \exp\left(\sum_{i=1}^{K} \theta_i f_i(y)\right), \tag{3}$$

where $f_i(y), i = 1, \ldots, K$, are called sufficient statistics [98, 102], 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 statistics used in the model). Network configurations, 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 [98] for more details about network configurations). The parameters of ERGMs reflect a wide variety of possible configurations in social networks [119]. In addition, Z is the normalization constant.

Some of the first proposed models, e.g., random graphs and p_1 models [47], used Bernoulli and dyadic dependence structures, which are generally overly-simplistic [120]. On the contrary, ERGMs are based on Markov dependence assumption [47] 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 [120] (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 [45, 98, 144]).

ERGMs have been widely employed to study the network and friendship formation [135] and global network structural using local structure of the observed network [146]. The observed network is considered as one realization from too many possible networks with similar important characteristics [120]. For example, Broekel *et al.* used ERGMs to identify factors determining the structure of inter-organizational networks based on the single observation [15]. Schaefer *et al.* 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 [128].

Morover, Goodreau *et al.* used ERGMs to examine the generative processes that give rise to widespread patterns in friendship networks [59]. Cranmer and Desmarais used ERGMs to model co-sponsorship networks in the U.S. Congress and conflict networks in the international system. They figured out that several previously unexplored network parameters are acceptable predictors of the U.S. House of Representatives legislative co-sponsorship network [32].

The ERGMs have also been utilized in modelling the changing communication network structure and classifying networks based on the occurrence of their local features [146] and to identify micro-level structural properties of physician collaboration network on hospitalisation cost and readmission rate [147]. Finally, a ERGM-based model of clustering nodes considering their role in the network has been reported [126].

5 Discussion

Mining social networks for knowledge and discovery has proven to be a very challenging and active research area [79]. This review focussed on PGMs, and motivated their use in social networks through a variety of diverse applications. An important consideration is the issue of scalability, which is a major challenge, not only for PGMs, but for SNA, in general. Structural and parameter learning in high-dimensions can be prohibitive. In practice, several different network structures may be plausible, and equally likely. Moreover, both greedy- and sampling-based search strategies can get stuck at local minima. These numerical caveats can give rise to misleading networks, generating models, and subsequent predictions. ERGMs can exhibit *degeneracy* [64], 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 [58, 71].

Social networks continuously evolve over time. The methods we have discussed either utilize a *static snapshot* of the social network at a given time, or a fixed template structure which captures the dynamics. Template-based dynamics have proven their utility in a few social network applications. However, they are overly simplistic in their assumptions. More realistically, social networks can give rise to several interrelated streams that contain complex overlapping relational data [83]. Moreover, communities drift as new members join, old members leave or becoming inactive, and activities change. PGMs are not equipped to model temporal models of this type. Data stream mining research is an active area of research that aims to analyze web data as a stream and upon arrival [86]. There are considerable challenges related not only to the sheer volume and speed in which data is processed, but also to the changes in the features or targets being processed. Another major challenge, which has been extensively studied, is the concept of *drift* [51]. This phenomena

occurs when the probability of features and targets change in time, in other words, probability distributions change in the stream. Estimation in posterior probabilities in DBNs is spirit to drift estimation, but much more severely constrained due to the Markov assumption.

Alternative methods to modeling dynamics of the network have been proposed, including latent modeling approaches and the adoption of smooth transition assumptions. Sarkar *et al.* proposed a latent space model which assumes smooth transitions between time-steps, i.e. networks that change drastically from one time step to the next are assigned a lower probability. They also adopt a standard Markov assumption which states that $t+1$ is conditionally independent of all previous time-steps given t, which is the assumption adopted in our discussion of DBNs. Hoff *et al.* describe a latent space approach that relies on mapping actors into a *social space* by leveraging assumed transitive tendencies in relationships in order to estimate proximity in the latent space [69]. The iterative Facetnet algorithm frames the dynamic problem in terms of a non-negative matrix factorization, and uses the Kubler-Leibler divergence measure to enforce temporal smoothness [95]. TESLA extends the well-known graphical LASSO method for sparse regression, and penalizes changes between time steps using l_1-regularization. [4]. The TESLA algorithm was tested on both biological and social networks.

In this review, we survey directed and undirected PGMs, and highlight 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.

6 Acknowledgements

AF and AN are supported by a Multidisciplinary University Research Initiative (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. RHB is supported through NSF DMS 1312250.

References

1. Afrasiabi, M.H., Guérin, R., Venkatesh, S.: Opinion formation in ising networks. In: Information Theory and Applications Workshop (ITA), 2013, pp. 1–10. IEEE (2013)
2. Aggarwal, C.C.: An introduction to social network data analytics. Springer (2011)
3. Agliari, E., Burioni, R., Contucci, P.: A diffusive strategic dynamics for social systems. Journal of Statistical Physics **139**(3), 478–491 (2010)
4. Ahmed, A., Xing, E.P.: Recovering time-varying networks of dependencies in social and biological studies. PNAS **106**(29) (2009)
5. Al Hasan, M., Zaki, M.J.: A survey of link prediction in social networks. In: Social network data analytics, pp. 243–275. Springer (2011)
6. Anderson, C.J., Wasserman, S., Crouch, B.: A¡ i¿ p¡/i¿* primer: logit models for social networks. Social Networks **21**(1), 37–66 (1999)
7. Anderson, R.M., May, R.M., et al.: Population biology of infectious diseases: Part i. Nature **280**(5721), 361–367 (1979)
8. Atkinson, R., Flint, J.: Accessing hidden and hard-to-reach populations: Snowball research strategies. Social research update **33**(1), 1–4 (2001)
9. Ayday, E., Fekri, F.: A belief propagation based recommender system for online services. In: Proceedings of the fourth ACM conference on Recommender systems, pp. 217–220. ACM (2010)
10. Bach, S.H., Broecheler, M., Getoor, L., O'Leary, D.P.: Scaling mpe inference for constrained continuous markov random fields with consensus optimization. In: NIPS, pp. 2663–2671 (2012)
11. Berk, R.A.: An introduction to sample selection bias in sociological data. American Sociological Review pp. 386–398 (1983)
12. Berry, M.J., Linoff, G.: Data mining techniques: For marketing, sales, and customer support. John Wiley & Sons, Inc. (1997)
13. Biernacki, P., Waldorf, D.: Snowball sampling: Problems and techniques of chain referral sampling. Sociological methods & research **10**(2), 141–163 (1981)
14. Bonchi, F., Castillo, C., Gionis, A., Jaimes, A.: Social network analysis and mining for business applications. ACM Transactions on Intelligent Systems and Technology (TIST) **2**(3), 22 (2011)
15. Broekel, T., Hartog, M.: Explaining the structure of inter-organizational networks using exponential random graph models. Industry and Innovation **20**(3), 277–295 (2013)
16. Bromberg, F., Margaritis, D., Honavar, V., et al.: Efficient markov network structure discovery using independence tests. Journal of Artificial Intelligence Research **35**(2), 449 (2009)
17. Van den Bulte, C., Lilien, G.L.: Medical innovation revisited: Social contagion versus marketing effort1. American Journal of Sociology **106**(5), 1409–1435 (2001)
18. Callaway, D.S., Newman, M.E.J., Strogatz, S.H., Watts, D.J.: Network robustness and fragility: Percolation on random graphs. Physical Review Letters **85**, 5468 (2000)
19. Canali, C., Colajanni, M., Lancellotti, R.: Data acquisition in social netowrks: Issues and proposals. In: Proc. of the International Workshop on Services and Open Sources (SOS11) (2011)
20. Cha, M., Mislove, A., Adams, B., Gummadi, K.P.: Characterizing social cascades in flickr. In: Proceedings of the 1st Workshop on Online Social Networks (WOSN'08),. Seattle, WA (2008)
21. Cha, M., Mislove, A., Gummadi, K.P.: 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 (2009)
22. Chapelle, O., Zhang, Y.: 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 (2009)
23. Chen, H., Ku, W.S., Wang, H., Tang, L., Sun, M.T.: Linkprobe: Probabilistic inference on large-scale social networks. In: Data Engineering (ICDE), 2013 IEEE 29th International Conference on, pp. 290–301. IEEE (2013)
24. Chen, X., Michael, K.: Privacy issues and solutions in social network sites. Technology and Society Magazine, IEEE **31**(4), 43–53 (2012)

25. Cheong, M., Lee, V.: Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base. In: Proceedings of the 2nd ACM workshop on Social web search and mining, pp. 1–8. ACM (2009)
26. Chib, S., Greenberg, E.: Understanding the metropolis-hastings algorithm. The American Statistician **49**(4), 327–335 (1995)
27. Chickering, D., Heckerman, D., Meek, C.: Large-sample learning of Bayesian networks in NP-hard. Computing Science and Statistics **33** (2001)
28. Cochran, W.G.: Sampling Techniques, 3rd edn. Wiley (1977)
29. Coleman, J.S., Katz, E., Menzel, H., et al.: Medical innovation: A diffusion study. Bobbs-Merrill Company Indianapolis (1966)
30. Cowan, R., Jonard, N.: Network structure and the diffusion of knowledge. Journal of economic dynamics and control **28**(8), 1557–1575 (2004)
31. Crane, R., McDowell, L.K.: 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 (2011)
32. Cranmer, S.J., Desmarais, B.A.: Inferential network analysis with exponential random graph models. Political Analysis **19**(1), 66–86 (2011)
33. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society. Series B (Methodological) pp. 1–38 (1977)
34. Dhurandhar, A., Dobra, A.: Collective vs independent classification in statistical relational learning. Submitted for publication (2010)
35. Dielmann, A., Renals, S.: Dynamic bayesian networks for meeting structuring. In: Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on, vol. 5, pp. V–629. IEEE (2004)
36. Dierkes, T., Bichler, M., Krishnan, R.: Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with markov logic networks. Decision Support Systems **51**(3), 361–371 (2011)
37. Ding, S.: Learning undirected graphical models with structure penalty. J. CoRR (2011)
38. Division, N.S.R.: Rand database of worldwide terrorism incidents (1948). URL http://www.rand.org/nsrd/projects/terrorism-incidents.html
39. Domingos, P., Kok, S., Lowd, D., Poon, H., Richardson, M., Singla, P.: Markov logic. In: Probabilistic inductive logic programming, pp. 92–117. Springer (2008)
40. Domingos, P., Lowd, D., Kok, S., Nath, A., Poon, H., Richardson, M., Singla, P.: Markov logic: A language and algorithms for link mining. In: Link Mining: Models, Algorithms, and Applications, pp. 135–161. Springer (2010)
41. Du, N., Wu, B., Pei, X., Wang, B., Xu, L.: Community detection in large-scale social networks. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp. 16–25. ACM (2007)
42. Efron, B.: Size, power and false discovery rates. The Annals of Statistics **35**(4), 1351–1377 (2007)
43. Fang, L., LeFevre, K.: Privacy wizards for social networking sites. In: Proceedings of the 19th international conference on World wide web, pp. 351–360. ACM (2010)
44. Faugier, J., Sargeant, M.: Sampling hard to reach populations. Journal of advanced nursing **26**(4), 790–797 (1997)
45. Fellows, I., Handcock, M.S.: Exponential-family random network models. arXiv preprint arXiv:1208.0121 (2012)
46. Fienberg, S.E.: A brief history of statistical models for network analysis and open challenges. Journal of Computational and Graphical Statistics **21**(4), 825–839 (2012)
47. Frank, O., Strauss, D.: Markov graphs. Journal of the american Statistical association **81**(395), 832–842 (1986)
48. Freeman, L.: The development of social network analysis. Empirical Press (2004)
49. Friedman, N., Murphy, K., Russell, S.: Learning the structure of dynamic probabilistic networks. In: Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence, pp. 139–147. Morgan Kaufmann Publishers Inc. (1998)
50. Galam, S.: Rational group decision making: A random field ising model at¡ i¿ t¡/i¿= 0. Physica A: Statistical Mechanics and its Applications **238**(1), 66–80 (1997)
51. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. ACM Computing Surveys (CSUR) **46**(4), 44 (2014)

52. Gelman, A., Shirley, K.: Inference from simulations and monitoring convergence. In: S. Brooks, A. Gelman, G.I. Jones, X.L. Meng (eds.) Handbook of Markov Chain Monte Carlo, pp. 163–174. Chapman & Hall: CRC Handbooks of Modern Statistical Methods (2011)
53. Getoor, L.: Social network datasets (2012). URL http://www.cs.umd.edu/ sen/lbc-proj/LBC.html
54. Gjoka, M., Kurant, M., Butts, C.T., Markopoulou, A.: Walking in facebook: A case study of unbiased sampling of osns. In: INFOCOM, 2010 Proceedings IEEE, pp. 1 –9 (2010)
55. Goldenberg, A., Moore, A.: Tractable learning of large bayes net structures from sparse data. In: Proceedings of the twenty-first international conference on Machine learning, p. 44. ACM (2004)
56. Goldenberg, A., Zheng, A.X., Fienberg, S.E., Airoldi, E.M.: A survey of statistical network models. Foundations and Trends® in Machine Learning **2**(2), 129–233 (2010)
57. Goodman, L.A.: Snowball sampling. Annals of Mathematical Statistics **32**(1), 148–170 (1961)
58. Goodreau, S.M.: Advances in exponential random graph (¡ i¿ p¡/i¿¡ sup¿*¡/sup¿) models applied to a large social network. Social Networks **29**(2), 231–248 (2007)
59. Goodreau, S.M., Kitts, J.A., Morris, M.: Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks*. Demography **46**(1), 103–125 (2009)
60. Grabowski, A., Kosiński, R.: Ising-based model of opinion formation in a complex network of interpersonal interactions. Physica A: Statistical Mechanics and its Applications **361**(2), 651–664 (2006)
61. Hammersley, J.M., Clifford, P.: Markov fields on finite graphs and lattices (1971)
62. Handcock, M., Hunter, D., Butts, C., Goodreau, S., Morris, M.: Statnet: An r package for the statistical analysis and simulation of social networks. manual. university of washington (2006)
63. Handcock, M.S., Gile, K.J., et al.: Modeling social networks from sampled data. The Annals of Applied Statistics **4**(1), 5–25 (2010)
64. Handcock, M.S., Robins, G., Snijders, T.A., Moody, J., Besag, J.: Assessing degeneracy in statistical models of social networks. Tech. rep., Working paper (2003)
65. He, J., Chu, W.W., Liu, Z.V.: Inferring privacy information from social networks. In: Intelligence and Security Informatics, pp. 154–165. Springer (2006)
66. Heckathorn, D.D.: Respondent-driven sampling: a new approach to the study of hidden populations. Social problems pp. 174–199 (1997)
67. Heckerman, D.: Bayesian networks for data mining. Data Mining and Knowledge Discovery **1**, 79–119 (1997)
68. Heckerman, D.: A tutorial on learning with Bayesian networks. Springer (2008)
69. Hoff, P.D., Raftery, A.E., Handcock, M.S.: Latent space approaches to social network analysis. Journal of the american Statistical association **97**(460), 1090–1098 (2002)
70. Humphreys, L.: Mobile social networks and social practice: A case study of dodgeball. Journal of Computer-Mediated Communication **13**(1), 341–360 (2007)
71. Hunter, D.R., Goodreau, S.M., Handcock, M.S.: Goodness of fit of social network models. Journal of the American Statistical Association **103**(481) (2008)
72. Jabeur, L.B., Tamine, L., Boughanem, M.: Featured tweet search: Modeling time and social influence for microblog retrieval. In: Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on, vol. 1, pp. 166–173. IEEE (2012)
73. Jabeur, L.B., Tamine, L., Boughanem, M.: 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 (2012)
74. Jansen, B.J., Zhang, M., Sobel, K., Chowdury, A.: Twitter power: Tweets as electronic word of mouth. Journal of the American society for information science and technology **60**(11), 2169–2188 (2009)
75. Java, A., Song, X., Finin, T., Tseng, B.: 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 (2007)

76. Karger, D., Srebro, N.: Learning Markov networks: maximum bounded tree-width graphs. In: Proc. SIAM-ACM Symposium on Discrete Algorithms, pp. pp. 392–401 (2001)
77. Kempe, D., Kleinberg, J., Tardos, É.: 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 (2003)
78. Kjaerulff, U.: A computational scheme for reasoning in dynamic probabilistic networks. In: Proceedings of the Eighth international conference on Uncertainty in artificial intelligence, pp. 121–129. Morgan Kaufmann Publishers Inc. (1992)
79. Kleinberg, J.M.: Challenges in mining social network data: processes, privacy, and paradoxes. In: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 4–5. ACM (2007)
80. Koelle, D., Pfautz, J., Farry, M., Cox, Z., Catto, G., Campolongo, J.: Applications of Bayesian belief networks in social network analysis. In: Proc. of the 4th Bayesian Modeling Applications Workshop, UAI Conference, p. pp. (2006)
81. Koller, D., Friedman, N.: Probabilistic graphical models: principles and techniques. Massachusetts Institute of Technology (2009)
82. Koller, D., Friedman, N.: Probabilistic graphical models: principles and techniques. MIT press (2009)
83. Koren, Y.: Collaborative filtering with temporal dynamics. Communications of the ACM **53**(4), 447–455 (2009)
84. Krause, S.M., Böttcher, P., Bornholdt, S.: Mean-field-like behavior of the generalized voter-model-class kinetic ising model. Physical Review E **85**(3), 031,126 (2012)
85. Krebs, V.E.: Mapping networks of terrorist cells. Connections **24**(3), 43–52 (2002)
86. Krempl, G., Žliobaite, I., Brzeziński, D., Hüllermeier, E., Last, M., Lemaire, V., Noack, T., Shaker, A., Sievi, S., Spiliopoulou, M., et al.: Open challenges for data stream mining research. ACM SIGKDD Explorations Newsletter **16**(1), 1–10 (2014)
87. Kurant, M., Markopoulou, A., Thiran, P.: On the bias of bfs (breadth first search) pp. 1 –8 (2010)
88. Kuter, U., Golbeck, J.: Sunny: A new algorithm for trust inference in social networks using probabilistic confidence models. In: AAAI, vol. 7, pp. 1377–1382 (2007)
89. Kwak, H., Lee, C., Park, H., Moon, S.: 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 (2010)
90. Laurila, J.K., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Do, T.M.T., Dousse, O., Eberle, J., Miettinen, M.: The mobile data challenge: Big data for mobile computing research. In: Proceedings of the Workshop on the Nokia Mobile Data Challenge, in Conjunction with the 10th International Conference on Pervasive Computing, pp. 1–8 (2012)
91. Lauritzen, S.L.: Graphical models. Oxford University Press (1996)
92. Lee, S.H., Kim, P.J., Hawoong, J.: Statistical properties of sampled networks. Phys. Rev. E **73**, 016,102 (2006). DOI 10.1103/PhysRevE.73.016102
93. Lee, S.I., Ganapathi, V., Koller, D.: Efficient structure learning of markov networks using l_1-regularization. In: Advances in neural Information processing systems, pp. 817–824 (2006)
94. Leenders, R.: Longitudinal behavior of network structure and actor attributes: modeling interdependence of contagion and selection. Evolution of social networks **1** (1997)
95. Lin, Y.R., Chi, Y., Zhu, S., Sundaram, H., Tseng, B.L.: Facetnet: a framework for analyzing communities and their evolutions in dynamic networks. In: Proceedings of the 17th international conference on World Wide Web, WWW '08, pp. 685–694. ACM (2008)
96. Lipford, H.R., Besmer, A., Watson, J.: Understanding privacy settings in facebook with an audience view. UPSEC **8**, 1–8 (2008)
97. Luo, Z., Tang, J., Wang, T.: Propagated opinion retrieval in twitter. In: Web Information Systems Engineering–WISE 2013, pp. 16–28. Springer (2013)
98. Lusher, D., Koskinen, J., Robins, G.: Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications. Cambridge University Press (2012)

99. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big data: The next frontier for innovation, competition, and productivity (2011)
100. McCreadie, R., Soboroff, I., Lin, J., Macdonald, C., Ounis, I., McCullough, D.: On building a reusable twitter corpus. In: Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, pp. 1113–1114. ACM (2012)
101. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. In: In Proceedings of the 5th ACM/USENIX Internet Measurement Conference (IMC'07). San Diego, CA (2007)
102. Morris, M., Handcock, M.S., Hunter, D.R.: Specification of exponential-family random graph models: terms and computational aspects. Journal of statistical software **24**(4), 1548 (2008)
103. Murphy, K.P.: Dynamic bayesian networks: representation, inference and learning. Ph.D. thesis, University of California (2002)
104. Murphy, K.P.: Machine learning: a probabilistic perspective. The MIT Press (2012)
105. Nagmoti, R., Teredesai, A., De Cock, M.: Ranking approaches for microblog search. In: Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on, vol. 1, pp. 153–157. IEEE (2010)
106. Netrapalli, P., Banerjee, S., Sanghavi, S., Shakkottai, S.: Greedy learning of markov network structure. In: Proc. of Allerton Conf. on Communication, Control and Computing, Monticello, USA (2010)
107. Neville, J., Jensen, D.: Relational dependency networks. The Journal of Machine Learning Research **8**, 653–692 (2007)
108. Newman, M.E.: Modularity and community structure in networks. Proceedings of the National Academy of Sciences **103**(23), 8577–8582 (2006)
109. Newman, M.E., Watts, D.J., Strogatz, S.H.: Random graph models of social networks. Proceedings of the National Academy of Sciences **99**(suppl 1), 2566–2572 (2002)
110. O'Connor, B., Krieger, M., Ahn, D.: Tweetmotif: Exploratory search and topic summarization for twitter. In: ICWSM (2010)
111. Ounis, I., Macdonald, C., Lin, J., Soboroff, I.: Overview of the trec-2011 microblog track. In: Proceeddings of the 20th Text REtrieval Conference (TREC 2011) (2011)
112. Park, J., Newman, M.E.: Statistical mechanics of networks. Physical Review E **70**(6), 066,117 (2004)
113. Pattison, P., Wasserman, S.: Logit models and logistic regressions for social networks: Ii. multivariate relations. British Journal of Mathematical and Statistical Psychology **52**(2), 169–193 (1999)
114. Pochampally, R., Varma, V.: User context as a source of topic retrieval in twitter. In: Workshop on Enriching Information Retrieval (with ACM SIGIR) (2011)
115. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE **77**(2), 257–286 (1989)
116. Ravikumar, P., Wainwright, M.J., Lafferty, J.D., et al.: High-dimensional ising model selection using l1-regularized logistic regression. The Annals of Statistics **38**(3), 1287–1319 (2010)
117. Richardson, M., Domingos, P.: Markov logic networks. Machine learning **62**(1-2), 107–136 (2006)
118. Rinaldo, A., Fienberg, S.E., Zhou, Y., et al.: On the geometry of discrete exponential families with application to exponential random graph models. Electronic Journal of Statistics **3**, 446–484 (2009)
119. Robins, G., Pattison, P., Elliott, P.: Network models for social influence processes. Psychometrika **66**(2), 161–189 (2001)
120. Robins, G., Pattison, P., Kalish, Y., Lusher, D.: An introduction to exponential random graph (¡ i¿ p¡/i¿¡ sup¿*¡/sup¿) models for social networks. Social networks **29**(2), 173–191 (2007)
121. Robins, G., Pattison, P., Wasserman, S.: Logit models and logistic regressions for social networks: Iii. valued relations. Psychometrika **64**(3), 371–394 (1999)
122. Robins, G., Snijders, T., Wang, P., Handcock, M., Pattison, P.: Recent developments in exponential random graph (¡ i¿ p¡/i¿*) models for social networks. Social networks **29**(2), 192–215 (2007)

123. Roy, S., Lane, T., Werner-Washburne, M.: Learning structurally consistent undirected probabilistic graphical models. In: Proc ICML (2009)
124. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web, pp. 851–860. ACM (2010)
125. Salganik, M.J., Heckathorn, D.D.: Sampling and estimation in hidden populations using respondent-driven sampling. Sociological Methodology **34**(1), 193–240 (2004)
126. Salter-Townshend, M., Murphy, T.B.: Role analysis in networks using mixtures of exponential random graph models. Journal of Computational and Graphical Statistics (just-accepted), 00–00 (2014)
127. Santos, F.C., Pacheco, J.M., Lenaerts, T.: Evolutionary dynamics of social dilemmas in structured heterogeneous populations. Proceedings of the National Academy of Sciences of the United States of America **103**(9), 3490–3494 (2006)
128. Schaefer, D.R., Simpkins, S.D.: Using social network analysis to clarify the role of obesity in selection of adolescent friends. American journal of public health (0), e1–e7 (2014)
129. Schmidt, M., Murphy, K., Fung, G., Rosales, R.: Structure learning in random fields for heart motion abnormality detection. In: CVPR (2010)
130. Scott, J., Carrington, P.J.: The SAGE handbook of social network analysis. SAGE publications (2011)
131. Singla, P., Domingos, P.: Lifted first-order belief propagation. In: AAAI, vol. 8, pp. 1094–1099 (2008)
132. Smyth, P., Heckerman, D., Jordan, M.I.: Probabilistic independence networks for hidden markov probability models. Neural computation **9**(2), 227–269 (1997)
133. Snijders, T.A.: Estimation on the basis of snowball samples: how to weight? Bulletin de méthodologie sociologique **36**(1), 59–70 (1992)
134. Snijders, T.A., Pattison, P.E., Robins, G.L., Handcock, M.S.: New specifications for exponential random graph models. Sociological methodology **36**(1), 99–153 (2006)
135. Song, X., Jiang, S., Yan, X., Chen, H.: Collaborative friendship networks in online healthcare communities: An exponential random graph model analysis. In: Smart Health, pp. 75–87. Springer (2014)
136. Sparrow, M.K.: The application of network analysis to criminal intelligence: An assessment of the prospects. Social networks **13**(3), 251–274 (1991)
137. Spiliopoulou, M.: Evolution in social networks: A survey. In: Social network data analytics, pp. 149–175. Springer (2011)
138. Stanford: Stanford network analysis package (snap) (2011). URL http://snap. stanford. edu
139. Stutzbach, D., Rejaie, R., Duffield, N., Sen, S., Willinger, W.: On unbiased sampling for unstructured peer-to-peer networks. Networking, IEEE/ACM Transactions on **17**(2), 377 –390 (2009)
140. Taskar, B., Abbeel, P., Koller, D.: Discriminative probabilistic models for relational data. In: Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence, pp. 485–492. Morgan Kaufmann Publishers Inc. (2002)
141. for the Study of Terrorism, N.C., to Terrorism (START), R.: International center for political violence and terrorism research (icpvtr) (2012). URL http://www.pvtr.org/ICPVTR/
142. for the Study of Terrorism, N.C., to Terrorism (START), R.: Gtd global terrorism database (2013). URL http://www.start.umd.edu/gtd/
143. Thi, D.B., Hoang, T.A.N.: Features extraction for link prediction in social networks. In: Computational Science and Its Applications (ICCSA), 2013 13th International Conference on, pp. 192–195. IEEE (2013)
144. Thiemichen, S., Friel, N., Caimo, A., Kauermann, G.: Bayesian exponential random graph models with nodal random effects. arXiv preprint arXiv:1407.6895 (2014)
145. Tresp, V., Nickel, M.: Relational models. Encyclopedia of Social Network Analysis and Mining. Ed. by J. Rokne and R. Alhajj. Heidelberg: Springer (2013)
146. Uddin, S., Hamra, J., Hossain, L.: Exploring communication networks to understand organizational crisis using exponential random graph models. Computational and Mathematical Organization Theory **19**(1), 25–41 (2013)

147. Uddin, S., Hossain, L., Hamra, J., Alam, A.: A study of physician collaborations through social network and exponential random graph. BMC health services research **13**(1), 234 (2013)
148. Vega-Redondo, F.: Complex social networks, vol. 44. Cambridge University Press (2007)
149. Viswanath, B., Mislove, A., Cha, M., Gummadi, K.P.: On the evolution of user interaction in facebook. In: Proceedings of the 2nd ACM SIGCOMM Workshop on Social Networks (WOSN'09). Barcelona, Spain (2009)
150. Wan, H., Lin, Y., Wu, Z., Huang, H.: A community-based pseudolikelihood approach for relationship labeling in social networks. In: Machine Learning and Knowledge Discovery in Databases, pp. 491–505. Springer (2011)
151. Wan, H.Y., Lin, Y.F., Wu, Z.H., Huang, H.K.: Discovering typed communities in mobile social networks. Journal of Computer Science and Technology **27**(3), 480–491 (2012)
152. Wang, D., Li, Z., Xie, G.: Towards unbiased sampling of online social networks. In: Communications (ICC), 2011 IEEE International Conference on, pp. 1 –5 (2011)
153. Wang, Y., Vassileva, J.: Bayesian network-based trust model. In: Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on, pp. 372–378. IEEE (2003)
154. Wasserman, S., Pattison, P.: Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp. Psychometrika **61**(3), 401–425 (1996)
155. Wortman, J.: Viral marketing and the diffusion of trends on social networks (2008)
156. Xiang, R., Neville, J.: 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 (2013)
157. Yang, T., Chi, Y., Zhu, S., Gong, Y., Jin, R.: Detecting communities and their evolutions in dynamic social networksa bayesian approach. Machine Learning **82**, 57189 (2011)
158. Yang, X., Guo, Y., Liu, Y.: Bayesian-inference based recommendation in online social networks. In: INFOCOM, 2011 Proceedings IEEE, pp. 551–555. IEEE (2011)
159. Yang, X., Guo, Y., Liu, Y.: Bayesian-inference-based recommendation in online social networks. Parallel and Distributed Systems, IEEE Transactions on **24**(4), 642–651 (2013)
160. Ye, S., Lang, J., Wu, S.F.: Crawling online social graphs. In: Proceedings of the 12th International Asia-Pacific Web Conference, pp. 236–242 (20010)
161. Zhu, J., Lao, N., Xing, E.P.: Grafting-light: fast, incremental feature selection and structure learning of markov random fields. In: Proc.16th ACM SIGKDD (2010)
162. Zweig, G., Russell, S.: Speech recognition with dynamic bayesian networks (1998)

7 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 [122] due to the lack of enough observed data. Robins *et al.* outline this and some other problems associated with ERGMs, e.g., degeneracy in model selection and bimodal distribution shapes [122] (see also [62, 64, 118, 134]).

The roots of ERGMs in the Principle of Maximum Entropy [112] and the Hammersley-Clifford theorem have been previously pointed out [56,119]. 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 [46]), 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 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 [120]. 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:

$$P(X) = \frac{1}{Z(\alpha)} \prod_{c=1}^{C} \exp\left(\beta_c \sum_{i=1}^{N} \alpha_{ci} g_{ci}\right) = \frac{1}{Z(\alpha)} \exp\left(\sum_{c=1}^{C} \beta_c \sum_{i=1}^{N} \alpha_{ci} g_{ci}\right). \quad (4)$$

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:

$$P(X) = \frac{1}{Z(\theta')} \exp\left(\sum_{i=1}^{N} \theta'_i \sum_{c=1}^{C} \beta_c g_{ci}\right). \quad (5)$$

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.

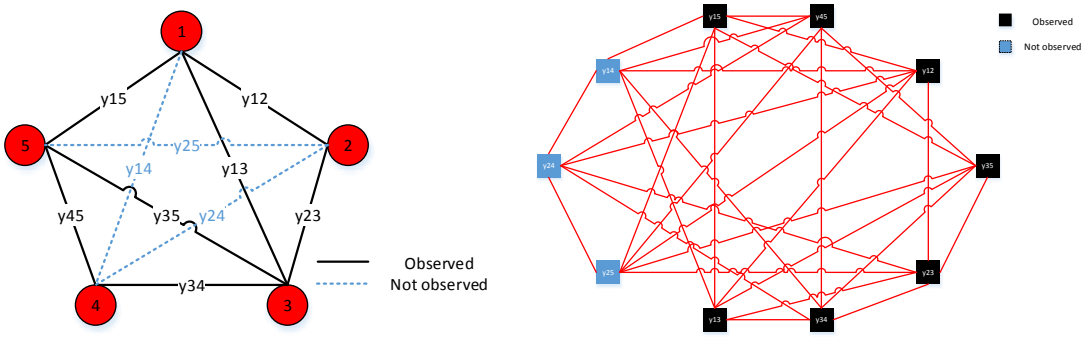

Fig. 5 A social network with five actors(left) and its corresponding Markov network (right).

Finally, if we replace f_i' in (5):

$$P(X) = \frac{1}{Z'} \exp\left(\sum_{i=1}^{N} \theta_i' f_i'\right). \tag{6}$$

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 depicts similarities between ERGMs and MNs. A social network with five actors, $N = 5$, is assumed (Figure 5 (left)). Considering Markov dependency assumption, there exists an unique corresponding Markov network shown in Figure 5 (right) 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.