# Social Capital in Online Communities

Matthew S. Smith
Department of Computer Science
Brigham Young University
smitty@byu.edu

## ABSTRACT

Online communities are connecting hordes of individuals and generating rich social network data. The social capital that resides within these networks is largely unknown. We propose to create a mathematical model of social capital that incorporates the mobilization of social resources through purposive actions. This includes evaluating nodes based not only on their relationships and attributes, but on their social resources as well. Investigating the costs associated with reciprocally connecting to individuals will also be assessed. The result is a quantitative model for characterizing and providing decision support on how to maximize participation within social networks.

## Categories and Subject Descriptors

I.6.0 [**Computing Methodologies**]: Simulation and Modeling; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*sociology*

## General Terms

Experimentation, Measurement

## Keywords

Social networks, social capital, implicit affinity networks

## 1. INTRODUCTION

The science of building and discovering communities is increasingly important as the Internet becomes the largest collection of ideas, personalities, and cultures in history. The continual emergence of new online communities requires better techniques for understanding these phenomena. Online communities, also referred to as neo-tribes [11], have proliferated the Internet. In particular, *the blogosphere*, the growing community of people that read and write Weblogs, has been more than doubling each year [25]. These communities represent groups of individuals connected by some well-defined, explicit relation, such as a shared medical condition in a health community, a trusted contact link in a business network, or an established friend or family relationship in a photo-sharing community. Online communities continue to rise in popularity by bringing people together to socialize, work together, and communicate.

The amount of data generated by these communities far exceeds everything collected previously. In the past, the available social network data has been limited and very static. For instance, it has been prohibitively expensive for researchers to survey individuals requesting each to name their friends, allowing a simple social network graph to be created for analysis. Nowadays, due to the increased ability to connect on the Internet, social network data is available, not only for static snapshots, but dynamically over time. The social graph that is now becoming available online is more comprehensive and pertinent than those generated from manual surveys.

The ideas surrounding social capital have been around for at least a century, however the surge of theory and research has been within the last two decades. Sociologists appear to have been most aggressive in studying the topic [17], while political scientists have made it popular [21]. The interest in social capital has since expanded to other areas including business, computer science, economics, organizational studies, and health.

Social capital within a community is grounded on:

1. relationships (e.g., see [7])
2. individuals' attributes (e.g., see [21, 9])
3. available social resources (e.g., see [17])

To exploit (1) and (2) above, we find it useful to distinguish between two types of connections among individuals, as follows.

- An *explicit* connection links individuals together based on a well-defined relationship, such as "is a friend of" or "collaborates with." Individuals thus linked are aware of the explicit connections among them.
- An *implicit* connection links individuals together based on loosely defined affinities, or inherent similarities, such as similar hobbies or shared interests. Individuals thus linked may not be aware of the similarities in attitudes and behaviors that exist among them.

We call *explicit social networks* (ESNs), social networks built from explicit connections and *implicit affinity networks* (IANs), social networks built from implicit connections. We have


*PIKM'08,* October 30, 2008, Napa Valley, California, USA.

shown elsewhere how to build IANs from individuals represented as collections of attributes and associated value sets, where links are created whenever two individuals share an attribute whose value sets overlap [26]. For example, the characterizations of Table 3.2 give rise to the IAN marked by dotted lines in Figure 1. The solid lines correspond to possible explicit connections that make up an ESN over the same set of individuals. We call a network that has both implicit and explicit links a *hybrid network*.

From the perspective of social capital, ESNs and IANs are complementary. Indeed, "social capital can be viewed as based on social similarity, the shared affiliations or activities that indicate *how* one knows someone." [1] (emphasis added). In this sense, social capital is naturally interested in implicit connections. On the other hand, social capital can really only accrue when individuals are aware of it, that is when they establish explicit connections among themselves.

In regards to (3), Lin suggests that accessing social resources within a network should consider the position of ego in hierarchical structures, the nature of the tie between ego and the other actors, and the location of the ties in the networks [17].

In our research, we propose to formalize the notion of social capital by creating a mathematical model of social capital that incorporates the mobilization of social resources through purposive actions. This includes evaluating nodes based not only on their relationships and attributes, but on their social resources as well. Investigating the costs associated with reciprocally connecting to individuals will also be assessed. The result is a quantitative model for characterizing and providing decision support on how to maximize participation within social networks.

# 2. RELATED WORK

We have organized the most relevant work related to this research under the three headings below.

## 2.1 Online Communities

The social networks generated by online communities are relationship-centered, and their analysis often assume that the network is static, or evolves in a sufficiently slow manner to make the study of snapshots relevant and meaningful. The science of building, discovering, and analyzing these communities is increasingly important as the Internet becomes the largest collection of ideas, personalities, and cultures in history.

Although useful from a practical (computational) standpoint, the assumption of a static network tends to limit the kinds of analyses that may be performed. Recently, some researchers have begun to study the actual dynamics of social network formation and evolution, leading to the discovery of several interesting patterns such as degree power laws and shrinking diameters (e.g., see [13, 15, 16, 23, 28]). It is possible to go even further by focusing on implicit affinities thus allowing the nature of the underlying relationship to vary over time. In this context, individuals are viewed as social actors characterized by a wide range of attributes and relationships among them emerge naturally as a result of commonalities across attributes.

## 2.2 Social Network Analysis

For decades, researchers have performed social network analysis. A plethora of structural properties and measures have been invented for social network analysis [31]. Interestingly, most properties and measures have been designed for static social networks. Some, however, such as nodal degree, diameter, and density, can easily be adapted to capture aspects of network evolution over time (e.g., see [31, 24]). Dynamic social network analysis techniques are increasingly important as the pertinent data becomes available.

Centrality measures have historically been used to determine the relative importance of a particular node within a network graph. For instance, Google's PageRank algorithm utilizes a form of centrality to provide ranked search results. Common centrality measures include degree, betweenness, closeness, and eigenvector centrality. *Degree centrality* consists simply of the in-degree or out-degree of a particular node. Often, high in-degree centrality represents popularity, while high out-degree centrality represents gregariousness. *Betweenness centrality* takes a different slant by calculating the shortest paths between every node within the network and assigning high values to nodes included in more shortest paths, thus signifying which nodes are most "central". *Closeness centrality* is the mean shortest path geodesic distance between the node and all other nodes reachable from it. Thus, the node with the lowest value is the closest to the most other nodes *Eigenvector centrality* is the principal eigenvector of the adjacency matrix defining the network. The eigenvector provides a score for each node within the network such that a high scoring node is one that is adjacent to nodes that are themselves high scoring. All of these centrality measures can provide a measure of node importance that is based solely on the connections in the network.

## 2.3 Social Capital

Social capital is a fundamental idea that originates in political science and sociology (e.g., see [17]). "Unlike other forms of capital, social capital is not possessed by individuals, but resides in the relationships individuals have with one another." [10]. Social capital fosters reciprocity, coordination, communication, and collaboration. It has been used to explain, for example, how certain individuals obtain more success through using their connections with other people. In an interesting study about CEO compensation, Belliveau and colleagues show that social capital plays a significant role in the level of compensation offered to CEOs [1]. In another study on social capital in the workplace, Erickson concludes that "good networks help people to get good jobs" [9].

Two main components of social capital have been defined: bonding social capital and bridging social capital [21, 22]. Bonding social capital refers to the value assigned to social networks among homogeneous groups of people. Bridging social capital refers to the value assigned to social networks among socially heterogeneous groups of people. Associations and clubs typically create bonding social capital; neighborhoods and choirs tend to create bridging social capital. Whereas bonding social capital increases through closure, as individuals strengthen existing links among themselves, bridging social capital increases through brokerage, as individuals establish new links across structural holes [6]. Erickson argues that network variety (i.e., bridging capital) is a form of social capital valuable to both employers and employees in the hiring process [9]. In order to create either bonding or bridging social capital, individuals must interact.

In general, bonding interactions are more likely to occur than bridging interactions [17]. Interacting homogeneously (i.e., bonding) "should be the expected pervasive pattern of interactions observed", because it requires the least effort [17]. On the other hand, interacting heterogeneously (i.e., bridging) demands effort due to resource differentials and the lack of shared sentiments and is therefore relatively less likely to occur [17].

As theorized by Lin, *personal* and *social resources* can be characterized for Individual actors. These resources are defined as either material goods (e.g. land, houses, car, and money) or symbolic goods (e.g., education, memberships in clubs, reputation, or fame). Personal resources (i.e., human capital) are in the possession of the individual, while social resources (i.e., social capital) are accessible through social connections [17]. Resources gained through bridging interactions are perceived to be of greater worth as they are more likely to be dissimilar than the resources already available.

Lin characterizes *access* and *mobilization* as theoretical approaches that describe how social capital is expected to produce returns [18]. Access estimates the amount of social capital (known to be) available to an individual. This approach is based on the assumption that the amount of accessible social capital largely determines the returns, without regard to the particular actions taken to use the social capital. Alternatively, the theoretical approach of mobilization reflects "a selection of one or more specific ties and their resources from the pool for a particular action at hand" [18]. For example, using a specific contact having certain resources (e.g., a highly trafficked blog, or domain-knowledge) to boost sales on an e-commerce site could be indicative of mobilized social capital.

# 3. PRELIMINARY WORK

We have begun formalizing the notion of social capital by building a mathematical model that reflects some of the main requirements (e.g., bonding and bridging) utilized in previous attempts (e.g., see [21]). There are several key features to our model, which we detail in the following sections.

1. The distinction between potential and actual social capital is clear.

2. Bonding and bridging social capital are de-coupled (in particular, they need not be reciprocal).

3. The model can be readily applied to available community data.

## 3.1 Potential vs. Actual Social Capital

Because individuals are complex entities whose attitudes and behaviors are prone to change over time, IANs are intrinsically dynamic, evolving with such things as their participants' age, occupation, interests, and life's circumstances (e.g., marriage, retirement). As new participants create and current ones update their profile, the network continually and automatically shifts. Indeed, small changes to one individual's profile may have many (unexpected) effects on the overall structure of the IAN.

Every time an individual's profile changes (e.g., by adding a new attribute or a new value to an existing attribute) the corresponding update creates an opportunity for new implicit connections to arise. Some are created immediately with individuals who share aspects of the updated profile, while others are established later as other individuals undergo related changes. In that sense, IANs capture the *potential* for social capital.

| | | **IAN Link** | |
|---|---|---|---|
| | | **Yes** | **No** |
| **ESN Link** | **Yes** | Actual Bonding | Actual Bridging |
| | **No** | Potential Bonding | Potential Bridging |

**Table 1: Potential vs. Actual Social Capital in Hybrid Networks**

On the other hand, changes to an ESN are more purposeful and localized. An individual chooses precisely which other individuals to connect with. Such changes have a direct impact on the social capital of the underlying community. Hence, we can interpret IANs as capturing the potential for social capital, and ESNs —overlayed on IANs— as measuring actual social capital. Moreover, depending on the kinds of implicit connections that may exist among the same individuals, one can determine what form of social capital is being affected and how.

Table 1 summarizes the relationship between potential and actual social capital based on the connections of a hybrid network.

The presence of both implicit and explicit connections between individuals indicates actual bonding social capital as like individuals (IAN links) are linked to one another (ESN links). When only implicit connections exist among individuals, one observes only potential for bonding social capital. For example, in Figure 1, Amy and Bob have linked only implicitly, indicating that there is a potential bond that would be realized if they were to become friends. The absence of implicit connections when explicit connections exist is an indicator of actual bridging capital as diverse individuals (no IAN links) are linked to one another (ESN links). Finally, the absence of either type of connections highlights the potential for bridging social capital, that would be realized when ESN links are established.[1]

Table 1 makes it clear that there is no *actual* bonding nor bridging social capital without explicit links. The amount of similarity implicit among individuals determines the amount of bridging and/or bonding that occurs within the network as explicit links are made or removed. Both implicit and explicit connections are therefore necessary to calculate the network's social capital.

## 3.2 Bonding and Bridging Social Capital

Recall that a hybrid social network consists of an implicit affinity network (IAN) and an explicit social network (ESN) defined over the same set of individuals. Hybrid networks can thus be visualized by overlaying ESNs onto corresponding IANs. In social network analysis terminology, a hybrid network is a multigraph having an explicit and implicit relation among actors (e.g., see Figure 1).

In [27], we showed how to derive an effective mathematical formulation of social capital by exploiting the complementarity of IANs and ESNs. We formalized aspects of social capital to show precisely when the community was either

[1]Note here that if IAN links were established first, this situation would of course turn into one of potential bonding social capital, rather than bridging social capital.

| Individual | Attribute Value Sets |
|---|---|
| Amy | {Cancer (C), Smoke (S)} |
| Bob | {Cancer (C), Bald (B)} |
| Cheryl | {Cancer (C), Smoke (S)} |
| Dan | {Smoke (S)} |
| Ed | {Bald (B)} |

**Table 2: Sample Individuals and Attributes**

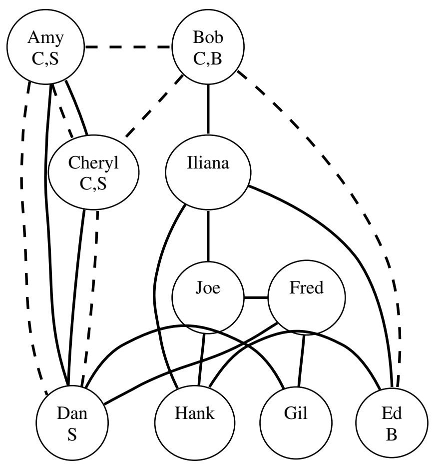


**Figure 1: Sample Hybrid Network**

bonding or bridging for the particular context. These analyses highlighted the effects that individual changes had on the community; the occurrence of an individual bridging out or showing their interest in new areas was of particular interest. We define *actual bonding social capital* between $i$ and $j$ as the product of the strength of the implicit edge (i.e., potential bonding social capital) by the strength of the explicit edge. That is,

$$bonding(i,j) = s_{ij}^{IAN} s_{ij}^{ESN}$$

Hence, as expected, if either the implicit strength or the explicit strength is 0, that is, if either $i$ and $j$ have nothing in common or they do not know about each other, then there is no bonding social capital. On the other hand, if both implicit and explicit strengths are 1, then bonding is also maximum at 1. Any other configuration reflects the amount of bonding social capital between $i$ and $j$.

Bonding social capital for an entire social network is the sum, over all edges, of the actual bonding social capital divided by the sum, over all edges, of the potential bonding social capital, as follows.

$$bonding = \frac{\sum_{i,j} bonding(i,j)}{\sum_{i,j} s_{ij}^{IAN}}$$

Conversely, potential bridging social capital between two nodes $i$ and $j$ is simply $1 - s_{ij}^{IAN}$. The more dissimilar the two nodes are the larger the potential for bridging. Then, actual bridging social capital between $i$ and $j$ can be defined as the product of the reciprocal of the strength of the implicit edge (i.e., potential bridging social capital) by the strength of the explicit edge. That is,

$$bridging(i,j) = (1 - s_{ij}^{IAN}) s_{ij}^{ESN}$$

If both implicit and explicit strengths are 0, then there is clearly no bridging social capital. However, potential bridging is maximum at 1, since the individuals have nothing in common. Similarly, if both implicit and explicit strengths are 1, then there is still no bridging social capital, as the individuals are homogeneous. Bridging social capital is maximum at 1 only when explicit strength is 1 but implicit strength is 0. Any other configuration reflects the amount of bridging social capital between $i$ and $j$.

Bridging social capital for an entire social network is the sum, over all edges, of the actual bridging social capital divided by the sum, over all edges, of the potential bridging social capital, as follows.

$$bridging = \frac{\sum_{i,j} bridging(i,j)}{\sum_{i,j} 1 - s_{ij}^{IAN}}$$

From the above formulation, we can see that bonding social capital and bridging social capital are not reciprocal of each other. Instead, their values are completely decoupled, allowing each to vary independently of the other. The motivation for such a decoupling is found in the following puzzle, posed by Putnam (Personal Communication).

> Too often, without really thinking about it, we assume that bridging social capital and bonding social capital are inversely correlated in a kind of zero-sum relationship —if I have lots of bonding ties, I must have few bridging ties, and vice versa. As an empirical matter, that assumption is often false. In the US, for example, whites who have more non-white friends also have more white friends. (This generalization is based on our extensive analysis of the 2000 Social Capital Community Benchmark Survey.) In other words, high bonding might well be compatible with high bridging, and low bonding with low bridging. Of course, one can artificially create a zero-sum relationship between bridging and bonding by asking what proportion of (say) friendships are bridging or bonding, or on relative trust of in-groups and out-groups, but the result is a mathematical trick, not an empirical finding.

Our formulation is not merely a mathematical trick, but is rooted in what we understand to be the nature of actual vs. potential bonding and bridging social capital.

In [27], we report on the construction of a large hybrid social network in the blogosphere and show how social capital may be used to highlight important properties of the network, as well as influence its behavior.

This allowed us to show how a hybrid network within the blogosphere is not only connected explicitly by the blogs they link to, but implicitly by the topics they choose to write about. We showed that these are not necessarily the same groups of blogs, suggesting the emergence of new sub-communities through bonding. Identifying these sub-communities has application in many domains. For example, the medical community could use the hybrid graph to help patient communities having implicit connections to connect explicitly, thus forming support groups. The political domain could use hybrid graphs to determine where political canidates should concentrate grass roots efforts online. Furthermore, the expanding blogosphere creates numerous social capital applications across many unique domains.

# 4. PROPOSED WORK

In this section, we provide an overview of the proposed work, along with areas where experiments will be conducted.

## 4.1 Overview

Up to this point, we have described social capital without discussing the role that specific social resources have within social networks. Recall that Lin characterizes how social capital is expected to produce returns through accessible social resources that can be mobilized [17, 18]. The bonding and bridging measures focused on in our preliminary work provide an intuitive sense of the homogeneity and connectedness of a community over time. However, these metrics alone fail to account for how social capital is expected to produce returns. In order to be able to give an accounting of how social capital is being used within a community, specific resources available through social connections must be considered.

Our research will address the following:

1. **Evaluate nodes based on their relationships, attributes, and social resources.** An important area of research is improving our understanding about how much social capital each individual within a given community has. Flow models [12] may be used to evaluate nodes to include social resources. Flow models incorporate a measure of prestige based on explicit links, however, they do not consider how similar nodes are (i.e., implicit affinities based on individuals' attributes).

2. **Identify a set of measurable social resources accessible within online communities.** These resources might include referring visitors, job information, and exposure to ideas or products. These social resources will be chosen within the context of a particular domain so that the results of this research can be directly applied to existing online social networks. Part of the challenge will be to make sure that many of these social resources are measurable within online communities. It is possible that some simple measures, such as unique visitors to a site, which are already being collected, may serve as a good starting point and could easily be used with existing data.

3. **Formalize the notion of accessible and mobilized social resources.** Our current social capital models will be extended to include social resources identified (in the previous point). These modifications will provide a measure based upon the social capital accessible to an individual over time and a mechanism for dynamically tracking social resources as they are mobilized. These additions will provide a more accurate assessment of the social capital available to each individual and within a given online community.

4. **Assess the cost associated with reciprocally connecting to an individual.** Expanding an individual's social network, to maximize their social capital, requires the knowledge of the social capital to be obtained and the cost associated with obtaining it. Bonding should be less inexpensive and less profitable, while bridging should be more expensive and have higher returns. PageRank [20], might be used to compare to, as it could be used heuristically to guess how valuable connecting to a blog might be. However, the cost of connecting with more prestigious nodes is not necessarily more expensive than connecting with less prestigious nodes.

5. **Run experiments to validate our formal models of social capital.** To validate the models above, experiments will need to be conducted that compare the estimated social capital to known values of social capital within publicly available community data sets. Our experiments are detailed in the next section.

## 4.2 Experiments

In this section, we describe two areas where online communities would be studied.

### *4.2.1 Blogosphere*

Experiments within the blogosphere can be conducted to increase our understanding of this important phenomenon. For these experiments, we extract an explicit network and generate an implicit affinity network based on blog links and entries. Rather than modeling blog communities based solely on explicit hyper-linked cross-references as in [14], we model them with an implicit overlay, based on blog content. We have performed preliminary experiments in this domain, which demonstrate promise [26].

Here, a *blog* refers to a single online journal, a *blog entry* refers to an entry in such a journal, and a *blogger* refers to an author of a blog. To build an IAN from the space of blogs, we represent each blogger as an individual (a single blog may have multiple authors), with attributes and associated values that we mine from the individual's blog entries.

Determining blogger's attributes is a significant sub-task that allows for various feature extraction techniques to be used. We propose to use probabilistic Dirichlet processes [30, 3, 4, 2] to discover attributes that represent the underlying concepts that bloggers tend to write about, rather than simply the terms they choose to use. We reserve using and the comparison of other IAN attribute extraction techniques for future work.

As the entire blogosphere is difficult, if not impossible, to capture and study, our experiments will focus on a sample of the blogosphere. For example, a community $C$ might be blogs from the public reading list such as Robert Scoble, an influential technology journalist. Another possibility, would be to randomly select blogs from one or more blog aggregators (e.g., Technorati, Google Reader, Bloglines). Yet another possibility, would be to choose a blog to start from and then spider all of the explicit links within the blog recursively up to some finite level.

Finally, we note another important potential benefit of IANs in the blogosphere. Explicit links, captured by hyperlinked cross-references, are "already known" to the bloggers, while our links are implicit and therefore may not be known to bloggers. In particular, bloggers may not realize how or where they fit within a particular community based on blog entry content. Our implicit links (i.e., affinities) are derived from text in a blogger's blog entries. Thus, an IAN might be used to inform bloggers as to where they reside in the implicit network. For example, are they blogging about things that few others in the community are (i.e., bridging) or are they blogging about the same things that many others are (i.e., bonding opportunity)? The notion of social capital is used to understand the state of the community.

#### 4.2.2 *Medical Communities*

Online medical communities are also becoming increasingly common. In general, they are designed to enable patients to discuss symptoms, treatments, and get support. Daily Strength [8], for example, is a community that offers support groups on many different medical conditions, including those that are less common. For some of the more common conditions, independent communities have been created, for example, there is a separate community for breast cancer [5], lung cancer [19], testicular cancer [29], and bladder cancer [5]. In addition, some of these communities incorporate doctors and other experts that provide advice and treatment options. Although, medical support groups have existed for some time, only recently have they become available online, thus offering many unexploited affinities among individuals.

We propose to conduct experiments within the medical domain to show the direct applicability to success of these communities. Individuals within these communities often share the challenges they face in hopes that they can find others in their same situation. However, sometimes these individuals are not able to find the desired support group and remain isolated even though others with related challenges exist within the community. The experiments we plan to conduct will measure the social capital within these communities, thus allowing groups with high or low social capital to be identified.

## 5. CONCLUSION

We have proposed to create a quantitative model for characterizing and providing decision support on how to maximize participation within social networks. This entails quantifying the cost of reciprocally connecting to other nodes and evaluating nodes based on their relationships, attributes, and social resources. The result is a mathematical model of social capital that incorporates the mobilization of social resources through purposive actions.

The results of this research will be useful in answering important questions in social network analysis, such as:

- Who should a newcomer to a community attempt to connect with?
- How much social capital does each individual within a social network have access to?
- How much social capital does a group (or community) of individuals have available?
- What social resources were mobilized within the community during the past month?
- Which individuals tend to mobilize the most social resources?

## 6. REFERENCES


[1] M. Belliveau, C. I. O'Reilly, and J. Wade. Social capital at the top: Effects of social similarity and status on CEO compensation. *Academy of Management Journal*, 39(6):1568–1593, 1996.

[2] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation, 2003.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. In *Neural Information Processing Systems 14*, 2001.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[5] breastcancer.org. Breast cancer support and community. Online at: http://www.breastcancer.org/support.html, March 2007.

[6] R. Burt. Network duality of social capital. In V. Bartkus and J. H. Davis, editors, *Reaching In, Reaching Out: Multidisciplinary Perspectives on Social Capital*. Edward Elgar Publishing, 2008.

[7] R. S. Burt. *Brokerage and Closure*. Oxford University Press, 2005.

[8] DailyStrength Inc. Health support groups at dailystrength. Online at: http://dailystrength.org/, March 2007.

[9] B. H. Erickson. Good networks and good jobs: The value of social capital to employers and employees. In N. Lin, K. S. Cook, and R. S. Burt, editors, *Social Capital: Theory and Research*, chapter 6, pages 127–158. Aldine Transaction, 2004.

[10] FAST. Social capital: Social capital as a theoretical construct. Families and Schools Together, Wisconsin Center for Education Research. Available online at http://fast.wceruw.org/theory/socialcap.htm, 2006.

[11] G. Johnson and P. Ambrose. Neo-tribes: The power and potential of online communities in health care. *Communications of the ACM*, 49(1):107–113, 2006.

[12] A. Josang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, 2007.

[13] J. Katz. Scale independent bibliometric indicators. *Measurement: Interdisciplinary Research and Perspectives*, 3:24–28, 2005.

[14] R. Kumar, J. Novak, P. Raghavan, and A. Tomkins. On the bursty evolution of blogspace. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 568–576, New York, NY, USA, 2003. ACM Press.

[15] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 611–617, 2006.

[16] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 177–187, 2005.

[17] N. Lin. *Social Capital: A Theory of Social Structure and Action*. NY: Cambridge University Press, 2001.

[18] N. Lin. A network theory of social capital. In

D. Castiglione, J. W. van Deth, and G. Wolleb, editors, *Handbook on Social Capital*. Oxford University Press, 2008.

[19] Lung Cancer Support Community. Lung cancer support community. Online at: http://lchelp.org/, March 2007.

[20] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.

[21] R. D. Putnam. *Bowling Alone: the Collapse and Revival of American Community*. Simon & Schuster, 2000.

[22] R. D. Putnam and L. M. Feldstein. *Better Together: Restoring the American Community*. Simon & Schuster, 2003.

[23] S. Redner. Citation statistics from 110 years of *Physical Review*. *Physics Today*, 58:49–54, 2005.

[24] J. P. Scott. *Social Network Analysis: A Handbook*. Sage, Thousand Oaks, CA, 2000.

[25] D. Sifry. State of the blogosphere. Online at: http://www.sifry.com/alerts/archives/000436.html, August 2006.

[26] M. Smith, C. Giraud-Carrier, and B. Judkins. Implicit Affinity Networks. In *Proceedings of Seventeenth Annual Workshop on Information Technologies and Systems*, pages 1–6, December 2007.

[27] M. Smith, N. Purser, and C. Giraud-Carrier. Social Capital in the Blogosphere: A Case Study. Number 06 in SS-08. AAAI Technical Report, The AAAI Press, Menlo Park, California, 2008.

[28] C. Tantipathananandh, T. Berger-Wolf, and D. Kempe. 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*, pages 717–726, New York, NY, USA, 2007. ACM.

[29] TC-Cancer.com. Testicular cancer support forum. Online at: http://www.tc-cancer.com/forum/, March 2007.

[30] Y. Teh, M. Jordan, M. Beal, and D. Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

[31] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.