Predictive Analytics of Social Networks

A Survey of Tasks and Techniques

Ming Yang, William H. Hsu, Surya Teja Kallumadi

Kansas State University

Abstract

In this article we survey the general problem of analyzing a social network in order to make predictions about its behavior, content, or the systems and phenomena that generated it. First, we begin by defining five basic tasks that can be performed using social networks: (1) link prediction; (2) pathway and community formation; (3) recommendation and decision support; (4) risk analysis, and (5) planning, especially intervention planning based on causal analysis. Next, we discuss frameworks for using predictive analytics, availability of annotation, text associated with (or produced within) a social network, information propagation history (*e.g.*, upvotes and shares), trust and reputation data. Meanwhile, we also review challenges such as imbalanced and partial data, concept drift especially as it manifests within social media, and the need for active learning, online learning, and transfer learning. We then discuss general methodologies for predictive analytics, involving network topology and dynamics, heterogeneous information network analysis, stochastic simulation, and topic modeling using the abovementioned text corpora. We continue by describing applications such as predicting "who will follow whom?" in a social network, making entity-to-entity recommendations (person-to-person, business-to-business *aka* B2B, consumer-to-business *aka* C2B, or business-to-consumer *aka* B2C), and analyzing big data (especially transactional data) for customer relationship management (CRM) applications. Finally, we examine a few specific recommender systems and systems for interaction discovery, as part of brief case studies.

1. Introduction: Prediction in Social Networks

Social networks provide a way to anticipate, build, and make use of links, by representating relationships and propagation of phenomena between pairs of entities that can be extended to large-scale dynamical systems. In its most general form, a social network can capture individuals, communities or other organizations, and propagation of everything from information (documents, memes, and rumors) to infectious pathogens. This representation facilitates the study of patterns in the formation, persistence, evolution, and decay of relationships, which in itself forms a type of dynamical system, and also supports modeling of temporal dynamics for events that propagate across a network.

In this first section, we survey goals of predictive analytics using a social network, outline the specific tasks that motivate the use of graph-based models of social networks, and discuss the general state-of-the-field in data science as applied to prediction.

1.1 Overview: Goals of Prediction

In general, time series prediction aims to generate estimates for variables of interest that are associated with future states of some domain. These variables frequently represent a continuation of the input data, modeled under some assumptions about how the future data are distributed as a function of the history of past input, plus exogenous factors such as noise. The term ***forecasting*** refers to this specific type of predictive task. (Gershenfeld & Weigend, 1994) Acquiring the information to support this operation is known as ***modeling*** and frequently involves the application of machine learning and statistical inference. A further goal of the analytical process that informs this model is ***understanding*** the way in which a generative process changes over time; in some scenarios, this means estimating high level parameters or especially structural elements of the time series model.

Getoor (2003) introduces the term *link mining* to describe a specialized form of data mining: analyzing a network structure to discover novel, useful, and comprehensible relationships that are often latent, *i.e.,* not explicitly described. Prototypical link mining tasks, as typified by the three domains that Getoor surveys, include modeling collections of web pages, bibliographies, and the spread of diseases. Each member of such a collection represents one entity. In the case of web page networks, links can be *outlinks* directed from a member page to another page, *inlinks* directed from another page to a member page, or *co-citation* links indicating that some page contains outlinks to both endpoints of a link. Bibliography or *citation networks* model paper-to-paper citations, co-author sets, author-to-institution links, and paper-to-publication relationships. Epidemiological domains are often represented using *contact networks*, which represent individual organisms (especially humans or other animals) using nodes and habitual or incidental contact using links. *Spread models* extend this graphical representation by adding information about incubation and other rates and time-dependent events.

Getoor and Diehl (2005) further survey the task of link mining, taxonomizing tasks into abstract categories such as *object-based*, *link-based*, and *graph-based*. Object-based tasks, used often in information retrieval and visualization, include ranking, classification, group detection (one instance of which is community detection), and identification (including disambiguation and deduplication). Link-based tasks, which we discuss in depth in this article, include the modeling task of *link prediction* – deducing or calculating the likelihood of a future link between two candidate entities, based on their individual attributes and mutual associations. Graph-based tasks include modeling tasks such as discovering subgraphs, as well as characterization or understanding tasks such as classifying an entire graph as a small-world network or being governed by a random generative model – *e.g.*, some type of Erdős–Rényi graph (Erdős & Rényi, 1960).

Social media have proliferated and gained in user population, bandwidth consumed, and volume of content produced since the early 2000s. A brief history and broad survey of social network sites is given by boyd and Ellison (2007), documenting different mechanisms by which online social identity is maintained and computer-mediated communication practiced. This article also introduces contemporary work on characterization and visualization of network structure, modeling offline and online social networks using a combined model, and preservation of privacy on social network sites (SNSs). Many of the modeling tools referenced in this survey paper admit direct application or extension to predictive analytics tasks for SNSs. (Yu, Han, & Faloutsos, 2010)

1.2 Tasks

Predictive analytics refers to the application of statistical and other computational tools to historical data, towards achieving goals of prediction listed in Section 1.1, with the purpose of identifying actionable patterns. These may be positive patterns that the end user wishes to promote and leverage, such as frequent web browsing sequences or social communities, or negative patterns representing phenomena to be counteracted, such as incipient epidemics and criminal networks. Link mining and prediction in social media exist as outgrowths and extensions of predictive analytics in general, but the applications thereof are formulated in service to specific goals. This section gives an overview of these goals and their supporting technical objectives. These are expressed in terms of task definitions: performance elements such as decision support, recommender systems, and risk management, to which the methodology of predictive analytics is applicable. In *decision support systems*, the goal is to provide assistive technology for generating and explaining a recommended course of action to a human user or users. We will see how decision support tasks for social networks can be understood in terms of link, pathway, and community prediction, giving rise to more specialized tasks such as the detection of at-risk groups and modeling the dynamics of information propagation to recognize and act on patterns in social networks.

1.2.1 Link Prediction

Liben-Nowell and Kleinberg (2007) formalize the *link prediction problem* as that of answering the question:

> *Given a snapshot of a social network, can we infer which new interactions among its members are likely to occur in the near future?*

They relate this task to the atemporal problem of inferring missing links from a partially observed network and characterize solution approaches as being based upon node neighborhoods, existing paths in the known network, and "meta-approaches" that are compatible with node and path-based methods. *Node-based scores* that are positively correlated with the existence of a link (u, v) between nodes u and v include: the count of their common neighbors in an undirected graph model; multiplicative or other nonlinear functions of their respective graph degrees, such as *preferential attachment*; and similarity measures used in information retrieval such as inverse log frequency of feature co-occurrence. *Path-based scores* are often based on the count or a parametrically weighted sum of alternative path lengths between u and v. Stochastic sampling-based variants of this type of score include the expected time for a random walk originating from u to reach v (the *hitting time*). Liben-Nowell and Kleinberg discuss Markov chain Monte Carlo approaches, including random walk with restart (Al Hasan, Ahmed, & Neville, 2013), towards estimation of path-based scores for the link prediction task. They also discuss how meta-level approximation methods, such as those resembling latent semantic analysis (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990), can be used to estimate the score for a candidate link (u, v).

Link prediction tasks may apply to graphs that are treated as unchanging over time (static) or dynamic. (Salem, 2009) In the case of static graphs, the task is to discover possibly hidden links from partial information. In the case of dynamic graphs, social network data is treated as a historical snapshot (Barabási, et al., 2002) and the task is to predict a continuation of the data, as in traditional time series. Recovering these missing links in the graph may be viewed as a reconstruction task for a hierarchical structure (Clauset, Moore, & Newman, 2008). This can be done by analyzing local structure or through transformations of the spectral density of edges across the graph (Kunegis & Lommatzsch, 2009). The local structure itself can represent general relationships in an entity-relational data model (Taskar, Wong, Abbeel, & Koller, 2004) or friendship and trust in social media (Hsu, Lancaster, Paradesi, & Weninger, 2007).

The existence of a relationship existing between two users in a social network can be identified by an inference process or by simple classification. Although the inference steps may be probabilistic, logical, or both, the links themselves tend to be categorical. They can be dependent purely on single nodes, local topology, or exogenous information. (Hsu, Lancaster, Paradesi, & Weninger, 2007) In addition to using the structure of the known graph, common features of candidates for link existence (friendship, trust, or mutual community membership) include text-based similarity measures such as the number of common interests or some semantically-weighted sum thereof. (Caragea, Bahirwani, Aljandal, & Hsu, 2009) Al Hasan and Zaki (2011) provide a broad survey of extant link prediction techniques, emphasizing the feature types surveyed earlier by Liben-Nowell and Kleinberg (2007). Recent work has focused on topic modelling approaches to the existence of friendship links in social networks (Parimi & Caragea, 2011) and to the development and use of spatial features in location-based social media (Scellato, Noulas, & Mascolo, 2011). Though these networks can theoretically contain hundreds of millions to billions of vertices, most empirical scientific studies to date have focused on data sets containing thousands to tens of millions of vertices. (Caragea, Bahirwani, Aljandal, & Hsu, 2009)

1.2.2 Degrees of Separation: Pathway and Community Formation

The task of link prediction can be extended to the general problem of finding paths and subgraphs (communities), a general class of problems which may involve the systematic application of local analytical techniques (USA Patent No. US 7069308, 2003; Backstrom, Huttenlocher, Kleinberg, & Lan, 2006) or holistic analysis of the entire social network. Nonlocal analysis is often based on having a global topic model that is used as a similarity measure between users to detect community structure, (Qian, Zhang, & Yang, 2006) This is a case of the general observed phenomenon and theory of *homophily*, the tendency of similar individuals to associate with one another. (McPherson, Smith-Lovin, & Cook, 2001) The purpose and typical performance elements of such a system are to understand the likely participation profiles of users: how long, how often, and with what frequency and volume of information propagation they are likely to participate. (Nov, Naaman, & Ye, 2009) Recent studies on computer-mediated communication have indicated that both internal observable factors such as a user's tenure (longevity of membership and role), and external ones such as a user's motivation for using a photo-sharing site, are relevant to these usage statistics. (Nov, Naaman, & Ye, 2010)

Meanwhile, techniques for predicting the formation of links such as follower linkage in social media are largely based on graph structure, *i.e.*, topology. (Romero & Kleinberg, 2010) Other features that are not purely topological may be overlaid on or otherwise combined with an existing social network. (Brown, Nicosia, Scellato, Noulas, & Mascolo, 2012) In addition, the basic tonal quality of posts by social network users, particularly sentiments expressed about topics of mutual interest, are a commonly-used basis for community formation. (Nguyen, Phung, Adams, & Venkatesh, 2012)

1.2.3 Prediction with Recommendation

One important function of a social network that is of particular importance to third-party providers of information and services is *recommendation*, the identification to users of information, services, and merchandise which they may be interested in. Recommender systems in social media are most often based on a model of intrinsic and tacit trust among users associated with the recipient of recommendations. Insofar as association is an indication of similarity of interests and preferences, this leads to a natural mechanism for collaborative filtering and ranking of recommendations. (O'Donovan & Smyth, 2005) This idea has subsequently been extended to systematic analyses of sparse user-item ratings matrices, weighted by similarity measures between users that are computed using these matrices, in order to make use of this normalization mechanism, a type of social regularization. (Ma, Zhou, Liu, Lyu, & King, 2011) This can also be used to recommend explicit association between users, especially a recommendation to follow another user's postings. (Ma, Yang, Wang, & Yuan, 2014)

In this article, we will introduce predictive methods based solely on this type of behavior within a social network and those that are based on, or augmented using, user profile information.

1.2.4 Risk Prediction and Identifying Risk Groups

Another important function of link prediction in social networks is the systematic identification of at-risk groups based on exposures that can be inferred from social contacts and known features. This often takes the form of contagious disease exposure, a topic that is heavily studied in the literature and which we will examine in this article; however, graph structure can reveal other information such as the support structure available to aged persons and their level of isolation. (Wenger, 1997)

This type of information can further be used to understand the intrinsic properties and identifying characteristics of risk groups. Furthermore, it can in some cases, such as in Wenger's study, provide some basic quantitative measures of risk levels and early-warning criteria for emergent problems, such as in elder care and community health.

1.2.5 Planning and Intervention

Once a mechanism exists for identifying groups that exhibit or admit elevated risk, it may be possible to use social network structure and content in order to plan for, and act during, emergencies. This includes disaster preparedness and intervention planning (National Research Council (S. L. Magsino, Rapporteur), 2009), as well as social mobilization for "time-critical feats, ranging from mapping crises in real time, to organizing mass rallies, to conducting search- and-rescue operations over large geographies". (Rutherford, et al., 2013) The limiting factor here is proximity, which differentiates most interventional models for social crisis management from flash mobs, which in turn constrains the recruitment potential and response time.

Predictive analytics also provides a data-driven basis for optimization of coordination strategies, such as assignment and scheduling of rescue units in a natural disaster scenario such as a flood or landslide. The prediction targets include early crisis warning metrics for such specific risks, extracted from social text and message propagation. (Wex, 2013) The emerging field of computational disaster management includes time-critical aspects of preparation, response, rescue, relief, and repair or cleanup effort in the aftermath of a disaster. Optimization and intelligent systems tasks exist at all stages of this process. (Van Hentenryck, 2013)

1.3 Approaches: Prediction in Data Science

Because of the large scale of social networks, the largest of which currently number in the hundreds of millions to low billions of users, each with as many as several thousand relationships, the general problem area of prediction in social media falls under the rubric of “analytics using big data”. Users of predictive analytics technology are often interested in decision support approaches that beyond the crisis intervention, risk management, and recommendation systems listed above. This presents new challenges to developers of analytics systems. (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011)

In this article, we will delve into the data sciences and specific methodologies (Davenport & Patil, 2012) behind presently used and emerging systems for prediction in social media. The methodologies for big data tend to be more enterprise-wide than limited to analytics or information technology units of the client organization. (Davenport, Barth, & Bean, 2012) Additionally, scalable data integration from heterogeneous sources, such as are prevalent in big data, presents more data management issues than traditional analytics – particularly with respect to data definitions (metadata and ontologies). (Chen, Chiang, & Storey, 2012; Davenport, Barth, & Bean, 2012)

2. Background

2.1 Predictive Analytics in Social Networks

The term *predictive analytics* generally refers to the development and federated display of models for the future state of a system based on observed data. As Wex (2013) notes, digital media outlets such as online news provide knowledge sources and a mechanism by which historical data (and text corpora) can help improve understanding on the emergence of crises:

> *With a main focus on online news and a ubiquitous information overload, crisis managers are constantly confronted to masses of publicly available, yet unstructured data sources. Online news cannot be clearly characterized as being “real-time” unlike e.g. ad-hoc messages, thus making it difficult to explain the latency between the occurrence of an event and its proclamation. Yet, news stories often possess meta-data such as geographical tagging, an accumulation of similar reports, keywords, or subjective author belief which calls for the application of superior analytical methods, i.e. text mining, to investigate hidden statistical relationships between the gradual emergence of a crisis and its medial proclamation. However, expertise and knowledge of how to transform this data into machine-readable information and how to engage in prediction methods is frequently non-existent.*

The challenges of predictive analytics applied to social media include data integration, cleaning, and visualization. (Thomas & Kielman, 2009) Thomas and Kielman identify the following ten needs of visual analytics technologies and systems:

1. **Whole-part relationships:** the ability to represent hierarchies in a scale-independent way
2. **Relationship discovery:** the capability to discover interrelationships among people, places, times, and other attributes and features, using techniques from information retrieval (indexing for phrase-based search), relational databases (query by example), and data mining (clustering and classification)
3. **Combined exploratory and confirmatory interaction**: a cognitive model for interactive hypothesis testing
4. **Multiple data types**: adaptive hypermedia and multimedia, task-adaptive views and representations, and a content repurposing capability
5. **Temporal views and interactions**: the ability to represent temporal dynamics of processes, including (causal) flows, timelines, and event and milestone visualizations
6. **Groupings and outlier identification:** labeling and annotation of clusters, and the application of clustering to outlier detection
7. **Multiple linked views**: materialized views supporting the application of data transformation actions committed on one view to data displayed using other views
8. **Labeling**: user-controllable, contextualized views for dynamic visualization and data modeling
9. **Reporting**: the ability to save and reproduce analytical operations and results for publication
10. **Interdisciplinary science**: accessibility by users and subject matter experts with different expertise and background

Predictive analytics applications that deal with social media are numerous, especially in business, and tend to focus around assistive technologies for customers as users, or decision support systems for customer relationship management and business intelligence. (Taylor, 2011)

2.1.1 Foundational Graph Theory and Link Analytics

Much of the methodology that supports link mining is based on analytical graph algorithms as the basis of predicting the existence or behavior of a link. (Washio & Motoda, 2003) The function of algorithms includes compilation of frequent pattern bases, in a manner similar to frequent itemset mining; *gSpan* (Yan & Han, 2002) is one of the earliest such algorithms and *CloseGraph* (Yan & Han, 2003) is its closed subgraph analogue. Some relational data mining systems use only such graph-based algorithms, while some also use logical constraints and properties as expressed using inductive logic programming. (Ketkar, Holder, & Cook, 2005) Ketkar *et al.* report results of experimental evaluation that illustrates tradeoffs: these results indicate that graph-based multirelational data mining algorithms perform better that logic-based ones on structurally complex networks, while the logic-based is best for semantically complex content, and the accuracy of relationship prediction is comparable for generic, semantically shallow, medium-sized networks.

Getoor (2003) introduced a catalog of link mining applications in the earliest survey of extant machine learning and data mining research applied to graph structure. Methodological advances since then have included both visual analytics and statistical analytics, and have focused on generalizing over both structure and content to recover network structure. (Shen, Ma, & Eliassi-Rad, 2006) Some of this early work on heterogeneous information network structure focuses semantic graphs with different entity types – hence the term "heterogeneous" – and the need to account for these semantics through some type of quantitative or formal ontology in estimating link strength.

Link analytics is an informal term that refers to the systematic analysis of graph topology and statistics in order to build a holistic predictive model. Algorithmic subtasks of link analytics include enumeration maximal cliques, a parallelizable task (Du, Wu, Xu, Wang, & Pei, 2006) and using large matrix factorization (Acar, Dunlavy, & Kolda, 2009). More recently Olsman, Roxendal, and Åkerblom (2013) have adapted network models of organizational theory to capture the behavior of organizational social networks.

2.1.2 Time Series Analysis: Forecasting, Modeling, and Understanding

Section 1 introduced the notions of forecasting, modeling, and understanding – terms from the literature of the time series analysis community, subareas such as signal identification, and related areas such as signal processing. We refer the interested reader to seminal references on the topic, especialy: Box, Jenkins, and Reinsel (2008), Chatfield (2004), and the introduction to time series modeling by Gershenfeld and Weigend (1994) in the anthology of time series analytics papers based on the Santa Fe Time Series Competition.

2.1.3 Statistical Modeling of Network Dynamics

Besides network topology, the task of predicting the behavior and output of a social network may involve network dynamics and thus fall under the purview of dynamical systems modeling. The mathematical foundations for this modeling include chaos theory (Gregersen & Sailer, 1993) and both statistical and graph-theoretic analysis of network topology as a determinant of dynamics (Borgatti, Mehra, Brass, & Labianca, 2009).

2.2 Using Network-Associated Content

Much of the early domain literature in content-based social network modeling and prediction originates from collaboration graphs and in particular citation graphs, where nodes represent papers and sometimes authors, and links represent citations. The *CiteSeer* system is one of the first of these (Giles, Bollacker, & Lawrence, 1998). The original system included an autonomous web agent that crawled and digested publication pages of academic authors in computer science (Bollacker, Lawrence, & Giles, 1998), extracting the local web of citations for all papers and linked preprints or reprints it identified. This autonomous citation indexing mechanism (Lawrence, Giles, & Bollacker, 1999) represents a broader class of document-based information federation (*aka* data integration or information integration) and information extraction systems based on web crawling and scraping. The resultant network modeling algorithms that emerged from this work support collaborative filtering of search results based on personality diagnosis of search engine users using memory-based and model-based inductive learning systems (Pennock, Horvitz, Lawrence, & Giles, 2000), and were later shown to be compatible with content-based filtering in data-sparse environments (Popescul, Ungar, Pennock, & Lawrence, 2001). Achieving high-recall link mining systems was shown to require more than simply maximizing scores in a system where publications competed for citations or other links (Pennock, Flake, Lawrence, Glover, & Giles, 2002).

More recently, the method of combining both content-based and collaborative filtering approaches has been applied to document categorization in scientific domains (Cao & Gao, 2005) and to online text in general (Angelova & Weikum, 2006). A profile of relevant work to date appears in a survey on web data mining research by Singh and Singh (2010).

2.2.1 User Profile Data

Web mining systems, going beyond simple link mining, often use user profile data (Cooley, Mobasher, & Srivastava, 1997). One application of this approach is towards user modeling, personalization, and adaptive synthesis of hypermedia (Mobasher, Cooley, & Srivastava, 2000); another is to use node-specific data (*i.e.,* single user data) in the link existence prediction task (Hsu, King, Paradesi, Pydimarri, & Weninger, 2006) and in understanding characteristic content-based features of their own accord, and in relation to subgraph patterns in social networks (Thelwall, 2008). This approach has been used to develop and refine data models for crawled social network data (Catanese, De Meo, Ferrara, Fiumara, & Provetti, 2011), to annotate user features with social relationships (Sun, Lin, Chen, & Liu, 2013), and to capture user self-description in microblogging sites such as *Twitter* (Semertzidis, Pitoura, & Tsaparas, 2013).

2.2.2 Temporal Event Data

Besides graph structure and content related to users or other individual nodes, we may consider identifiable events associated with originating nodes, which may propagate through the network structure. These events, which are often spatiotemporal in nature or can be associated with location and time, can in turn yield detected social network structure (Lauw, Lim, Pang, & Tan, 2011), and can be used to predict subsequent events and rank predictions (O'Madadhain, Hutchins, & Smyth, 2005). We refer the reader to Aggarwal (2011) for a general introduction to this type of data analytics.

2.2.3 Free Text Corpora

The idea of a semantic network dates back to work by Woods (1975), who considered the heterogeneity of nodes and links, and discussed ways in which link structure can be inferred from data and observations. Later researchers explored the use of free text corpora – *i.e.*, collections of documents for natural language text in unrestricted form – in this task (Mladenic, 1999). Blog and forum posts are often published online but are sometimes available only to registered users, limiting their general use as test beds. By the middle of the 2000s, a surge in formation of social network sites (SNSs) led to increased interest in e-mail corpora (Culotta, Bekkerman, & McCallum, 2004), especially author-recipient-topic models that could be built from public corpora such as the Enron e-mail corpus (McCallum, Corrada-Emmanuel, & Wang, 2005). The problem of discovering the roles of author and recipient in a behavioral rubric or schema then gained interest (McCallum, Wang, & Corrada-Emmanuel, 2007), both as a means of extracting social computing models from free text corpora but as a way to augment topic modeling (Chang, Boyd-Graber, & Blei, 2009). By 2010, the use of social media as a test bed for political sentiment analysis had become a popular topic of experimental and applied research (Tumasjan, Sprenger, Sandner, & Welpe, 2010).

Meanwhile, the named entities and relationships in free text corpora facilitated the development of document-topic hierarchies that used the organizational structure of the document collection itself. (Weninger, Bisk, & Han, 2012) This augmented the methodology of identifying and tracking actors to form a timeline of relationships using social networks and text (Danowski & Cepela, 2010) and the "news feed" approach of selecting items of relevance to a user (Berkovsky, Freyne, Kimani, & Smith, 2011).

The current state of the field in free text-augmented predictive analytics using social networks makes use of self-description (Semertzidis, Pitoura, & Tsaparas, 2013) and integrates sentiment, which may be dynamic, with detected topics (He, Lin, Gao, & Wong, 2013).

2.2.4 Decision History: Selections and Ratings

One key application of analytics using social data is helping users track their communications and history of relationships. Recently this has taken the form of a decision history, which users occasionally search through, but often review only for recent or significant events. The technique of personality diagnosis has been used to provide a topical filter for recommendation of such items (Pennock, Horvitz, Lawrence, & Giles, 2000). For more information, we refer the reader to a survey by Cacheda, Carneiro, Fernández, and Formoso (2011) on collaborative filtering systems.

2.2.5 Trust and Reputation

Trust and reputation systems can be based on ties indicated by social networks (Levin & Cross, 2004) and other forms of computer-mediated communication (CMC). A KDD 2006 panel (Piatetsky-Shapiro, et al., 2006) indicated that link mining applications such as trust networks represented one of the grand challenge problems in knowledge discovery in databases. More recent work has yielded approaches based on mutual influence within communities (Matsuo & Yamamoto, 2009). King, Li, and Chan (2009) survey these techniques in relation to data mining problems in social computing.

Some research on collaborative recommendation has looked at adversarial and remedial aspects (Mobasher, Burke, Bhaumik, & Sandvig, 2007). The evolution of trust networks is an important recent topic of interest to which supervised machine learning techniques have been applied (Zolfaghar & Aghaie, 2011).

2.3 Challenges

Link mining, especially social network mining, has frequently come up in the past decade as a major challenge in data mining research. (Yang & Wu, 2006) Challenging aspects of prediction in social networks include their large scale, the existence of partial data (especially due to having incomplete graphs at the time of application), hidden user information, heterogeneous structure and semantics, and multiple types and provenance of data such as free text that are associated with the network. (Gao, 2012)

2.3.1 Imbalanced Data

An additional challenge is that social networks can be sparse, which means that a candidate link may be unlikely to exist, and presenting imbalanced class labels in prediction data. Differences in link distribution have been observed in real-world applications (Al Hasan, Chaoji, Salem, & Zaki, 2006), including large-scale networks (Wakita & Tsurumi, 2007) and time series compiled by periodically crawling social network sites (Hsu, Weninger, & Paradesi, 2008). This property is particularly challenging in scale-free networks and in some multirelational, large-scale global networks (Szell, Lambiotte, & Thurner, 2010).

2.3.2 Partial Data

Partial data is also particularly challenging due to nuanced properties of social structure in some application domains such as prisons and other closed societies. (Zheng, Salganik, & Gelman, 2006) This is borne out in looking at the topology of huge SNSs (Ahn, Han, Kwak, Moon, & Jeong, 2007) and those that emerge episodically based on both online and in-person friend recommendations (Caragea, Bahirwani, Aljandal, & Hsu, 2009).

Links can also propagate gradually in a social network that publishes upvotes and likes. Modeling this in phenomenon in a fast and scalable way can be highly challenging. (Kashima, Kato, Yamanishi, Sugiyama, & Tsuda, 2009; Raymond & Kashima, 2010)

2.3.3 Social Concept Drift

In addition to partial observability, social networks are subject to other changes, such as gradual or sudden episodic changes in organizational structure. This can be thought of as a graph-based variation of concept drift. The problem of community detection is made more challenging by this temporal aspect. (Zhou, Councill, Zha, & Giles, 2007); recent work on time-aware link prediction has focused on this problem (Tylenda, Angelova, & Bedathur, 2009). The general problem of learning under concept drift is surveyed by Žliobaitė (2009).

2.3.4 Active Learning

Link prediction may be seen as a passive or active machine learning and classification task. The active variant may involve querying a source of information such as eliciting user annotation; this is appropriate for some tasks, such as deduplication, that are already intrinsically interactive. (Sarawagi & Bhamidipaty, 2002) Active learning can help to focus and speed up search over candidate network structures when the hypothesis space is very complex (Newman, 2003) and especially when a statistical risk minimization criterion is available (Macskassy, 2009). Recent applications of active learning to networked data include some clustering algorithms in both unsupervised and semi-supervised learning scenarios (Bilgic, Mihalkova, & Getoor, 2010).

Active learning in structured domains is an active research topic (Dietterich, Domingos, Getoor, Muggleton, & Tadepalli, 2008). One application of particular interest is interactivity in community formation, where users can provide individual feedback on the fly. (Amershi, Fogarty, & Weld, 2012)

2.3.5 Online Learning and Incrementality

Online learning is that which occurs during the application of the performance element of learning, such as recommendation or classification. The requirement that this be achieved without significant retraining is called *incremental learning*. Apart from potentially having an interactive setting, online learning can make use of conversations in progress (Glance, et al., 2005), search (Davitz, Yu, Basu, Gutelius, & Harris, 2007), semantic social media such as status updates and tweets (Barbieri, et al., 2010), and events between which some measure of similarity exists (Becker, Naaman, & Gravano, 2010).

2.3.6 Transfer Learning

Transfer learning allows learned models, produced using training data from one domain, to be applied to another domain that is either sufficiently similar or admits some analogical transformation or derivation. Pan and Yang (2010) provide a general survey of the task. In social networks, collective link prediction can be a very data-intensive problem, and so the application of multiple heterogeneous domains is one way to bootstrap the data mining task, when feasible. (Cao, Liu, & Yang, 2010) Some text-based approaches to transfer learning look at inferring implicit opinion from linguistic biases (Guerra, Veloso, Meira Jr., & Almeida, 2011). Finally, some social network mining systems attempt to infer social ties across heterogeneous networks. (Tang, Lou, & Kleinberg, 2012)

3. Techniques

3.1 Topological Analysis

As outlined in Section 1.2.1 and 2.1.1, the most basic and general link prediction algorithms are based purely on graph structure. (Huang, 2006) This has given rise to models of hierarchical structure (Clauset, Moore, & Newman, 2008), collective classification (Sen, et al., 2008), clique structure in social networks (Caragea, Bahirwani, Aljandal, & Hsu, 2009), and parameter estimation tasks that depend on the underlying graph structure, such as: learning influence probabilities between nodes (Goyal, Bonchi, & Lakshmanan, 2010); predicting positive and negative links (Leskovec, Huttenlocher, & Kleinberg, 2010); using supervised random walks over the network to predict and recommend links (Backstrom & Leskovec, 2011); performing feature construction using low-level topological features to predict links (Fire, et al., 2011); and performing structural role extraction and mining in large graphs (Henderson, et al., 2012).

3.2 Network Dynamics

Beyond topological structure lie statistics on the dynamics of networks: frequencies of link formation, deletion, and transformation, plus information propagation events. The latent structure of a social network can be learned in part using data about this dynamic behavior. (Myers & Leskovec, 2010) Some models are based on analysis of the network formation process according to a predictive process, such a power law process. (Clauset, Shalizi, & Newman, 2009)

3.3 Information Propagation in Heterogeneous Information Network

A central approach to social network mining is to examine the way that information propagates through events. In SNSs, these include sharing events such as posts, comments, shares (and retweets), likes (upvotes), and dislikes (downvotes).

Some organizational social networks, especially multiunit organizations, admit a hybrid cooperation and competition model of knowledge sharing. (Tsai, 2002) Some knowledge propagates purely by the sharing of links (Han, 2009). In networks where link evolution leads to a fixed structure such as a star schema, ranking information can be captured using graph clustering algorithms (Sun, Yu, & Han, 2009) More generally, heterogeneous information networks can be used to perform collective classification by graph-based transduction. (Ji, Sun, Danilevsky, Han, & Gao, 2010)

This general approach of tracking information propagation can be implemented using simple author-recipient tagging, even on public social media such as Twitter (Wu, Hofman, Mason, & Watts, 2011). The result of this type of addressing, as well as that of social tagging, can be applied to detect communities (Murata, 2011). By a similar token, observed sharing events can be used to infer topological features (Ohara, Saito, Kimura, & Motoda, 2011) and predict catastrophic propagation events in the case of an epidemiological spread model for a contact network (Boman, 2011). When teams in an organizational network are large and decentralized, heterogeneous information network analysis provides a means of performing efficient sharing and resolution of conflicting opinions (Pryymak, Rogers, & Jennings, 2011), biased propagation (Deng, Han, Zhao, Yu, & Lin, 2011) and ranking-based collective classification (Ji, Han, & Danilevsky, 2011), and top-*k* similarity search (Sun, Han, Yan, Yu, & Wu, 2011).

Network-wide dynamic properties can be estimated from similar observation data (Sycara, 2012), as can clustering of nodes based on relation strength (Sun, Aggarwal, & Han, 2012), and time-localized prediction of relationship links (Sun, Han, Aggarwal, & Chawla, 2012). This approach to clustering also supports path selection, *i.e.*, selection of entity type sequences in a directed graph, with user guidance (Gupta, Gao, Sun, & Han, 2012), giving rise to a general mechanism for using structural analysis to mine heterogeneous information networks (Sun & Han, 2012).

Specific applications of topological and statistical analysis of heterogeneous graphs include identification of important individuals (Schulte, Riahi, & Li, 2013), weighted collaborative filtering based on entity similarity (Yu, Ren, Gu, Sun, & Han, 2013), development of topic modeling similarity measures for collaborative classification and filtering (Hsu, Koduru, & Zhai, 2013), and identifying individuals in a cybercriminal network by probable role (Lau, 2013). The applicable clustering mechanisms include large matrix factorization (Liu & Han, 2013) and statistical relational learning (Schulte & Qian, 2013).

In addition, pairwise co-occurrence relationships such as label and instance correlations can add additional information for collective classification (Wang & Sukthankar, 2013), as can meta-path selection, *e.g.*, sequences such as **$\textbf{Author}_1$-$\textbf{Paper}_1$-Conference-$\textbf{Paper}_2$-$\textbf{Author}_2$**, when combined with the abovementioned user guidance (Sun, et al., 2013).

3.4 Predicting Events: Stochastic Processes and other Time Series Models

Berendt *et al.* (2003) give a roadmap to web mining methods that were extant or under development at the time when the Semantic Web was first emerging. Since then, developments have been published that focus specifically on prediction of events. These include prediction and ranking algorithms (O'Madadhain, Hutchins, & Smyth, 2005), learning similarity metrics (Becker, Naaman, & Gravano, 2010), monitoring sharing events and user tags (Mathioudakis & Koudas, 2010), relating social media events to physical-world behavior (Abbasi, Chai, Liu, & Sagoo, 2012), analyzing responses to microblog posts (Artzi, Pantel, & Gamon, 2012), and performing aggregate sentiment analysis of Twitter (Hu, Wang, & Kambhampati, 2013).

3.5 Topic Modeling and Text Analytics

Text analytics adds a further dimension of complexity and richness to prediction in social networks. The literature is sparser on topic modeling as related to prediction in social networks than text mining *from* social network posts. A general survey is provided by Hu and Liu (2012), the first author of whom has published recent work on topic modeling to align events with Twitter feedback (Hu, John, Wang, &

Kambhampati, 2012) and perform aggregate sentiment analysis over tweets from multiple users (Hu, Wang, & Kambhampati, 2013).

4. Applications

This section surveys selected popular applications of predictive analytics in social network domains. These are: predicting which users will follow whom; making entity-to-entity recommendations (product-to-user, business-to-business, *etc.*); managing risk; analyzing big data in a scientific or engineering setting; and performing information management.

4.1 Follow Prediction

The problem of predicting and recommending followers can be addressed using scoring functions and latent factor analysis (Zhao, 2012), and may further involve the semantics of links and nodes by capturing which users are paying attention to what information from others (Rowe, Stankovic, & Alani, 2012). Factor models over social graphs may be asymmetric. (Ma, Yang, Wang, & Yuan, 2014)

4.2 Recommender Systems

Entity-to-entity recommendation in social media is based on a tacit model of trust between associated users, representing the idea that if another user or company is directly connected to the user or the user's company, then its recommendations should carry some positive weight. In social networks where trust and reputation (which puts a nonlocal bias on trust) is not only fluid but can affect whether links persist or are removed, the dynamics of trust in the network itself are important. Aula (2010) examines mechanisms of reputation risk and publicity management for users making recommendations. In trust networks, quantitative indicators of trust can be transmitted, by word of mouth or the equivalent in CMC. Studies of trust propagation have shown that it can be modeled predictively from observations just as recommendations themselves are: using large matrix factorization. (Jamali & Ester, 2010)

As discussed in Section 1.2.3, social networks can provide a regularization mechanism for recommendations that reweights them based on connectivity. (Ma, Zhou, Liu, Lyu, & King, 2011) A similar mechanism can be used to generate news article or topic recommendations based on retweets (Abel, Gao, Houben, & Tao, 2013) or other reshares. Analogous mechanisms extend this property to the general domain of social recommendation. (Sun, Lin, Chen, & Liu, 2013)

4.2.1 Collaborative Filtering and Collaborative Recommendation

Konstas, Stathopoulos, and Jose (2009) review the state of the field in collaborative recommendation using social networks. Some of the more recent work focuses on temporal dynamics as mentioned in Sections 2.2.2 and 3.2 (Koren, 2009); other approaches incorporate content-based personalization (Berkovsky, Freyne, & Smith, 2012) and attentional factors (Rowe, Stankovic, & Alani, 2012).

4.2.2 Coping with and Using Time

Fluctuations and the frequency and amplitude of concept drift are critical factors in predictive systems. In some intelligent systems, such as recommender systems, variation is deliberately induced to keep material fresh; some recommender systems are specifically designed to avoid redundant or stale (over-similar) recommendations within a set period of time. (Lathia, Hailes, Capra, & Amatriain, 2010) Many such systems (Koenigstein, Dror, & Koren, 2011) use an explicit "avoid this item for a while" or "avoid this category for a while" control; others (Baltrunas & Amatriain, 2009) use implicit feedback and perform time-dependent recommendation. Time-awareness is also important in link prediction. (Tylenda, Angelova, & Bedathur, 2009)

4.2.3 Using Location in Spatiotemporal Prediction

Some SNSs use geolocation as a feature for predicting user behavior such as visiting physical locales or frequenting local establishments. (Scellato, Noulas, & Mascolo, 2011; Brown, Nicosia, Scellato, Noulas, & Mascolo, 2012) Other social network-based user modeling systems maintain actual proximity networks to predict interactions between users. (Do & Gatica-Perez, 2013)

4.3 Risk Management and Assistive Technologies

Risk management is a key application of social networks, and includes concrete applications ranging from large scale disaster management (Wex, 2013; Van Hentenryck, 2013) to predicting runaways and isolated seniors at greater risk in case of medical emergencies.

4.3.1 Identifying At-Risk Groups in Epidemology and Health Care

Social networks for epidemiology date back to long before the current generation of contact network models. Berkman and Syme (1978) conducted a nine-year study of host resistance and mortality among Alameda County residents. Since then, graphical models of disease spread, including causal interventional models, have been used to predict the propagation of foot-and-mouth disease (Ferguson, Donnelly, & Anderson, 2001), tuberculosis (Getoor, Rhee, Koller, & Small, 2004), SARS (Meyers, Pourbohloul, Newman, Skowronski, & Brunham, 2005), and more generally individuals at higher risk of infection of communicable diseases, especially airborne diseases, due to social interactions (Christley, et al., 2005).

Risk management modeling using social networks is not limited to epidemiology; interaction networks have been used in applications such as antiterrorism and intelligence analytics (McCue, 2005), or combating obesity (Arteaga, Kudeki, & Woodworth, 2009). Also, not all epidemiological models built using social media are based on contact networks; some are based on geotagged mentions (*e.g.*, of influenza) in social media (Corley, Cook, Mikler, & Singh, 2009) and hybrid mining of free text and network structure for the same (Corley, Cook, Mikler, & Singh, 2010). Some epidemiological models admit mitigation strategies (Hsu, Roy Chowdhury, & Scoglio, 2011), while others are purely predictive and provide decision support or visual analytics capabilities.

The state of the field in predictive analytics for consumer health focuses on usability and accessibility of social media (Goldberg, et al., 2011), the mining of electronic health records (Silow-Carroll, Edwards, & Rodin, 2012), accounting for seasonal trends in abnormal event detection (Chae, et al., 2012), and spatiotemporal anomaly detection (Thom, Bosch, Koch, Worner, & Ertl, 2012).

4.3.2 Fraud Detection, Anomaly Detection, and Prediction of Risks

Anomaly detection is another major risk management area where social networks and prediction can often be of use. High-visibility applications include fraud detection, especially in internet banking (Aggelis, 2006), econometrics and market analytics (May, Levin, & Sugihara, 2008), supply chain integration (Cruz, Nagurney, & Wakolbinger, 2006), predictive analytics for marketing (Hair Jr, , 2007), decision management (Taylor, 2011), and automated power systems or "smart grid" systems (Li, Fang, Mahatma, & Hampapur, 2011).

4.4 Big Data Analytics

The term *big data analytics* currently refers to terascale to petascale computation applied to generate novel, valid, useful (actionable), and human-comprehensible models from data. As of 2014, "big data" usually connotes datasets that are multiple terabytes to petabytes in size or of a commensurately high complexity in terms of the complexity of analysis. Many issues facing analysts who work with big data, such as distributed backup, high bandwidth consumption for data federation and warehousing, and reliability (Russom, 2011) are beyond the scope of this article. Russom refers to the size as *volume* and bandwidth as *velocity* in reference to the "three Vs: volume, velocity, and variety". The issues of variety (semantic heterogeneity, especially a mixture of unstructured and structured data), on the other hand, have been highly relevant to this article, and figure prominently in the list of new challenges presented by terascale to petascale data sets, new metadata standards, and heterogeneous data sources. (Davenport, Barth, & Bean, 2012)

4.4.1 Business Intelligence (BI) and Decision Support

Business intelligence (BI), the application of data sciences to decision making, planning, and automation in domains such as management, marketing, and e-commerce, is another critical and very commonlyencountered application of predictive analytics, especially in organizational management (Horn, 2005). Recent work on monetizing BI has focused on using visual analytics and web analytics in an enterprise-wide context (Chen, Chiang, & Storey, 2012), on leveraging the results of data mining, and on decision support (Zaraté, 2012).

4.4.2 Customer Relationship Management (CRM)

Customer relationship management (CRM) refers to the use of technology to manage and support a company's interaction with its current and potential customers. It is facilitated directly by means of customer analytics (Stodder, 2012). In particular, recruitment, retention, and recommendation are supported by social media analytics, especially predictive analytics applied to customer feedback (USA Patent No. US 20130282430, 2013) and collective sentiment analysis (Hu, Wang, & Kambhampati, 2013).

4.4.3 Scientific Applications

While social networks are used in sociological and anthropological domains to model human behavior – a very narrow and mostly commercial aspect of which is discussed in this article – there are numerous applications of quasi-social networks to scientific domains. These include general analysis of covariance over events using autocorrelation matrices (Krackhardt, 1988), text mining of scientific literature (Bollacker, Lawrence, & Giles, 2000), and annotation of learning conversations using argumentation theory, discourse analytics based on text and social graphs (De Liddo, Shum, Quinto, Bachler, & Cannavacciuolo, 2011).

4.5 Computational Information and Knowledge Management (CIKM)

Computational information and knowledge management (CIKM) refers to data sciences applied to the development, collection, maintenance, annotation, and distribution of information resources such as document collections, data sets, audiovisual recordings, multimedia, and their associated metadata. One way in which CIKM is augmented by the use of predictive analytics is through user analytics for personalized and user-centered resource management, such as in the domains of education (Zhang, et al., 2010), public health and security informatics (Boulos, Sanfilippo, Corley, & Wheeler, 2010), information systems research (Shmueli & Koppius, 2011), and to categorize and present human resource (HR) data such as performance assessment data (USA Patent No. US 20130246339, 2012).

5. Systems

Fielded systems for social network-augmented decision support technology, including recommender systems and analytics systems for policymakers, are beginning to make increasing use of prediction techniques. This section surveys current and future research, fielded applications, and systems engineering competitions for three main purposes: for discovery informatics, for item-to-person and B2C (business-to-consumer *aka* targeted marketing) recommendation, and for citation and follow prediction, which is also a form of person-to-person (P2P) recommendation.

5.1 Interaction Discovery

Interaction discovery is a form of *discovery informatics* – that is, analyzing observed data to discover patterns and relationships that are interesting according to some specifiable criterion. It is frequently equated with discovery of interpersonal relationships in social networks, but is not limited to domains of human interaction.

For example, protein-protein interaction (PPI) is an important modeling problem in bioinformatics, specifically, proteomics (Rual, et al., 2005). PPI networks are quasi-social in the sense of Section 4.4.3, and as such have been represented using graphical models of probability such as Bayesian networks (Jansen, et al., 2003). More generally, some subfields of systems biology are centered around interacting members of some family of large molecules or molecular sequences (nucleotide sequences in the case of genomics, proteins in the case of proteomics, cellular lipids in the case of lipidomics, metabolic products and enzymes in the case of metabolomics); sets of mutually interacting molecules across one of these categories, and especially across multiple categories, are referred to as *interactomes* and present a general quasi-social network modeling and prediction problem. (Soler-López, Zanzoni, Lluís, Stelzl, & Aloy, 2011) Domains to which this type of predictive analytics is applicable include drug discovery. (Pujol, Mosca, Farrés, & Aloy, 2010)

Outside of systems biology but within the realm of computational sociology, classification of human interactions based on smartphone proximity data (Do & Gatica-Perez, 2011), localized prediction of human mobile phone-based interactions (Do & Gatica-Perez, 2013), and classifying social relationships by applicable life-facet based on smartphone data (Min, Wiese, Hong, & Zimmerman, 2013) also constitute interaction discovery applications.

5.2 Entity-to-Entity Recommendation

Entity-to-entity recommendation generally falls under the rubric of item-to-person, or business-to-consumer (B2C), or person-to-person (P2P). Most recommender systems that are reported as being for e-commerce (Driskill & Riedl, 1999) fall under the rubric of item-to-person or B2C. As such, they are subject to criteria such as trust and reputation that are commonly ascribed to vendor brands or the companies they represent, or less frequently to marketing agencies. Wang and Vassileva (2007) survey trust and reputation systems such as *Epinions*, *eBay*, *Amazon*, and *Google* for XML-based web services in a service-oriented architecture (SOA).

P2P recommendations include social network follow recommendation and friendship and dating sites (Pizzato, et al., 2013). In the next section, we focus on follow prediction and recommendation (a frequent side effect) as a special case of P2P prediction.

5.3 KDD Cup Competitions

5.3.1 KDD Cup 2003: Citation Prediction in arXiv

Section 2.2 introduces the concept of link mining using citation graphs and indexing data, an approach first popularized by systems such as *CiteSeer* (Giles, Bollacker, & Lawrence, 1998). Since the debut of the "web of citation" and the advent of methods for constructing and analyzing it, data mining competitions have been created to encourage the advancement of these methods and evaluate their effectiveness. One of the most prominent of these was the *KDD Cup 2003* competition (Gehrke, Ginsparg, & Kleinberg, 2003), which challenged competitors to predict citations between articles in the reprint repository *arXiv*.

One of the top-ranked systems (Manjunatha, Sivaramakrishnan, Pandey, & Murthy, 2003) processed the timestamped citation data to form a time series and used *k*-nearest neighbor regression-based classification to predict the existence of a candidate citation in a paper. The authors reported comparable results using *k*-nearest neighbor (an instance-based learning method) and support vector machines. A decade later, meta-path selection methods such as the type described by Sun *et al.* (2013), and for which an example is given in Section 3.3, were shown to outperform classification-based approaches. (Yu, Gu, Zhou, & Han, 2012)

5.3.2 KDD Cup 2012: Follow Prediction in Tencent Weibo

One of the more recent P2P prediction tasks is *follow prediction*, the problem of identifying, for a celebrity or other user of interest in a social network, whether a candidate follower will in fact follow them. This is related to the problem of developing a recommender system for suggesting people to follow, but such a system is just one possible application; these tasks are not identical. The acceptance rate of suggestions is typically low (less than 10%). (Wu, Sorathia, & Prasanna, 2012)

The top-performing system (Chen, et al., 2012) in Track 1 of the *KDD Cup 2012* competition, which was organized around the Chinese microblogging service *Tencent Weibo*, used a hybrid approach combining matrix factorization with an additive forest. In related work, the authors relate a technique for applying a multifaceted version of the factorization model part to a general recommender system task (Chen, et al., 2012).

Another system that achieves good performance on this task (Wu, Sorathia, & Prasanna, 2012) uses a five-term scoring function for ranking:

1. recommended item category
2. recommended item popularity
3. followee acceptance (how many people one follows who in turn follow the recommended item or person)
4. semantic keyword matching score
5. a normalized additive bonus for information sharing events

References

Abbasi, M.-A., Chai, S.-K., Liu, H., & Sagoo, K. (2012). Real-World Behavior Analysis through a Social Media Lens. In S. J. Yang, A. M. Greenberg, & M. Endsley (Ed.), *Proceedings of the 5th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction (SBP 2012) - College Park, MD, USA, April 3-5, 2012. Lecture Notes in Computer Science (LNCS) 7227*, pp. 18-26. New York, NY, USA: Springer. doi:10.1007/978-3-642-29047-3_3

Abel, F., Gao, Q., Houben, G.-J., & Tao, K. (2013). Twitter-Based User Modeling for News Recommendations. In F. Rossi (Ed.), *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013) - Beijing, China, August 3-9, 2013.* Menlo Park, CA, USA: AAAI Press / International Joint Conferences on Artificial Intelligence.

Abrams, J. H. (2003). *USA Patent No. US 7069308.* Retrieved March 9, 2014, from http://www.google.com/patents/US7069308

Acar, E., Dunlavy, D. M., & Kolda, T. G. (2009). Link Prediction on Evolving Data Using Matrix and Tensor Factorizations. In Y. Saygin, J. X. Yu, H. Kargupta, W. Wang, S. Ranka, P. S. Yu, & X. Wu (Ed.), *Workshops Proceedings of the 9th IEEE International Conference on Data Mining (ICDM 2009)* (pp. 262-269). Los Alamitos, CA, USA: IEEE Computer Society.

Aggarwal, C. C. (2011). An Introduction to Social Network Data Analytics. In C. C. Aggarwal (Ed.), *Social Network Data Analytics* (pp. 1-15). New York, NY, USA: Springer.

Aggelis, V. (2006). Offline Internet banking fraud detection. *Proceedings of the 1st International Conference on Availability, Reliability and Security (ARES 2006): The International Dependability Conference, Bridging Theory and Practice - Vienna, Austria, April 20-22, 2006* (pp. 904-905). Los Alamitos, CA, USA: IEEE Computer Society. Retrieved from 10.1109/ARES.2006.89

Ahn, Y.-Y., Han, S., Kwak, H., Moon, S. B., & Jeong, H. (2007). Analysis of topological characteristics of huge online social networking services. In C. L. Williamson, M. E. Zurko, P. F. Patel-Schneider, & P. J. Shenoy (Ed.), *Proceedings of the 16th International World Wide Web Conference (WWW 2007) - Banff, AB, Canada, May 8-12, 2007* (pp. 835-844). New York, NY, USA: ACM Press. doi:10.1145/1242572.1242685

Al Hasan, M., & Zaki, M. J. (2011). A Survey of Link Prediction in Social Networks. In C. C. Aggarwal (Ed.), *Social Network Data Analytics* (pp. 243-275). New York, NY, USA: Springer.

Al Hasan, M., Ahmed, N., & Neville, J. (2013). Network Sampling: Methods and Applications - KDD 2013 Tutorial. In I. S. Dhillon, Y. Koren, R. Ghani, T. E. Senator, P. Bradley, R. Parekh, . . . R. Uthurusamy (Ed.), *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2013) - Chicago, IL, USA, August 11-14, 2013* (p. 1528). New York, NY, USA: ACM Press. doi:10.1145/2487575.2506180

Al Hasan, M., Chaoji, V., Salem, S., & Zaki, M. J. (2006). Link Prediction using Supervised Learning. In C. C. Aggarwal, *Social Network Data Analytics* (pp. 243-275). New York, NY, USA: Springer.

Amershi, S., Fogarty, J., & Weld, D. S. (2012). Regroup: interactive machine learning for on-demand group creation in social networks. In J. A. Konstan, E. H. Chi, & K. Höök (Ed.), *Proceedings of the 2012 ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2012) - Austin, TX, USA, May 5-10, 2012* (pp. 21-30). New York, NY, USA: ACM Press. doi:10.1145/2207676.2207680

Angelova, R., & Weikum, G. (2006). Graph-based text classification: learn from your neighbors. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2006) - Seattle, Washington, USA, August 6-11, 2006* (pp. 485-492). New York, NY, USA: ACM Press.

Arteaga, S. M., Kudeki, M., & Woodworth, A. (2009). Combating obesity trends in teenagers through persuasive mobile technology. *ACM SIGACCESS Accessibility and Computing* (94), 17-25. doi:10.1145/1595061.1595064

Artzi, Y., Pantel, P., & Gamon, M. (2012). Predicting Responses to Microblog Posts. In E. Fosler-Lussier, E. Riloff, & S. Bangalore (Ed.), *Proceedings of the 2012 Human Language Technologies Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL 2012) - Montréal, Canada, June 3-8, 2012* (pp. 602-606). Pittsburgh, PA, USA: NAACL.

Aula, P. (2010). Social media, reputation risk and ambient publicity management. *Strategy & Leadership, 38*(6), 43-49. doi:10.1108/10878571011088069

Backstrom, L., & Leskovec, J. (2011). Supervised random walks: predicting and recommending links in social networks. In I. King, W. Nejdl, & H. Li (Ed.), *Proceedings of the 4th International Conference on Web Search and Web Data Mining (WSDM 2011)* (pp. 635-644). New York, NY, USA: ACM Press. doi:10.1145/1935826.1935914

Backstrom, L., Huttenlocher, D. P., Kleinberg, J. M., & Lan, X. (2006). Group formation in large social networks: membership, growth, and evolution. In T. Eliassi-Rad, L. H. Ungar, M. Craven, & D. Gunopulos (Ed.), *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006)* (pp. 44-54). New York, NY, USA: ACM Press.

Baltrunas, L., & Amatriain, X. (2009). Towards Time-Dependant Recommendation based on Implicit Feedback. In G. Adomavicius, & F. Ricci (Ed.), *Proceedings of the Workshop on Context-Aware Recommender Systems (CARS 2009), 3rd ACM Conference on Recommender Systems (RecSys 2009) - New York, NY, USA, October 23-25, 2009.* New York, NY, USA: ACM Press. Retrieved from http://ids.csom.umn.edu/faculty/gedas/cars2009/

Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications, 31*(3-4), 590-6114.

Barbieri, D., Braga, D., Ceri, S., Valle, E. D., Huang, Y., Tresp, V., . . . Wermser, H. (2010). Deductive and Inductive Stream Reasoning for Semantic Social Media Analytics. *IEEE Intelligent Systems, 25*(6), 32-41. doi:10.1109/MIS.2010.142

Becker, H., Naaman, M., & Gravano, L. (2010). Learning similarity metrics for event identification in social media. In B. D. Davison, T. Suel, N. Craswell, & B. Liu (Ed.), *Proceedings of the 3rd International Conference on Web Search and Web Data Mining (WSDM 2010) - New York, NY, USA, February 4-6, 2010* (pp. 291-300). New York, NY, USA: ACM Press. doi:10.1145/1718487.1718524

Berendt, B., Hotho, A., Mladenič, D., Someren, M. v., Spiliopoulou, M., & Stumme, G. (2003). A Roadmap for Web Mining: From Web to Semantic Web. In B. Berendt, A. Hotho, D. Mladenič, M. v. Someren, M. Spiliopoulou, & G. Stumme (Ed.), *Proceedings of the 1st European Web Mining Forum (EWMF 2003) - Cavtat-Dubrovnik, Croatia, September 22, 2003. Lecture Notes in Computer Science (LNCS) 3209*, pp. 1-22. New York, NY, USA: Springer.

Berkman, L. F., & Syme, S. L. (1978). Social networks, host resistance, and mortality: a nine-year follow-up study of. *American Journal of Epidemiology, 109*(2), 186-204.

Berkovsky, S., Freyne, J., & Smith, G. (2012). Personalized Network Updates: Increasing Social Interactions and Contributions in Social Networks. In J. Masthoff, B. Mobasher, M. C. Desmarais, & R. Nkambou (Ed.), *Proceedings of the 20th International Conference on User Modeling, Adaption, and Personalization (UMAP 2012) - Montreal, Canada, July 16-20, 2012. Lecture Notes in Computer Science (LNCS) 7379*, pp. 1-13. New York, NY, USA: Springer. doi:10.1007/978-3-642-31454-4_1

Berkovsky, S., Freyne, J., Kimani, S., & Smith, G. (2011). Selecting Items of Relevance in Social Network Feeds. In J. A. Konstan, R. Conejo, J.-L. Marzo, & N. Oliver (Ed.), *Proceedings of the 19th International Conference on User Modeling, Adaption, and Personalization (UMAP 2011) - Girona, Spain, July 11-15, 2011. Lecture Notes in Computer Science (LNCS) 6787*, pp. 329-334. New York, NY, USA: Springer.

Bilgic, M., Mihalkova, L., & Getoor, L. (2010). Active Learning for Networked Data. In J. Fürnkranz, & T. Joachims (Ed.), *Proceedings of the 27th International Conference on Machine Learning (ICML 2010) - Haifa, Israel, June 21-24, 2010* (pp. 79-86). Madison, WI, USA: Omnipress.

Bollacker, K. D., Lawrence, S., & Giles, C. L. (1998). CiteSeer: An Autonous Web Agent for Automatic Retrieval and Identification of Interesting Publications. In K. P. Sycara, & M. Wooldridge (Ed.), *Proceedings of the 2nd International Conference on Autonomous Agents (Agents 1998)* (pp. 116-123). New York, NY, USA: ACM Press.

Bollacker, K. D., Lawrence, S., & Giles, C. L. (2000). Discovering relevant scientific literature on the Web. *IEEE Intelligent Systems and Their Applications, 15*(2), 42-47. doi:10.1109/5254.850826

Boman, M. (2011). On Understanding Catastrophe: The case of highly severe influenza-like illness (Position Paper). In W. H. Hsu, S. Kallumadi, & J. Han (Ed.), *Working notes of Workshop W-16, 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011): Heterogeneous Information Network Analysis (HINA 2011) - Barcelona, Spain, July 16, 2011.* Retrieved March 9, 2014, from http://www.kddresearch.org/KDD/Workshops/IJCAI-2011-HINA/Papers/

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science, 323*(5916), 892-895. doi:10.1126/science.1165821

Boulos, M. N., Sanfilippo, A. P., Corley, C. D., & Wheeler, S. (2010). Social Web mining and exploitation for serious applications: Technosocial Predictive Analytics and related technologies for public health, environmental and national security surveillance. *Computer Methods and Programs in Biomedicine, 100*(1), 16–23. doi:10.1016/j.cmpb.2010.02.007

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control* (4th ed.). Hoboken, NJ, USA: John Wiley & Sons.

boyd, d. m., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210-230. doi:10.1111/j.1083-6101.2007.00393.x

Brown, C., Nicosia, V., Scellato, S., Noulas, A., & Mascolo, C. (2012). Where Online Friends Meet: Social Communities in Location-Based Networks. In J. G. Breslin, N. B. Ellison, J. G. Shanahan, & Z. Tufekci (Ed.), *Proceedings of the 6th International Conference on Weblogs and Social Media (ICWSM 2012).* Menlo Park, CA, USA: AAAI Press.

Cacheda, F., Carneiro, V., Fernández, D., & Formoso, V. (2011). Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Transactions on the Web (TWEB), 5*(1), 2:1-2:33. doi:10.1145/1921591.1921593

Cao, B., Liu, N. N., & Yang, Q. (2010). Transfer Learning for Collective Link Prediction in Multiple Heterogenous Domains. In J. Fürnkranz, & T. Joachims (Ed.), *Proceedings of the 27th International Conference on Machine Learning (ICML 2010) - Haifa, Israel, June 21-24, 2010* (pp. 159-166). Madison, WI, USA: Omnipress.

Cao, M. D., & Gao, X. (2005). Combining Contents and Citations for Scientific Document Classification. In S. Zhang, & R. Jarvis (Ed.), *Advances in Artificial Intelligence: Proceedings of the 18th Australian Joint Conference on Artificial Intelligence (AI 2005), Sydney, Australia, December 5-9, 2005. Lecture Notes in Computer Science (LNCS) 3809*, pp. 143-152. New York, NY, USA: Springer.

Caragea, D., Bahirwani, V., Aljandal, W., & Hsu, W. H. (2009). Ontology-based link prediction in the LiveJournal social network. In V. Bulitko, & J. C. Beck (Ed.), *Proceedings of the 8th Symposium on Abstraction, Reformulation and Approximation (SARA 2009).* Lake Arrowhead, CA, USA.

Catanese, S., De Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. (2011). Crawling Facebook for social network analysis purposes. In R. Akerkar (Ed.), *Proceedings of the 1st International Conference on Web Intelligence, Mining and Semantics (WIMS 2011) - Sogndal, Norway, May 25-27, 2011* (p. 52). New York, NY, USA: ACM Press. doi:10.1145/1988688.1988749

Chae, J., Thom, D., Bosch, H., Jang, Y., Maciejewski, R., Ebert, D., & Ertl, T. (2012). Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. *Proceedings of the 7th IEEE Conference on Visual Analytics Science and Technology (VAST 2012) - Seattle, WA, USA, October 14-19, 2012* (pp. 143-152). Los Alamitos, CA, USA: IEEE Computer Society. doi:10.1109/VAST.2012.6400557

Chang, J., Boyd-Graber, J., & Blei, D. M. (2009). Connections between the lines: augmenting social networks with text. In J. F. Elder IV, F. Fogelman-Soulié, P. A. Flach, & M. J. Zaki (Ed.), *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009) - Paris, France, June 28 - July 1, 2009* (pp. 169-178). New York, NY, USA: ACM Press. doi:10.1145/1557019.1557044

Chatfield, C. (2004). *The Analysis of Time Series: An Introduction* (6th ed.). Boca Raton, FL, USA: CRC Press.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly, 36*(4), 1165-1188.

Chen, T., Tang, L., Liu, Q., Yang, D., Xie, S., Cao, X., & Wu, C. (2012). Combining Factorization Model and Additive Forest for Collaborative Followee Recommendation. *KDD Cup 2012.* New York, NY, USA: ACM Press.

Chen, Y., Liu, Z., Ji, D., Xin, Y., Wang, W., Yao, L., & Zou, Y. (2012). Context-aware Ensemble of Multifaceted Factorization Models for Recommendation Prediction in Social Networks. *Proceedings of the KDD Cup 2012 Workshop, 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2012) - Beijing, China, August 12-16, 2012.* San Francisco, USA: Kaggle.

Christley, R. M., Pinchbeck, G. L., Bowers, R. G., Clancy, D., French, N. P., Bennett, R., & Turner, J. (2005). Infection in Social Networks: Using Network Analysis to Identify High-Risk Individuals. *American Journal of Epidemiology, 162*(10), 1024-1031. doi:10.1093/aje/kwi308

Clauset, A., Moore, C., & Newman, M. E. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature, 453*, 98-101. doi:10.1038/nature06830

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-Law Distributions in Empirical Data. *SIAM Review, 51*(4), 661-703. doi:10.1137/070710111

Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: information and pattern discovery on the World Wide Web. *Proceedings of the 9th International Conference on Tools with Artificial*

Intelligence (ICTAI 1997) - Newport Beach, CA, USA, November 3-8, 1997 (pp. 558-567). Los Alamitos, CA, USA: IEEE Computer Society. doi:10.1109/TAI.1997.632303

Corley, C. D., Cook, D. J., Mikler, A. R., & Singh, K. P. (2009). Monitoring Influenza Trends through Mining Social Media. In H. R. Arabnia, & M. Q. Yang (Ed.), *Proceedings of the 4th International Conference on Bioinformatics & Computational Biology (BIOCOMP 2009) - Las Vegas Nevada, USA, July 13-16, 2009* (pp. 340-346). Atlanta, GA, USA: CSREA Press.

Corley, C. D., Cook, D. J., Mikler, A. R., & Singh, K. P. (2010). Text and Structural Data Mining of Influenza Mentions in Web and Social Media. *International Journal of Environmental Research and Public Health, 7*(2), 596-615. doi:10.3390/ijerph7020596

Cruz, J. M., Nagurney, A., & Wakolbinger, T. (2006). Financial engineering of the integration of global supply chain networks and social networks with risk management. *Naval Research Logistics, 53*(7), 674-696. doi:10.1002/nav.20179

Culotta, A., Bekkerman, R., & McCallum, A. (2004). Extracting social networks and contact information from email and the Web. *Proceedings of the 1st Conference on Email and Anti-Spam (CEAS 2004) - Mountain View, CA, USA, July 30-31, 2004.* Retrieved March 9, 2014, from http://scholarworks.umass.edu/cs_faculty_pubs/33/

Danowski, J. A., & Cepela, N. (2010). Automatic Mapping of Social Networks of Actors from Text Corpora: Time Series Analysis. In N. Memon, J. J. Xu, D. L. Hicks, & H. Chen, *Data Mining for Social Network Data. Annals of Information Systems Volume 12* (pp. 31-46). New York, NY, USA: Springer.

Davenport, T. H., & Patil, D. J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review* (October 2012), pp. 70-76.

Davenport, T. H., Barth, P., & Bean, R. (2012). How `Big Data' is Different. *MIT Sloan Management Review, 54*(1), 22-24.

Davitz, J., Yu, J., Basu, S., Gutelius, D., & Harris, A. (2007). iLink: search and routing in social networks. In P. Berkhin, R. Caruana, & X. Wu (Ed.), *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2007) - San Jose, CA, USA, August 12-15, 2007* (pp. 931-940). New York, NY, USA: ACM Press. doi:10.1145/1281192.1281292

De Liddo, A., Shum, S. B., Quinto, I., Bachler, M., & Cannavacciuolo, L. (2011). Discourse-centric learning analytics. In P. Long, G. Siemens, G. Conole, & D. Gasevic (Ed.), *Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK 2011) - Banff, AB, Canada, February 27 - March 1, 2011* (pp. 23-33). New York, NY, USA: ACM Press. doi:10.1145/2090116.2090120

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science and Technology, 41*(6), 391-407. doi:10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Deng, H., Han, J., Zhao, B., Yu, Y., & Lin, C. X. (2011). : Probabilistic topic models with biased propagation on heterogeneous information networks. In C. Apté, J. Ghosh, & P. Smyth (Ed.), *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2011)* (pp. 1271-1279). New York, NY, USA: ACM Press. doi:10.1145/2020408.2020600

Dietterich, T. G., Domingos, P., Getoor, L., Muggleton, S., & Tadepalli, P. (2008). Structured machine learning: the next ten years. (H. Blockeel, & J. Shavlik, Eds.) *Machine Learning, 73*(1), 3-23.

Do, T. M., & Gatica-Perez, D. (2011). GroupUs: Smartphone Proximity Data and Human Interaction Type Mining. *Proceedings of the 15th Annual International Symposium on Wearable Computers (ISWC 2011) - San Francisco, CA, USA, June 12-15, 2011* (pp. 21-28). Piscataway, NJ. USA: IEEE Press. doi:10.1109/ISWC.2011.28

Do, T. M., & Gatica-Perez, D. (2013). Human interaction discovery in smartphone proximity networks. *Personal and Ubiquitous Computing, 13*(7), 413-431. doi:10.1007/s00779-011-0489-7

Driskill, R., & Riedl, J. (1999). Recommender Systems for E-Commerce: Challenges and Opportunities. *Working Notes of the Workshop on Artificial Intelligence for Electronic Commerce (AAAI Technical Report WS-99-01)* (pp. 73-76). Menlo Park, CA, USA: AAAI Press.

Du, N., Wu, B., Xu, L., Wang, B., & Pei, X. (2006). A Parallel Algorithm for Enumerating All Maximal Cliques in Complex Network. *Workshops Proceedings of the 6th IEEE International Conference on Data Mining (ICDM 2006)* (pp. 320-324). Los Alamitos, CA, USA: IEEE Computer Society.

Erdős, P., & Rényi, A. (1960). On the Evolution of Random Graphs. *Magyar Tudományos Akadémia Matematikai Kutató Intézete Közlemények (Bulletins of the Mathematical Research Institute of the Hungarian Academy of Sciences), 5*, 17-61.

Ferguson, N. M., Donnelly, C. A., & Anderson, R. M. (2001). The Foot-and-Mouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions. *Science, 292*(5519), 1155-1160 . doi:10.1126/science.1061020

Fire, M., Tenenboim, L., Lesser, O., Puzis, R., Rokach, L., & Elovici, Y. (2011). Link Prediction in Social Networks Using Computationally Efficient Topological Features. *Proceedings of the 2011 IEEE International Conference on Privacy, Security, Risk and Trust and 3rd IEEE International Conference on Social Computing (PASSAT/SocialCom 2011) - Boston, MA, USA, October 9-11, 2011* (pp. 73-80). Piscataway, NJ, USA: IEEE Press. doi:10.1109/PASSAT/SocialCom.2011.20

Gao, J. (2012). *Exploring the power of heterogeneous information sources (Doctoral dissertation).* Computer Science. Urbana, IL, USA: University of Illinois at Urbana-Champaign. Retrieved March 9, 2014, from http://hdl.handle.net/2142/29822

Gehrke, J., Ginsparg, P., & Kleinberg, J. (2003). Overview of the 2003 KDD Cup. *ACM SIGKDD Explorations Newsletter, 5*(2), 149-151. doi:10.1145/980972.980992

Gershenfeld, N. A., & Weigend, A. S. (1994). The Future of Time Series. In N. A. Gershenfeld, & A. S. Weigend (Eds.), *Time Series Prediction: Forecasting the Future and Understanding the Past (Santa Fe Institute Series, Book 15)* (pp. 1-70). Boulder, CO, USA: Westview Press.

Getoor, L. (2003). Link Mining: A New Data Mining Challenge. *ACM SIGKDD Explorations Newsletter, 5*(1), 84-89. doi:10.1145/959242.959253

Getoor, L., & Diehl, C. P. (2005). Link Mining: A Survey. *ACM SIGKDD Explorations Newsletter, 7*(2), 3-12.

Getoor, L., Rhee, J. T., Koller, D., & Small, P. (2004). Understanding tuberculosis epidemiology using structured statistical models. *Artificial Intelligence in Medicine, 30*(3), 233-256. doi:10.1016/j.artmed.2003.11.003

Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998). CiteSeer: An Automatic Citation Indexing System. *Proceedings of the 3rd ACM International Conference on Digital Libraries (DL 1998)* (pp. 89-98). New York, NY, USA: ACM Press.

Glance, N. S., Hurst, M., Nigam, K., Siegler, M., Stockton, R., & Tomokiyo, T. (2005). Deriving marketing intelligence from online discussion. In R. Grossman, R. J. Bayardo, & K. P. Bennett (Ed.), *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2005) - Chicago, IL, USA, August 21-24, 2005* (pp. 419-428). New York, NY, USA: ACM Press. doi:10.1145/1081870.1081919

Goldberg, L., Lide, B., Lowry, S., Massett, H. A., O'Connell, T., Preece, J., . . . Shneiderman, B. (2011). Usability and Accessibility in Consumer Health Informatics: Current Trends and Future Challenges. *American Journal of Preventive Medicine - Supplement on Cyberinfrastructure for Consumer Health, 40*(5), S187–S197. doi:10.1016/j.amepre.2011.01.009

Goyal, A., Bonchi, F., & Lakshmanan, L. V. (2010). Learning influence probabilities in social networks. In B. D. Davison, T. Suel, N. Craswell, & B. Liu (Ed.), *Proceedings of the 3rd International Conference on Web Search and Web Data Mining (WSDM 2010) - New York, NY, USA, February 4-6, 2010* (pp. 241-250). New York, NY, USA: ACM Press. doi:10.1145/1718487.1718518

Gregersen, H., & Sailer, L. (1993). Chaos Theory and Its Implications for Social Science Research. *Human Relations, 46*(7), 777-802. doi:10.1177/001872679304600701

Guerra, P. H., Veloso, A., Meira Jr., W., & Almeida, V. (2011). From bias to opinion: a transfer-learning approach to real-time sentiment analysis. In C. Apté, J. Ghosh, & P. Smyth (Ed.), *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2011)* (pp. 150-158). New York, NY, USA: ACM Press. doi:10.1145/2020408.2020438

Gupta, M., Gao, J., Sun, Y., & Han, J. (2012). Integrating meta-path selection with user-guided object clustering in heterogeneous information networks. In Q. Yang, D. Agarwal, & J. Pei (Ed.), *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2012) - Beijing, China, August 12-16, 2012* (pp. 859-867). New York, NY, USA: ACM Press. doi:10.1145/2339530.2339738

Hair Jr, , J. F. (2007). Knowledge creation in marketing: the role of predictive analytics. *European Business Review, 9*(4), 303-315. doi:10.1108/09555340710760134

Han, J. (2009). Mining Heterogeneous Information Networks by Exploring the Power of Links. In J. Gama, V. S. Costa, A. M. Jorge, & P. Brazdil (Ed.), *Proceedings of the 12th International Conference on Discovery Science (DS 2009) - Porto, Portugal, October 3-5, 2009. Lecture Notes in Computer Science (LNCS) 5808*, pp. 13-30. New York, NY, USA: Springer.

He, Y., Lin, C., Gao, W., & Wong, K.-F. (2013). Dynamic joint sentiment-topic model. *ACM Transactions on Intelligent Systems and Technology (TIST) - Special Section on Intelligent Mobile Knowledge Discovery and Management Systems and Special Issue on Social Web Mining, 5*(1), 6:1-6:21. doi:10.1145/2542182.2542188

Henderson, K., Gallagher, B., Eliassi-Rad, T., Tong, H., Basu, S., & Leman Akoglu, D. K. (2012). RolX: Structural Role Extraction & Mining in Large Graphs. In Q. Yang, D. Agarwal, & J. Pei (Ed.), *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2012) - Beijing, China, August 12-16, 2012* (pp. 1231-1239). New York, NY, USA: ACM Press.

Horn, J. K. (2005). Parameters for sustained orderly growth in learning organizations. In K. Richardson, *Managing Organizational Complexity: Philosophy, Theory and Application* (pp. 473-492). Charlotte, NC, USA: Information Age Publishing.

Hsu, W. H., King, A. L., Paradesi, M. S., Pydimarri, T., & Weninger, T. (2006). Collaborative and Structural Recommendation of Friends using Weblog-based Social Network Analysis. In N. Nicolov, F. Salvetti, M. Liberman, & J. H. Martin (Ed.), *Computational Approaches to Analyzing Weblogs, Papers from the 2006 AAAI Spring Symposium, Technical Report SS-06-03 - Stanford, CA, USA, March 27-29, 2006* (pp. 55-60). Menlo Park, CA, USA: AAAI Press.

Hsu, W. H., Koduru, P., & Zhai, C. (2013). Heterogeneous Information Networks for Text-Based Link Mining: A Position Paper on Visualization and Structure Learning Methods. In W. H. Hsu, S. Kallumadi, & Y. Sun (Ed.), *Working notes of the 2nd International Workshop on Heterogeneous Information Network Analysis (HINA 2013). Workshop W36, 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013) - Beijing, China, August 5, 2013.*

Hsu, W. H., Lancaster, J. P., Paradesi, M. S., & Weninger, T. (2007). Structural Link Analysis from User Profiles and Friends Networks: A Feature Construction Approach. In N. S. Glance, N. Nicolov, E.

Adar, M. Hurst, M. Liberman, & F. Salvetti (Ed.), *Proceedings of the 1st International Conference on Weblogs and Social Media (ICWSM 2007)*, (pp. 75-80). Boulder, CO, USA.

Hsu, W. H., Roy Chowdhury, S., & Scoglio, C. (2011). Mitigation Strategies for Foot and Mouth Disease: A Learning-Based Approach. *International Journal of Artificial Life Research, 2*(2), 42-76. doi:10.4018/jalr.2011040103

Hsu, W. H., Weninger, T., & Paradesi, M. S. (2008). Predicting links and link change in friends networks: supervised time series learning with imbalanced data. In C. H. Dagli, D. L. Enke, K. M. Bryden, H. Ceylan, & M. Gen (Ed.), *Intelligent Engineering Systems Through Artificial Neural Networks: Proceedings of the 18th International Conference on Artificial Neural Networks in Engineering (ANNIE 2008) - St. Louis, MO, USA, November 9-12, 2008. 18.* New York, NY, USA: ASME Press. doi:10.1115/1.802823.paper68

Hu, X., & Liu, H. (2012). Text Analytics in Social Media. In C. C. Aggarwal, & C. Zhai, *Mining Text Data* (pp. 385-414). New York, NY, USA: Springer.

Hu, Y., John, A., Wang, F., & Kambhampati, S. (2012). ET-LDA: Joint Topic Modeling for Aligning Events and Their Twitter Feedback. In J. Hoffmann, & B. Selman (Ed.), *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI 2012) - July 22-26, 2012, Toronto, Ontario, Canada.* Menlo Park, CA, USA: AAAI Press.

Hu, Y., Wang, F., & Kambhampati, S. (2013). Listening to the crowd: automated analysis of events via aggregated twitter sentiment. *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013) - Beijing, China, August 3-9, 2013.* Menlo Park, CA, USA: AAAI Press / International Joint Conferences on Artificial Intelligence.

Huang, Z. (2006). Link Prediction Based on Graph Topology: The Predictive Value of Generalized Clustering Coefficient. In M. Grobelnik, J. Adibi, N. Milic-Frayling, D. Mladenic, & P. Pantel (Ed.), *Working Notes of the Workshop on Link Analysis, 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (LinkKDD 2006) - Philadelphia, PA, USA, August 20–23, 2006.* New York, NY, USA: ACM Press. Retrieved March 9, 2014, from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.67.2613

Jamali, M., & Ester, M. (2010). A matrix factorization technique with trust propagation for recommendation in social networks. In X. Amatriain, M. Torrens, P. Resnick, & M. Zanker (Ed.), *Proceedings of the 2010 ACM Conference on Recommender Systems (RecSys 2010) - Barcelona, Spain, September 26-30, 2010* (pp. 135-142). New York, NY, USA: ACM Press. doi:10.1145/1864708.1864736

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S., . . . Gerstein, M. (2003). A Bayesian Networks Approach for Predicting Protein-Protein Interactions from Genomic Data. *Science, 302*(5644), 449-453. doi:10.1126/science.1087361

Ji, M., Han, J., & Danilevsky, M. (2011). Ranking-based classification of heterogeneous information networks. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2011)* (pp. 1298-1306). New York, NY, USA: ACM Press.

Ji, M., Sun, Y., Danilevsky, M., Han, J., & Gao, J. (2010). Graph Regularized Transductive Classification on Heterogeneous Information Networks. In J. L. Balcázar, F. Bonchi, A. Gionis, & M. Sebag (Ed.), *Machine Learning and Knowledge Discovery in Databases: European Conference (ECML PKDD 2010) - Barcelona, Spain, September 20-24, 2010, Proceedings, Part I* (pp. 570-586). New York, NY, USA: Springer.

Kannan, P. V., & Vijayaraghavan, R. (2013). *USA Patent No. US 20130282430.* Retrieved March 9, 2014, from www.google.com/patents/US20130282430

Kashima, H., Kato, T., Yamanishi, Y., Sugiyama, M., & Tsuda, K. (2009). Link Propagation: A Fast Semi-supervised Learning Algorithm for Link Prediction. *Proceedings of the 9th SIAM International Conference on Data Mining (SDM 2009) - Sparks, NV, USA, April 30 - May 2, 2009* (pp. 1100-1111). Philadelphia, PA, USA: SIAM Press.

Kendrena, K. R., Donchetz, J. G., & Istre, R. R. (2012). *USA Patent No. US 20130246339.* Retrieved March 9, 2014, from http://www.google.com/patents/US20130246339

Ketkar, N. S., Holder, L. B., & Cook, D. J. (2005). Comparison of graph-based and logic-based multi-relational data mining. *ACM SIGKDD Explorations Newsletter, 7*(2), 64-71.

King, I., Li, J., & Chan:, K. T. (2009). A brief survey of computational approaches in Social Computing. *Proceedings of the 2009 International Joint Conference on Neural Networks (IJCNN 2009) - Atlanta, Georgia, USA, 14-19 June 2009* (pp. 1625-1632). Piscataway, NJ, USA: IEEE Press.

Koenigstein, N., Dror, G., & Koren, Y. (2011). Yahoo! music recommendations: modeling music ratings with temporal dynamics and item taxonomy. In B. Mobasher, R. D. Burke, D. Jannach, & G. Adomavicius (Ed.), *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys 2011) - Chicago, IL, USA, October 23-27, 2011* (pp. 165-172). New York, NY, USA: ACM Press. doi:10.1145/2043932.2043964

Konstas, I., Stathopoulos, V., & Jose, J. M. (2009). On social networks and collaborative recommendation. In J. Allan, J. A. Aslam, M. Sanderson, C. Zhai, & J. Zobel (Ed.), *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2009) - Boston, MA, USA, July 19-23, 2009* (pp. 195-202). New York, NY, USA: ACM Press. doi:10.1145/1571941.1571977

Koren, Y. (2009). Collaborative Filtering with Temporal Dynamics. In J. F. Elder IV, F. Fogelman-Soulié, P. A. Flach, & M. J. Zaki (Ed.), *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009) - Paris, France, June 28 - July 1, 2009* (pp. 447-456). New York, NY, USA: ACM Press.

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks, 10*(4), 359–381. doi:10.1016/0378-8733(88)90004-4

Kunegis, J., & Lommatzsch, A. (2009). Learning Spectral Graph Transformations for Link Prediction. In A. P. Danyluk, L. Bottou, & M. L. Littman (Ed.), *Proceedings of the 26th Annual International Conference on Machine Learning (ICML 2009)* (p. 71). New York, NY, USA: ACM Press.

Lathia, N., Hailes, S., Capra, L., & Amatriain, X. (2010). Temporal diversity in recommender systems. In F. Crestani, S. Marchand-Maillet, H.-H. Chen, E. N. Efthimiadis, & J. Savoy (Ed.), *Proceedings of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2010) - Geneva, Switzerland, July 19-23, 2010* (pp. 210-217). New York, NY, USA: ACM Press. doi:10.1145/1835449.1835486

Lau, R. Y. (2013). A Probabilistic Generative Model for Mining Collaborative Cybercriminal Networks. In W. H. Hsu, S. Kallumadi, & Y. Sun (Ed.), *Working notes of the 2nd International Workshop on Heterogeneous Information Network Analysis (HINA 2013). Workshop W36, 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013) - Beijing, China, August 5, 2013.*

Lauw, H. W., Lim, E.-P., Pang, H., & Tan, T.-T. (2011). Social Network Discovery by Mining Spatio-Temporal Events. *Computational & Mathematical Organization Theory, 11*(2), 97-118.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review, 52*(2), 21-31.

Lawrence, S., Giles, C. L., & Bollacker, K. D. (1999). Digital Libraries and Autonomous Citation Indexing. *IEEE Computer, 32*(6), 67-71.

Leskovec, J., Huttenlocher, D. P., & Kleinberg, J. M. (2010). Predicting positive and negative links in online social networks. In M. Rappa, P. Jones, J. Freire, & S. Chakrabarti (Ed.), *Proceedings of the 19th International World Wide Web Conference (WWW 2010) - Raleigh, NC, USA, April 26-30, 2010* (pp. 641-650). New York, NY, USA: ACM Press. doi:10.1145/1772690.1772756

Levin, D. Z., & Cross, R. (2004). The Strength of Weak Ties You Can Trust: The Mediating Role of Trust in Effective Knowledge Transfer. *Management Science, 50*(11), 1477-1490. doi:10.1287/mnsc.1030.0136

Li, H., Fang, D., Mahatma, S., & Hampapur, A. (2011). Usage analysis for smart meter management. *Proceedings of the 8th International Conference & Expo on Emerging Technologies for a Smarter World (CEWIT 2011) - Long Island, NY, USA, November 2-3, 2011* (pp. 1-6). Stony Brook, NY, USA: CEWIT. doi:10.1109/CEWIT.2011.6135871

Liben-Nowell, D., & Kleinberg, J. (2007). The Link-Prediction Problem for Social Networks. *Journal of the American Society for Information Science and Technology, 58*(7), 1019-1031.

Liu, J., & Han, J. (2013). HINMF: A Matrix Factorization Method for Clustering in Heterogeneous Information Networks. In W. H. Hsu, S. Kallumadi, & Y. Sun (Ed.), *Working notes of the 2nd International Workshop on Heterogeneous Information Network Analysis (HINA 2013). Workshop W36, 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013) - Beijing, China, August 5, 2013.*

Ma, H., Zhou, D., Liu, C., Lyu, M. R., & King, I. (2011). Recommender systems with social regularization. In I. King, W. Nejdl, & H. Li (Ed.), *Proceedings of the 4th International Conference on Web Search and Web Data Mining (WSDM 2011)* (pp. 287-296). New York, NY, USA: ACM Press.

Ma, T., Yang, Y., Wang, L., & Yuan, B. (2014). Recommending People to Follow Using Asymmetric Factor Models with Social Graphs. In V. Snášel, P. Krömer, M. Köppen, & G. Schaefer (Eds.), *Soft Computing in Industrial Applications: Proceedings of the 17th Online World Conference on Soft Computing in Industrial Applications. Advances in Intelligent Systems and Computing 223* (pp. 265-276). New York, NY, USA: Springer.

Macskassy, S. A. (2009). Using graph-based metrics with empirical risk minimization to speed up active learning on networked data. In J. F. Elder IV, F. Fogelman-Soulié, P. A. Flach, & M. J. Zaki (Ed.), *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009) - Paris, France, June 28 - July 1, 2009* (pp. 597-606). New York, NY, USA: ACM Press. doi:10.1145/1557019.1557087

Manjunatha, J. N., Sivaramakrishnan, K. R., Pandey, R. K., & Murthy, M. N. (2003). Citation Prediction Using Time Series Approach: KDD Cup 2003 (Task 1). (pp. 152-153). New York, NY, USA: ACM Press.

Mathioudakis, M., & Koudas, N. (2010). TwitterMonitor: trend detection over the twitter stream. In A. K. Elmagarmid, & D. Agrawal (Ed.), *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 2010) - Indianapolis, IN, USA, June 6-10, 2010* (pp. 1155-1158). New York, NY, USA: ACM Press. doi:10.1145/1807167.1807306

Matsuo, Y., & Yamamoto, H. (2009). Community gravity: measuring bidirectional effects by trust and rating on online social networks. In J. Quemada, G. León, Y. S. Maarek, & W. Nejdl (Ed.), *Proceedings of the 18th International World Wide Web Conference (WWW 2009) - Madrid, Spain, April 20-24, 2009* (pp. 751-760). New York, NY, USA: ACM Press.

May, R. M., Levin, S. A., & Sugihara, G. (2008). Complex systems: Ecology for bankers. *Nature, 451*, 893-895. doi:10.1038/451893a

McCallum, A., Corrada-Emmanuel, A., & Wang, X. (2005). *The Author-Recipient-Topic Model for Topic and Role Discovery in Social Networks: Experiments with Enron and Academic Email (Computer Science Department Faculty Publication Series, Paper 44).* University of Massachusetts - Amherst, Department of Computer Science. Amherst, MA, USA: ScholarWorks@UMass Amherst. Retrieved March 9, 2014, from http://scholarworks.umass.edu/cs_faculty_pubs/44/

McCallum, A., Wang, X., & Corrada-Emmanuel, A. (2007). Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email. *Journal of Artificial Intelligence Research, 30*, 249-272.

McCue, C. (2005). Data Mining and Predictive Analytics: Battlespace Awareness for the War on Terrorism. *Defense Intelligence Journal, 13*(1-2), 47-63.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology, 27*, 415-444. doi:10.1146/annurev.soc.27.1.415

Meyers, L. A., Pourbohloul, B., Newman, M. E., Skowronski, D. M., & Brunham, R. C. (2005). Network theory and SARS: predicting outbreak diversity. *Journal of Theoretical Biology, 232*(1), 71-81. doi:10.1016/j.jtbi.2004.07.026

Min, J.-K., Wiese, J., Hong, J. I., & Zimmerman, J. (2013). Mining smartphone data to classify life-facets of social relationships. In A. Bruckman, S. Counts, C. Lampe, & L. G. Terveen (Ed.), *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work (CSCW 2013) - San Antonio, TX, USA, February 23-27, 2013* (pp. 285-294). New York, NY, USA: ACM press. doi:10.1145/2441776.2441810

Mladenic, D. (1999). Text-Learning and Related Intelligent Agents: A Survey. *IEEE Intelligent Systems & Their Applications, 14*(4), 44-54.

Mobasher, B., Burke, R., Bhaumik, R., & Sandvig, J. (2007). Attacks and Remedies in Collaborative Recommendation. *IEEE Intelligent Systems, 22*(3), 56-63. doi:10.1109/MIS.2007.45

Mobasher, B., Cooley, R., & Srivastava, J. (2000). Automatic personalization based on Web usage mining. *Communications of the ACM, 43*(8), 142-151. doi:10.1145/345124.345169

Murata, T. (2011). Detecting Communities from Social Tagging Networks Based on Tripartite Modularity. In W. H. Hsu, S. Kallumadi, & J. Han (Ed.), *Working notes of Workshop W-16, 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011): Heterogeneous Information Network Analysis (HINA 2011) - Barcelona, Spain, July 16, 2011.* Retrieved March 9, 2014, from http://www.kddresearch.org/KDD/Workshops/IJCAI-2011-HINA/Papers/

Myers, S. A., & Leskovec, J. (2010). On the Convexity of Latent Social Network Inference. In J. D. Lafferty, C. K. Williams, J. Shawe-Taylor, R. S. Zemel, & A. Culotta (Ed.), *Advances in Neural Information Processing Systems 23: Proceedings of the 24th Annual Conference on Neural Information Processing Systems (NIPS 2010) - Vancouver, BC, Canada, December 6-9, 2010* (pp. 1741-1749). Red Hook, NY, USA: Curran & Associates.

National Research Council (S. L. Magsino, Rapporteur). (2009). Social Network Analysis for Improved Disaster Preparedness and Intervention Planning. In S. L. Magsino, *Applications of Social Network Analysis for Building Community Disaster Resilience: Workshop Summary* (pp. 33-42). Washington, DC, USA: The National Academies.

Newman, M. E. (2003). The Structure and Function of Complex Networks. *SIAM Review, 45*(2), 167–256. doi:10.1137/S003614450342480

Nguyen, T., Phung, D. Q., Adams, B., & Venkatesh, S. (2012). A Sentiment-Aware Approach to Community Formation in Social Media. In J. G. Breslin, N. B. Ellison, J. G. Shanahan, & Z. Tufekci (Ed.), *Proceedings of the 6th International Conference on Weblogs and Social Media (ICWSM 2012).* Menlo Park, CA, USA: AAAI Press.

Nov, O., Naaman, M., & Ye, C. (2009). Motivational, Structural and Tenure Factors that Impact Online Community Photo Sharing. In E. Adar, M. Hurs, T. Finin, N. S. Glance, N. Nicolov, & B. L. Tseng (Ed.), *Proceedings of the 3rd International Conference on Weblogs and Social Media (ICWSM 2009).* Menlo Park, CA, USA: AAAI Press.

Nov, O., Naaman, M., & Ye, C. (2010). Analysis of participation in an online photo-sharing community: A multidimensional perspective. *Journal of the American Society for Information Science and Technology, 61*(3), 555–566.

O'Donovan, J., & Smyth, B. (2005). Trust in recommender systems. In R. St. Amant, J. Riedl, & A. Jameson (Ed.), *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI 2005)* (pp. 167-174). New York, NY, USA: ACM Press.

Ohara, K., Saito, K., Kimura, M., & Motoda, H. (2011). Efficient Detection of Hot Span in Information Diffusion from Observation. In W. H. Hsu, S. Kallumadi, & J. Han (Ed.), *Working notes of Workshop W-16, 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011): Heterogeneous Information Network Analysis (HINA 2011) - Barcelona, Spain, July 16, 2011.* Retrieved March 9, 2014, from http://www.kddresearch.org/KDD/Workshops/IJCAI-2011-HINA/Papers/

Olsman, K., Roxendal, O., & Åkerblom, V. (2013). *Designing for Collaboration Using Social Network Analysis - Towards a Conceptual Method to Understand Organisational Interaction.* Lund University , Department of Informatics. Lund, Sweden: Lund University School of Economics and Management.

O'Madadhain, J., Hutchins, J., & Smyth, P. (2005). Prediction and ranking algorithms for event-based network data. *ACM SIGKDD Explorations Newsletter, 7*(2), 23-30. doi:10.1145/1117454.1117458

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering (TKDE), 22*(10), 1345-1359. doi:10.1109/TKDE.2009.191

Parimi, R., & Caragea, D. (2011). Predicting Friendship Links in Social Networks Using a Topic Modeling Approach. In J. Z. Huang, L. Cao, & J. Srivastava (Ed.), *Advances in Knowledge Discovery and Data Mining - 15th Pacific-Asia Conference, PAKDD 2011, Shenzhen, China, May 24-27, 2011, Proceedings, Part II. Lecture Notes in Computer Science (LNCS) 6635*, pp. 75-86. New York, NY, USA: Springer.

Pennock, D. M., Flake, G. W., Lawrence, S., Glover, E. J., & Giles, C. L. (2002). Winners don't take all: Characterizing the competition for links on the web. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 99*(8), 5207–5211. doi:10.1073/pnas.032085699

Pennock, D. M., Horvitz, E., Lawrence, S., & Giles, C. L. (2000). Collaborative filtering by personality diagnosis: a hybrid memory- and model-based approach. In C. Boutilier, & M. Goldszmidt (Ed.), *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI 2000) - Stanford, California, USA, June 30 - July 3, 2000* (pp. 473-480). San Francisco, CA, USA: Morgan Kaufmann.

Piatetsky-Shapiro, G., Djeraba, C., Getoor, L., Grossman, R., Feldman, R., & Zaki, M. J. (2006). What are the grand challenges for data mining?: KDD-2006 panel report. *ACM SIGKDD Explorations Newsletter, 8*(2), 70-77. doi:10.1145/1233321.1233330

Pizzato, L., Rej, T., Akehurst, J., Koprinska, I., Yacef, K., & Kay, J. (2013). Recommending people to people: the nature of reciprocal recommenders with a case study in online dating. *User Modeling and User-Adapted Interaction, 23*(5), 447-488. doi:10.1007/s11257-012-9125-0

Popescul, A., Ungar, L. H., Pennock, D. M., & Lawrence, S. (2001). Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In J. S. Breese, & D. Koller (Ed.), *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence (UAI 2001) - Seattle, Washington, USA, August 2-5, 2001* (pp. 437-444). San Francisco, CA, USA: Morgan Kaufmann.

Pryymak, O., Rogers, A., & Jennings, N. R. (2011). Efficient Sharing of Conflicting Opinions with Minimal Communication in Large Decentralised Teams. In W. H. Hsu, S. Kallumadi, & J. Han (Ed.), *Working notes of Workshop W-16, 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011): Heterogeneous Information Network Analysis (HINA 2011) - Barcelona, Spain, July 16, 2011.* Retrieved March 9, 2014, from http://www.kddresearch.org/Workshops/IJCAI-2011-HINA/Papers/

Pujol, A., Mosca, R., Farrés, J., & Aloy, P. (2010). Unveiling the role of network and systems biology in drug discovery. *Trends in the Pharmacological Sciences, 31*(3), 115–123. doi:10.1016/j.tips.2009.11.006

Qian, R., Zhang, W., & Yang, B. (2006). Detect community structure from the Enron Email Corpus Based on Link Mining. *Proceedings of the 6th International Conference on Intelligent Systems Design and Applications (ISDA 2006). 2*, pp. 850-855. Los Alamitos, CA, USA: IEEE Computer Society.

Raymond, R., & Kashima, H. (2010). Fast and Scalable Algorithms for Semi-supervised Link Prediction on Static and Dynamic Graphs. In J. L. Balcázar, F. Bonchi, A. Gionis, & M. Sebag (Ed.), *Machine Learning and Knowledge Discovery in Databases: European Conference (ECML PKDD 2010) - Barcelona, Spain, September 20-24, 2010, Proceedings, Part III. Lecture Notes in Computer Science (LNCS) 6323*, pp. 131-147. New York, NY, USA: Springer.

Romero, D. M., & Kleinberg, J. M. (2010). The Directed Closure Process in Hybrid Social-Information Networks. In S. G. William W. Cohen (Ed.), *Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM 2010).* Menlo Park, CA, USA: AAAI Press.

Rowe, M., Stankovic, M., & Alani, H. (2012). Who Will Follow Whom? Exploiting Semantics for Link Prediction in Attention-Information Networks. In P. Cudré-Mauroux, J. Heflin, E. Sirin, T. Tudorache, J. Euzenat, M. Hauswirth, . . . E. Blomqvist (Ed.), *The Semantic Web: Proceedings of the 11th International Semantic Web Conference (ISWC 2012) - Boston, MA, USA, November 11-15, 2012, Proceedings, Part II. Lecture Notes in Computer Science (LNCS) 7649.* New York, NY, USA: Springer.

Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., . . . Vidal, M. (2005). Towards a proteome-scale map of the human protein–protein interaction network. *Nature, 437*, 1173-1178. doi:10.1038/nature04209

Russom, P. (2011). *Big Data Analytics: TDWI Best Practices Report, 4th Quarter 2011.* The Data Warehousing Institute (TWDI). Renton, WA, USA: TDWI Research. Retrieved March 9, 2014, from http://public.dhe.ibm.com/common/ssi/ecm/en/iml14293usen/IML14293USEN.PDF

Rutherford, A., Cebrian, M., Dsouza, S., Moro, E., Pentland, A., & Rahwan, I. (2013). Limits of social mobilization. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 110*(16), 6281–6286. doi:10.1073/pnas.1216338110

Salem, N. J. (2009). *Link Prediction in Static and Dynamic Graphs.* Retrieved March 9, 2014, from Graph Mining: Washington State University AI Lab: http://ailab.wsu.edu/graphmining/node/14

Sarawagi, S., & Bhamidipaty, A. (2002). Interactive deduplication using active learning. *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2002) - Edmonton, AB, Canada, July 23-26, 2002* (pp. 269-278). New York, NY, USA: ACM Press.

Scellato, S., Noulas, A., & Mascolo, C. (2011). Exploiting place features in link prediction on location-based social networks. In C. Apté, J. Ghosh, & P. Smyth (Ed.), *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2011)* (pp. 1046-1054). New York, NY, USA: ACM Press.

Schulte, O., & Qian, Z. (2013). Learning Bayes Nets for Relational Data With Link Uncertainty. In W. H. Hsu, S. Kallumadi, & Y. Sun (Ed.), *Working notes of the 2nd International Workshop on Heterogeneous Information Network Analysis (HINA 2013). Workshop W36, 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013) - Beijing, China, August 5, 2013.*

Schulte, O., Riahi, F., & Li, Q. (2013). Identifying Important Nodes in Heterogenous Networks. In W. H. Hsu, S. Kallumadi, & Y. Sun (Ed.), *Working notes of the 2nd International Workshop on Heterogeneous Information Network Analysis (HINA 2013). Workshop W36, 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013) - Beijing, China, August 5, 2013.*

Semertzidis, K., Pitoura, E., & Tsaparas, P. (2013). How people describe themselves on Twitter. In K. LeFevre, A. Machanavajjhala, & A. Silberstein (Ed.), *Proceedings of the 3rd ACM SIGMOD Workshop on Databases and Social Networks (DBSocial 2013) - New York, NY, USA, June, 23, 2013* (pp. 25-30). New York, NY, USA: ACM Press.

Sen, P., Namata, G., Bilgic, M., Getoor, L., Galligher, B., & Eliassi-Rad, T. (2008). Collective Classification in Network Data. *AI Magazine, 29*(3), 93-106. doi:10.1609/aimag.v29i3.2157

Shen, Z., Ma, K.-L., & Eliassi-Rad, T. (2006). Visual Analysis of Large Heterogeneous Social Networks by Semantic and Structural Abstraction. *IEEE Transactions on Visualization and Computer Graphics, Special Issue on Visual Analytics, 12*(6), 1427-1439.

Shmueli, G., & Koppius, O. R. (2011). Predictive Analytics in Information Systems Research. *MIS Quarterly, 35*(3), 553-572.

Silow-Carroll, S., Edwards, J. N., & Rodin, D. (2012). *Using Electronic Health Records to Improve Quality and Efficiency: The Experiences of Leading Hospitals (Publication 1608).* New York, NY, USA: The Commonwealth Fund.

Singh, B., & Singh, H. K. (2010). Web Data Mining Research: A Survey. In N. Krishnan, & M. Karthikeyan (Ed.), *Proceedings of the 1st IEEE International Conference on Computational Intelligence and Computing Research (ICCIC 2010) - Coimbatore, India, December 28-29, 2010* (pp. 1-10). Piscataway, NJ, USA: IEEE Press. doi:10.1109/ICCIC.2010.5705856

Soler-López, M., Zanzoni, A., Lluís, R., Stelzl, U., & Aloy, P. (2011). Interactome mapping suggests new mechanistic details underlying Alzheimer's disease. *Genome Research, 21*, 364-376. doi:10.1101/gr.114280.110

Stodder, D. (2012). *Customer Analytics in The Age of Social Media: TDWI Best Practices Report, 3rd Quarter 2012.* The Data Warehousing Institute (TWDI). Renton, WA, USA: TDWI Research. Retrieved March 9, 2014, from http://www.businesstimes.com.sg/archive/monday/sites/businesstimes.com.sg/files/Customer%20Analytics%20in%20the%20Age%20of%20Social%20Media.pdf

Sun, C., Lin, L., Chen, Y., & Liu, B. (2013). Expanding User Features with Social Relationships in Social Recommender Systems. In G. Zhou, J. Li, D. Zhao, & Y. Feng (Ed.), *Natural Language Processing and Chinese Computing: Proceedings of the 2nd CCF Conference (NLPCC 2013) - Chongqing, China, November 15-19, 2013. Communications in Computer and Information Science (CCIS) 400*, pp. 247-254. New York: Springer.

Sun, Y., & Han, J. (2012). Mining heterogeneous information networks: a structural analysis approach. *ACM SIGKDD Explorations Newsletter, 14*(2), 20-28. doi:10.1145/2481244.2481248

Sun, Y., Aggarwal, C. C., & Han, J. (2012). Relation strength-aware clustering of heterogeneous information networks with incomplete attributes. (Z. M. Ozsoyoglu, Ed.) *Proceedings of the VLDB Endowment, 5*(5), 394-405.

Sun, Y., Han, J., Aggarwal, C. C., & Chawla, N. V. (2012). When will it happen?: relationship prediction in heterogeneous information networks. In E. Adar, J. Teevan, E. Agichtein, & Y. Maarek (Ed.), *Proceedings of the 5th International Conference on Web Search and Web Data Mining (WSDM 2012) - Seattle, WA, USA, February 8-12, 2012* (pp. 663-672). New York, NY, USA: ACM Press. doi:10.1145/2124295.2124373

Sun, Y., Han, J., Yan, X., Yu, P. S., & Wu, T. (2011). Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, Tianyi Wu: PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks. (H. V. Jagadish, Ed.) *Proceedings of the VLDB Endowment, 4*(11), 992-1003.

Sun, Y., Norick, B., Han, J., Yan, X., Yu, P. S., & Yu, X. (2013). PathSelClus: Integrating Meta-Path Selection with User-Guided Object Clustering in Heterogeneous Information Networks. *ACM Transactions on Knowledge Discovery from Data (TKDD) - Special Issue on ACM SIGKDD 2012, 7*(3), 11:1-11:23. doi:10.1145/2500492

Sun, Y., Yu, Y., & Han, J. (2009). Ranking-based clustering of heterogeneous information networks with star network schema. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009) - Paris, France, June 28 - July 1, 2009* (pp. 797-806). New York, NY, USA: ACM Press. doi:10.1145/1557019.1557107

Sycara, K. (2012). Dynamics of Information Propagation in Large Heterogeneous Networks. In F. M. Brazier, K. Nieuwenhuis, G. Pavlin, M. Warnier, & C. Badica (Ed.), *Intelligent Distributed Computing V: Proceedings of the 5th International Symposium on Intelligent Distributed Computing (IDC 2011) - Delft, The Netherlands, October 2011. Studies in Computational Intelligence 382*, p. 3. New York, NY, USA: Springer.

Szell, M., Lambiotte, R., & Thurner, S. (2010). Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 107*(31), 13636–13641. doi:10.1073/pnas.1004008107

Tang, J., Lou, T., & Kleinberg, J. M. (2012). Inferring social ties across heterogenous networks. In E. Adar, J. Teevan, E. Agichtein, & Y. Maarek (Ed.), *Proceedings of the 5th International Conference on Web Search and Web Data Mining (WSDM 2012) - Seattle, WA, USA, February 8-12, 2012* (pp. 743-752). New York, NY, USA: ACM Press. doi:10.1145/2124295.2124382

Taskar, B., Wong, M.-F., Abbeel, P., & Koller, D. (2004). Link Prediction in Relational Data. In S. Thrun, L. K. Saul, & B. Schölkopf (Ed.), *Advances in Neural Information Processing Systems 16: Proceedings of the 17th Annual Conference on Neural Information Processing Systems (NIPS 2003) - Vancouver and Whistler, BC, Canada, December 8-13, 2003.* Cambridge, MA, USA: MIT Press.

Taylor, J. (2011). *Decision Management Systems: A Practical Guide to Using Business Rules and Predictive.* Upper Saddle River, NJ, USA: Pearson Education.

Thelwall, M. (2008). Social networks, gender, and friending: An analysis of MySpace member profiles. *Journal of the American Society for Information Science and Technology, 59*(8), 1321–1330. doi:10.1002/asi.20835

Thom, D., Bosch, H., Koch, S., Worner, M., & Ertl, T. (2012). Spatiotemporal anomaly detection through visual analysis of geolocated Twitter messages. In H. Hauser, S. G. Kobourov, & H. Qu (Ed.), *Proceedings of the 2012 IEEE Pacific Visualization Symposium (PacificVis 2012) - Songdo, South Korea, February 28 - March 2, 2012* (pp. 41-48). Piscataway, NJ. USA: IEEE Press. doi:10.1109/PacificVis.2012.6183572

Thomas, J., & Kielman, J. (2009). Challenges for Visual Analytics. *Information Visualization, 8*(4), 309-314. doi:doi: 10.1057/ivs.2009.26

Tsai, W. (2002). Social Structure of “Coopetition” Within a Multiunit Organization: Coordination, Competition, and Intraorganizational Knowledge Sharing. *Organization Science, 13*(2), 179-190. doi:10.1287/orsc.13.2.179.536

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In W. W. Cohen, & S. Gosling (Ed.), *Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM 2010)* (pp. 178-185). Menlo Park, CA, USA: AAAI Press.

Tylenda, T., Angelova, R., & Bedathur, S. J. (2009). Towards time-aware link prediction in evolving social networks. In C. L. Giles, P. Mitra, I. Perisic, J. Yen, & H. Zhang (Ed.), *Proceedings of the 3rd Workshop on Social Network Mining and Analysis (SNAKDD 2009), 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009) - Paris, France, June 28, 2009* (p. 9). New York, NY, USA: ACM Press. doi:10.1145/1731011.1731020

Van Hentenryck, P. (2013). Computational Disaster Management. In F. Rossi (Ed.), *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013) - Beijing, China, August 3-9, 2013.* Menlo Park, CA, USA: AAAI Press / International Joint Conferences on Artificial Intelligence.

Wakita, K., & Tsurumi, T. (2007). Finding community structure in mega-scale social networks: [extended abstract]. In C. L. Williamson, M. E. Zurko, P. F. Patel-Schneider, & P. J. Shenoy (Ed.), *Proceedings of the 16th International World Wide Web Conference (WWW 2007) - Banff, AB, Canada, May 8-12, 2007* (pp. 1275-1276). New York, NY, USA: ACM Press. doi:10.1145/1242572.1242805

Wang, X., & Sukthankar, G. (2013). Multi-label classification by mining label and instance correlations from heterogeneous information networks. In I. S. Dhillon, Y. Koren, R. Ghani, T. E. Senator, P. Bradley, R. Parekh, . . . R. Uthurusamy (Ed.), *Proceedings of the 19th ACM SIGKDD International*

Conference on Knowledge Discovery and Data Mining (KDD 2013) - Chicago, IL, USA, August 11-14, 2013 (pp. 464-472). New York, NY, USA: ACM Press. doi:10.1145/2487575.2487577

Wang, Y., & Vassileva, J. (2007). A Review on Trust and Reputation for Web Service Selection. *Proceedings of the 27th International Conference on Distributed Computing Systems Workshops (ICDCSW 2007) - Toronto, Ontario, Canada, June 25-29, 2007* (p. 25). Los Alamitos, CA, USA: IEEE Computer Society. doi:10.1109/ICDCSW.2007.16

Washio, T., & Motoda, H. (2003). State of the art of graph-based data mining. *ACM SIGKDD Explorations Newsletter, 5*(1), 59-68. doi:10.1145/959242.959249

Wenger, G. C. (1997). Social networks and the prediction of elderly people at risk. *Aging & Mental Health, 1*(4), 311-320. doi:10.1080/13607869757001

Weninger, T., Bisk, Y., & Han, J. (2012). Document-topic hierarchies from document graphs. In X.-W. Chen, G. Lebanon, H. Wang, & M. J. Zaki (Ed.), *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM 2012)* (pp. 635-644). New York, NY, USA: ACM Press.

Wex, F. (2013). *Coordination Strategies and Predictive Analytics in Crisis Management.* Albert-Ludwigs-Universität Freiburg im Breisgau, Department of Economic Behavior and Sciences (Wirtschafts- und Verhaltenswissenschaftlichen Fakultät). Breisgau, Germany: University of Freiburg.

Woods, W. A. (1975). What's in a Link: Foundations for Semantic Networks. In D. G. Bobrow, & A. Collins (Eds.), *Representation and Understanding: Studies in Cognitive Science (Language, Thought & Culture)* (pp. 35-82). New York, NY, USA: Academic Press. Retrieved March 9, 2009, from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA022584

Wu, H., Sorathia, V., & Prasanna, V. (2012). Predict Whom One Will Follow: Followee Recommendation in Microblogs. In K. Aberer, A. Flache, W. Jager, L. Liu, J. Tang, & C. Guéret (Ed.), *Proceedings of the 4th International Conference on Social Informatics (SocInfo 2012) - Lausanne, Switzerland, December 5-7, 2012* (pp. 260-264). New York, NY, USA: Springer. doi:10.1109/SocialInformatics.2012.74

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Who says what to whom on twitter. In S. Srinivasan, K. Ramamritham, A. Kumar, M. P. Ravindra, E. Bertino, & R. Kumar (Ed.), *Proceedings of the 20th International World Wide Web Conference (WWW 2011) - Hyderabad, India, March 28 - April 1, 2011* (pp. 705-714). New York, NY, USA: ACM Press. doi:10.1145/1963405.1963504

Yan, X., & Han, J. (2002). gSpan: Graph-Based Substructure Pattern Mining. *Proceedings of the 2nd IEEE International Conference on Data Mining (ICDM 2002)* (pp. 721-724). Los Alamitos, CA, USA: IEEE Computer Society.

Yan, X., & Han, J. (2003). CloseGraph: Mining Closed Frequent Graph Patterns. In L. Getoor, T. E. Senator, P. Domingos, & C. Faloutsos (Ed.), *Proceedings of the 9th ACM SIGKDD International Conference* *on Knowledge Discovery and Data Mining (KDD 2003)* (pp. 286-295). New York, NY, USA: ACM Press.

Yang, Q., & Wu, X. (2006). 10 challenging problems in data mining. *International Journal of Information Technology & Decision Making, 5*(4), 597-604. doi:10.1142/S0219622006002258

Yu, P. S., Han, J., & Faloutsos, C. (2010). *Link Mining: Models, Algorithms, and Applications.* New York, NY, USA: Springer.

Yu, X., Gu, Q., Zhou, M., & Han, J. (2012). Citation Prediction in Heterogeneous Bibliographic Networks. *Proceedings of the 12th SIAM International Conference on Data Mining (SDM 2012) - Anaheim, California, USA, April 26-28, 2012* (pp. 1119-1130). Philadelphia, PA, USA: SIAM Press.

Yu, X., Ren, X., Gu, Q., Sun, Y., & Han, J. (2013). Collaborative Filtering with Entity Similarity Regularization in Heterogeneous Information Networks. In W. H. Hsu, S. Kallumadi, & Y. Sun (Ed.), *Working notes of the 2nd International Workshop on Heterogeneous Information Network Analysis (HINA 2013). Workshop W36, 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013) - Beijing, China, August 5, 2013.*

Zaraté, P. (2012). *Integrated and Strategic Advancements in Decision Making Support Systems.* Hershey, PA, USA: IGI Global.

Zhang, J., HI, W. C., Lee, B. S., Lee, K. K., Vassileva, J., & Looi, C. K. (2010). A Framework of User-Driven Data Analytics in the Cloud for Course Management. In S. L. Wong, S. C. Kong, & F. -Y. Yu (Ed.), *Proceedings of the 18th International Conference on Computers in Education (ICCE 2010): Enhancing and Sustaining New Knowledge through the Use of Digital Technology in Education - Putrajaya, Malaysia, November 29 - December 3, 2010* (pp. 698-702). Jhongli City, Taiwan: Asia-Pacific Society For Computers in Education.

Zhao, X. (2012). Scorecard with Latent Factor Models for User Follow Prediction Problem. In B. Dalessandro, C. Perlich, G. Sun, & A. Yue (Ed.), *Proceedings of the KDD Cup 2012 Workshop, 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2012) - Beijing, China, August 12-16, 2012. Track 1, 3rd Place.* San Francisco, CA, USA: Kaggle. Retrieved March 9, 2014, from https://kaggle2.blob.core.windows.net/competitions/kddcup2012/2748/media/FICO.pdf

Zheng, T., Salganik, M. J., & Gelman, A. (2006). How Many People Do You Know in Prison? Using Overdispersion in Count Data to Estimate Social Structure in Networks. *Journal of the American Statistical Association, 101*(474), 409-423. doi:10.1198/016214505000001168

Zhou, D., Councill, I. G., Zha, H., & Giles, C. L. (2007). Discovering Temporal Communities from Social Network Documents. *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM 2007) - Omaha, NE, USA, October 28-31, 2007* (pp. 745-750). Los Alamitos, CA, USA: IEEE Computer Society.

Žliobaitė, I. (2009). *Learning under Concept Drift: an Overview.* Vilnius University, Faculty of Mathematics and Informatics. Ithaca, NY, USA: Cornell University Library. Retrieved March 9, 2014, from http://arxiv.org/abs/1010.4784

Zolfaghar, K., & Aghaie, A. (2011). Evolution of trust networks in social web applications using supervised learning. (A. Karahoca, & S. Kanbul, Eds.) *Procedia Computer Science: World Conference on Information Technology - Istanbul, Turkey, October 6–10, 2010, 3*, 833-839.