Detecting Communities in Science Blogs

Christina K. Pikas
College of Information Studies
University of Maryland
College Park, MD, USA
cpikas@umd.edu

Abstract—**Many scientists maintain blogs and participate in online communities through their blogs and other scientists' blogs. This study used social network analysis methods to locate and describe online communities in science blogs. The structure of the science blogosphere was examined using links between blogs in blogrolls and in comments. By blogroll, the blogs are densely connected and cohesive subgroups are not easily found. Using spin glass community detection, six cohesive subgroups loosely corresponding to subject area were found. By commenter links, the blogs form into more easily findable general subject area or interest clusters.**

Keywords - blogs; community detection; social network analysis

I. INTRODUCTION

In social studies of science, a great deal of attention has rested on how scientists organize in groups. Science is fundamentally a social activity. Scientists collaborate to tackle big science [1] and little science [2]. They form into invisible colleges [3], paradigms [4], thought collectives [5], social circles [6], and core sets [7] around research subjects, scientific practices, and ways of making new scientific knowledge. Practically, scientists are employed together in research centers and universities.

Social scientists have long studied scientists' group membership to learn about information diffusion, how scientists create new knowledge, and for reasons related to science policy and understanding the structure of science. Typical research methods include examining authoring patterns, citation patterns, collaboration patterns (as evidenced by co-authorship on grants and on reports of completed work), employment, association membership, and more recently, participation in online communities.

Blogs, or weblogs, are a social computing technology used to publish information on the web. They are identified by their format: reverse chronologically arranged collections of discrete posts [8]. Blogs are used for many purposes including as personal diaries, research logs, for political speech, and as news filters [9-12]. In a recent survey, 13% of scientists reported reading blogs and 3%

reported maintaining a blog [13]. Bonetta suggested that there are more than one thousand science blogs [14]. Little is known about what these science communities look like, on what basis they form and continue, and what the nature of communication is within and between science blogs.

In this study, social network analysis (SNA) is used to yield an understanding of the topology and the structure of the science blogosphere. The structure is revealed through identification of central actors and cohesive subgroups. This structure has implications for information diffusion, influence, and the ability to learn from and get help from other scientists. Connections through the blogosphere might lead to future collaborations.

II. BACKGROUND: THE STRUCTURE OF BLOGS

A. Internal Structure

As mentioned above, blogs are identified by their internal structure. They use standard web protocols to manage and display content. In general, a blog consists of a standard template page that is either two or three columns. In the outside columns, or either of the two columns in a two column design, there is a collection of static or automatically generated links such as links to previous posts, to search the blog, to advertisements, and to other blogs or websites the author endorses. These collections of links to other blogs are called blogrolls. Blogrolls can be generated automatically by web tools used to read and aggregate RSS feeds such as Bloglines and Google Reader, or can be manually generated by typing in a series of URLs and page names. The center section of the blog, or the other of the two columns, contains a list of the posts in reverse chronological order. Each of these posts has a unique URL (permalink), can be assigned keyword tags, and can receive comments and links back from other blogs (trackbacks). The URL generally points to a post page that has the same outside columns, but just contains the single post and its associated comments and trackbacks in the other column or center.

To leave a comment, the visitor completes a short form with a name, an e-mail, and optionally a URL. The commenter's e-mail is generally not displayed on the page, but the comment will be signed with the name and the link to the URL the commenter provided. Various methods are

employed to prevent spam comments; that is, irrelevant comments used to link to a site to boost that site's web traffic. These methods include typing in characters that appear in a picture that is not machine readable, requiring an account with the host (a TypeKey account for MovableType or TypePad or a Google account for Blogger), or requiring approval before the comment appears on the site. Sites requiring an account to comment often do not link back to the commenter's blog, but instead to a profile page that links to all of the blogs that commenter has permission to post on at that host.

There are three dominant software tools used for science blogs: Blogger, WordPress, and MovableType or Typepad. Each has a hosted version that allows a blogger to visit the site and set up a blog with a few clicks. For MovableType and WordPress, there is also a version of the software that can be downloaded and hosted on another server. Each of the tools allows the bloggers to select a theme, or set of colors, layouts, and design features, and to select the content that will appear in the more static columns. Some web-savvy bloggers create new or significantly modify system templates, but certain markup is consistent to allow the software to display the content. In addition to these software tools, ScienceBlogs.com, a site run by the science culture magazine *Seed*, has what appears to be a custom-built software that is similar to MovableType. These regularities of markup on blogs enable the harvesting of content, but it is the links between blogs that create the structure that is of particular interest for this study.

B. Blogosphere Structure

A large group of blogs, including the group of all blogs, is called the blogosphere. Of primary interest for this study is the structure of the blogosphere created through links between blogs and bloggers. There are four main categories of links between blogs: blogroll, commenter, within posts, and trackback [15]. The selection of a blog in a blogroll implies some endorsement of that blog. Similarly, the choice to comment on a post, regardless of the content of the comment, and the choice to sign that comment with a URL implies a link between the two websites. Another category of link between blogs that is not addressed in this study is the link within the content of a post to a post on another blog. This link is a weaker endorsement as it might only be a passing mention. An author might specifically choose not to link to another blog to withhold endorsement based on knowledge of how links boost web traffic, visibility in search engines, and sometimes income from advertising [16]. Trackbacks are ways to track mentions of a post on other blogs.

III. LITERATURE REVIEW

This study is based on empirical and theoretical work blogs, online communities, and on SNA.

A. Science Blogs

In a previous study of how and why chemists and physicists use blogs, Pikas found that some of the participants thought of their blogs as a new form of scholarly communication useful for publishing scientific results that were either not enough or too much for journal publication. Others thought of their blogs as a personal choice or hobby not really relevant to their work life. Many of the participants expressed a sense of community with other science bloggers and their commenters. The participants frequently asked for and received help and reported feeling a sense of membership in the physics, chemistry, or science blogging community[17].

B. Online Communities

The term *online community* is sometimes used to describe nothing more meaningful than nodes connected on a graph; but it can also mean a sense of belonging, shared emotional connection, the possibility of influence, and the fulfillment of needs [18]. In what sense can blogs be part of or constitute online community? Blanchard [19] addressed this issue in her study of the commenters who participated in the Julie/Julia Project blog. She determined that blogs could be virtual settlements because of the connections between them through blogrolls, comments, and links within posts, and that some participants in the Julie/Julie Project do experience a sense of community. Overall the average score on the sense of community scale was moderate, but the participants who used the comments as a forum or bulletin board felt that it was a community. In the science blogosphere, however, we are not studying the comments on a single blog but the community formed across blogs that Blanchard [19] named a virtual settlement.

C. Social Network Analysis

Unlike other research methods, SNA focuses on the connections or relationships between actors, not the attributes of the actors themselves [20]. The collection of connections and actors or nodes forms a network. The connections between actors can be analyzed using graph theoretic methods to describe the nature of the network including its density and diameter, as well as to determine the centrality and prestige of actors, cohesive subgroups of actors, and roles and structural equivalence of actors. Centrality and prestige measures can be useful to understand the potential flows of information or resources or the opportunities or constraints on actors in a network [20]. Once interesting actors or groups are identified through SNA, quantitative or qualitative methods can be used to explore or make inferences about the attributes of the actors.

1) Measures of Centrality, Prestige, and Power. An important use of SNA is to identify important or prominent actors in a network. Prominent actors are more visible to other actors in the network [20]. Central actors are more

96

visible because of their links to other actors. Prestigious actors are more visible because, for directional networks, they were more frequently chosen. Standard measures of centrality include degree, betweenness, and closeness. Degree centrality simply measures how many connections that actor has. For a directed network, in degree measures how many times that node has been chosen while out degree measures the number of nodes the actor has chosen. Betweenness centrality measures the number of shortest paths between nodes that must go through that node. The node is between other nodes on the network and can therefore influence the network. Closeness measures the shortest path to other nodes on the network. Closeness is only appropriate for connected networks [20].

Power, such as Bonacich's measure, measures not only the choices, but the prestige of the actors who are linked to the actor of interest [21]. In other words, a link to a more prestigious actor is valued differently than a link to a less prestigious actor.

2) Identification of Cohesive Subgroups. Cohesive subgroups are collections of actors in the network that are more connected to each other than to the rest of the graph. Subgroups can be determined by links between actors, by structural similarity, or by clustering. In this study link and clustering methods are used.

Link methods. The first way to find subgroups in a network is to locate components. Components are groups of nodes connected to each other, but not connected to the remainder of the network. Within components, the most strict definition of a cohesive subgroup is a clique. A clique is a "maximally complete subgraph of three of more nodes;" that is, it is an area of the graph where all of the nodes are adjacent to each other and there are no other nodes that are adjacent to all of these nodes [20]. Cliques may overlap and may not be too meaningful. By allowing additional links between nodes instead of requiring them to be adjacent, we get n-cliques. A measure similar to the n-clique is the k-core. K-cores are areas in which each node is adjacent to at least k other nodes in the subgraph.

Several link methods to identify cohesive subgroups rely on the degree of the nodes. For example, a k-plex is requires that the degree of each node in the subgraph is at least equal to the number of nodes in the subgraph minus k [20].

Lambda sets maximize line connectivity within the group and minimized line connectivity between any nodes within the group and nodes elsewhere on the graph. Line connectivity, $\lambda(i,j)$ is the minimum number of connections that must be removed from the graph to leave no path between the two nodes [20]. To identify lambda sets within a graph, the UCInet algorithm uses progressively increasing numbers for λ [22].

Clustering methods. Clustering methods are computational ways to iteratively select nodes that are more closely linked to each other than to others in the network.

Hierarchical clustering can be either top down (divisive) or bottom up (agglomerative); that is, the graph can be split into increasingly small clusters or individual nodes can be grouped into larger clusters. Another method starts with partitions of the network prior to clustering instead of starting with either individual nodes or the entire network.

In the past few years there has been increased attention in the physics community to "community detection" in large network graphs. Physicists have used analogs to statistical mechanics, bridge circuits, and other physical systems to detect cohesive subgroups or clusters. Newman and Girvan [23] used a divisive clustering method, but removed edges based on their edge betweenness centrality. Similar to Freeman's node betweenness centrality discussed above, edge betweenness is a measure of how many of the shortest paths between nodes traverse that edge. Newman and Girvan determine the stopping point for clustering by locating peaks in the modularity of the graph. The modularity "measures the fraction of the edges in the network that connect vertices of the same type (i.e., within community edges) minus the expected value of the same quantity in a network with the same community divisions but random connections between the vertices"[23].

Another method proposed as an analog to physical systems is the Spin Glass method developed by Reichardt and Bornholdt [24] which is based on methods from statistical mechanics. The Spin Glass method uses simulated annealing techniques to optimize cluster assignment based on the modularity of the network.

D. SNA Methods for Community Detection

Several recent studies have used SNA methods including community detection methods to identify cohesive subgroups or communities within social computing sites online.

Adamic and Glance [25] studied the political blogosphere during the run up to the 2004 presidential election. They first studied 40 of the most highly linked political bloggers over a period of two months prior to the election and then studied 1,000 political bloggers over a single day. Political blogs were located by searching several web directories and then marked conservative or liberal using the categorization of the directory. For the study of the larger set of bloggers, Adamic and Glance crawled all blogs found in the directories, and then retrieved all links from the front page of each blog. They found that the majority of the links stayed within either the conservative or liberal community and did not cross to the other community. Looking at the top 20 conservative and top 20 liberal blogs, Adamic and Glance found that the conservatives had more links per post and the links were more evenly distributed among the top blogs whereas the liberals cited the top few blogs very heavily and others in the top 20 much less frequently.

Chin and Chignell [26-28] published a series of articles discussing the use of SNA to find communities in social computing technologies. In a study of blogs linked to an independent music blog, k-plexes and n-cliques were used to locate cohesive subgroups [28]. To build the network around the independent music blog, they crawled two links away from the starting point, but only retained blogs that were also hosted on MSN Spaces. They then posted a Sense of Virtual Community survey based on Blanchard [19]. Unfortunately they only received 15 responses so could not validate their SNA findings but the survey did support the roles and community membership they attributed to certain nodes.

The developers of the spin glass community detection algorithm tested the algorithm to locate markets in the users of the German eBay site in December 2004 [29]. They created links based on bids placed within the time period; users bidding on the same item were adjacent. The spin glass method successfully clustered 85% of users bidding on similar items. Analysis of the categories assigned to the auctions by the sellers verified that the resulting clusters were logical groupings of users with similar hobbies and interests.

IV. METHODS

This study employed SNA to examine the structure of the science blogosphere. Links between blogs established through blogroll lists and through signed comments were used to establish the network.

This section describes sample selection, data gathering and management, and analysis methods.

A. Sample Selection

The definition of science blog for this study is quite broad: blogs maintained by scientists that deal with any aspect of being a scientist or science or blogs about scientific topics by non-scientists. This definition was developed to address research questions regarding the use and value of blogs in science.

It is difficult to classify many blogs as properly part of the political blogosphere or science blogosphere. For example, I generally did not include blogs written by journalists discussing the Dover trial or the teaching of evolution in schools if these were part of larger discussions of politics; however, I did include blogs written by amateur astronomers who were reporting on science news and their own scientific observations as well as blogs by scientists that had multiple posts on political issues. Future studies should use multiple coders and check classification of blogs for inter-rater reliability to provide a more reliable sample.

Any selection of blogs cannot be comprehensive and is more likely to include popular or more connected blogs because finding the blogs depends on their being linked from other blogs or from search engine results.

B. Data Gathering and Management

1) Blogroll. Data gathering started with an initial listing of blogs maintained by scientists gathered for the purpose of a previous study. Each of the blogs previously located was visited using Firefox web browser. Using the menu command, "page info" was displayed. The urls from the blogroll were copied and pasted into a database created using a template downloaded from the de Nooy, Mrvar, and Batagelj [30] companion site. If the blogroll was divided into "science blogs" and "personal" or other categories, only science blogs were chosen. Blogs identified through blogrolls were then visited to extract their blogrolls. Blogs not updated for more than a year, maintained by a corporate entity, or not discussing science were not used as starting points, but may still appear in the list of blogs. Blogs not written in English were omitted.

The database was exported as a report and edited for input into Pajek [31] and UCInet [22]. Editing included removing duplicate entries and consolidating entries for blogs that had changed hosts and web addresses including bloggers who had recently moved to ScienceBlogs.com from Blogspot.com or Wordpress.com.

2) Commenters. Interesting blogs were selected from the results of the blogroll analysis. The top 20 blogs for in degree, betweenness and closeness were chosen (n=46). For each of these, perl scripts was used to extract commenter URLs signing comments on the last 10 posts. These URLs may refer to blogs, personal web pages, or other web pages. Perl scripts were used to associate Blogger or Typekey profile URLs to science blogs. The data were gathered in April 2008. Ten posts covered a period of one week to a few months.

C. Analysis

Network files were exported from the database and imported into Pajek [31], Ucinet [22], R [32], and NetDraw [33] for visualization and analysis. Within R, the igraph [34] package was used to determine power centrality and for community detection.

Components were identified, and then analyzed to determine density and diameter. Cohesive subgroups were identified using link and clustering methods described above. Within selected groups, prestige and centrality measures were used to identify important actors.

V. RESULTS

A. Blogroll Network

The links between blogs in blogrolls were used to create a blogroll network. This network was very densely connected and formed only one component (see fig. 1). There were 1091 nodes, 6621 arcs (directional connections such that links from a > b and from b > a count as 2 arcs, but only one edge), and density 0.005563. The diameter is 9. In degree ranges from 1 to 292, with the median in degree of 3, and mean 6. The mean in degree of blogs

98

hosted on the ScienceBlogs platform is 16. Ten of the top 20 blogs by in degree are authored by or co-authored by women. Four of the top five blogs by closeness are authored or co-authored by women.

Several methods were used to attempt to locate cohesive subgroups. K-plexes, LS sets, and hierarchical clustering methods that started with a binary split or agglomeration such as Newman-Girvan found only one group. Spin glass community detection using an undirected version of the graph and the igraph package in R yielded six communities (see fig. 1). The communities ranged in size from 47 to 299 blogs. The modularity was 0.64 while the modularity for a random graph with the same number of nodes and lines is 0.30 [35].

Spin glass techniques require the specification of γ. Setting $\gamma=1$ gives equal importance to existing and non-existing links. Higher values of γ yield smaller communities while smaller values yield larger communities [24]. Running the spin glass algorithm with $\gamma > 1$ lowered the modularity and created new communities with 1-5 nodes while the majority of the communities found using $\gamma=1$ remained in tact; therefore, the six communities seem to be the best division using this method.

The blogs identified in the clusters were examined to determine common features. Four of the clusters had scientific discipline or interest area in common (see table I). Cluster six was predominantly composed of female bloggers. No common feature could be determined for cluster four. The most central blog in this cluster is a Singapore-based blog that discusses various aspects of science and politics. Other blogs in the cluster are from Canada, China, the United Kingdom and are on biology, cosmology, evolution, and other diverse topics.

B. Commenters

The second network was formed using links left by commenters on selected blogs. The arrows point to the blog on which the comment was made and loops are created when a blogger comments on his own blog. The number of comments by each commenter are not tracked, but might be interesting for future work.
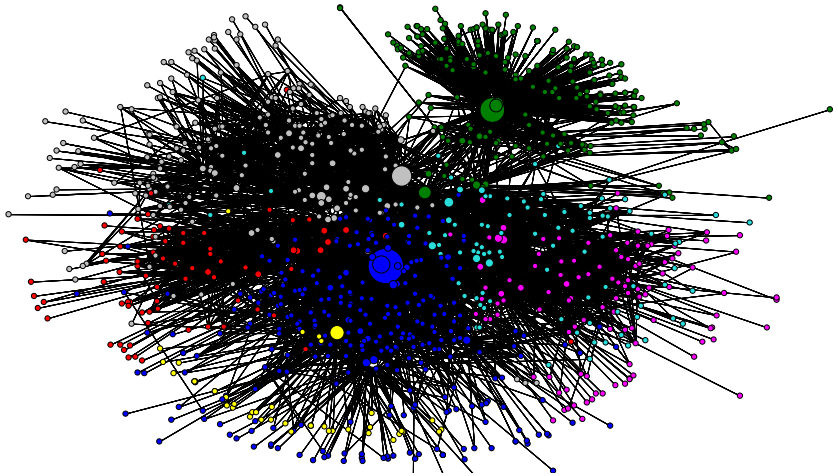

Figure 1 Blogroll network, nodes sized by betweenness centrality, colored by spin glass community assignment $\gamma=1$

The graph has 938 nodes, 1152 arcs, and a density of 0.0013. The graph breaks down in to five components. The largest has 911 nodes while the others have 11 or fewer nodes. The largest component has a diameter of 5.

For this network, out degree centrality marks bloggers who have commented on multiple blogs. Several high out degree nodes on this chart comment regularly on multiple blogs, but are not in the top 20 in any measure of centrality using blogroll measures (see table II). Seven of the top ten blogs in betweenness centrality are written by women and 14 of the top 21 blogs by commenter out degree are written by women.
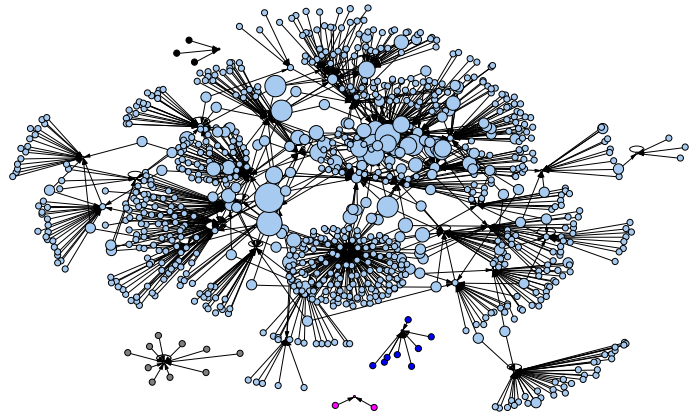

Figure 3 Commenter network, nodes sized by out degree

TABLE I: BLOGROLL NETWORK COMMUNITIES

n	Subject	Color	Example
114	chemistry	pink	Blog.chembark.com
78	geosciences	red	Clasticdetritus.com
299	biological sciences	blue	Scienceblogs.com/clock
232	astronomy	green	Astroprofspage.com
220	physics, math, computer science	gray	Cosmicvariance.com
47	?	yellow	Freshbrainz.blogspot.com
101	female bloggers	light blue	Scienceblogs.com/ sciencewoman

99

TABLE II: TOP BLOGS BY COMMENTER OUT DEGREE

	Commenter Out Degree	Commenter In Degree	Blogroll Out Degree	Blogroll In Degree
lablemminglounge.blogspot.com/	7	n/a	23	28
scienceblogs.com/sciencewoman/	6	35	44	39
www.sunclipse.org	6	n/a	22	14
academiccrossroads.blogspot.com	6	n/a	0	2
scienceblogs.com/terrasig	5	n/a	35	16
amadtea-party.blogspot.com/	5	n/a	0	9
physioprof.wordpress.com/	5	n/a[a]	n/a	n/a
orgprepdaily.wordpress.com	5	n/a	8	24
www.mazepath.com/uncleal/	5	n/a	n/a	n/a
vwxynot.blogspot.com/	5	n/a	37	9
propterdoc.blogspot.com/	4	22	72	44
www.coronene.com/blog/	4	21	n/a[b]	32
mareserinitatis.livejournal.com	4	n/a	9	2
biochemgradstudent.blogspot.com	4	n/a	0	9
shearsensibility.blogspot.com	4	n/a	24	25
drjekyllandmrshyde.blogspot.com	4	n/a	0	1
rabett.blogspot.com	4	n/a	n/a	n/a
candidengineer.blogspot.com/ [c]	4	n/a	n/a	n/a
unbalanced-reaction.blogspot.com/	4	n/a	0	7
girlyscientist.blogspot.com/	4	n/a	24	4
hypoglycemiagirl.blogspot.com/	4	n/a	38	6

Blogs authored or co-authored by women are in italics.
[a] deemed out of scope for blogroll data gathering
[b] this blog has an extensive blogroll as of May 25, 2008, but was not included as a starting point
[c] new since May 2008

VI. DISCUSSION

A. Blogroll Network

The strength of the endorsement of including a blog on a blogroll is taken as a constant; however, from the data it is clear that the blogrolls are more or less meaningful for different bloggers. For example, several of the astronomy bloggers appear to share a common blogroll, errors and duplicates included. Also, some bloggers have several hundred blogs listed on their blogrolls. It seems unlikely that the blogger frequently reads all of these blogs. It appears that some bloggers use their blogs as a central place to store bookmarks. They list journals, department information, and other frequently visited sites. Several of the most highly linked to blogs do not maintain blogrolls at all. Further work examining the various uses of blogrolls and determining ways to weight the strength of the tie would be useful in understanding the blogroll network.

Using the most common methods for community detection, the blogroll network remained tightly clumped into one cluster. This might be an artifact of the method used to locate science blogs. However, it seems more likely that blog usage diffused in a manner similar to other interactive communication technologies – early adopters trialed the software, and later adopters linked to early adopters [36]. It is unsurprising that blogs hosted by ScienceBlogs win in degree popularity contests, as the blogs were selected to move to ScienceBlogs.com based on their topic and audience. Also, the blogs are more easily findable because they are listed in a directory and highlighted regularly on the ScienceBlogs.com home page.

B. Commenter Network

The commenter network studied is just a small part of the entire science blogosphere. Only the most recent posts from the most central blogs were studied. More central blogs are more likely to have comments, but it is unclear what bias is introduced with this sampling method. Additionally, the number of comments and their distribution in the blogosphere changes over time as controversies arise, people get busy in their lives offline, and bloggers come and go, so different snapshot in time might show different connections and communities.

In a previous study, bloggers reported not discussing controversial topics because these posts drew unwelcome and sometimes inflammatory comments from non-regular commenters [17]. On the other hand, several blogs in this study included many posts on controversial topics. These blogs can be expected to receive many comments from outside the like-minded community. Collins' (1985) core sets explicitly group partisans on various sides of a scientific controversy; however, studying this grouping of science blogs is less useful for learning about informal scholarly communication, information diffusion, and collaboration. Future work should look more broadly at the science blogosphere and should look for ways to separate comment links that are more political or inflammatory from those that are furthering scientific information sharing.

Along these same lines, it is interesting to look in more detail at the blogs of high out degree centrality nodes who are not also high in centrality using blogroll data. One of these leaves outrageous and inflammatory comments on chemistry and physics blogs. Playing the community role of troll [37] might be one reason a node would be high in centrality by commenter out degree and not by blogroll.

A majority of the blogs with the highest centrality are authored or co-authored by women. This might point to a different type of community with different norms of behavior. Looking at content of these blogs, many of the women are sharing information on how to run a lab, how to

100

get a job, and how to get tenure. They post frequently, link to other blogs on their blogrolls, and comment on each other's posts. This group really appears to be a virtual community in the sense Blanchard [38] proposes instead of just a virtual settlement.

VII. CONCLUSIONS

This study used SNA methods to examine the structure of the science blogosphere. Links on two dimensions, blogroll and commenter, were used to construct two networks of science blogs. Within the two networks, cohesive subgroups were formed around subject areas including areas of science and women in science.

Future Work. Structural analysis of the science blogosphere proved useful in understanding the topology and clustering of the blogs, but more quantitative and qualitative research is needed to fully understand how scientists can and do use blogs as well as blogs' value or potential value to science. Further, we can't know about the sense of community or the culture of the identified communities based on the links alone. Qualitative research with the bloggers as participants or co-investigators would shed more light on the value the bloggers get from their community involvement.

Throughout this study, I have referred to clusters of blogs maintained by women scientists. Many of these are anonymous or pseudonymous. These blogs might be more attentive to issues of life in the lab and mentoring of younger scientists. They also appear to be a different type of community. It would be valuable to do a case study on this cluster of blogs including more detailed content analysis, administration of a sense of virtual community instrument [38], and group interviews.

REFERENCES

[1] D.J.D. Price, Little Science, Big Science…and Beyond. New York: Columbia University Press, 1986.

[2] K. Knorr-Cetina, Epistemic Cultures: How the Sciences make Knowledge. Cambridge, MA: Harvard University Press, 1999.

[3] D. Crane, Invisible Colleges: Diffusion of Knowledge in Scientific Communities. Chicago: University of Chicago Press, 1972.

[4] T. S. Kuhn, The Structure of Scientific Revolutions. ,3rd ed. Chicago, IL: University of Chicago Press, 1996.

[5] L. Fleck, Genesis and Development of a Scientific Fact. Chicago: University of Chicago Press, 1979.

[6] M. Polanyi, "The Republic of Science: Its Political and Economic Theory," Minerva, vol. 38, pp. 1-21, 2000.

[7] H. M. Collins, "The scientist in the network," in Changing Order: Replication and Induction in Scientific Practice. Beverly Hills, CA: Sage Publications, 1985, pp. 129-157.

[8] T. Mortensen and J. Walker, " Blogging thoughts: Personal publication as an online research tool," in SKIKT-Researchers' Conference 2002: Researching ICTs in Context, 2002, pp. 249-279.

[9] J. Walker, "Blogging from inside the ivory tower," in Uses of Blogs, A. Bruns and J. Jacobs, Eds. New York: Peter Lang, 2006, pp. 127-138.

[10] A. Halavais, "Scholarly blogging: Moving toward the visible college," in Uses of Blogs, A. Bruns and J. Jacobs, Eds. New York: Peter Lang, 2006, pp. 117-126.

[11] S. C. Herring, L. A. Scheidt, I. Kouper and E. Wright. "Longitudinal content analysis of blogs: 2003-2004," in Blogging, Citizenship, and the Future of Media, M. Tremayne, Ed., London: Routledge. 2007.

[12] S. C. Herring, L. A. Scheidt, S. Bonus and E. Wright, "Bridging the gap: A genre analysis of weblogs," in Proceedings of the 37th Hawaii International Conference on System Sciences- 2004, 2004.

[13] Mark Ware Consulting, "Peer review in scholarly journals: Perspective of the scholarly community – an international study," Author, Bristol, UK, 2008.

[14] L. Bonetta, "Scientists Enter the Blogosphere," Cell, vol. 129, pp. 443-445, 2007.

[15] C. Marlow. Audience, structure and authority in the weblog community. Presented at International Communication Association Conference. Available: http://alumni.media.mit.edu/~cameron/cv/pubs/04-01.pdf

[16] L. Efimova, S. Hendrick and A. Anjewierden. Finding'the life between buildings': An approach for defining a weblog community. Presented at Internet Research 6.0. Available: https://doc.freeband.nl/dsweb/Get/Document-55092/AOIR_blog_communities.pdf

[17] C. K. Pikas, "How and why chemists and physicists use blogs," unpublished.

[18] A. L. Blanchard and M. L. Markus, "Sense of virtual community - maintaining the experience of belonging," in HICSS. Proceedings of the 35th Annual Hawaii International Conference on System Sciences, 2002. 2002, pp. 3566-3575.

[19] A. L. Blanchard. Blogs as virtual communities: Identifying a sense of community in the Julie/Julia project. Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs, 2004. Available: http://blog.lib.umn.edu/blogosphere/blogs_as_virtual.html

[20] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications. New York: Cambridge University Press, 1997.

[21] P. Bonacich, "Power and Centrality: A Family of Measures," American Journal of Sociology, vol. 92, pp. 1170-1182, 1987.

[22] S. P. Borgatti, M. G. Everett and L. C. Freeman. (2006). Ucinet 6 for windows: Software for social network analysis. 6.143 Available: http://www.analytictech.com/downloaduc6.htm

[23] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), vol. 69, pp. 26113-21. 2004.

[24] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), vol. 74, pp. 16110-11, 2006.

[25] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 US election: Divided they blog," in LinkKDD '05: Proceedings of the 3rd International Workshop on Link Discovery, 2005, pp. 36-43.

[26] A. Chin and M. Chignell, "Identifying subcommunities using cohesive subgroups in social hypertext," in Hypertext 2007: 18th ACM Conference on Hypertext and Hypermedia, HT'07, 2007, pp. 175-178.

[27] A. Chin and M. Chignell, "Identifying communities in blogs: Roles for social network analysis and survey instruments," International Journal of Web Based Communities, vol. 3, pp. 345-363, 2007.

[28] A. Chin and M. Chignell, "A social hypertext model for finding community in blogs," in Seventeenth ACM Conference on Hypertext and Hypermedia, HT'06, 2006, pp. 11-22.

[29] J. Reichardt and S. Bornholdt. (2007). Clustering of sparse data via network communities-a prototype study of a large online market. Journal of Statistical Mechanics: Theory and Experiment, P06016, 10.1088/1742-5468/2007/06/P06016.

[30] W. d. Nooy, A. Mrvar and V. Batagelj, Exploratory Social Network Analysis with Pajek. New York: Cambridge University Press, 2005.

[31] V. Batagelj and A. Mrvar. (2006). Pajek. 1.17 Available: http://vlado.fmf.uni-lj.si/pub/networks/pajek/default.htm

[32] R Development Core Team. (2008) R: A language and environment for statistical computing. Available: http://www.R-project.org

[33] S. P. Borgatti. (2005) NetDraw: Graph visualization software. 2.046 Available: http://www.analytictech.com/downloaduc6.htm

[34] G. Csardi and T. Nepusz. (2008) The igraph software package for complex network research. 0.5 Available: http://igraph.sf.net

[35] J. Reichardt and S. Bornholdt, "When are networks truly modular?" Physica D, vol. 224, pp. 20-26. 2006.

[36] E. M. Rogers. Diffusion of Innovations, 5th ed. New York: Free Press, 2003.

[37] S. A. Golder and J. Donath. Social roles in electronic communities. Presented at Association of Internet Researchers (AoIR) Conference Internet Research 5.0. Available: http://web.media.mit.edu/~golder/projects/roles/golder2004.pdf

[38] A. L. Blanchard, "Developing a Sense of Virtual Community Measure," CyberPsychology & Behavior, vol. 10, pp. 827-830. 2007.