International Journal of Machine Learning and Cybernetics
https://doi.org/10.1007/s13042-018-0823-8

ORIGINAL ARTICLE

On relational learning and discovery in social networks: a survey

Ji Zhang[1] · Leonard Tan[1] · Xiaohui Tao[1]

Received: 14 August 2017 / Accepted: 27 April 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract
The social networking scene has evolved tremendously over the years. It has grown in relational complexities that extend a vast presence onto popular social media platforms on the internet. With the advance of sentimental computing and social complexity, relationships which were once thought to be simple have now become multi-dimensional and widespread in the online scene. This explosion in the online social scene has attracted much research attention. The main aims of this work revolve around the knowledge discovery and datamining processes of these feature-rich relations. In this paper, we provide a survey of relational learning and discovery through popular social analysis of different structure types which are integral to applications within the emerging field of sentimental and affective computing. It is hoped that this contribution will add to the clarity of how social networks are analyzed with the latest groundbreaking methods and provide certain directions for future improvements.

Keywords Online social networks · Social internetworking scenarios · Homogeneous networks · Heterogeneous networks · Hybrid networks · Multi-dimensional relational learning

1 Introduction

Networks today span wide areas of interest. These include, and are not limited to online social relationships, biological networks, marketing, politics, etc [1]. Information networks are formed from nodes with interconnecting links [1]. Complex schemas provide a realistic representation [e.g. Bespoke(star), multi-relation, bipartite, edge-node (multi-hub), etc] of how they have evolved over a temporal space [2–4]. Unique relationships between nodes are represented by high-dimensional, complex structures [1, 5]. Such complexities pre-define the co-existence of sub-structures commonly known as communities within these networks [1]. On a functional scale, these networks are capable of social and biological inferences by uncovering latent relational intelligence established between actors of a community. These inferences contain a rich source of detected emotions and feelings such as collaboration, prediction, reciprocities, status, etc. The multi-disciplinary applications of social networks have gained substantial recognition and scholarly interest over the past years in a vast variety of areas especially in the emerging field of sentimental and affective computing. Some of the more popular ones include link prediction, community detection, recommender systems, outlier and fraud detection, evolutionary processes, mood identification, depression detection, emotional disorder identification, etc [1, 6].

Social Networks are made up of metadata entities connected to each other through relational dyads [7]. As a way to represent complexities between different types of networks, various approaches have been developed to identify the depth of this dimensionality [8–10]. Relationally, the answer to the research question of how nodes are connected to each other, and to what degree of depth has a deep impact and significance towards how social schemas are structured. In the chapters that follow this survey, each unique structure will be given a proper introduction and investigation of its own.

Nucleus structures represent the most fundamental, basic building blocks which constitute the agglomeration of predicate structures. These atomic structures determine

✉ Ji Zhang
Ji.Zhang@usq.edu.au

Leonard Tan
Leonard.Tan@usq.edu.au

Xiaohui Tao
Xiaohui.Tao@usq.edu.au

[1] Faculty of Health Engineering and Sciences, University of Southern Queensland, Darling Heights, QLD, Australia

Published online: 17 May 2018

how networks evolve over time [11]. They are usually constrained to a maximum of no more than 10 nodes which are tightly interconnected with each other [1]. As nucleus relational structures agglomerate, they form Cliques which are slightly larger clusters of closely interconnected relations. These small clustered subsets of social relations between selected actors establish a semi-independence related to the larger network from which they are formed. Cliques in turn subsequently evolve into communities [6], which are then characterized by larger, more randomized relational hypergraphs which are grounded by a firm set of belief states [12].

1.1 Homogeneous networks

Homogeneous Networks refer to distinct network types of super-structures containing one similar type of node in multiple instances. These instances are interconnected to each other through a single, identical relational structure type [1]. The kinds of networks formed in theory are limited in both dimensionality and complexity, and oftentimes referred to as "flat" networks [13]. This "limitation" in literature and traditional study is not due to the natural (un)realistic aspect of how these networks are formed/represented [14]. Instead, it is due more to the early graph models and to how these networks were, intuitively "thought" to represent the (study) information of interest.

1.2 Heterogeneous networks

Heterogeneous networks (HNs) on the other hand, are intended to generally reflect on high levels of abstraction representative of actual structural data [1]. They can perhaps be considered by some research to represent a working analysis of the closest next generation framework to a real-life social network [13]. As an emerging research area however, this architecture still contains limitations which prevent it from being able to effectively model a given realistic social structure accurately [15]. Heterogeneous networks remain largely a conceptual rather than a functional model of implementation [16]. Recently, there have been some developments in common work-arounds for these attempts. These include:

1. The wide use of Markov chain logic to quantify state spaces by which HNs transition probabilistically across time frames [17].
2. Area under curve (AUC) analysis to characterize the efficiency of predictive case models [18].
3. Energy models like Markov logic networks (MLNs) to predict the evolutionary state behavior of HNs [6, 19].
4. Multivariate auto-regressive (AR) models that attempt to explain the regressive state dimensionalities of highly correlated HN data [20].
5. Clustering Ensemble methods that selectively extract information of interest from a highly dimensional data sub-space.
6. The inference to the stability of features and associated metrics for further computational analysis [13, 21, 22].
7. Shapley value based variants to predict confidence scores of given structural patterns identified from baseline models [23].
8. Markov Thermal percolative models (MTPM) that emulate heat transfer mechanisms of influence, trust, sentiment, etc., in the diffusivity of information through a HN [24].
9. EM Wave Theory analyses such as divisive and spectral methods to detect communities in HNs through identification and detection of feature fundamental frequencies of interest which are obtained from Fourier transformations [25].

1.3 Hybrid networks

Hybrid heterogeneous networks (HHNs) are smaller subproblem sets of the HNs superset. They are directly modeled from HNs to satisfy the need for a more accurate definition of the social network in question. This is done from a more focused perspective that directly addresses the problems of lossy generalization and computational overload, derived from their parent superstructures [1, 26]. For example, Facebook is a fully functional Heterogeneous Network which includes multiple entity types like user profiles, events, news, ads, groups, etc. Several HHN models exist in literatures that have been widely used for study in the areas of datamining and knowledge discovery [1]. For completeness sake, the study of HHNs is also worth mentioning in the later chapters, as one of our more robust and novel initiatives of this survey paper. Additionally, this helps the general research community gain deeper insights into the structural characteristics of the functioning aspects of the entire HN.

1.4 Social internetworking scenarios

Social internetworking scenarios (SIS) emerged from the farsighted vision of a unified social interactivity across various online platforms. Current popular social networks include but are not limited to: Facebook, Yahoo, OrCHiD, LinkedIn, Twitter, Instagram, Google, etc., with more niche online communities like research gate, SRJ, etc., establishing their growing online presence daily. The postulation is based on the principle of users and groups from various OSNs in today's social marketplace interacting with each other through high density linked pathways known as bridges [27]. The order of complexity which SIS seeks to solve is undoubtedly several times that of HNs. Today, SIS remains largely an untrodden area of research due to its sparse but

large volumes of highly complex data which requires management and handling [27].

2 Community detection

Community detection is an important research topic of structural analysis within OSNs and SISs. This research tackles the problem from two significant viewpoints. One mainstream approach involves analyzing node attributes and the similarity indices between them. This method is based on the postulation that a community is made up of nodes with high similarity indexes between them. These networks are modeled after a modular organization, revealing the existence of special affinities among the vertices within the group [28]. The other mainstream approach involves examining the relational attributes between the actors within group. This method is based on the social study that interconnected nodes within the community will have high node degrees between one another. Translated into a friend network like Facebook, this means that if node A has a high degree of mutual friends (directly or indirectly connected) with node B, then node A and B must belong to the same community [28]. Such a common indicative measure is known as modularity; and is simply given as:

$$ModX = \frac{\Phi_A \cup \Phi_B}{\Phi_A \cap \Phi_B} \tag{1}$$

where Φ_A and Φ_B are immediate neighbor relations of nodes A and B respectively. Therefore, if modularity is high, then relationally nodes A and B belong to different communities. If, however, modularity is low, then similarly nodes A and B belong to the same community. It is a common measure to determine if a node is strongly connected to other similar nodes within the network. This method too is based on a similarity measure like the first. For example, in social networks, communities correspond with a group of friends/colleagues/people who attend the same institution or who come from the same hometown. In protein interacting networks, communities represent functional groups of peptides that evolved within a superstructure. In co-authorship (heterogeneous) networks, communities refer to highly correlated disciplines between the authorship of similar papers [29, 30].

2.1 Community detection techniques

Community detection techniques are broadly confined to three major structural scopes of study: homogeneous networks, heterogeneous networks, and social internetworking scenarios [31, 32]. Listed in ascending orders of complexity and size, these methods suffer from a limited scope of study and utility. In a homogeneous network, only actors

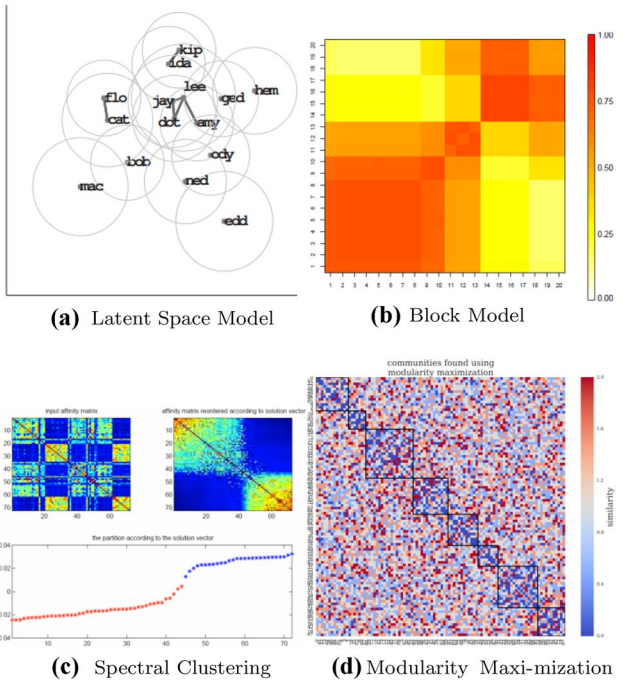


Fig. 1 The four most popular approaches to the community detection problem

of a single type exist, and this feature extends into unidimensional relationships with other actors from the same social circle [33]. This type of network is easy to analyze since structurally, both node and links are confined to one dimension [34, 35]. Many approaches have been developed to detect the community membership of each actor under this study. Some developments include latent space models (Fig. 1a), block model representation (Fig. 1b), spectral clustering (Fig. 1c) and modularity maximization (Fig. 1d).

Intuitively, a stable community is recognized by strong internal connections and weak external links. Strong and Weak here represent established relational densities of varying scales. Thus, most clearly defined and stable communities are often characterized by dense intra-community bonds and sparse intercommunity links at node edges [34, 36, 37].

Nguyen et al. in [11] presents a framework for detecting stable communities by first enriching the links within a chosen network with weighted stability indices based on a mutual friends similarity measure. The main aim of their research is to seek a community partition that remains socially wealthy over time. However, this method suffers limitations of stationary distributions where the random walker is only allowed to walk within and not between communities. Furthermore, their proposed solution essentially reduces to a similar problem in modularity where the wealth of a network is measured as the difference between the sampled and apparent strength of node degree expectations within a community. As an NP hard problem, this

stochastic estimation based solution suffers from resolution when the network size grows arbitrarily large.

Charkraborty et al. in [12] provides a study into when vertex orderings stay invariant and how this changes the results of the community detection algorithms. They show that using invariant orderings of vertices, the variation of the community detection results can be significantly reduced. Through a comparison of their results, they observed that constant communities correlates to a strength metric that is determined by the number of different external communities to which it is connected.

Cheng et al. in [25] proposed a network sparsification pre-processing step to selectively eliminate weak edges in the network (characterized by inter-community links). In their paper, they claim that this approach improves the cluster partitioning capabilities of a spectral method that dissects the bipartite network repeatedly into clusters which are closer to the ground truth. However, their methods follow data reduction techniques that remove information which may be apparent and influential towards structural changes in the organization of communities. This can cause errors in detecting large communities of low edge similarity measures. Again, such a solution faces the tricky problem of electing suitable thresholds for graph partitioning and edge removal.

Lancichinetti et. al in [38] provides insight into combining consensus clustering techniques with existing community detection methods in a consistent manner to detect communities both accurately and with high stability as reflected in the resulting partitions. Additionally, they argue that their framework is also useful in determining the evolutionary nature of community structures in temporal networks.

Seffi et. al in [39] proposes a method to utilize the non-determinism of a community detection process (e.g. modularity, spectral clustering, etc) to identify invariant nodes called community cores that are more similar to ground truths than communities in real and artificial networks. Their method however, suffers from detection accuracy problems as known null models are not scalable to large network sizes. Hence, the prediction errors of calculated significance scores of cores grow exponentially in proportion to network size.

Li et. al in [40] develops a mutuality tendency theory that builds on the intuition that nodes establishing a two-way dyadic reciprocity feature have a tendency to establish mutual connections between themselves more frequently than those which occur by chance. Their proposed algorithm is developed from a spectral clustering framework but with the added capability of being mutuality tendency aware. However, as a spectral method their pre-processing step involves "flattening" a high level multi-dimensional network like slashdot by nonlinear data reduction techniques that could contain truth information related to the actual community structure in real life scenarios.

Delvenne et al. in [41] proposes a measure of quality for the spectral clustering methods based on the clustered auto-covariance of a dynamic Markov process over time taking place in the network. They applied their results over constructive and real networks, including hierarchical (structured) graphs, social (unstructured) networks and a peptide network to obtain reduced descriptions of protein interaction over different time scales.

Heterogeneous networks on the other hand contain rich information embedded in the form of meta nodes and meta paths through hyperlinks [36, 42]. Simplifications by data reduction techniques will lead to inaccurate detection of community structures [43–45]. In a highly complex space, uni-dimensional measures are inadequate at describing the detection process. Currently, one of the more popular and intuitive methods of handling such a problem is to firstly reduce dimensions of the presented heterogeneous network into single dimension layers and apply the homogeneous networks community detection algorithms to them. Secondly, detected community positives are then integrated across the manifold [36, 46, 47].

Liu et al. in [48] tackles the problem of high dimensional heterogeneous networks by first building a bipartite network of node and link vertices reflected in the original network. Their method succeeds in detecting communities in their constructed bipartite network, but is highly inaccurate and not suitable for modularity based approaches for community detection. Furthermore, their method's computational efficiency does not scale well with large network sizes.

In a similar vein, Liu et al. in [49] proposed a method of detecting communities for a tripartite network (3 types of nodes in a 3-uniform hypernetwork) and extending it up to k-dimensionality superstructures for community detection. Their solution relies on information compression which is another popular data reduction technique borrowed from the disciplines of information theory. Essentially, most compression algorithms are lossy and as a major drawback, this strategy eliminates some of the valuable information of the original network during the reduction process which is non-recoverable at the later reconstruction phases.

Meng et al. in [50] suggests a combination of different semantic (meta) social information paths which use a matrix decomposition method similar to [51–53] to extract the similarity weighted measures of interconnected pathways between source and ending objects (which they declare have to be of the same type). They show that their results fair better than spectral and path-selection clustering. As an agglomerative type clustering method however, the method is memoryless and does not keep track of previous transitions. The semantic path type matrix is also highly sensitive to noise and outliers and can also break up large clusters. Additionally, no objective function is sought to be minimized through the community detection process. Therefore,

this method also suffers from inaccuracies when identifying cluster numbers and their correct sizes.

Tang et al. in [24, 54] proposes unified models and the learning of sparse social dimensions for detecting community structures through two steps. The first step is achieved by integrating through multiple dimensions to discover hidden community structures shared by heterogeneous interactions. The second step is to cluster them through an edge-centric clustering scheme. In their experimentation and results, this method appears to be the most robust because the extraction of features eliminates noise and outliers which might otherwise interfere with the tabulated results. Lastly, partition integration is the accumulation of all k-means partitions of individual dimensions within the HN. A summation of all such partitions also sums up the uncertainties which are likely to yield results with relatively high uncertainties.

Mucha et al. in [55] developed a generalized model of network quality functions in each node from one Laplacian slice which connects to itself in other Laplacian slices. Their framework allows the study of community structure in a very general setting encompassing networks that evolve over time, addressing hyperlinks (multiplicity) and multiple scales (time).

Liu et al. in [56] proposes a method of detecting communities through a measure of composite modularity which is made up of weighted component modularity against the fraction of edges in the subnetworks of individual dimensions within the hypergraph. This method is superior to the one proposed in [57] because it does not require a priori knowledge about the number of communities as ground truths for their computation and detection processes. However, a major drawback lies in their assumption that all links are treated equally when assigning weights. This means that edges which contain noise are amplified as a result of adding high weights. This consequently results in inaccurate detection of communities because of the presence of noisy edges.

Zhen et al. in [58] develops a framework based on regularized joint non-negative matrix factorization (RJNMF) that utilizes hyperlink and content information to improve the accuracy of community detection methods. Content information is derived from matrix factorization and structure similarities are obtained directly from combining multiple link and content information, similar to methods used in [59, 60]. As a topic content probability technique, NMF qualitatively leads to worse mixtures over time because of the lack of a Dirichlet prior (which helps control sparsity) on top of the data generation process. This means that NMF will be prone to giving incoherent topics when used with content mining that affects the accuracy of detected communities using their proposed RJNMF.

Aggarwal et al. in [16] proposes the use of a single local succinctness property to extract compressed descriptions of the underlying community representation in a social network with the use of a min-hash approach. However, as an agglomerative link clustering technique, it suffers from the same drawbacks as [50] in being highly susceptible to noise, computationally inefficient and diffractive (cluster-wise).

Social Internetworking Scenarios are a developing area of interest within the OSN community and propose to consider even higher levels of dimensionality (and hence complexity) at even larger scales of data volumes [31, 61, 62].

Buccafurri et al. in [27] proposes a conceptual framework for constructing stereotypical maps in a social internetworking scenario to cluster users of similar stereotypical projections together. Their proposed method (SISO) contains three layers of processing: the stereotype detection layer, the user stereotypical map construction layer, and the SIS stereotypical map construction layer. The aim of this conceptual framework is to detect communities of hyper structures based on the stereotypical behavior metric of nodes/actors within these topological graphs.

Pasquale et al. in [63] adopts the use of k-path edge centrality measures to rank edges based on their centrality scores. With the added advantage of maximizing modularity, their results show improvement over traditional methods like spectral and/or modularity clustering because they consider both the global and local information of the network. However, their methods still require a priori knowledge of existing community structures as ground truths to work effectively. This a priori knowledge in large heterogeneous architectures may not always be available and can remain sparse over long periods of time [64].

3 Link prediction

As a fundamental measure of structural strength, link prediction provides a wide variety of predictive scores to estimate the quality of multi-dimensional relationships between meta nodes [65]. Link prediction is a well-defined research area which tries to solve the problem of predicting missing future inferences of relationships between nodes [14, 66, 67]. The fundamental problem of link prediction centers on likelihoods of the inferred link existence given a priori knowledge of node attributes and structure [68, 69]. As a wide field of research, several methods attempting to justify such inferences have been developed. The first use of link predictive methods on social networks in [70–72] were done on flat homogeneous networks. Largely used as baseline models for network structures of higher dimensionalities and complexities, link estimation and analysis fall under two broad categories of approaches [13, 14, 23]. The first is static node attribute based while the second is structural based. Static node attribute based methods infer the metrics associated with links from identified node features which produce high correlations between each other. These striking

Table 1 Popular social measurement metrics

Graph distance	(Negated) length of shortest path between x and y
Common neighbors	$\lvert\Gamma(x) \cap \Gamma(y)\rvert$
Jaccard's coefficient	$\frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\lvert\Gamma(x)\cup\Gamma(y)\rvert}$
Adamic/Adar	$\sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{1}{\log\lvert\Gamma(z)\rvert}$
Preferential attachment	$\lvert\Gamma(x)\rvert \cdot \lvert\Gamma(y)\rvert$

feature similarities are then used to infer the "strength" of an interconnecting relationship [73, 74]. Some of the popular measurement metrics used within this field of study include the Jaccard's coefficient (JC), preferential attachment (PA), Adamic/Adar (AA), SimRank (SR), and common neighbors (CN) [20, 75]. A formal definition and explanation of these popular terms are given in Table 1 below. Structural based methods are concerned with estimating the existence of links by drawing upon features identified within the node structure of the network [76–79]. Popular estimators include Katz constant (K), centrality measure (CM), shortest path distance (SPD), rooted page rank (RPR) and hitting time (HT). A formal definition and explanation of these popular measures is given in Table 2

3.1 The problem of link prediction

Link prediction models generalize the methods of logically inferring missing links from present feature observations [80–82]. The problem of link prediction extends beyond the mere inference of relationships and similarities between the connecting actors when network complexity increases [73, 83, 84]. The ultimate goal of link prediction is to make this intuitive notion of inference precise and accurate [85, 86]. Several measures and methods are used as indicators for the inference algorithms and performance index of the implementations respectively. In a homogeneous network, link prediction can be reduced to solving a one-dimensional problem of a flat layered relational structure between single node types. In a heterogeneous network, the link prediction problem grows in complexity across multiple logical structural layers of node connections. Each layer describes a unique relational structure between the nodes themselves [87–89]. In a social internetworking scenario (SIS) scheme, link prediction becomes a problem that is extremely complicated to solve [90–92]. This is because solution manifolds increase exponentially in relation to the dimensionality of heterogeneous networks. In social internetworking scenarios, links among different social networks assume a fundamental role of relationally connecting the same user across different social circles together. These missing edges contain very high levels of information that functions to bridge relational structures across different complex hypergraphs. In predicting missing links of such edges, a new discovery of bridges must first be made which overlaps across the problem of link prediction.

3.2 Methods of link prediction

Leskovec et al. in [70] presents a link predictive model that infers the presence of missing links through an adjacency matrix of signed relationships between nodes within a network. In their paper, the authors fused the link prediction problem with the sign prediction problem. Their studies

Table 2 Popular social estimators

Katz	$\sum_{l=1}^{\infty} \beta^l \left\lvert \text{paths}_{x,y}^{\langle l\rangle} \right\rvert$
	Where $\text{paths}_{x,y}^{\langle l\rangle} :=$ {paths of length exactly l from x to y }
	Weighted: $\text{paths}_{x,y}^{\langle 1\rangle} :=$ number of collaborations between x, y
	Unweighted: $\text{paths}_{x,y}^{\langle 1\rangle} := 1$ iff x and y collaborate
Hitting time	$H_{x,y}$
Stationary-normed	$-H_{x,y} \cdot \pi_y$
Commute time	$-(H_{x,y} + H_{y,x})$
Stationary-normed	$-(H_{x,y} \cdot \pi_y + H_{y,x} \cdot \pi_x)$
	Where $H_{x,y} :=$ expected time for random walk from x to reach y
	$\pi_y :=$ stationary distribution weight of y
	(proportion of time the random walk is at node y)
Rooted PageRank_α	Stationary distribution weight of y under the following random walk:
	With probability a, jump to x
	With probability $1-\alpha$, go to random neighbor of current node
SimRank_γ	$\begin{cases} 1 & \text{if} \quad x = y \\ \gamma \cdot \frac{\sum_{a\in\Gamma(x)} \sum_{b\in\Gamma(y)} \text{score(a,b)}}{\lvert\Gamma(x)\rvert\cdot\lvert\Gamma(y)\rvert} & \text{otherwise} \end{cases}$

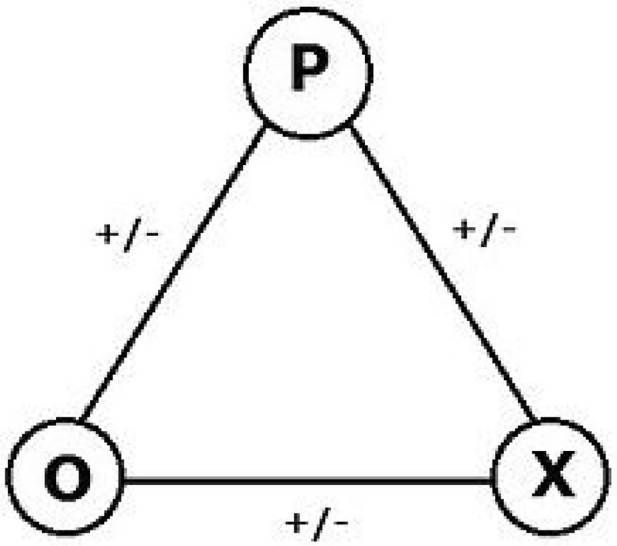


Fig. 2 Basic triadic friendships of three interconnected node actors P, O, and X

of link sign prediction on three datasets of popular OSNs (Epinions, Slashdot, and Wikipedia), using logistic regression classification of individual node features to predict the sign conventions significantly improves performance over previous univariate unsigned (undirected) graph approaches for link prediction (Fig. 2).

Backstrom et al. in [93] tackles the problem of link prediction by effectively combining meta node and hyper-edge attribute information based on their developed method of supervised random walks. The goal of their algorithm is to learn a function that assigns strengths to edges based on likelihood ratios that the random walker will transition through a probabilistic path between two arbitrary nodes which do not have an established relationship. They performed experiments on Facebook and co-authorship networks and their results show large improvements over random walks with restarts (their baseline comparison) and supervised machine learning techniques which require feature extraction and generalization [51, 94, 95].

Fire et al. [96] proposes a method of identifying missing links through the use of simple structural attributes and features. They show through their implementation, that a machine-learning classifier is capable of performing much better at the link prediction task than conventional attribute based classifiers. The experimental results demonstrated that using the friends-measure for link prediction gave better results when compared to the use of more popular measures like JC, CN, AA, etc. However, their research currently considers link prediction using only graph topology features. As a result, if the given graphical network is sparse and contains missing link information, such an approach will quickly diverge from the actual ground truth in the prediction task.

Liden et al. [18] elaborates on the problem formulation of link prediction and establishes an intuitive notion of optimizing proximity measures that theoretically leads to more accurate predictions of missing and unobserved links. They advocate that both network topology and node attributes can work together to infer missing links more accurately than conventional attribute-based or structural based singular approaches. Their listed variables for the actual consideration include finite sums over paths in the network like shortest path distances and number of shared common neighbors.

Chen et al. [71] in their study evaluated four friend recommender systems that qualify the problem of link prediction by establishing links between two or more people who have similar interests and attributes. Their studies discovered that algorithms based on structural network information were able to produce better-received recommendations and find more known contacts for users, while algorithms using object similarity of the relational structure were better at discovering new friends. Results tabulated from their study show that the algorithms they have used to compare performance of the link prediction problem are effective in making people recommendations from social network structures.

4 Structural evolution

4.1 Markov logic networks

Markov logic is an area of research that focuses on stochastically predicting the evolution of networks using a predefined set of behavior beliefs [97]. Such beliefs may be established independently or learned from past node and structure state transition behavior. Independent beliefs are also known as ground truth primitives [98–100]. In a Markovian sense, cliques within a network are associated with a potential energy function that dictates their ground truth states [101–103]. Such functions are also a factor affecting the transitional probabilities from one state to another temporally [104]. In short, a Markov logic network commonly referred to as Markov random fields (MRF) of a graph G(V,E) is given by:

$$P(X = x) = \frac{1}{Z} \exp(\Sigma_j w_j f_j(x_j)) \tag{2}$$

where

$$Z = \Sigma_{x \in \rho} \exp(\Sigma_j w_j f_j(x_j)) \tag{3}$$

which essentially states that the joint probability distribution (which is feature conditional) of all nodes (variables) of the graph is a log linear curve of the weighted exponentiated sum of all features of the clique states [100, 105, 106]. In

the above mentioned mathematical relations, ρ is defined as the superset of all possible assignments to the network structure's random variables. f_j denotes the clique structural features at all j states and x_j corresponds to the probabilistic distribution of the jth posterior feature state of each clique in question, where, in the most direct translation, there is one feature corresponding to each possible state x_k of each clique. Its weight is given as:

$$w_k = \log \phi_k(x_k) \tag{4}$$

where ϕ_k is referenced as the clique potential at any given state in its transitions. Markov logic has been used as an evolutionary inference mechanism to predict the structural behavior of any given knowledge domain based relational schema [97]. It does this by creating a tailored template from combining probabilities with first-order logic specific to the relations and variable features of the respective graph entities in question [97, 107]. Yet, its methods are NP-complete and it scales poorly when the size of the network grows arbitrarily large [99, 108, 109]. To circumvent exponentially increasing complexity with structural growth, MLNs often rely on other disciplines of Markovian statistical learning like Markov chain Monte Carlo (MCMC), Gibbs sampling, and up until recently, Hamiltonian dynamics (HMC) [100, 110, 111]. The true performance measures of the latter Markovian processes are often far superior to MLNs. This arises from the fact that Markovian stochastic processes uses a softer set of first order knowledge base (KB) constraints. Statistical logic-based approaches however, derive from direct brute force Markovian inferences which require hard First Order KB constraints in a manifold. The trade-off from such a compromise is accuracy for speed [99]. MCMC, Gibbs sampling and HMC are all capable of converging faster than a strict first-order logic inference mechanism, towards an approximation of the actual ground truth distribution (assuming no local minimas or maximas within the search solution space) [111]. MCMC with Gibbs sampling and HMC all provide for an accurate enough approximation of the structural schema given enough samples of the posterior distribution surfaces which it builds between each transitional state. A first order knowledge base (KB) is a set of sentences or formulas in first order logic [19]. These formulas are constructed using four types of symbols: constants, variables, functions and predicates. Symbols represent object entities within the logic network, variables refer to any one of such objects that may change during the course of time, functions (mostly probabilistic) negotiate the transitions through which these state changes occur, and predicates are existing relational constraints (links) between object entities (nodes). Together, they form a complete description of the Markovian logic network and its internalized working mechanisms [100, 108].

Markov chain models have been used in a variety of scenarios to predict the behavior of networks over a time frame. In [17] Ching et al. extended Markov chain models into a multi-dimensional space to model multi-variate change with high complexity multi-categorical data. Based on a proposed linear programming method of parameter estimation with a complexity of $O(n^3L)$, their method scales poorly with increasing dimensionality (variables). Their method presents two problems which detection and prediction algorithms are faced with. Firstly, if the patterns of observable data are short and suffer from data sparsity, then state transitioning probabilities will be erroneous given poor a priori estimates. Secondly, convergence to the real structural distribution will be near impossible given the high error rates accumulated through cubic complexity correlations. The log probability weights W_j will have to be carefully chosen and are often done through trial and error. For high order complexities with sparse data observables, the convergence rate of the first-order Markovian logic process functionals entropies. A higher order Markov chain may be required to tackle the problem of high dimensionality metadata.

Piccadi et al. in [112] uses a Markov chain random walk to determine a community quality measure identified in their paper as persistence probability to identify clusters with comparatively strong internal connectivity. An analysis of their study was presented on four networks which showed that their results produce a sharper definition of quality in the detection of communities in comparison to traditional community detection algorithms. However, like [17], this method suffers from a requirement to pre-determine a threshold parameterization of community quality which it assumes as a priori information to the random walk process. Also, a lumped Markov process requires a clustering of original communities into node clusters which are independent distributions from other node clusters. This means that as a single node can only belong to one cluster and not the other, detecting overlapping communities will become a challenge. Furthermore, the modular approach of their model methods lacks granularity and overlooks the time varying nature of edge and hyper-edge attributes within the node cluster. This could impact the accuracy of an identified community given a predefined quality threshold to satisfy.

To address the problem of inaccurate community detections from a predefined quality threshold through lumped Markov models in [112], Hoffman et al. in [113] proposed bounding the errors that could lead to inaccurate observations from lumping of the microstates of the Markov chain into coarser meso-state levels of prediction. An important observation the authors have made through their work is that in a mesoscopic scale, deviations from a microscopic scale are linearly related by uniform bounds through pairwise state transitions. Their results show that their bounds are consistently uniform and not asymptotic. However, their

method requires an ergodic, time homogeneous, and reversible Markov Chain, which necessitates the instantiation of a stationary distribution of the lumped model. The functional flaw is that social networks are a mixture of lumped communities that do not always establish these initial working prerequisites. This will lead to widening inaccuracies through successive iterations of the power-law structure as ergodicity breaks down. Additionally, this method also requires the careful selection of a weighted norm that preserves detailed balance. If the parameterization choice fails, lumping Markov chain models as node clusters will not scale well to prediction performance of a predefined quality threshold.

In [114], Bunescu et al. performs information extraction through Relational Markov networks (RMN) which considers arbitrary dependencies between extractions. The time complexity of their method in computing messages from a potential node to a label node however, is $O(n^2L)$. Their tests on two datasets of Yapex[1] and Aimed[2] show that their method incorporating the use of global entity influencing logical OR templates for cliques significantly improves performance of protein (information) extraction over localized independent extractions through an F-measure. However, this method suffers from a problem of accuracy which limits the performance of the system.

In [115] Taskar et al. presents an in-depth exploration of Relational Markov Networks as a design of a representation language that allows for flexible modeling of complex relational interactions in a heterogeneous structural domain. The authors focus on supervised learning as a motivation for their framework. They argue that structurally, an important source of information is often captured through links. In their results, the authors discussed two particular templates which they have found to be of beneficial use in several applications through RMN models: similarity templates which relate to the classification of links or objects that share a certain graph based property and transitivity templates that relate tuples of objects and links in an organized triangular tessellation on the hypergraph's structural schema.

In [116] Domingos et al. proposes the use of Markov logic to combine probabilistic distributions with first-order logic to address the paradigm of competing tenets between uncertainty and complexity in machine learning applications. They argue that Markov logic outperforms traditional approaches like stochastic logic programs, probabilistic relational models (PRMs), Bayesian logic programs, relational dependency networks, etc. It achieves this through the use of well-defined mathematical relations of extending first order logic by attaching weights to transitional states of graphical relational entities.

[1] URL: http://www.sics.se/humle/projects/prothalt/.

[2] URL: http://ftp.cs.utexas.edu/mooney/bio-data/.

Richardson et. el. in [19] implemented a single representation of both first-order logic and probabilistic graphical models into a ground Markov network for inference and learning of stochastic evolutionary processes of real-world networks. They propose a novel implementation, MaxWalkSat, which performs inference by grounding the minimal subset of the network descriptors required for answering an information query through Gibbs sampling over identified initial states of this subnetwork. Preliminary tests done on real-world university domain data are good and illustrate the promise of MLNs.

In [117] Pedro et al. establishes the use of MLNs in the domain of structured databases to address the problem of data entity resolution in the presence of outliers and noise; usually in the form of similar (corrupted) duplicate records. They show that MLNs are able to efficiently support the data cleaning and preparation stages of the data mining process. The authors propose the use of hybrid measures like predicate and reverse predicate learning to improve performance over pure word-based and string-based approaches for entity resolution. As the entity resolution problem is domain dependent, such relations have to be taken into account when constructing formulas for the first-order logic with ground truths established in the first order KB.

Singla et al. in [118] explores the discriminative learning aspects of Markov Logic ground structures to predict information about co-authorship within a meta schema of hyper nodes and links. They developed an algorithm for discriminative learning of MLN parameters by combining the voted perceptron with a weighted satisfiability solver. The results of conducted experiments on two real-world domains show much promise in this general direction of approach.

Lowd et al. in [119] explores several alternatives to weight learning ranging from per-weight learning rates to second-order methods. Their efforts were focused on solving the problem of inconsistencies due to the ill-conditioning of widely varying clauses which causes a slow gradient descent from state of the art discriminative learning methods, the voted perceptron algorithm. Their developed MLN weight learning methods outperformed the perceptron because of its effective use of second-order information. A major flaw of weight learning for Markov networks is that they can be extremely ill-conditioned, making simple gradient descent-style algorithms very slow to converge.

Poon et al. in [120] presented the first unsupervised approach to co-reference resolution with performance accuracies, rivaling traditional supervised approaches through the joint inference across mentions. Their approach is based on Markov logic to powerfully describe uncertainty in their joint inferences. Their leverages on apposition, predicate nominals, and transitivity paid off in their results when their system achieved an F1 score 7–9% higher than ***Haghighi and Klein (2007). However, co-reference, being an inference

mechanism itself, is still susceptible to ill-conditioning of performance if training with noisy data.

4.2 Trust analysis

Trust computing spreads across many applications in today's digital age. Essentially, trust evolves as a result of stable bonds shared between entities. Such bonds contain a rich set of multi-dimensional attributes and similarities that correlate to the behavior of node entities directly and indirectly connected to it. Although trust is an inferred property of node behavior derived from latent relational constraints, its measure is highly complex and vast. Essentially, it is based on a set of mutually identified predicates that define a set of ground truths which constitute towards a set of belief states that node behavior will be constrained to expected outcomes. Broadly speaking, trust is defined in three main domains: transitivity, asymmetry and personalization. It is also precisely the relational constraining forms of these feature attributes that leads to the complex structure of any trust network.

Trust analysis is a relational learning technique that empowers applications with data integrity protection mechanisms, whenever there are exchanges of information. A problem that trust computing grapples constantly with is the question of how to provide for data requested without sacrificing the disclosure of sensitive fields. As a relational approach, trust analysis has profound implications on data security and privacy. Essentially, through the learning of trust relations between entities, modern-day applications are able to provide for more secure and granular access to sensitive data and information.

In [121], Wang et al. emphasizes the use of trust relations through the correlation of application purpose as a measure of providing the level of access to data. Their work provides an information management framework based on relational integrity between requester and requestee. Relationally, trust varies in proportion to the reputation of a person/entity/requesting source. Most privacy models however, provide sharing of data through implementation of security polices only, which in turn restricts the sharing of information on a needs based requirement irrespective of the purpose of the information request. Additionally, most privacy management policies seek to generalize information towards high levels of anonymity that are costly to decipher at the receiving end, thus preventing the efficient sharing of low level data and information of interest. In this vein, the authors thus propose associating data access through an application-wise intended purpose" for the IR process.

A fine balance exists between privacy and utility. Many social information transactions work on the principle that some amount of privacy will have to be sacrificed in order to provide for the information utility requested by the other party. In practice, many applications have yet to find a suitable fine-tuned solution based on this balance for IR applications. In a bold attempt to passively empower applications with such security features, Sun et al. in [122] extends on [121] by incorporating purpose and trust into anonymizing data so that both privacy of data is preserved for the trustor while at the same time, an adequate utility is provided for the trustee requesting the IR. The authors achieve the above mentioned balance of data security to adequate utility through the use of entropy to prioritize the importance of attributes required by the IR process. In information theory, entropy is defined as a probabilistic measure of the likeliness (or unlikeliness) that a piece of information received at a destination was directed from an interesting source in question.

Although entropy as a predicate function seeks to assimilate sensitive information into anonymity through randomizing (generalizing) multi-layered hierarchy attributes such that information becomes encoded and indiscernible, it suffers from major drawbacks. One such drawback is the problem of overfitting (as in [123]). Firstly, as information entropies the conditional entropy function tends to overfit the data presented in its current state which creates high computational overheads for the degree decomposition of a presented attribute A_i. As a result, the complexity of the decomposition task increases exponentially in proportion to the distribution degrees of entropy. This is given in the equation below:

$$Deg(A_i) = \frac{P_i}{\sum_{i=1}^{n} P_i} \quad \times \quad Deg(T) \quad \text{for } 1 \leq i \leq n \tag{5}$$

which is defined so that the degree of decomposition of an attribute A_i is the weighted ratio of the attribute priority P_i against all attribute priorities in the generalization hierarchy of the degree of table anonymizations, T corresponding to the height of the generalization hierarchy m_i.

Secondly, entropy is not truly infinitely random in a Gaussian manifold. It is only as random as the bounds its distribution is contained in. Li et al. in [124] extends the prior developments in [122] by furthering this constraint in their developed entropy model on l-diversity anonymized dimensions to tackle the problem of overfitting. The authors define two types of disclosures in their work: the identity disclosure and the attribute disclosure and argue that k-anonymity is not enough to protect the integrity of data. Although effective in tackling the problem of overfitting, constraining the problem set to l-diverse dimensions presents serious limitations to the concept and requirement of information entropy. Even though it may be adequate to prevent similarity attacks by implementing the α weighted categorical sum of sensitive information fields, reducing the solution space provides a major advantage to patterned hacks. As a probabilistic

measure, entropy ensures that anonymity is maintained at any given time-aware state of generalized data representation, as long as that presented data independency is above a certain threshold of randomness. Although it is resilient against similarity attacks, it may not prevent patterned hacks from occurring. These kinds of intrusions depend largely on the search space in question. If the space is small, patterned attacks have the advantage of reducing complexity in their identification schema and are thus able to decode information accurately within a shorter amount of time.

The reason why similarity attacks fail is simply because the former tries to find a one dimensional answer in a multi-dimensional solution space, and in so doing arrives at indeterministic results because the solution space overfits the search. However, the latter instead bridges this gap by considering high-complexity data space manifolds of patterned co-occurrences and solves for them simultaneously. This thereby produces more accurate and deterministic results with every single search (however, at the prime cost of speed if the search/solution space is large). Thus, while entropy-governed anonymization methods increases the probabilistic scores of randomness to measures above negative log-linear distribution characteristics of entropy models; De-anonymization techniques therefore, learn these relations which seek to reduce that possibility to zero. Increasing the solution space however, provides a more resilient protection scheme against relational-learning based compromises; however, at the prime cost of utility measurement. In Section 3.1 of [124], the authors argue that it is necessary to balance the opposing goals of data privacy and utility. Their evidence is provided by the discernibility metric (DM) given by the equation:

$$DM = \sum_{QI-groupG} |G^2| \tag{6}$$

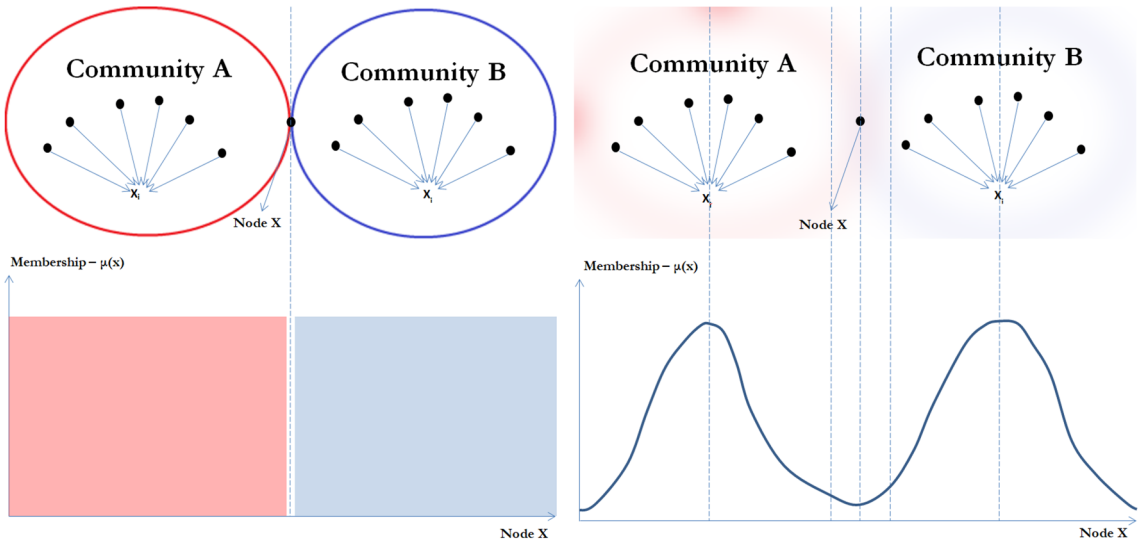
In the equation, $|G|$ is the size of the QI-group G. As can be easily derived from the above mentioned relation, the complexity of discernibility is therefore $O(m^2)$ where m denotes the size of the QI group in question. An objective function therefore, would be to minimize this cost from the direct process of anonymizing discernable data into smaller l-diversity dimensions however, again, at the jeopardy of the security drawbacks mentioned above.

4.3 Sentiment computing and folksonomy

Sentiment computing is an emerging but well-defined area of relational learning applied techniques. Essentially, numerous efforts have been made to develop intelligent algorithms which are capable of detecting emotions in the form of sentiments expressed over the occurrences of key events. Sentiment as an expressive form of emotions over short texts is a highly complex but fully evolved hyperlinked network of social relations between actor node entities. They define the underlying scores over the formation of links between nodes and how they may evolve over time. The methods include the use of natural language processing techniques, text analysis and computational linguistics to identify and extract subjective information from source materials [123]. Its approaches of identifying social subjectivity in emotions are wide and varied. They include, but are not limited to, signed relational graphs, transitivity pattern detection and identification, term frequency-inverse document frequency (TF-IDF) bag-of-words, cosine similarity measure, other TF variants (e.g. sublinear scaling, maximum normalization, query weighting, pivoted normalization, etc), latent Direchlet allocation (LDA), K-means clustering, hierarchical modeling, agglomeration, etc [125].

Folksonomy on the other hand is a derived system of classifying emotions based on their identified tags and their frequency of occurrences. There are two branches of scientific thought approaches to the problem of classification in this domain. One approach is based on intrinsic tag characteristics and pattern recognition schemes while the other focuses on application based tagging techniques. Several algorithms developed in this area include but are not limited to: social Sim rank (SSR), social page rank (SPR), and ranking-based multi-correlation tensor factorization (RMTF) [126]. Both sentiment analysis and folksonomy have impactful implications on the process of information retrieval.

In [123], Rao et al. explores the use of maximum entropy techniques constrained to topic levels for modeling of social emotion classification over short and sparse text. The challenge of sentiment identification over short text as compared to full document pages is the lack of relational subjective word co-occurrences for sentiment analysis methods to effectively reconstruct the contextual topic of the text. This means that in short texts, topic-level identification models suffer from severe data sparsity problems. To overcome the problem of data sparsification inherent in short texts, the authors propose the use of maximum entropy (ME) methods to identify probabilistic thresholds beyond which the contextual relevance of the sentiments expressed within short texts assimilates into generality. The drawback of using classical ME as explained before is the problem of overfitting. The problem with sparsification of words is that their underlying meanings and latent information get displaced over a large superset of inter-correlated contexts of use. The authors argue that it is difficult to generalize the sparsity problem of short texts with a single topic and propose to extract fine topics in social information exchanges through the breakdown of documents into bitterms and learning the global distributions to identify the contextual topic in question. This method of modeling is referred to as the Bitterm topic model (BTM). In short texts, grammar words such as pronouns, articles, conjunctions and prepositions are often



(a) Crisp Community Sets Model **(b)** Fuzzy Community Sets Model

Fig. 3 Differences in membership distributions between well-defined and fuzzy community boundaries

omitted. This can be quite challenging for the sentiment analysis algorithm to detect.

Through their conducted experiments, the authors have shown that topic-level based maximum entropy modeling outperforms other unsupervised sentiment classification and tagging approaches of real-world short text documents and remains competitive in comparison to various other baseline techniques. However, the constraint of topics within a corpus solution space is still vast and may result in indeterministic outcomes if the parameters of latent sentiment word co-occurrences are not tuned properly. This requires the entropy distribution to be further constrained to the specific sentiments tagged within the social information exchange. Which again gives rise to another statistical question: how can we determine the actual distribution of sentiments expressed in a short text given only a sparse a priori co-occurrence of data points?

In a similar vein, Xie et al. in [127] explores the relevancy of contextual data to the corpus tagging of sentiments in folksonomy of OSNs. Their work involves modeling of contextual data in three phases: pre-filter, post-filter and implementation. The authors also define two kinds of contexts which are used for tagging of sentiments expressed over a series of social information exchanges. They are the social context model and the verbal context models.

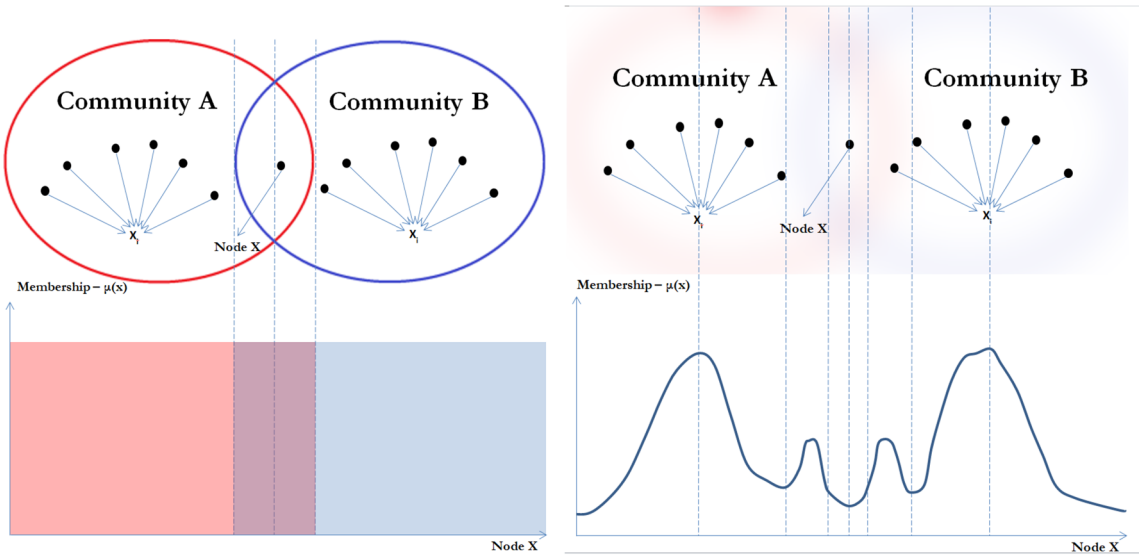
The experiments conducted on the Movielens dataset with their context-aware search optimizations against baseline search methods (which are not context aware and only matches query Q with resource profiles R), show that a dominant context incorporating both verbal and social contextual data outperforms baselines and other contextual-aware variants of search optimizations.

4.4 Fuzzy logic and artificial intelligence

Fuzzy logic is a branch of artificial intelligence that associates itself with quantifying the spread of uncertainties at decision boundaries. Some popular forms of artificial intelligence research include but are not limited to: fuzzy logic, artificial neural networks (ANN), support vector machines (SVMs), K-nearest neighbor (KNN), and genetic algorithms (GA). In almost all branches of stochastic approaches (and therefore online social network analysis), a major flaw of most classification (and therefore, detection, labeling, etc) approaches are that of nodes at the edges of the identified communities and groups. In order to effectively learn from an observation pattern to uncover pivotal information of interest, social relational learning approaches will have to deal with decision ambiguities which are further amplified at the boundaries of such groups of interest. This "vagueness" is then consequently expressed as a probabilistic distribution of likelihoods over predefined quantized (mostly binary) thresholds.

While some AI methods adopt a certain degree of "fuzziness" in their approaches at attempting to solve a theoretical problem logically (and hence, therefore practically), there are others still well received within the research community that adopt a hard decision-boundary, rule-based approach. The concept of "fuzziness" is nothing more than simply a softening the "strict" conditions which exist at the boundaries of most rule-based models (SVM included). As an illustration, refer to Fig. 3a, b below.

As can be seen from the Fig. 3a, at the decision boundary, if an edge node is detected, the probabilities associated with



(a) Crisp Overlapping Communities Model **(b)** Fuzzy Overlapping Communities Model

Fig. 4 Differences in membership distributions between well-defined and fuzzy overlapping communities

its membership of either community A or B is 0. Community membership then, is given by a rectangular function as:

$$rect(\mu_{A/B}) = \Pi(\mu_{A/B}) \tag{7}$$

where

$$\Pi(\mu_{A/B}) = \begin{cases} 0 & \text{where } \mu_{A/B} \leq \mu_{A/B_b} \\ 1 & \text{where } \mu_{A/B} \geq \mu_{A/B_b} \end{cases} \tag{8}$$

where A/B_b are nodes defined at the boundaries of either community A and/or B. However, with the softening of hard conditions defined at the boundaries in Fig. 3b, the membership functions becomes $\mu_{\tilde{A}} : X \rightarrow [0, 1]$ and $\mu_{\tilde{B}} : X \rightarrow [0, 1]$ respectively. Where, $\mu_{\tilde{A}}$ and $\mu_{\tilde{B}}$ denote the fuzziness of memberships in either community groups A or B respectively.

In fact, for realistic online social networks, such crisp boundaries are rarely defined, if ever. Community edge nodes are oftentimes, members of many other such social groups in question, thereby creating regions of decision ambiguities where more than one community overlaps. If these regions were to be defined within enclosed hard boundary conditions (e.g. $A \cap B$), then all nodes which lie within these intersections are definite misclassifications. However, if the boundaries governing such intersections were "blurred" out by a "vagueness" (most conveniently described in most cases as a Gaussian Kernel Mask), then a corresponding distribution filtered out from the softening of these hard constraints will be described as in Fig. 4a, b below.

As can be seen, the trade-off from "weakening" the strict boundary conditions is such that membership of nodes detected from the exact center of the community structure can never be inferred with absolute certainty. It is always lesser than the probability of "1." However, the benefit from this trade-off is acquired at the boundaries, where certain discriminative measures can be used to weight the membership of the edge nodes in question. For such a discriminator as K-NN for example, the decision problem then reduces to a minimization of the proximities between the edge node and its K-nearest neighbors the larger the "K" the more granular (and therefore accurate) the measure. In Fig. 4b, the smaller "peaks" in the distribution are artefacts (outliers) which may be effectively frequency filtered. Essentially, Fuzzy Logic as an intuitive logical method imposes a spread of ambiguity over the entire model subspace of interest, which moderates expectations between absolutes of certainties and uncertainties often defined by hard boundary constraints.

In [128] the authors explore the correlations between generalization capabilities of classifiers and their fuzzy counterparts. The authors argue that uncertainties exist in every aspect of the classification process. This includes both the learning and the reference phases of the classifier. Their observation led to a key discovery of methods of training a classifier to classify samples within the fuzzy set effectively.

Training of a classifier becomes a task which is specific to the type of classifier used to solve the classification problem in question. There is a clear distinction made between crisp classifiers like SVM (which maximizes the Euclidean distance between points of two or more groups) and vague classifiers like K-NN (which classify nodes based on the localized spread of its neighbors). While different classifiers are adept at making decisions under fuzzy conditions, uncertainties are not

only restricted to the algorithms employed under such conditions. Such uncertainties may be reflected as observations of partial information and feature ambiguities.

The fuzziness of fuzzy refers to such pre-existing ambiguities already prevalent in an existing ill-defined set of observables. As a measure of indefiniteness associated with the retrieval of information of interest, the authors mentioned that entropy, as a non-probabilistic approach, can be used to measure the likelihood of making a confidently correct distinction between observed and inferred data sets.

In [129], the authors again expanded on their study on generalization and fuzziness towards high dimensional data structures. From their experimentation, they conclude their discovery with the consensus that generalization of a classifier improves stochastically, when fuzzy randomness of a highly complex data set increases. However, when data dimensionality is reduced, complexity decreases as classifier generalization performance decreases. In this paper, the authors propose the use of an extreme learning machine (ELM) as a single layer feed forward neural network (SLFN) classification approach. Their results and experiments show that there are existing correlations between the fuzziness and generalizations of the classifier. Their graphical representations on 12 data sets calculated from the Pearsons correlation between changing accuracies and fuzziness levels indicate that generally, given a level of relative complexity in the existing data set, the level of fuzziness reaches a threshold where generalization performance of the classifier dips.

In [130], the authors apply the logic of fuzziness onto multiple instance learning (MIL) algorithms. They argue that current research done on multiple instance learning algorithms involves only the informativeness of a bag of training samples in order to make a decision on whether to select it for learning. In their paper, the authors developed the multiple instance active learning (MIAL) algorithm which automates the selection of training bags which contain both high informativeness and diversity. As a measure of fuzziness, the authors study the sparsity of instances within the bag of training data. Their experimental results show that bags with greater internal instance sparsity tend to score a higher dissimilarity degree, which in turn correlates well to a higher diversity index. Such bags can then be chosen by the active learning classifier for multiple instance learning. However, the performance of MIL in diversity-based approaches suffers when instance distributions are not as irregular and its occurrences are not very differentiable from each other.

5 Conclusion

Social networks have evolved from the pre-digital ages to the online internet scene. With the explosion of data prevalent in every aspect of social function in today's social structure, more efficient and robust methods of discovering knowledge from data mining approaches are required to organize and infer useful social relationships for various real-world applications like recommendation, topic modeling, trust reciprocity, etc. In this paper, we have covered a wide scope of current research literature on various relational learning techniques employed in the online social network context. The widely used methods covered in this survey are involved in inferring and identifying relationships co-referenced from structures and include but are not limited to: detection, prediction and Markovian logic approaches from their related core social perspectives of communities, links and networks. It is hoped that this survey will provide both a wider and deeper perspective on the various methods and techniques used to learn from relational structures so that new knowledge can be uncovered from highly complex and voluminous data of the online social networks scene. In conclusion, relational structures are the building blocks of social networks that relate to how actual relationships evolve in real-life scenarios.

References

1. Shi C, Li Y, Zhang J, Sun Y, Philip SY (2017) A survey of heterogeneous information network analysis. IEEE Trans Knowl Data Eng 29(1):17–37
2. Kashima H, Kato T, Yamanishi Y, Sugiyama M, Tsuda K (2009) Link propagation: a fast semi-supervised learning algorithm for link prediction. In: Proceedings of the 2009 SIAM international conference on data mining. SIAM, Philadelphia, pp 1100–1111
3. Liu F, Liu B, Sun C, Liu M, Wang X (2015) Deep belief network-based approaches for link prediction in signed social networks. Entropy 17(4):2140–2169 **(Multidisciplinary Digital Publishing Institute)**
4. Song D, Meyer DA (2015) Link sign prediction and ranking in signed directed social networks. Soc Netw Anal Min 5(1):52 **(Springer)**
5. Lü L, Zhou T (2011) Link prediction in complex networks: a survey. Phys A Stat Mech Appl 390(6):1150–1170 **(Elsevier)**
6. Wang X, He D, Chen D, Xu J (2015) Clustering-based collaborative filtering for link prediction. AAAI, Menlo Park, pp 332–338
7. Taskar B, Abbeel P, Wong M-F, Koller D (2007) Relational Markov networks. In: Introduction to statistical relational learning. MIT Press, New York, pp 175–200
8. Sun Y, Yu Y, Han J (2009) Ranking-based clustering of heterogeneous information networks with star network schema. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 797–806
9. Lin Y-R, Sun J, Castro P, Konuru R, Sundaram H, Kelliher A (2009) Metafac: community discovery via relational hypergraph factorization. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 527–536
10. Ruan Y, Fuhry D, Parthasarathy S (2013) Efficient community detection in large networks using content and links. In: Proceedings of the 22nd international conference on world wide web. ACM, New York, pp 1089–1098

11. Nguyen NP, Alim MA, Dinh TN, Thai MT (2014) A method to detect communities with stability in social networks. In: Social network analysis and mining
12. Chakraborty T, Srinivasan S, Ganguly N, Bhowmick S, Mukherjee A (2013) Constant communities in complex networks. arXiv :1302.5794 (arXiv preprint)
13. Yang F, Zhang F (2017) Community detection for multilayer heterogeneous network. arXiv:1705.05967 **(arXiv preprint)**
14. Tang J, Wang X, Liu H (2012) Integrating social media data for community detection. In: Modeling and mining ubiquitous social media. Springer, New York, pp 1–20
15. Getoor L, Diehl CP (2005) Link mining: a survey. SIGKDD Explor 7:3–12
16. Aggarwal CC, Xie Y, Yu PS (2011) Towards community detection in locally heterogeneous networks. In: Proceedings of the 2011 SIAM international conference on data mining, pp 391–402. SIAM, Philadelphia
17. Cheng W, Zhang X, Guo Z, Wu Y, Sullivan PF, Wang W (2013) Flexible and robust co-regularized multi-domain graph clustering. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 320–328
18. De Meo P, Nocera A, Quattrone G, Rosaci D, Ursino D (2009) Finding reliable users and social networks in a social internetworking system. In: Proceedings of the 2009 international database engineering & applications symposium. ACM, New York, pp 173–181
19. Tao Z, Liu H, Li S, Fu Y (2016) Robust spectral ensemble clustering. In: Proceedings of the 25th ACM international on conference on information and knowledge management. ACM, New York, pp 367–476
20. Sengupta S, Chen Y (2015) Spectral clustering in heterogeneous networks. Stat Sin 2015:1081–1106 **(JSTOR)**
21. Junuthula RR, Xu KS, Devabhaktuni VK (2016) Evaluating link prediction accuracy in dynamic networks with added and removed edges. In: 2016 IEEE international conferences on big data and cloud computing (BDCloud), social computing and networking (SocialCom), sustainable computing and communications (SustainCom) (BDCloud–SocialCom–SustainCom). IEEE, New York, pp 377–384
22. Jalili M, Orouskhani Y, Asgari M, Alipourfard N, Perc M (2017) Link prediction in multiplex online social networks. R Soc Open Sci 4(2):160–863
23. Radicchi F (2013) Detectability of communities in heterogeneous networks. Phys Rev E 88(1):101801 (APS)
24. Tang L, Wang X, Liu H (2012) Scalable learning of collective behavior. In: IEEE transactions on knowledge and data engineering, vol 24(6), pp 1080–1091. IEEE, New York
25. Cheng J, Li L, Leng M, Lu W, Yao Y, Chen X (2016) A divisive spectral method for network community detection. J Stat Mech Theory Exp **(IOP Publishing)**
26. Domingos P, Lowd D (2009) Markov logic: an interface layer for artificial intelligence. In: Synthesis lectures on artificial intelligence and machine learning, vol 3(1). Morgan & Claypool Publishers, San Rafael, pp 1–155
27. Buccafurri F, Lax G, Nocera A, Ursino D (2012) SISO: a conceptual framework for the construction of stereotypical maps in a social internetworking scenario. In: Proceedings of the international workshop on new frontiers in mining complex knowledge patterns at ECML/PKDD
28. Feng X, Zhao JC, Xu K (2012) Link prediction in complex networks: a clustering perspective. Eur Phys J B 85(1):3 **(Springer)**
29. Munasinghe L, Ichise R (2012) Time score: a new feature for link prediction in social networks. In: IEICE transactions on information and systems, vol 95. The Institute of Electronics, Information and Communication Engineers, pp 821–828
30. Munasinghe L, Ichise R (2011) Time aware index for link prediction in social networks. In: DaWaK. Springer, New York, pp 342–353
31. Chen J, Geyer W, Dugan C, Muller M, Guy I (2009) Make new friends, but keep the old: recommending people on social networking sites. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 201–210
32. Tylenda T, Angelova R, Bedathur S (2009) Towards time-aware link prediction in evolving social networks. In: Proceedings of the 3rd workshop on social network mining and analysis. ACM, New York, p 9
33. Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. J Assoc Inf Sci Technol 58(7):1019–1031 **(Wiley Online Library)**
34. Mengshoel OJ, Desai R, Chen A, Tran B (2013) Will we connect again? Machine learning for link prediction in mobile social networks. In: Eleventh workshop on mining and learning with graphs
35. Munasinghe, Lankeshwara and Ichise, Ryutaro *"Multi-class link prediction in social networks"*. proceedings of 27th Annual Conference of the Japanese Society for Artificial Intelligence, 2013
36. Schall D (2014) Link prediction in directed social networks. Soc Netw Anal Min 4(1):157 **(Springer)**
37. Lichtenwalter RN, Lussier JT, Chawla NV (2010) New perspectives and methods in link prediction. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 243–252
38. Lancichinetti A, Fortunato S (2012) Consensus clustering in complex networks. Scientific reports. Nature Publishing Group, London
39. Seifi M, Junier I, Rouquier J-B, Iskrov S, Guillaume J-L (2013) Stable community cores in complex networks. Complex networks. Springer, New York
40. Li Y, Zhang Z-L, Bao J (2012) Mutual or unrequited love: identifying stable clusters in social networks with uni-and bi-directional links. In: International workshop on algorithms and models for the web-graph. Springer, New York, pp 113–125
41. Delvenne J-C, Yaliraki SN, Barahona M (2010) Stability of graph communities across time scales. In: Proceedings of the National Academy of Sciences, vol 107(29), pp 12755–12760. National Academy of Sciences, New York
42. Huang Z, Lin DKJ (2009) The time-series link prediction problem with applications in communication surveillance. INFORMS J Comput 21(2):286–303 **(INFORMS)**
43. Potgieter A, April KA, Cooke RJE, Osunmakinde IO (2009) Temporality in link prediction: understanding social complexity. In: Emergence: complexity and organization, vol 11. ISCE Publishing, Mansfield
44. Choudhury N, Uddin S (2016) Time-aware link prediction to explore network effects on temporal knowledge evolution. Scientometrics 108(2):745–776 **(Springer)**
45. Taskar B, Wong M-F, Abbeel P, Koller D (2004) Link prediction in relational data. In: Advances in neural information processing systems, pp 659–666
46. Munasinghe L, Ichise R (2013) Link prediction in social networks using information flow via active links. In: IEICE transactions on information and systems, vol 96(7). The Institute of Electronics, Information and Communication Engineers, pp 1495–1502
47. Scripps J, Tan P-N, Chen F, Esfahanian A-H (2008) A matrix alignment approach for link prediction. In: 19th international conference on pattern recognition, ICPR 2008. IEEE, New York, pp 1–4

48. Liu W, Murata T, Liu X (2013) Community detection on heterogeneous networks. In: The 27th annual conference of the Japanese Society for Artificial Intelligence
49. Liu X, Murata T (2011) Detecting communities in *k*-partite *k* -uniform (hyper) networks. J Comput Sci Technol 26(5):788–791 **(Springer)**
50. Meng Q, Tafavogh S, Kennedy PJ (2014) Community detection on heterogeneous networks by multiple semantic-path clustering. In: 6th international conference on computational aspects of social networks (CASoN), pp 7–12. IEEE, New York
51. Ahmad MA, Borbora Z, Srivastava J, Contractor N (2010) Link prediction across multiple social networks. In: 2010 IEEE international conference on data mining workshops (ICDMW). IEEE, New York, pp 911–918
52. Tang J, Lou T, Kleinberg J, Wu S (2010) Transfer link prediction across heterogeneous social networks. ACM Trans Inf Syst 9(4)
53. Gao F, Musial K, Cooper C, Tsoka S (2015) Link prediction methods and their accuracy for different social networks and network metrics. In: Scientific programming, vol 2. Hindawi Publishing Corporation, p 1
54. Tang L, Wang X, Liu H (2012) Community detection via heterogeneous interaction analysis. In: Data mining and knowledge discovery, vol 25(1), pp 1–33. Springer, New York
55. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela J-P (2010) Community structure in time-dependent, multiscale, and multiplex networks. Science 328(5980):876–878 **(American Association for the Advancement of Science)**
56. Liu X, Liu W, Murata T, Wakita K (2014) A framework for community detection in heterogeneous multi-relational networks. Adv Complex Syst 17(6):1450018 **(World Scientific)**
57. Yang J, McAuley J, Leskovec J (2013) Community detection in networks with node attributes. In: IEEE 13th international conference on data mining (ICDM), pp 1151–1156. IEEE, New York
58. Li Z, Pan Z, Zhang Y, Li G, Hu G (2016) Efficient community detection in heterogeneous social networks. In: Mathematical problems in engineering. Hindawi Publishing Corporation, Cairo
59. Hisano R (2016) Semi-supervised graph embedding approach to dynamic link prediction. arXiv:1610.04351 **(arXiv preprint)**
60. Vishwanathan SVN, Borgwardt KM, Schraudolph NN (2006) Fast computation of graph kernels. In: Proceedings of the 19th international conference on neural information processing systems. MIT Press, New York, pp 1449–1456
61. Aggarwal CC (2011) An introduction to social network data analytics. In: Social network data analytics. Springer, New York, pp 1–15
62. Song HH, Cho TW, Dave V, Zhang Y, Qiu L (2009) Scalable proximity estimation and link prediction in online social networks. In: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference. ACM, New York, pp 322–335
63. De Meo P, Ferrara E, Fiumara G, Provetti A (2011) Generalized Louvain method for community detection in large networks. In: 11th international conference on intelligent systems design and applications (ISDA), pp 88–93. IEEE, New York
64. De Meo P, Nocera A, Quattrone G, Ursino D (2014) A conceptual framework for community detection, characterisation and membership in a social internetworking scenario. Int J Data Min Model Manag 6(1):22–48 **(Inderscience Publishers Ltd)**
65. Liu W, Lü L (2010) Link prediction based on local random walk. EPL (Europhys Lett) 89(5):58–63 **(IOP Publishing)**
66. Singla P, Domingos P (2006) Entity resolution with Markov logic. In: Sixth international conference on data mining, ICDM'06. IEEE, New York, pp 572–582
67. Kok S, Domingos P (2005) Learning the structure of Markov logic networks. In: Proceedings of the 22nd international conference on machine learning. ACM, New York, pp 441–448
68. Guimerà R, Sales-Pardo M (2009) Missing and spurious interactions and the reconstruction of complex networks. In: Proceedings of the National Academy of Sciences, vol 106(52), pp 22073–22078. National Academy of Sciences, New York
69. Leung CW, Lim E-P, Lo D, Weng J (2010) Mining interesting link formation rules in social networks. In: Proceedings of the 19th ACM international conference on information and knowledge management. ACM, New York, pp 209–218
70. De Meo P, Ferrara E, Fiumara G, Provetti A (2014) Mixing local and global information for community detection in large networks. J Comput Syst Sci 80(1):72–87 **(Elsevier)**
71. He D, Liu D, Jin D, Zhang W (2015) A stochastic model for detecting heterogeneous link communities in complex networks. AAAI, New York, pp 130–136
72. Gupta M, Gao J, Han J (2013) Community distribution outlier detection in heterogeneous information networks. In: Joint European conference on machine learning and knowledge discovery in databases. Springer, New York, pp 557–573
73. Zhou T, Lü L, Zhang Y-C (2009) Predicting missing links via local information. Eur Phys J B Condens Matter Complex Syst 71(4):623–630 **(Springer)**
74. Al Hasan M, Chaoji V, Salem S, Zaki M (2006) Link prediction using supervised learning. In: SDM06: workshop on link analysis, counter-terrorism and security
75. Bickel S, Scheffer T (2004) Multi-view clustering, vol 4. ICDM, Atlantic City, pp 19–26
76. De Sa VR (2005) Spectral clustering with two views. In: ICML workshop on learning with multiple views, pp 20–27
77. Long B, Yu PS, Zhang Z (2008) A general model for multiple view unsupervised learning. In: Proceedings of the 2008 SIAM international conference on data mining. SIAM, Philadelphia, pp 822–833
78. Tang W, Lu Z, Dhillon IS (2009) Clustering with multiple graphs. In: Ninth IEEE international conference on data mining, ICDM'09. IEEE, New York, pp 1016–1021
79. Argyriou A, Herbster M, Pontil M (2006) Combining graph Laplacians for semi-supervised learning. In: Advances in neural information processing systems, pp 67–74
80. Ching W, Zhang S, Ng M (2007) On multi-dimensional Markov chain models. Pac J Optim 3(2) (Yokohama Publishers). http://www.ybook.co.jp/pjo.html
81. Piccardi C (2011) Finding and testing network communities by lumped Markov chains. PLoS One 6(11):e27028 **(Public Library of Science)**
82. Richardson M, Domingos P (2006) Markov logic networks. Mach Learn 62(1):107–136 **(Springer)**
83. Romero DM, Kleinberg JM (2010) The directed closure process in hybrid social-information networks, with an analysis of link formation on Twitter. In: ICWSM
84. Hoffmann KH, Salamon P (2009) Bounding the lumping error in Markov chain dynamics. Appl Math Lett 22(9):471–1475 **(Elsevier)**
85. Lin Y-R, Chi Y, Zhu S, Sundaram H, Tseng BL (2009) Analyzing communities and their evolutions in dynamic social networks. In: ACM transactions on knowledge discovery from data (TKDD), vol 3(2). ACM, New York, p 8
86. Sun Y, Aggarwal CC, Han J (2012) Relation strength-aware clustering of heterogeneous information networks with incomplete attributes. In: Proceedings of the VLDB endowment, vol 5(5). VLDB Endowment Inc, pp 394–405
87. Murata T, Moriyasu S (2007) Link prediction of social networks based on weighted proximity measures. In: Proceedings

of the IEEE/WIC/ACM international conference on web intelligence. IEEE Computer Society, New York, pp 85–88
88. Wang P, Xu B, Wu YR, Zhou XY (2015) Link prediction in social networks: the state-of-the-art. Sci China Inf Sci 58(1):1–38 **(Springer)**
89. Buccafurri F, Lax G, Nocera A, Ursino D (2012) Discovering links among social networks. J Mach Learn Knowl Discov Databases 467–482 (Springer)
90. Singla P, Domingos P (2005) Discriminative training of Markov logic networks, vol 5. AAAI, Menlo Park, pp 868–873
91. Lowd D, Domingos P (2007) Efficient weight learning for Markov logic networks. In: Knowledge discovery in databases: PKDD 2007. Springer, New York, pp 200–211
92. Poon H, Domingos P (2008) Joint unsupervised coreference resolution with Markov logic. In: Proceedings of the conference on empirical methods in natural language processing. Association For Computational Linguistics, pp 650–659
93. De Meo P, Ferrara E, Fiumara G, Provetti A (2013) Enhancing community detection using a network weighting strategy. Inf Sci 222:648–688
94. Buccafurri F, Lax G, Nocera A, Ursino D (2012) Discovering hidden me edges in a social internetworking scenario. In: SEBD, pp 15–26
95. Bunescu R, Mooney RJ (2004) Relational Markov networks for collective information extraction. In: ICML-2004 workshop on statistical relational learning
96. Buccafurri F, Lax G (2009) A probabilistic framework for localizing people with K-anonymity. In: SEBD, pp 349–360
97. Getoor L, Friedman N, Koller D, Taskar B (2001) Learning probabilistic models of relational structure, vol 1. ICML, Stockholm, pp 170–177
98. Domingos P, Richardson M (2007) Markov logic: a unifying framework for statistical relational learning. Stat Relat Learn 2007:339–347
99. Ching W-K, Ng MM, Fung ES, Akutsu T (2005) On construction of stochastic genetic networks based on gene expression sequences. Int J Neural Syst 15(4):297–310 **(World Scientific)**
100. Siu T, Ching W, Ng M, Fung E (2005) On multivariate credibility approach for portfolio credit risk measurement. Quant Financ 5:543–556
101. Tran S, Davis L (2008) Event modeling and recognition using Markov logic networks. In: Computer vision—ECCV 2008. Springer, New York, pp 610–623
102. Domingos P (2005) Mining social networks for viral marketing. IEEE Intell Syst 20(1):80–82
103. Lafferty J, McCallum A, Pereira F et al (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: 18th international conference on machine learning, Williamstown
104. Snijders TAB (2002) Markov chain Monte Carlo estimation of exponential random graph models. J Soc Struct 3(2):1–48
105. Kok S, Domingos P (2010) Learning Markov logic networks using structural motifs. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp 551–558
106. Ching WK, Fung ES, Ng MK (2003) A higher-order Markov model for the Newsboy's problem. J Oper Res Soc 54(3):291–298 **(Palgrave Macmillan)**
107. 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 **(Elsevier)**
108. Staab S, Domingos P, Mike P, Golbeck J, Ding L, Finin T, Joshi A, Nowak A, Vallacher RR (2005) Social networks applied. In: IEEE intelligent systems, vol 20(1). IEEE, New York, pp 80–93
109. Ching W-K, Fung ES, Ng MK (2002) A multivariate Markov chain model for categorical data sequences and its applications in demand predictions. IMA J Manag Math 13(3):187–199 **(Oxford University Press)**
110. Davis J, Domingos P (2009) Deep transfer via second-order Markov logic. In: Proceedings of the 26th annual international conference on machine learning. ACM, New York, pp 217–224
111. Taskar B, Segal E, Koller D (2001) Probabilistic classification and clustering in relational data. In: International joint conference on artificial intelligence, vol 17(1). Lawrence Erlbaum Associates Ltd, pp 870–878
112. Deng C, Lv Z, Liu W, Huang J, Tao D, Gao X (2015) Multi-view matrix decomposition: a new scheme for exploring discriminative information. In: IJCAI, pp 3438–3444
113. Cheng B, Liu G, Wang J, Huang Z, Yan S (2011) Multi-task low-rank affinity pursuit for image segmentation. In: IEEE international conference on computer vision (ICCV). IEEE, New York, pp 2439–2446
114. Guo X, Liu D, Jou B, Zhu M, Cai A, Chang S-F (2013) Robust object co-detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3206–3213
115. Xia R, Pan Y, Du L, Yin J (2014) Robust multi-view spectral clustering via low-rank and sparse decomposition. In: AAAI, pp 2149–2155
116. Zheng X, Zhu S, Gao J, Mamitsuka H (2015) Instance-wise weighted nonnegative matrix factorization for aggregating partitions with locally reliable clusters. In: IJCAI. AAAI Press, pp 4091–4097
117. Wang F, Li T, Wang X, Zhu S, Ding C (2011) Community discovery using nonnegative matrix factorization. Data Min Knowl Discov 22(3):493–521 **(Springer)**
118. Leskovec J, Huttenlocher D, Kleinberg J (2010) Predicting positive and negative links in online social networks. In: Proceedings of the 19th international conference on world wide web. ACM, New York, pp 641–650
119. Backstrom L, Leskovec J (2011) Supervised random walks: predicting and recommending links in social networks. In: Proceedings of the fourth ACM international conference on web search and data mining. ACM, New York, pp 635–644
120. Backstrom L, Leskovec J (2011) Link prediction in social networks using computationally efficient topological features. In: 2011 IEEE third international conference on privacy, security, risk and trust (PASSAT) and 2011 IEEE third international conference on social computing (SocialCom). IEEE, New York, pp 73–80
121. Wang H, Sun L (2010) Trust-involved access control in collaborative open social networks. In: 4th international conference on network and system security (NSS). IEEE, New York, pp 239–246
122. Sun X, Wang H, Li J, Zhang Y (2011) Injecting purpose and trust into data anonymisation. Comput Secur 30(5):332–345 **(Elsevier)**
123. Rao Y, Xie H, Li J, Jin F, Wang FL, Li Q (2016) Social emotion classification of short text via topic-level maximum entropy model. Inf Manag 53(8):978–986 **(Elsevier)**
124. Sun X, Li M, Wang H (2011) A family of enhanced (L, α)-diversity models for privacy preserving data publishing. Future Gener Comput Syst 27(3):348–356 **(Elsevier)**
125. Li X, Xie H, Chen L, Wang J, Deng X (2014) News impact on stock price return via sentiment analysis. Knowl Based Syst 69:14–23 **(Elsevier)**
126. Xie H, Li Q, Mao X, Li X, Cai Y, Rao Y (2014) Community-aware user profile enrichment in folksonomy. Neural Netw 58:111–121 **(Elsevier)**
127. Xie H, Li X, Wang T, Chen L, Li K, Wang FL, Cai Y, Li Q, Min H (2016) Personalized search for social media via dominating verbal context. Neurocomputing 172:27–37 **(Elsevier)**

128. Wang X-Z, Xing H-J, Li Y, Hua Q, Dong C-R, Pedrycz W (2015) A study on relationship between generalization abilities and fuzziness of base classifiers in ensemble learning. In: IEEE transactions on fuzzy systems, vol 23(5). IEEE, New York, pp 1638–1654
129. Wang X-Z, Wang R, Xu C (2017) Discovering the relationship between generalization and uncertainty by incorporating complexity of classification. In: IEEE transactions on cybernetics. IEEE, New York
130. Wang R, Wang X-Z, Kwong S, Xu C (2017) Incorporating diversity and informativeness in multiple-instance active learning. In: IEEE transactions on fuzzy systems, vol 25(6). IEEE, New York, , pp 1460–1475
131. Zhou D, Burges CJC (2007) Spectral clustering and transductive learning with multiple views. In: Proceedings of the 24th international conference on machine learning, pp 1159–1166. ACM, New York
132. Tang L, Liu H, Zhang J, Nazeri Z (2008) Community evolution in dynamic multi-mode networks. In: Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 677–685
133. Özcan A, Öğüdücü ŞG (2015) Multivariate temporal link prediction in evolving social networks. In: IEEE/ACIS 14th international conference on computer and information science (ICIS). IEEE, New York, pp 185–190
134. Jiang S, He D, Rao J (2001) A prediction-based link availability estimation for mobile ad hoc networks. In: Proceedings of the 20th annual joint conference of the IEEE computer and communications societies, INFOCOM 2001, vol 3. IEEE, New York, pp 1745–1752

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.