Identifying communities in blogs: roles for social network analysis and survey instruments

Alvin Chin*

Department of Computer Science
University of Toronto
10 King's College Road
Toronto, Ontario M5S 3G4, Canada
Fax: 416–978–4765
E-mail: achin@cs.toronto.edu
*Corresponding author

Mark Chignell

Department of Mechanical and Industrial Engineering
University of Toronto
5 King's College Road
Toronto, Ontario M5S 3G8, Canada
Fax: 416–978–3453
E-mail: chignel@mie.utoronto.ca

Abstract: We examine the problem of identifying, measuring and evaluating communities in blogs. In our approach, we use sense of community (McMillan and Chavis, 1986) and centrality measures derived from social network analysis. We then describe a method for identifying communities in blogs using both sense of community measurement and Social Network Analysis (SNA). We apply this framework to a blog on Canadian independent music. Hypotheses are developed concerning the specific structural properties of social networks formed by blogging relationships and how those properties are related to the sense of community experienced by the participants (authors and readers of the blogs). Finally, we discuss the implications for blog implementation and use.

Keywords: blogs; virtual communities; sense of community; social hypertext; Social Network Analysis; SNA.

Biographical notes: Alvin Chin is currently a 4th year PhD student in Computer Science at the University of Toronto in the Interactive Media Lab, working under Professor Mark Chignell. He graduated with a Bachelor's degree in Computer Engineering and a Master's degree in Electrical and Computer Engineering from the University of Waterloo. He has worked two-and-a-half years full time in industry with researching and demonstrating new and emerging technologies in the wireless and pervasive computing area, especially with Bluetooth and 802.11. His current research interests include

social networking, computer-supported collaborative work, human-computer interaction, context-aware computing, and mobile and wireless computing. He can be blogged at http://www.the-gadgetman.com/blog.

Mark Chignell is a Professor of Mechanical and Industrial Engineering at the University of Toronto and Director of the Interactive Media Laboratory. He is also an Associate Director of the Knowledge Media Design Institute at the University of Toronto, and a Visiting Scientist both at the IBM Centre for Advanced Studies (Toronto Lab) and Keio University in Tokyo (Shonan-Fujisawa campus). He holds a PhD in Psychology from the University of Canterbury (New Zealand) and an MASc in Industrial and Systems Engineering from the Ohio State University. He conducts research in a variety of areas with a current focus on interfaces and technologies for communication and collaboration.

1   Introduction

Weblogs (blogs) adapt web technology to allow for instant, updated and frequent communication of information such as events, personal interests, thoughts, and news. The conversational nature of blogs can be used to uniquely identify topics and similar areas of interest (Herring *et al.*, 2005). These similarities can then be used to build (virtual) community, and specifically identify virtual community, in blogs.

Physical communities provide an imperfect model of virtual community. As Rheingold (1993) noted: "People who use computers to communicate, form friendships that sometimes form the basis of communities, but you have to be careful to not mistake the tool for the task and think that just writing words on a screen is the same thing as real community". Wellman and Guilia (1999) remarked that online and 'real-world' communities are not synonymous, and neither are they a 'zero-sum' game (participation in one does not automatically reduce participation in the other). Clearly, virtual communities differ in some ways from real communities, yet how can and should they be characterised?

A common theme that has emerged from past research is that the concepts of sense of community and virtual settlement are prerequisites to finding virtual community (McMillan and Chavis, 1986; Jones, 1997). These prerequisites can then be supplemented with links from blogs and clustering algorithms to indicate the shape or structures of potentially overlapping communities.

Our goal in this paper is to provide theoretical and methodological infrastructure for measuring online (virtual) community. The approach is based on a social hypertext model of blogging, where social networks of bloggers can be inferred from the comments that the bloggers post on each other's blogs. We propose a framework and preliminary model for finding community in blogs that uses sense of community along with network and link analysis approaches.

The paper is organised as follows. After this introduction, Section 2 provides background and related work. Section 3 presents our framework and model for finding and measuring communities in blogs. Section 4 applies the framework to a case study of an independent music blog. Conclusions and recommendations are made in Section 5.

2 Background: virtual community, sense of community, and social network analysis

This section reviews literature from the domains of virtual community, sense of community and social network analysis, describing briefly how each can be used to find community in blogs.

2.1 Virtual community

Community is an ambiguous term with over 120 definitions noted by Poplin (1979). According to the Wikipedia (2006) entry as of this writing, "A community is an amalgamation of living things that share an environment… In human communities, intent, belief, resources, preferences, needs…may be present and common…, but the definitive driver of community is that all individual subjects in the mix have something in common". With the growth of the internet, online groups have emerged and communities have begun to form from social interactions in an electronic environment. This type of community is known as a virtual community.

Finding community in virtual environments, such as e-mail and newsgroups, is difficult because there is no physical boundary and no physical, human interaction to observe. However, there is a wealth of literature demonstrating that forms of communities exist in virtual environments (Allen and Dillman, 1994; Kim, 2000; Kumar *et al.*, 1999; Kleinberg, 1998; Rheingold, 1993; ten Thij *et al.*, 2006; Wellman and Guilia, 1999). Virtual communities have been described by Rheingold (1993) as "social aggregations that emerge from the Net when enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace". Thus we may infer that a virtual community is a collection of online mediated personal relationships that are motivated or facilitated by a common interest that is shared across the members of that virtual community. With the web becoming a ubiquitous medium, virtual communities can now be found in web-based environments and they are called web-based communities (Martins, 2006; ten Thij *et al.*, 2006; Preece *et al.*, 2004). Korica *et al.* (2006) demonstrate the importance of electronic or virtual communities on the web for forming relationships, solving problems, searching for information, or conversing with others. Virtual communities can be formed on the web because the web is a form of social hypertext (Erickson, 1996) where the hyperlinks in the hypertext can indicate a social relationship between web pages.

Blogs are an excellent example of social hypertext because the comments from blogs are hyperlinks to other sources on the web and to the commenter's blog. Since the comments are associated with a particular blog's post and are in chronological order, blogs can facilitate members' social interactions (Nardi *et al.*, 2004a; Blanchard, 2004) and provide conversation as studied by Herring *et al.* (2005).

Prior to this writing, a few studies have looked into identifying virtual communities in blogs. Methods such as content analysis examine the content of blog posts and comments to find norms that indicate membership rules (Wei, 2004) or elicit feelings of community (Nardi *et al.*, 2004b, pp.41–46). Participant interviews and surveys (Nardi *et al.*, 2004a, pp.222–231) have also been used to identify virtual community. However, the difficulty

with previous methods for identifying community is that they involve opinions from bloggers themselves, which are highly subjective, time consuming to collect, and prone to erroneous interpretation. Since blogs are hypertext, algorithms from searching on the web (Kleinberg, 1998; Kumar *et al.*, 1999; Flake *et al.*, 2002) can be adapted for the task of identifying communities in blogs. The time structure of blogs can also be examined using methods such as Kleinberg's bursty algorithm (Kumar *et al.*, 2003) and clustering techniques such as Kohonen's Self-Organising Map (Merelo-Guervos *et al.*, 2004). However, while there are many algorithms that can in principle discover properties related to the construct of community, they need to be validated and calibrated against the amount of community actually experienced by participants. In addition, since community may be a multidimensional construct, mappings need to be identified between relevant algorithms and the subscales or dimensions within the overall community construct that they actually measure.

2.2   *Sense of community*

To measure and calibrate community, we can examine the wealth of literature on established methods from psychology and sociology for finding community, whether it be in physical communities (where face-to-face meetings and phone conversations are the norm) or virtual communities, such as e-mail and newsgroups. One of the major themes to emerge from this literature is that members of a community experience a Sense of Community (SOC) which is "a feeling that members [of a group] have of belonging, a feeling that members matter to one another and to the group, and a shared faith the members' needs will be met through their commitment to be together" (McMillan and Chavis, 1986). According to McMillan and Chavis, there are four features of sense of community which are:

1   Feelings of membership – feelings of belonging to, and identification with the group arising from community boundaries.

2   Feelings of influence – feelings of having influence on, and being influenced by, the community that emerge from enforcing and challenging norms within the group.

3   Reinforcement of needs – feelings of support from others, status in the group, and meeting other's needs while having one's own needs met.

4   Shared emotional connection – feelings of having a relationship and shared connection with others due to frequent and high quality interaction.

Studies of sense of community have not been widely applied to virtual communities. However, researchers have identified evidence of membership, influence, reinforcement of needs, and shared emotional connection, as reviewed by Blanchard and Markus (2004). Sense of community has been studied in blogs by researchers such as Blanchard (2003) and Efimova and Hendrick (2005). Blanchard suggested that in order for a blog to be a virtual community, it has to exhibit the characteristics of a 'virtual settlement', that is, a virtual place in which people interact by examining artefacts. Efimova and Hendrick (2005) applied this technique and used blog reading patterns, linking patterns, blog conversations and blogger directories as artefacts (indicators) of blog community.

Chavis (1986) created a score for measuring sense of community which involves issuing a survey of 12 questions and computing scores along the features of sense of community according to the SOC index and subscales. While the SOC index for sense of community is based on observing artefacts in a virtual settlement, the above techniques do not take into account the structural linking behaviour that is inherent in the commenting features of blogs. This raises the possibility (considered in Section 2.3) that analysis of the social networks formed by groups of blogs that link to and comment on each other, may also generate useful insights into the formation of community in blogs.

2.3 Social network analysis

The interlinking between blogs is used as a way to indicate relationships and determine conversations that can potentially identify communities (Herring *et al.*, 2005; Efimova and Hendrick, 2005). Social network analysis (or network analysis) exploits the link structure and examines the roles and behaviour of nodes on other nodes in the network, and on the network as a whole. Using tools such as UCINET (Borgatti *et al.*, 2002) and Pajek (de Nooy *et al.*, 2005), social networks can be visualised, and measures potentially related to aspects of community can be calculated.

Network analysis has been used to study the relationships and structures of networks in e-mail, newsgroups and the web. Network centrality (Freeman, 1978) can be used to measure community in networks such as a blog, since it is a measure of how important or central an individual node is to the network. Equations for calculating each of these types of centrality may be found in the paper by Freeman (1978). Degree centrality measures the network activity of a node and gives an indication as to which people are influential and may be useful nodes to focus on in community building. Closeness centrality measures how many steps on average it takes to reach every other node in the network. A node with high closeness centrality can most efficiently make contact with other nodes in the network. Betweenness centrality measures the extent to which a node can act as an intermediary or broker to other nodes in the network. For example if node A connects to node C (and node C has high betweenness centrality), then node A can reach other nodes through node C, thus extending node A's community to include node C's community.

A number of researchers have proposed using network analysis as a method for identifying community. Given the large number of social network analytic measures that can be calculated, there is as yet no consensus on which measures may be most closely related to different aspects of community. Tyler *et al.* (2005) used an algorithm based on betweenness centrality to identify possible communities, while Fisher (2005) used ego-centric networks and degree distribution to look for social structure in Usenet newsgroups. Other researchers have also used network analysis to identify communities in blogs (Efimova *et al.*, 2005; Herring *et al.*, 2005). However, after a number of promising forays into this topic, a more formal approach is needed that looks for explicit and well-calibrated mappings between social network analytic measures and the corresponding sense of community that is experienced by the readers and authors of the associated blogs. In this paper, we will take a first step towards developing this more formal approach, using a framework and model that is described in Section 3.

## 3   Framework and model for identifying and measuring virtual community in blogs

In order to identify and measure virtual community in blogs, we need to look for patterns of community activity in the structures and conversations formed by blogs, so that we can formalise properties of a blog community, and then provide measures of community that link formal properties (such as social network analytic centrality) to experienced sense of community. These measures can then be used as diagnostic tools for comparing blogs and for building better communities in blogs.

Based on this methodological motivation, we propose a framework and model for measuring and evaluating communities in existing and newly created blogs.

3.1   Framework

Our framework is illustrated in Figure 1 and consists of three processes. The first main step in this framework is to *identify community* in the blogs, typically represented not by a single community, but as a set of blog communities. Visualisation of network structure is one way to accomplish this step, as explained in Section 3.2.2. Second, one can *measure community* in the set of blog communities by creating a community score. Third, one may compare the community score for the blog and the set of blog communities with others in order to *evaluate community*. The resulting feedback and recommendations from evaluating the community may then be used to identify strategies for further improving the amount of community in the blogs.

Figure 1   Framework for measuring and evaluating community in blogs
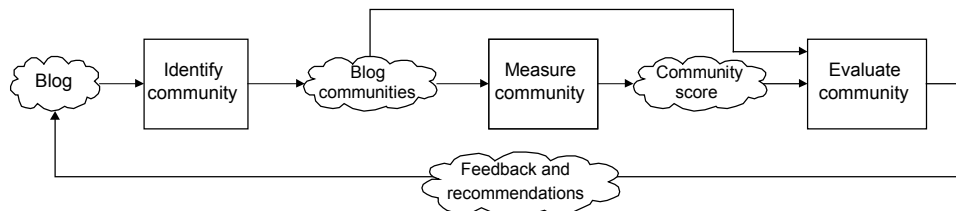

3.1.1   Identify community

The first process is to discover community in the blog and we use sense of community as our concept of community, as followed by Nardi *et al*. (2004b, pp.41–46), Wei (2004), and Efimova and Hendrick (2005). We illustrate this process in Figure 2.

We distribute a SOC survey to visitors of the blogs, following the approach of Chavis (1986) and Blanchard (2004). A blog crawler then mines the blogs and outputs the links between blogs as a social network. The survey results and output from the blog crawler are then used to identify blog communities.

Figure 2   Identifying the community
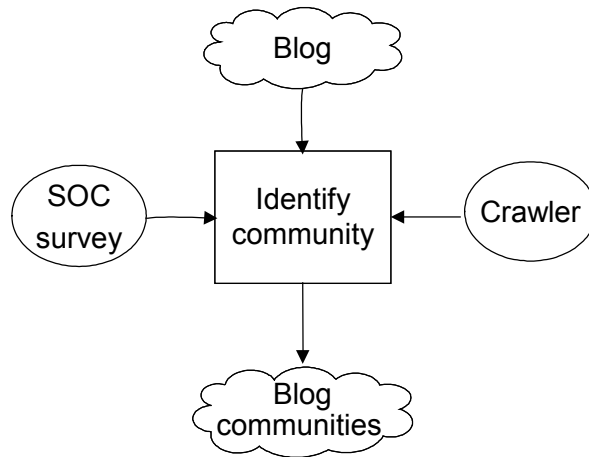
3.1.2 *Measure community*

In this process as shown in Figure 3, we create a measure of community for the blog communities identified in the previous step. We calculate the SOC measures from the SOC survey according to Chavis (1986) and compute the Social Network Analysis (SNA) measures of network centrality from the blog social network. We combine network centrality with the SOC measures to create an overall score of a measure of community for a particular blog and the blog network.

Figure 3   Measuring the community
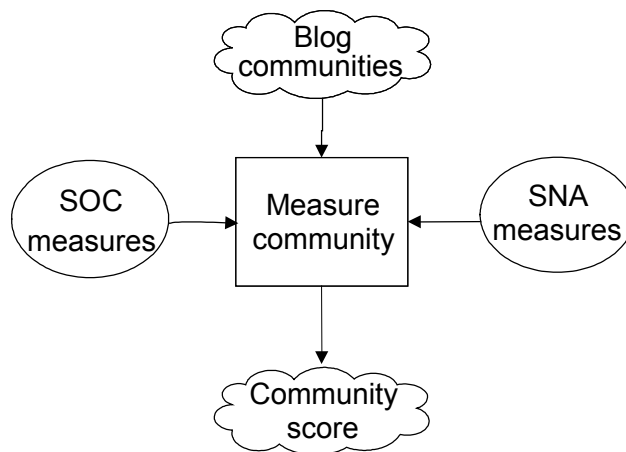
3.1.3 Evaluate community

After creating a community score for the blog, we can then evaluate community as in
Figure 4. We compare the blog's community score and the set of blog communities
identified from the *Identify Community* process, with the community score and the set of
blog communities from other blogs, in order to create feedback and recommendations on
how to improve the building of community in the blog.

Figure 4   Evaluating the community

3.2   A sense of community model for finding community in blogs

Having described our framework in the previous subsection, we now describe our
methodology for identifying community in blogs. Then, we use McMillan and Chavis'
(1986) sense of community to create a model for identifying blogs that form a
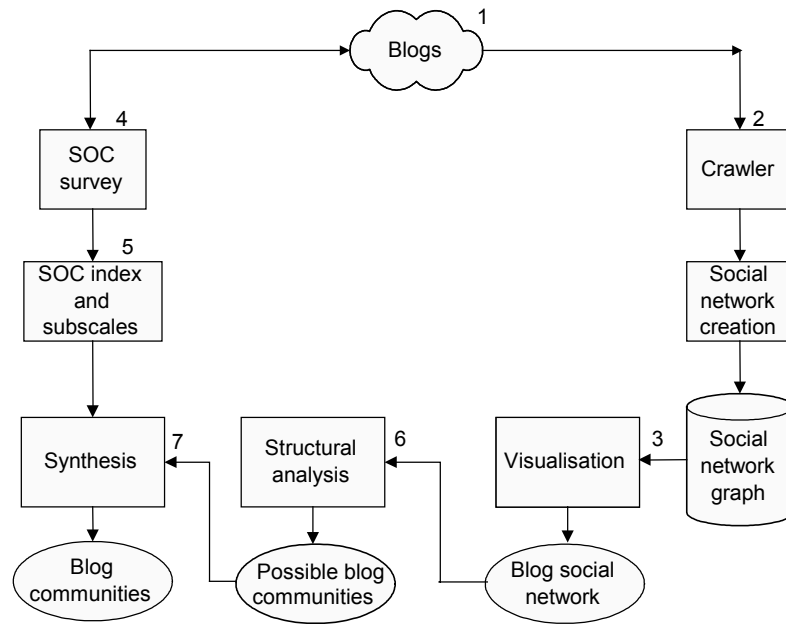blog community.

3.2.1 Methodology

Our methodology is divided into seven steps as illustrated in Figure 5.

Step 1   Involves selecting the blog or blogs that we want to find community in.

Step 2   Involves crawling those blogs and recording the links between blogs through
          posts and comments to create a graph that forms the social network.

Step 3   We visualise the social network as a directed graph.

Step 4   We administer a SOC survey.

Step 5   We compute the total SOC index and SOC subscales.

Step 6   We use structural analysis to filter out those blogs that are not part of
          any community and identify structures of possible blog communities from
          our model.

Step 7   We analyse the SOC scores and compute measures of network centrality in order to highlight those combinations of blogs that seem to have the strongest degree of community associated with them.

We discuss the details of our model for finding community in blogs below.

Figure 5   Methodology for identifying community in blogs

3.2.2 Model

We build a model as a template to identify structures from the blog social network that indicate evidence of community. For each characteristic of sense of community, we identify patterns from visualising the blog social network (we call visualisation indicators) and correlate that with measures from social network analysis (social network analysis indicators). This can provide us with the first step for creating a quantitative measure for community. Table 1 shows the indicators that we propose for identifying sense of community in blogs.

Membership can be detected in a blog network by reciprocal links (Efimova and Hendrick, 2005). Figure 6(a) shows the structure for links found in the blog social network. Figure 6(b) shows the reciprocal link where blogger B comments on blogger A's blog *x* times and blogger A comments on blogger B's blog *y* times. The star network can be used as a structure for membership, as illustrated in Figure 7. From social network analysis, we can use degree centrality as an indicator for membership because this gives an indication as to which people are influential and could be possible nodes to be included in a community.

Table 1   Model for finding sense of community in blogs

SOC characteristic	Visualisation indicator	Social network analysis indicator
Membership	Reciprocal links with each directed link having frequency $\geq k$	Degree centrality
	Star network	
Influence	Reciprocal links	Betweenness centrality
	Brokers or bridges	
Reinforcement of needs	High sum of frequencies in the link compared to others	Closeness centrality
Shared emotional connection	Triangles	k-cores
	Completely connected graphs	

Figure 6   Reciprocal link as structure for membership where (a) shows the individual links and (b) shows the reciprocal link

**Figure 7**   Star network as structure for membership in sense of community
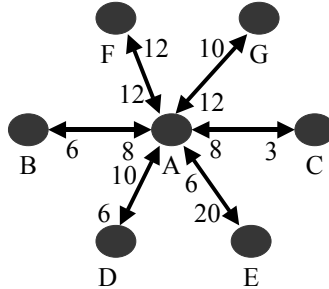
Influence can be detected in the network by finding reciprocal links in blogs because according to McMillan and Chavis (1986), influence in a community is bidirectional in nature. Blogs acting as brokers or bridges like blogger B in Figure 8, for example, are structural indicators of influence because blogger B can influence blogger A to allow A's community to reach B's community. Betweenness centrality can be used to measure the capacity of a blog's brokerage role such as B in Figure 8 as a social network analysis indicator. A blog with a high betweenness centrality can exert a high degree of influence on others in the community.

Finding structures for *reinforcement of needs* in the blog social network is difficult since the reciprocal links do not reveal the content of the blog comment on the blog post from which the link is created. However, the high frequency that is inherent in certain links in the network, as compared to others, can be a good indication that there is an increasing acknowledged interdependence between the two bloggers in the link. In fact,

frequencies of interaction can be used to indicate weak or strong ties in a social network (Paolillo, 1999). From social network analysis, closeness centrality is related to the construct of interdependence and reinforcement of needs because, for example, if blog C has high closeness centrality and blog A has a need, then A can contact C who can refer A to one of the members of C's community.

Figure 8 Example structure of influence where B can influence A

A's community   B's community
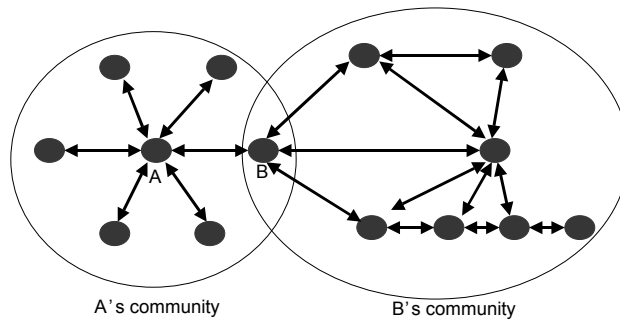
To detect *shared emotional connection*, one can identify intense clusters of interaction and identify groups in the network where every blog connects to every other blog in the group, as in the completely connected graph shown in Figure 9, forming a triangle. From social network analysis, intense clusters can be identified by k-cores in which the minimum degree of each node in the k-core is at least k. Figure 9 shows an example of a 2-core where *k* is equal to 2.

Figure 9 Example of a completed connected graph and 2-core as structure for shared emotional connection
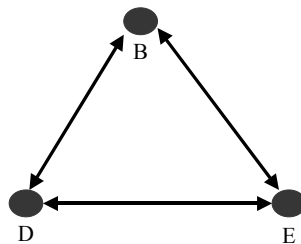
4 Case study: indie music blog

In this section, we apply our framework from Section 3 to an independent (indie) music blog which we created on MSN Spaces (now called Windows Live Spaces), a popular blog website hosted by Microsoft. This blog showcases upcoming indie bands from an indie music website and is designed to promote and add value to the indie music virtual community. We first describe the design of the blog and then describe the first two steps of the framework. The third step (evaluation) is left for future work.

4.1 Design of the blog

In the indie music blog, we added features to stimulate readers to visit and comment on the blog. We added a media player, photo album, custom lists, rating scheme, and song of the week to the blog, in addition to making regular posts and comments on particular topics. A media player automatically loaded the featured song of the day and allowed users to stop and re-play the song as desired. A photo album displayed new and exciting pictures of artists, tours, and concerts. Custom lists categorised and indexed reviewed music by genre (*i.e.*, rock, pop, folk, *etc*.) with links to download the reviewed songs. A rating scheme rated reviewed songs on a 5-point scale and encouraged readers to post their impressions of the music in the comments. One particular song was selected as the 'song of the week' providing an easy way for visitors to sample indie music. A screenshot of the blog was shown in our previous work (Chin and Chignell, 2006).

4.2 Identify community

We follow the methodology outlined in Section 3.2.1 and Figure 5, to identify community in the indie music blog. We crawled the indie music blog to two degrees of separation (the 'first-degree' people/blogs who directly made comments on the indie music blog and the 'second-degree' people/blogs who made comments on the first-degree people/blogs that were directly linked to the indie music blog) using RSS feeds from MSN Spaces, then created a social network using UCINET (Borgatti *et al*., 2002). Really Simple Syndication (RSS) is an XML format that aggregates entries from blogs and other syndicated web content made available for others to use. We chose MSN Spaces blogs to crawl because we wanted to extract implicit communities from MSN Spaces. Then, we visualised the social network in Pajek (de Nooy *et al*., 2005). Figure 10 illustrates the network surrounding the indie music blog crawled on 8 November 2005 from the above and contains 604 blogs.

Figure 10 Social network for the indie music blog
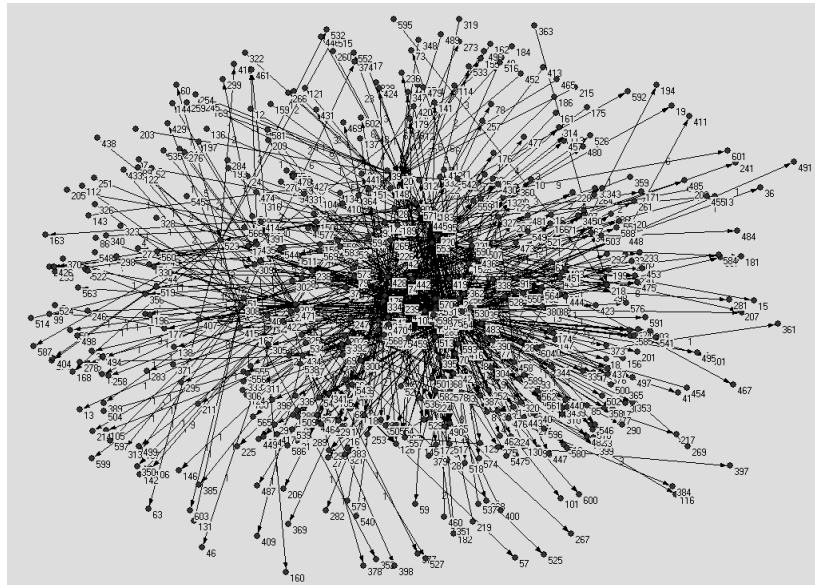
Social network analysis of the graph formed by the links/comments between blogs was supplemented by collecting questionnaire data concerning strength of community associated with the blog. The second column of Table 2 shows the list of questions that we asked visitors in the SOC survey which we adapted from the original SOC questions from Chavis (1986), as shown in the third column of Table 2. A total of 15 people completed the survey. Most of the respondents were between 21 and 30 years of age, had frequent access to the internet, and were blog readers who did not make comments frequently.

Table 2 Questions for sense of community survey in blogs

Question code	SOC question for blogs (adapted from Blanchard (2004))	Original SOC question from Chavis (1986)
Q1	I think this blog is a good one to read	I think my [block] is a good place for me to live
Q2	Readers of this blog do not share the same values	People on this [block] do not share the same values
Q3	Other readers and I want the same thing from this blog	My [neighbours] and I want the same things from the [block]
Q4	I can recognise the names most readers who post comments in this blog	I can recognise most of the people who live on my [block]
Q5	I feel at home in this blog	I feel at home on this [block]
Q6	Very few other readers of this blog know me	Very few of my [neighbours] know me
Q7	I care about what other blog readers think of my actions	I care about what my [neighbours] think of my actions
Q8	I have no influence over what this blog is like	I have no influence over what this [block] is like
Q9	If there is a problem in this blog, there are people here who can solve it	If there is a problem on this [block] people who live here can get it solved
Q10	It is very important to me to be a reader of this blog	It is very important to me to live on this particular [block]
Q11	Readers of this blog generally do not get along with each other	People on this [block] generally do not get along with each other
Q12	I expect to stay a reader here for as long as I can	I expect to live on this [block] for a long time

Possible communities were identified from the network in Figure 10 using the approach described in Section 3.2 and summarised in Table 1. The following data processing steps were used. First, all blogs with only a single directed link (*i.e.*, no indication of community), were filtered out. The frequencies of each directed link in the reciprocal link were then summed, and a hierarchical reduction was performed to recursively remove all blogs that had only zero or one neighbours. The resulting visualisation of the network is shown in Figure 11 with 54 nodes, centred on the indie music blog (node 29), and the set of possible blog communities being highlighted. Table 3 shows the blogs that satisfy each SOC feature according to inspection of the filtered and visualised network, along with the corresponding social network analysis indicators of network centrality, as calculated using the Pajek software.

Figure 11   Possible sense of communities from the indie music blog social network
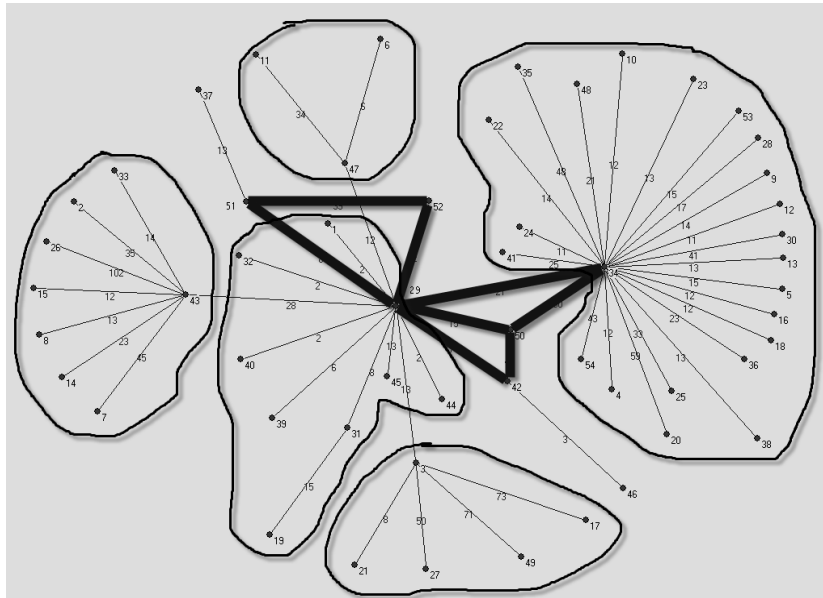

Table 3   Applying the SOC model for blogs to the blog network of Figure 11

SOC feature	Visualisation indicator	Blogs found from network of Figure 11	Social network analysis indicator
Membership	Reciprocal links with each directed link having frequency ≥ 2 Star network	blogs connected to 43	0.0016 < degree centrality < 0.038
		blogs connected to 47	0.0016 < degree centrality < 0.017
		blogs connected to 29	0.005 < degree centrality < 0.05
		blogs connected to 3	0.0016 < degree centrality < 0.027
		blogs connected to 34	0.0016 < degree centrality < 0.081
Influence	Reciprocal links Brokers or bridges	43, 34, 3, 47, 29	0.0086 < betweenness centrality < 0.17
Reinforcement of needs	High sum of frequencies (from individual links in reciprocal link) > 20	(29,34), (34,50), (50,42), (34,54), (20,34), (25,34), (36,34), (30,34), (35,34), (41,34), (51,52), (29,52), (43,29), (2,43), (26,43), (14,43), (7,43), (11,47), (3,27), (3,49), (3,17)	0.256 < closeness centrality < 0.50
Shared emotional connection	Triangles Completely connected graphs	(51,52,29), (29,50,34), (29,42,50)	2-cores

4.3 Measure community

The network centrality measures calculated in Table 3 show that there may be a range of centrality measures that correspond to situations where a blog forms part of a community. We explore these further and compare them with the SOC scores computed from the SOC survey in Table 4 for only those blogs whose bloggers responded to the survey and who had their blogs hosted on MSN Spaces. Four of these MSN Spaces blogs (45, 50, 34 and 343) were in the MSN Spaces blog network surrounding the indie music blog (Figure 11), while the remaining MSN Spaces blogs (605 and 606) were not located in the network because they were only blog readers and did not make comments.

Table 4 Results for sample blogs from sense of community survey and structural analysis

	Blog number					
Analysis step	*45*	*50*	*34*	*343*	*605*	*606*
Step #5: SOC index						
Membership	7	10	11	5	7	4
Influence	9	12	11	5	9	5
Need	9	13	12	8	9	10
Shared emotion	11	13	10	8	9	7
Total SOC index	36	48	44	26	34	26
Step #6: Structural analysis						
Visualisation indicator						
Membership	Star network	Star network	Star network	None	None	None
Influence	Not a broker	Broker	Broker	None	None	None
Need						
min f	13	7	11	None	None	None
max f	13	41	59	None	None	None
Shared emotion	None	Part of two triangles which are completely connected graphs	Part of one triangle which is completely connected graph	None	None	None
SNA indicator						
Degree centrality (normalised)	0.007	0.044	0.081	0.001		
Betweenness centrality (normalised)	0.001	0.016	0.076	0		
Closeness centrality (normalised)	0.353	0.380	0.388	0.291		
k-core	None	Part of two 2-cores	Part of one 2-core	None		

For the blogs found in the set of possible communities of Figure 11 (45, 50, 34), we discovered that the total SOC index was higher (36, 48, and 44, respectively) than the other blogs, as shown from Table 4. This suggests that there may be a correlation between the total SOC index measure of strength of community and the social network analytic indicators of community. Since the individual SOC subscales also vary, we hypothesise that virtual community, in contrast to physical community, relies more on a subset only of the four indicators (subscales) of sense of community. The degree, betweenness and closeness centrality for blogs 45, 50 and 34 (shown in Table 4), tended to be higher than the corresponding measures for other blogs. Thus, as for the SOC scores, the variation in centrality measures may indicate strength of community.

We propose that comparing the SOC scores, the visualisation indicators, and social network analytic indicators with each other is a means of assessing whether or not blogs form communities. In the indie music case study that we conducted, blogs 50 and 34 were deemed to be part of communities because their positions in the visualised network were symptomatic of community membership (they were at the centre of star networks and had high degree centrality), they influenced others because they were brokers with high betweenness centrality, they had relatively high frequency of connections with high closeness centrality, and they had high interdependence (shared emotional connection) with other blogs because they were part of k-cores (in this case triangle-shaped 2-cores). On the other hand, blog 45 was not part of a community because it had low feelings of membership (second lowest degree centrality), low feelings of influence (second lowest betweenness centrality) and was not a broker, had low feelings of need (second lowest closeness centrality) and had a low frequency of connections, and no evidence of interdependence (not part of any completely connected graphs or k-cores). Blogs 343, 605 and 606 were also not part of communities because they had no structures of membership, influence, need, and shared emotional connection, as indicated by the corresponding structural indicators defined above. In addition, they had the lowest degree, betweenness, and closeness centrality and had no k-cores.

4.4   Types of communities

We hypothesise that communities can be categorised into different types. The first type is *core or directed community* in which a community is developed around the target blog (the indie music blog in the case study presented here). The second type is *developed community* where related community arises as a result from conversations that the target blog participates in with nearby blogs (shown as triangles in Figure 11). Finally, there is the *extended community* consisting of nearby communities that are not currently connected. In Figure 11, the extended community for the target blog consisted of star networks that were centred around the broker blogs (blogs that connect sub-groups of blogs to each other). We anticipate that the framework introduced here (involving social network analysis, visualisation, and direct assessment of sense of community) could be refined to assess distinctions between core or directed, developed and extended communities. Extended communities, in particular, offer useful targets for extending the sphere of influence of a blog, since a few well chosen links may introduce the blog into a number of other communities.

5 Conclusion

Blogs are a form of social hypertext, functioning as a one-to-one mapping between a network of web pages and a network of people, which can be represented as a social network and from which communities can emerge. In this paper, we have presented our framework for identifying and measuring community in blogs. Our framework is based on measures of centrality, visualisation, and psychological measures of community using sense of community, as indicators of community. The assumption is that blogs which are strongly centred within a (sub)network of surrounding blogs, function as centres within a community.

Using a case study involving an indie music blog as a motivating example, we illustrated the use of this multi-faceted approach to measuring community. Hypotheses were developed concerning possible alignments between structural network measures, visualised properties and rated sense of community. In addition, a distinction was drawn between three levels of community which may be conceptualised as representing different regions around a blog from:

1 its direct connections

2 through its indirect connections

3 to the nearby communities to which it could be connected to with one or two
   well-placed links.

We provide three recommendations from our study. First, given that blogs have different goals, the types of community and role within each community that a blog shares in, should be evaluated according to the goals of the blog. Second, measures of community can be used to target strategies for promoting and marketing blogs. Relevant extended communities around a blog can be targeted as new links that can allow the blog to participate in the surrounding communities, thereby strengthening and building its own community. Third, and from a more academic perspective, measures of community can be created as a research tool for studying the properties and uses of blogs in general and in response to various design interventions. For example, what happens when interactive tools such as a chatterbox (interactive text chat system) are added to a blog, and how does the response to different design interventions change depending on the type of blog or readership?

Better methods for measuring and diagnosing community promise to revolutionise the study and use of social hypertext as implemented in blogs. In this paper, we outlined some interesting directions for research in this area, organised around an initial framework and model for assessing community. Further research is needed using this framework and others to come up with validated models of community in blogs that can be used for blogging research and in the design, development, and use of blogs.

References

Allen, J. and Dillman, D. (1994) *Against All Odds: Rural Community in the Information Age*, Boulder, CO: Westview Press.

Blanchard, A. (2003) 'The effects of dispersed virtual communities on face-to-face social capital', in M. Huysman and V. Wulf (Eds.) *Social Capital and Information Technology*, Cambridge: MIT Press.

Blanchard, A. (2004) 'Blogs as virtual communities: identifying a sense of community in the Julie/Julia project', *Into the Blogosphere: Rhetoric, Community and Culture*, http://blog .lib.umn.edu/blogosphere/ (accessed 3 March 2006).

Blanchard, A. and Markus, M. (2004) 'The experienced "sense" of a virtual community: characteristics and processes', *The DATA BASE for Advances in Information Systems*, Vol. 35, No. 1.

Borgatti, S.P., *et al.* (2002) 'Ucinet for windows: software for social network analysis', *Analytic Technologies*, Harvard, USA.

Chavis, D. (1986) 'Sense of community index', *Association for the Study and Development of Community*, http://www.capablecommunity.com/pubs/SCIndex.PDF (accessed 14 April 2006).

Chin, A. and Chignell, M. (2006) 'Finding evidence of community from blogging co-citations: a social network analytic approach', *Proceedings of 3rd IADIS International Conference Web Based Communities 2006*, San Sebastian, Spain, pp.191–200.

de Nooy, W., *et al.* (2005) *Exploratory Social Network Analysis with Pajek*, New York: Cambridge University Press.

Efimova, L. and Hendrick, S. (2005) 'In search for a virtual settlement: an exploration of Weblog community boundaries', https://doc.telin.nl/dscgi/ds.py/Get/File-46041/weblog_community _boundaries.pdf (accessed 3 March 2006).

Efimova, L., *et al.* (2005) 'Finding "the life between buildings": an approach for defining a weblog community', *AOIR Internet Research 6.0: Internet Generations*, Chicago.

Erickson, T. (1996) 'WWW as Social Hypertext', *Communications of the ACM*, Vol. 39, No. 1, pp.15–17.

Fisher, D. (2005) 'Using egocentric networks to understand communication', *IEEE Internet Computing*, Vol. 9, No. 5, pp.20–28.

Flake, G., *et al.* (2002) 'Self-organization and identification of web communities', *IEEE Computer*, Vol. 35, No. 3, pp.66–71.

Freeman, L.C. (1978) 'Centrality in social networks: conceptual clarification', *Social Networks 1*, pp.215–239.

Herring, S.C., *et al.* (2005) 'Conversations in the blogosphere: an analysis "from the bottom up"', *Proceedings of the 38th Hawaii International Conference on System Sciences*, Los Alamitos, USA, pp.107–118.

Jones, Q. (1997) 'Virtual-communities, virtual settlements and cyber-archaeology: a theoretical outline', *Journal of Computer Supported Cooperative Work*, Vol. 3, No. 3.

Kim, A.J. (2000) *Community Building on the Web: Secret Strategies for Successful Online Communities*, Berkeley, CA: Peachpit Press.

Kleinberg, J. (1998) 'Authoritative sources in a hyperlinked environment', *Proceedings 9th ACM-SIAM Symposium on Discrete Algorithms*.

Korica, P., *et al.* (2006) 'The growing importance of e-communities on the web', *Proceedings of the IADIS International Conference on Web Based Communities 2006*, San Sebastian, Spain, pp.165–174.

Kumar, R., *et al.* (1999) 'Trawling the web for emerging cyber-communities', *Computer Networks*, Amsterdam, the Netherlands.

Kumar, R., *et al*. (2003) 'On the bursty evolution of blogspace', *Proceedings of the 12th International Conference on World Wide Web*, New York, USA, pp.568–576.

Martins, H.F. (2006) 'Social issues as success factors for web-based communities', *Proceedings of the IADIS International Conference on Web Based Communities 2006*, San Sebastian, Spain, pp.284–288.

McMillan, D.W. and Chavis, D.M. (1986) 'Sense of community: a definition and theory', *Journal of Community Psychology*, Vol. 14, No. 1, pp.6–23.

Merelo-Guervos, J., *et al*. (2004) 'Mapping weblog communities', http://arxiv.org/pdf/cs.NE/0312047 (accessed 16 December 2005).

Nardi, B.A., *et al*. (2004a) 'Blogging as social activity, or, would you let 900 million people read your diary?', *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, New York, USA, pp.222–231.

Nardi, B.A., *et al*. (2004b) 'Why we blog', *Communications of the ACM*, Vol. 47, No. 12, pp.41–46.

Paolillo, J. (1999) 'The virtual speech community: social network and language variation on IRC', *Journal of Computer Mediated Communication*, Vol. 4, No. 4.

Poplin, D.E. (1979) *Communities: A Survey of Theories and Methods of Research*, 2nd ed., New York: MacMillan Publishing Co., Inc.

Preece, J., *et al*. (2004) 'Designing and evaluating online communities: research speaks to emerging practice', *International Journal of Web Based Communities*, Vol. 1, No. 1, pp.2–18.

Rheingold, H. (1993) *The Virtual Community: Homesteading on the Electronic Frontier*, Toronto, Canada: Addison-Wesley.

ten Thij, E., *et al*. (2006) 'Success factors of (Dutch) online dance communities: a validation of website features for social interaction', *Proceedings of the IADIS International Conference on Web Based Communities 2006*, San Sebastian, Spain, pp.284–288.

Tyler, J.R., *et al*. (2005) 'E-mail as spectroscopy: automated discovery of community structure within organizations', *The Information Society*, Vol. 21, No. 2, pp.143–153.

Wei, C. (2004) 'Formation of norms in a blog community', *Into the Blogosphere: Rhetoric, Community and Culture of Weblogs*, University of Minnesota, Minnesota, USA.

Wellman, B. and Guilia, M. (1999) 'Net surfers don't ride alone: virtual communities as communities', in B. Wellman (Ed.) *Networks in the Global Village: Life in Contemporary Communities*, Boulder: Westview Press.

Wikipedia (2006) 'Blogs', http://en.wikipedia.org/wiki/Blogs (accessed 16 April 2006).