

Data in Social Network Analysis

Anu Vaidyanathan¹, Malcolm Shore², and Mark Billingham¹

¹ University of Canterbury, Christchurch, New Zealand,
anuradha.vaidyanathan@canterbury.ac.nz,
mark.billinghurst@canterbury.ac.nz,

² Telecom New Zealand, malcolm.shore@telecom.co.nz

Abstract. Social Network research relies on a variety of data-sources, depending on the problem-scenario and the questions which the research is trying to answer or inform. In this paper, we analyze some of the data-sources indexed by the sizes of these data-sets and relating them back to the research question, which the data-set is used for. In carrying out such an exercise, our goal is to assign a confidence metric to the data-set when tied to the field within the Social Network analysis that the data is used for. This will lay the foundation for benchmarking the results of any social networking study by means of concrete requirements for the data-sets used in research.

1 Introduction

Social Network analysis is used to understand the social structure, which exists amongst entities in an organization. The size, diversity and ubiquity of social networks act in combination to necessitate understanding these networks in a systematic manner. Several aspects of social network analysis are currently the subject of academic research. Some of the themes of research in social networks includes processes in on-line social networks relating to communication [38], the formation of communities [23], visualizing social network data [1], extracting social network metrics [21] and enabling various functionalities within social networks [29]. The data-sets used in these analyses are important to understand for the following reasons:

- a. Data-sets used in any line of research ultimately lead to the formation of benchmarks which are used to evaluate new proposals to address known bottlenecks.
- b. Data-sets have to be accurate and representative of the problem being addressed, in order to provide confidence in the research being conducted.
- c. Ultimately, understanding the nature of data-sets is required to perform accurate social-network simulation.

Social Network research has a multitude of experts participating from fields as diverse as Sociology, Anthropology, Computer Sciences, Library Sciences, Engineering and Information Technology. Bringing together such diverse expertise is not without challenges, especially when trying to understand which data-sets

II

can provide results with the most confidence. In our initial survey of Social Network literature, we found a variety data-sets used to validate research, ranging from 14-25 user interviews [?] to a 9-month survey of the users of Friendster ranging in the order of several thousands [1]. Depending on the question being answered, which varies across several categories, the nature of the data is bound to vary. This paper proposes a basic outline of the characteristics, which the data-sets might need in order to make room for discussion across the board on the various research topics.

At first, we propose that four simple characteristics be taken into consideration. These characteristics include:

Temporal nature of the data: Social networks are experiencing growth similar to that of the internet. Over time, the growth experienced by social network in terms of raw size gives rise to new issues and perspectives, when it comes to the proposed solutions. Understanding the nature of this growth and having the data-sets reflect the temporal component of such growth is therefore imperative in this line of study.

Expertise of the participants: In the set of research papers, which used participant surveys or input, the expertise of the participants varied from being random participants [2] to extremely focussed work-groups of GPs [3], AI researchers [4], etc. Once again, depending on the research question being answered by the research, we propose that the expertise of the participants is a factor, which will impact the confidence of the results produced.

Sample size of the data: When studying metrics relating to social networks, in particular, the sample size of the data-sets is an important factor. This extends to other analyses in privacy and trust, collaboration etc. as the size of the typical Social Network is always increasing through the addition of a global demographic of users, who wish to stay connected.

Source of the data: This refers to the background within which the data collected initially resided. The diversity of data-sets spans portions of popular Social Networking applications such as Flickr, Yahoo! 360 and Friendster [5] to a collection of conference papers [6] to e-mail lists [7] to wikis [8] to simply users carrying a certain type of cell-phone [9], understanding the source of the data is important in assigning a confidence metric.

The sum-total of this proposal is the evolution of a framework that incorporates the desirable characteristics T(emporal), E(xpertise), S(ample-sizes), (S)ource (TESS), which summarizes how well the data-sets used in a particular study relating to Social Networks. By analysing the data-sets used from the focal point of these characteristics, we go on to assign a confidence metric with TESS. An additional characteristic, which we hypothesize as being important, in certain areas of research, with varying definitions of metrics, such as centrality, trust etc. is the definition itself.

2 Related Work

Network data, in particular Social Network data is available from many different sources. For example, some of the data-sets used in network analysis include; Zachary's karate club [10] defines a social network of friendships between 34 members of a karate club at a US university. The coappearance network of characters in the novel *Les Misérables* [11] has been created and the adjacency network of common adjectives and nouns in the book *David Copperfield* has also been studied [12]. A network of American football games between Division 1A colleges during the Fall 2000 season [13], an undirected social network of 62 dolphins in a community from Doubtful Sound New Zealand [14], a directed network of hyperlinks between blogs on politics in the United States recorded in the year 2004 [15], a network of books about US politics sold online by Amazon.com [16], and a network of co-authorships of scientist working on network theory and experiment [17].

Several benchmarking schemes exist in the area of Knowledge Based Systems, which can be extrapolated to the Semantic Web and further to Social Networks. The Resource Description Framework (RDF) is a family of W3C specifications, which has become an accepted form of metadata, extending itself to Semantic Web applications in such manifestations as RSS and FOAF ontologies. The FOAF ontology is machine-readable and used to describe people and their interrelations with other entities. This ontology decentralizes the data used in Social Networks by allowing users to create and describe social networks, without referring back to a central database. FOAF extends the RDF specification and is described using OWL. The Leigh University Benchmark (LUBM) is used to benchmark the Semantic Web with respect to use in large OWL applications. The LUBM uses a uniform ontology and can be applied to various scales and configurations. Benchmarks further exist for specifications such as Web2.0, used in extending the social semantic web. For example, del.icio.us can be considered as one such benchmark for Web2.0.

Since Social Network research embodies a range of expertise from anthropology to Computer Sciences, it is difficult to find benchmarks for social networks, per se. This paper aims at analyzing the data-sets used in various fields of Social Network research to perform the groundwork for such benchmarking in the future. Social Networks have been measured in many ways and the measurements have been carried out on various data-sets, from on-line social networks [18] to sexual transmission networks [19].

Some benchmarks are known in social-networking literature including the Southern Women data from 1985 [20]. This particular data-set dates back to the 1930s and was used to understand inter-personal relationships. Using sociological definitions, researchers such as Roetlisberger and Dickson (1939) and Davis, Gardner and Gardner (1941) segregate the data into core and peripheral group members. In the 1930s, five ethnographers collected data pertaining to stratification, from Natchez, Mississippi. The aim of this study was to understand social class in a mixed-race society. Eighteen women were picked for this study and a systematic analysis was carried out on their social activities, over a nine month

period. During that time, subsets of this set of women participated in social events and their participation was monitored by means of interviews, recording the observation of other participant observers, guest lists and news paper reports. The size of this data-set was small and the source was not definitive, given two contrasting studies. The size of social networks has grown considerably over the last seven odd decades, making the size of the sample set more relevant. The temporal aspect of the Southern Women data-set is to be noted. The researchers did not simply sample one point in time, in order to carry out their analysis, but in fact sampled the data over a nin-month period. The expertise of the data-set here was not limited to the participants themselves but included that of the observers and the press. This data-set, while from the 1930s, certainly took into account the inclusion of the basic characteristics, which we propose as part of this work, in the TESS framework.

3 Methodology

In this section, we discuss the methods we use to define and assess confidence in the data-sets used in Social Network research. In section 4.1, we present the data-sets found in various lines of research. The major themes we have encountered include extracting social network metrics, community formation, visualization, trust and privacy. We go on to present the metrics, which are a part of TESS, ultimately used to assign confidence to the answers proposed with the use of these data-sets.

3.1 Data-Sets vs. Problems Solved

A number of themes exist within Social Network research. Examples of these themes include, extracting social network metrics, community formation, visualization, understanding trust and privacy, collaboration, wearable computing and value-added services such as tagging, for online Social Network applications. In this section, we explore the broad categories of academic work within Social Network research and provide the characteristics of the data-sets used. This will lay the foundation for assigning ratings using the TESS framework in later sections.

Extracting Social Network Metrics Social network metrics such as degree, between-ness, closeness and network centrality are often the subject of academic research. Understanding social networks and their metrics is important as these networks form the underlying structure, which allows for rapid information distribution [21]. A preliminary analysis of research includes data-sets from a variety of resources including email lists [7], the world-wide-wed [7] and Instant Messaging Populations [22]. Further, Social Network Mining using Google and data-sets from conferences [4] have been proposed to extract relations between people and identify groups. Table 1 presents sample data-sets, which are used in these studies. [23] suggests the use of computer-generated networks, to perform a controlled study of metric extraction and the use of bibliographies from arxiv.com to study

this problem. Zachary’s karate club network [10] is used in this work to understand the real-world applications of these ideas. Link topologies have been used [24] to predict social connections and extract metrics indicating connectedness. Information sharing has been proposed with the use of Saori [25], in order to enable information dissemination. A relationship algebra [26] has been proposed, in order to understand and analyze social connections using data-sets from publication bibliographies and parts of the online network Orkut. The Citeseer dataset is used in [27] to understand how the social actors, in this case authors of various papers, affect the lines of research, which are observed. A new research paper search engine, Rexa.info, is proposed in [28], in order to organize publications for effective retrieval, enabling social network analysis. Event and place semantics are extracted using Flickr tags in [29] to extract usage patterns of people sharing photographs. Table 1 summarizes the data-sets used in these studies.

Table 1. Data-sets in analyzing Social Network Metrics

Year of publication	Data Set
2005	4 academic conferences, 500 participants, 3 years
2004	53 e-mail participants, 229 web-pages
2004	Buddy lists from LiveJournal, 25 days
2006	1 academic conference, 503 attendees,
2000	145 scientists, bibliography over 3 years
2000	1265 people, Friends listed on personal homepages in Stanford and MIT
2007	49897 photos from Flickr.com, 1015 days worth data
2000	108,676 academic papers from Citeseer, 13 years worth of data

Community Formation Community formation is important to understand within Social Network analysis, in order to understand patterns of collaboration. BitTorrent communities were studied in [30], in order to understand the factors affecting the participant’s co-operative behavior. The Iris and DPLB datasets are used in [31] to mine communities within social networks. Group Formation is studied using data-sets from LiveJournal and DBLP [32] to understand the evolution of communities. User experiences at Open Office [33] were discussed wherein an open source office suite with nearly 62,000 mailing list subscribers was analyzed. Data from a hundred mobile phones were analyzed in [9], over a period of nine months, in order to understand and reflect on social patterns. A user-group of older people [34] was used to understand the accessibility and inclu-

sion of this demographic, in online social interaction. An online community in a suburban town was studied [35] to investigate means to stimulate social engagement. New information interfaces are proposed in [36] to provide hypermedia capabilities for information sharing and collaboration. Blog entries are mined [37] to discover stories within the data found in blogosphere. Digital Libraries act in unison [38] to create a common learning substrate accessible by a variety of learners with a proper interface to stimulate learning. [39] uses two sets of data, both synthetic and real-world data, to identify communities, while proposing heuristics to analyze what could be NP-Hard problems. Table 2 summarizes the data-sets, which are used in these studies.

Table 2. Data-sets in analyzing Community Formation

Year of publication	Data Set
2006	875 LiveJournal communities over 10 days, 71,618
2005	70 conferences
2007	62,000 registered
2005	100 people using Nokia6600
2002	280 individual visitors
1997	3 colleges, 15 teachers and admin
2007	1200 on-line n
2007	Two synthetic data-sets (Assembly Line, Dutiful Children), two real-world data-sets (Southern W

Visualization Visualizing social networks assists researchers in understanding new ways to present and manage data and effectively convert that data into meaningful information [40]. A number of tools have been proposed for this task of visualizing social networks including Pajek [41], NetVis , Krackplot, I-Know, InFlow, Visone, JUNG and Prefuse, to name a few. Discussion forums are considered to be another source of online collaboration and these have been visualized to better understand interactions [42]. Visualizing tasks for better collaboration during software development has been proposed [42] to address issues of co-ordination and geographical distribution of developer teams. Visualizing social networks using Query interfaces for wikis and blogs [8] are used to provide the end-users with more user-friendly alternatives. Weblink graphs were used in [43] to extract hierarchies of complex networks. Over a year-long period, individual and team use of tablet PCs was studied [44] to understand the process of learning, within a group of students. Business Intelligence search [45] was facil-

itated by looking at the agreement between participants on certain statement and visualizing the same. The network value of customers is visualized [46] in order to enable direct marketing more effectively. Visualization further finds its use in law-enforcement [47], wherein crime-pattern recognition and criminal associations were mined and visualized for the Tuscon PD. Bibliographic data has been visualized, which finds its end application in summarizing scientific fields. Simple techniques for visualizing social graphs [48] have been proposed. Table 3 summarizes the data-sets, which are used in these studies.

Table 3. Data-sets in visualizing Social Networks

Year of publication	Data Set
2007	Discussion forums, 16 participants
2006	Web link graph of 51,497 internet pages, An empirical set of 8210 word-associations
2006	7 participants using tablet PCs over a 12 month period
2007	30 undergraduate researchers assessing two websites to gain business intelligence
2005	Incident reports and GIS tools from the Tuscon Police Department
2007	2.8 million movie ratings of 1628 movies by 72916 users over 18 months
2002	Dataset from the 2001 Graph Drawing Contest with papers from 1994-2000
2004	5 students creating messages for one another

Trust and Privacy As the size and ubiquity of social networks grow, trust and privacy become very important issues for both designers and users to address and understand. The wordpress blogging engine is used in conjunction with Mozilla Firefox, in order to provide signature-based architectures for secure communication on the Social Web [49]. The MovieLens data set is used in [50] to deconstruct recommender systems. Extensions to the RDF framework to incorporate mechanisms to enable trust have also been proposed [51]. Other examples of data-sets include MBA students from colleges in the USA [52], supply chain data [53] and agents [54]. Over 1200 people from EU countries were studied [55] in order to analyze the value of location privacy. Community connectedness as understood by analyzing the privacy requirements [56] was studied in order to reduce detachment within online social communities. Sybil attacks in distributed systems wherein several fake identities are utilized to start attacks are studied [57] and countered using trust, which is introduced by adhering to social networks amongst user identities. The value of creating on-line identities has been explored [58] in order to understand identity theft in online communi-

ties, which actively encourage users to create profiles, share personal information and network socially. Table 4 presents data-sets used in this line of research.

Table 4. Data-sets in Trust and Privacy studies

Year of publication	Data Set
2005	18 to 706 profiles of movie ratings from MovieLens
2000	239 students evaluating online shopping system
2002	53 e-mail participants, 229 web-pages
2002	Supply Chain vendors
2006	2000 participants from law, sciences, ten European countries carrying mobile phones, 1 year
2007	18 users of environment services, 80 percent expert users
2006	A synthetic social network model with 10,000 nodes
2006	41 respondents in a classroom setting

4 A new confidence metric - TESS

The Southern Women’s data [20], albeit being a relatively small and old data-set, encompasses some of the necessary characteristics in data-sets, which can be used to confidently assess new proposals within Social Network research. These characteristics include the following:

4.1 Temporal Characteristics

The Southern Women’s data was collected over a 9-month period. Since the size and popularity of Social Networks is growing in leaps and bounds, in order to address the problems or roadblocks in the development of these networks, it would be essential to have data-sets which are characterizing the networks over a period of time. A single snapshot in time may not be the most effective way to gather data, to assess new proposals for research problems.

4.2 Expertise

Sampling the data from various points of expertise is important because this normalizes the confidence in the data. For example, some datasets are procured from researchers with several publications [4] while some other datasets are procured from members of a fraternity house [59]. These ranges of data need a

meeting point by assigning a certain degree to the expertise characteristic, in order to enable fair evaluation. In the Southern Women’s data, the expertise of participants also varied between three peer-groups, the women who participated in the survey, observers and members of the press.

4.3 Source

The source of data is important to understand with the rapid proliferation of information. All data sources from Wikis to proper bibliographies are represented in the data-sets popularly used in Social Network Research. There needs to be a clear understanding of the source of the data, in order to understand whether the data is indeed reliable.

4.4 Sample Size

The Souther Women’s data set had only 14 participants. The data-sets analyzed in this paper vary between a few participants to several thousand participants. In order to be confident in the results of any proposal, the sample size of the data needs to be analyzed and discussed, in order to realize its suitability for the analysis.

The TESS framework is simply aiming to assign a confidence metric, based on the data-sets used in the research. We are using a simple assignment of ratings between 1-5, 1 being the lowest and 5 being the highest confidence. This creates a confidence-vector, which is TESS. It is desirable to have a confidence-vector, which balances all the four characteristics equally or pin-points exactly why a certain rating for any of the characteristics is where it is.

4.5 Assigning Confidence

Using the metrics outlined in table 5, we assign a confidence factor to the literature pertaining to Scoial Networks, where a diverse set of data is employed to address several existing challenges. Fig. 1-3 shows the confidence-vector, TESS, for the fields of social network extraction, community formation research and trust and privacy research. We choose to leave out visualization as the basis of this study would be better performed if juxtaposed with the multiple actual tools [41], which exist, several of them being open source. The x-axis pertains to the literature whose data-sets we analyzed and the y-axis shows the actual confidence rating, in each of the characteristic fields within TESS. In Fig. 1, the average rating varies between 3-4 in this field of Social Network analysis, that of metric extraction. The expertise of the participants is rated the highest in this field of work as the data-sets mostly pertain to academic conferences.

In Fig. 2, the average rating varies between 2.375 to 3.35, with the source being rated the lowest. This is because the source of the data-sets in this field of study seem to pertain to a single demographic of users, such as users of a certain model of phones or mailing-list subscribers. There are no peers from

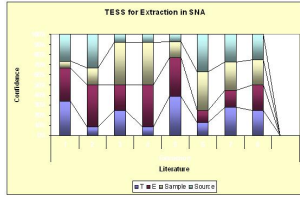


Fig. 1. TESS ratings for Social Network Extraction

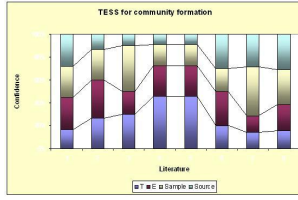


Fig. 2. TESS ratings for Social Network Community Formation

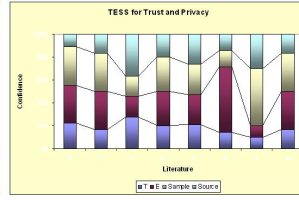


Fig. 3. TESS ratings for Trust and Privacy within Social Networks

different demographics to even the ratings out. In the Southern Women’s data, for example, besides the participants themselves, the sources included the press and observers, representing different peer groups. Within studies on community formation, it seems that the analyses could extend to providing analogies of the utility of the proposed solution within different demographics.

In Fig. 3, we see that the ratings are the poorest across all characteristics, ranging between 1.8-2.8. The temporal aspect of the data-set is valued at the lowest in this set of data because most studies seem to consider a single snapshot in time. This observation could also be attributed to the fact that research pertaining to trust and privacy, mostly propose alternate models [53] for trust or improved security in terms of protocols or alternate specification [51]. In this case, implementing the proposals on data-sets spanning time may or may not be relevant for initial analyses. This leads us to a discussion of factoring in the definitions of various metrics, into our assignment of confidence, which is the subject of future work. This implies that the actual definition of trust, privacy and other metrics such as centrality, might affect the assignment of a confidence metric to data-sets.

5 Future Work

In this paper, we have presented the diversity in data-sets used in several sub-fields within Social Network research, such as extracting metrics, understanding community formation, visualization and understanding trust and privacy within these structures. Data-sets are interesting and important because they ultimately lead to the creation of useful benchmarks [60], which can be used to evaluate new proposals in research. Data sets further have to represent the problem space accurately in order to validate the utility of the solutions. Social Network simulation also requires a robust understanding of these data sets. In this paper, we have presented our confidence metric, TESS, whose various ratings elucidate whether the data used has the desired attributes. We see some variability in the average ratings, across various sub-sections of research, specifically social network extraction, community formation and trust and privacy studies.

Future work would include placing side-by-side the data-sets used in visualization tools built to be utilized in Social Network analysis, to see if there is

accurate representation of these characteristics. Since the visualization section of this research is pretty advanced, with several existing tools such as Vizter, JUNG, it would validate whether the metrics presented here are adhered to well. We contend that the actual definition of the metrics such as trust, privacy, degree of centrality etc. will affect how these ratings are assigned to the data-sets. For example, if the definition of trust is Context-Specific [50] wherein a user is required to trust another user in a specific situation, a single snapshot of data (i.e, one which need not be sampled over a period of time as arguably the situation expires, past that point in time) might still be assigned a high rating, in the T(emporal) aspect.

6 Conclusion

References

1. Marchionini, G., Nolet, V., Williams, H., Ding, W., Josephus Beale, J., Rose, A., Gordon, A., Enomoto, E., Harbinson, L.: Content + connectivity = community : *digitalresourcesforalearningcommunity*. In : *DL'97 : ProceedingsofthesecondACMinternationalconferenceonDigitallibraries, NewYork, NY, USA, ACM(1997)212-220*
2. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. *Phys. Rev. E* **69**(2) (Feb 2004) 026113
3. Heer, J., Boyd, D.: Vizster: Visualizing online social networks. In: *INFOVIS '05: Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization, Washington, DC, USA, IEEE Computer Society (2005)* 5
4. Hamasaki, M.: Community focussed social network extraction. In: *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work, New York, NY, USA, ACM (2004)* 328–331
5. Ortiz, J.A., Tapia, A., Rangel, E.M.: Recruiting diverse, high-skilled it employees through existing virtual social networks. In: *SIGMIS CPR '06: Proceedings of the 2006 ACM SIGMIS CPR conference on computer personnel research, New York, NY, USA, ACM (2006)* 4–11
6. Counts, S., Geraci, J., Geraci, J.: Incorporating physical co-presence at events into digital social networking. In: *CHI '05: CHI '05 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM (2005)* 1308–1311
7. Chung, K.S.K., Hossain, L., Davis, J.: Individual performance in knowledge intensive work through social networks. In: *SIGMIS-CPR '07: Proceedings of the 2007 ACM SIGMIS CPR conference on 2007 computer personnel doctoral consortium and research conference, New York, NY, USA, ACM (2007)* 159–167
8. Matsuo, Y., Mori, J., Hamasaki, M., Ishida, K., Nishimura, T., Takeda, H., Hasida, K., Ishizuka, M.: Polyphonet: an advanced social network extraction system from the web. In: *WWW '06: Proceedings of the 15th international conference on World Wide Web, New York, NY, USA, ACM (2006)* 397–406
9. danah michele boyd: Friendster and publicly articulated social networking. In: *CHI '04: CHI '04 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM (2004)* 1279–1282
10. Yu, B., Singh, M.P.: Searching social networks. In: *AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems, New York, NY, USA, ACM (2003)* 65–72

11. Culotta, A., Bekkerman, R., McCallum, A.: Extracting social networks and contact information from email and the web. In: Proceedings of CEAS-04, the 1st Conference on Email and Anti-Spam. (2004)
12. Gardner, T.: Visual query interfaces for wiki's and blogs. In: SIGGRAPH '04: ACM SIGGRAPH 2004 Web graphics, New York, NY, USA, ACM (2004) 5
13. Eagle, N., Pentland, A.S.: Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.* **10**(4) (2006) 255–268
14. Zachary, W.W.: A survey of augmented reality. *Journal of Anthropological Research* **33** (1977) 452–473
15. Knuth, D.E.: *The Stanford GraphBase: A Platform for Combinatorial Computing.* Addison-Wesley (1993)
16. Newman, M.E.J.: Mixture models and exploratory analysis in networks. *Phys. Reviews* **74** (2006)
17. Girvan, M., Newman, M.E.J. *Proc. Natl. Acad. Sci. USA* **99** (2002) 7821–7826
18. D. Lusseau, K. Schneider, O.J.B.P.H.E.S., Dawson, S.M.: *Behavioral Ecology and Sociobiology* **54**. Addison-Wesley (2003)
19. Adamic, L.A., Glance, N.: The political blogosphere and the 2004 us election. *WWW-2005 Workshop on the Weblogging Ecosystem* (2005)
20. Krebs, V.: Books on us politics sold by amazon.com. V. Kerbs, personal website, www.orgnet.com (2008)
21. Newman, M.E.J.: Community centrality. *Phys. Reviews* **74** (2006)
22. et. al, A.M.: Measurement and analysis of online social networks. *Internet Measurement Conference* (2007)
23. Breiger, R.L., M, K., Pattison, P.: *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers.* National Academies Press (2003)
24. Freeman, L.: Finding Social Groups: A Meta-Analysis of the Southern Women Data. (2003) 39–45
25. John Resig, Santosh Dawara, C.M.H., Teredesai, A.: Extracting social networks from instant messaging populations. In: *Workshop on Link Analysis and Group Detection (LinkKDD2004)*. (2004)
26. Adamic, L.A.: Friends and neighbors on the web. <http://www.sciencedirect.com/science/article/B6VD1-48Y6SYK-1/2/8216f71528567bedb3a194828fe68c34>
27. Goecks, J., Mynatt, E.D.: Leveraging social networks for information sharing. In: *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, New York, NY, USA, ACM (2004) 328–331
28. Khan, J.I., Shaikh, S.: Relationship algebra for computing in social networks and social network based applications. In: *WI '06: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, Washington, DC, USA, IEEE Computer Society (2006) 113–116
29. Zhou, D., Ji, X., Zha, H., Giles, C.L.: Topic evolution and social interactions: how authors effect research. In: *CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management*, New York, NY, USA, ACM (2006) 248–257
30. McCallum, A.: Information extraction, data mining and joint inference. In: *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, ACM (2006) 835–835
31. Rattenbury, T., Good, N., Naaman, M.: Towards automatic extraction of event and place semantics from flickr tags. In: *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, ACM (2007) 103–110

32. Andrade, N., Mowbray, M., Lima, A., Wagner, G., Ripeanu, M.: Influences on cooperation in bittorrent communities. In: P2PECON '05: Proceeding of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems, New York, NY, USA, ACM (2005) 111–115
33. Cai, D., Shao, Z., He, X., Yan, X., Han, J.: Mining hidden community in heterogeneous social networks. In: LinkKDD '05: Proceedings of the 3rd international workshop on Link discovery, New York, NY, USA, ACM (2005) 58–65
34. Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X.: Group formation in large social networks: membership, growth, and evolution. In: KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, ACM (2006) 44–54
35. Müller-Prove, M.: Community experience at openoffice.org. *interactions* 14(6) (2007) 47–48
36. Pfeil, U.: Online social support for older people. *SIGACCESS Access. Comput.* (88) (2007) 3–8
37. Millen, D.R., Patterson, J.F.: Stimulating social engagement in a community network. In: CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work, New York, NY, USA, ACM (2002) 306–313
38. Gr, K.: Ubiquitous hypermedia and social interaction in physical environments. In: HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia, New York, NY, USA, ACM (2006) 119–120
39. Qamra, A., Tseng, B., Chang, E.Y.: Mining blog stories using community-based and temporal clustering. In: CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management, New York, NY, USA, ACM (2006) 58–67
40. Tantipathananandh, C., Berger-Wolf, T., Kempe, D.: A framework for community identification in dynamic social networks. In: KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, ACM (2007) 717–726
41. Freeman, L.: Visualizing social networks. *Journal of Social Structure*
42. de Nooy, W., Mrvar, A., Batagelj, V.: *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, New York, NY, USA (2004)
43. Álvaro Reis Figueira, Laranjeiro, J.B.: Interaction visualization in web-based learning using igraphs. In: HT '07: Proceedings of the 18th conference on Hypertext and hypermedia, New York, NY, USA, ACM (2007) 45–46
44. Schulz, H.J., Nocke, T., Schumann, H.: A framework for visual data mining of structures. In: ACSC '06: Proceedings of the 29th Australasian Computer Science Conference, Darlinghurst, Australia, Australia, Australian Computer Society, Inc. (2006) 157–166
45. Berry, M., Hamilton, M.: Mobile computing, visual diaries, learning and communication: changes to the communicative ecology of design students through mobile computing. In: ACE '06: Proceedings of the 8th Australian conference on Computing education, Darlinghurst, Australia, Australia, Australian Computer Society, Inc. (2006) 35–44
46. Chung, W., Leung, A.: Supporting web searching of business intelligence with information visualization. In: WI '07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Washington, DC, USA, IEEE Computer Society (2007) 807–811
47. Domingos, P., Richardson, M.: Mining the network value of customers. In: KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, ACM (2001) 57–66

48. Chen, H., Atabakhsh, H., Tseng, C., Marshall, B., Kaza, S., Eggers, S., Gowda, H., Shah, A., Petersen, T., Violette, C.: Visualization in law enforcement. In: dg.o2005: Proceedings of the 2005 national conference on Digital government research, Digital Government Research Center (2005) 229–230
49. Saltz, J.S., Hiltz, S.R., Turoff, M.: Student social graphs: visualizing a student’s online social network. In: CSCW ’04: Proceedings of the 2004 ACM conference on Computer supported cooperative work, New York, NY, USA, ACM (2004) 596–599
50. Quasthoff, M., Sack, H., Meinel, C.: Why https is not enough – a signature-based architecture for trusted content on the social web. In: WI ’07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Washington, DC, USA, IEEE Computer Society (2007) 820–824
51. O’Donovan, J., Smyth, B.: Trust in recommender systems. In: IUI ’05: Proceedings of the 10th international conference on Intelligent user interfaces, New York, NY, USA, ACM (2005) 167–174
52. Kaufman, J.H., Edlund, S., Ford, D.A., Powers, C.: The social contract core. In: WWW ’02: Proceedings of the 11th international conference on World Wide Web, New York, NY, USA, ACM (2002) 210–220
53. Gefen, D.: Reflections on the dimensions of trust and trustworthiness among online consumers. *SIGMIS Database* **33**(3) (2002) 38–53
54. Sabater, J., Sierra, C.: Social regret, a reputation model based on social relations. *SIGecom Exch.* **3**(1) (2002) 44–56
55. Fullam, K.K., Barber, K.S.: Learning trust strategies in reputation exchange networks. In: AAMAS ’06: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems, New York, NY, USA, ACM (2006) 1241–1248
56. Cvrcek, D., Kumpost, M., Matyas, V., Danezis, G.: A study on the value of location privacy. In: WPES ’06: Proceedings of the 5th ACM workshop on Privacy in electronic society, New York, NY, USA, ACM (2006) 109–118
57. Chatfield, C., Hexel, R.: Privacy and community connectedness: designing intelligent environments for our cities. In: OZCHI ’07: Proceedings of the 2007 conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artifacts and environments, New York, NY, USA, ACM (2007) 265–272
58. Yu, H., Kaminsky, M., Gibbons, P.B., Flaxman, A.: Sybilguard: defending against sybil attacks via social networks. In: SIGCOMM ’06: Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications, New York, NY, USA, ACM (2006) 267–278
59. Gibson, R.: Who’s really in your top 8: network security in the age of social networking. In: SIGUCCS ’07: Proceedings of the 35th annual ACM SIGUCCS conference on User services, New York, NY, USA, ACM (2007) 131–134
60. Narayanan, R., Özisikylmaz, B., Zambreno, J., Memik, G., Choudhary, A.N.: Minebench: A benchmark suite for data mining workloads. In: IISWC. (2006) 182–188