# Is There Online Collective Identity within Social Bookmarking Systems?

**Juan D. Borrero\***

*Departamento de Dirección de Empresas y Marketing, Universidad de Huelva, Plaza de la Merced, Spain*

## Abstract

Social tagging systems attract researchers in information systems and social sciences because they offer an enormous quantity of user-generated annotations that reveal the interest of millions of people. In this paper, we develop two hypotheses around the idea of online collective identity. We present a model in which agents exchange practical and symbolic resources via opinion and frame networks. The results discussed in the paper come from the analysis of a sample of a sample of 3,668 users, 2,148 URLs and 4,776 tags from the Delicious classification system on the subject of globalization of agriculture. Our social network analysis has identified differences between networks that reflect different degrees of explicitness within the Delicious online community and it has provided an understanding of social bookmarking systems that was not previously available. This research is one of the first to apply information to the frame components in an attempt to establish the implicit links within these networks.

**Keywords:** Online collective identity; Opinion network; Tagging; Social bookmarking systems; Social tagging

## Introduction

Social Network Sites (SNS) have been so popular since their inception that they are now a routine feature in the everyday lives of millions of users, facilitating connections on the basis of common interests or activities [1-3].

For this paper we have chosen to study a specific type of SNS, Social Bookmarking Services, specifically Delicious. These SNS enable the user to send identifiers of interest, such as URLs with a short tag attached. Tagging in systems like Delicious is an important change in the way web bookmarks are organized and shared [4].

This paper aims to contribute to the growing research into the analysis of SNS [5]. Several recent studies have analyzed the main social networks that form as a result of user interaction inside social bookmarking services. Our work is one of the first to attempt to use information from frame components to infer implicit relations between users in these networks. In line with new social movement theory [6,7], our model emphasizes the importance of a shared sense of identity among social actors in forging collective behaviour [8].

Thus, the paper is about finding out if these Delicious users exhibit a collective identity. We use the data analysis to reveal that users of these kind of networks exhibit a shared sense of identity, as predicted by the literature.

We focus our study, specifically, in its user community around the issue of globalization of agriculture. Globalization essentially implicates the extension and deepening of markets as a result of the reduction of the transaction costs associated with trading internationally. The globalization of agriculture is at the center of this debate because many of the poor depend on agriculture as a source of income, and because the poor spend a large proportion of their resources on food. The topic is a popular point of discussion on the Internet.

Due to the importance of this topic and its attendant human concerns, it is easy to find information about it on Web 2.0 sites, where public opinion about the globalization of agriculture takes place. People are sharing knowledge through social bookmarking sites as Delicious, creating a sort of international structuration of information in this area. In our model, the actors who form part of the *Delicious* globalization network share URLs and tags.

We posit the following hypotheses on network actors' use of the Web. The first refers to homophily. People tend to associate with those with whom they share some degree of similarity [9]. The second hypothesis relates to the distinction between explicit expressive behaviour (hyperlinks) and implicit expressive behaviour (implicit links between users).

The paper discusses the approach of collective online identity models paying specific attention to socio-semantic and network perspectives via Social Network Analysis (SNA).

## Literature Review

### Social tagging systems, online networks and socio-semantic networks

Social tagging is a common feature of shared web content applications that enable users to share favourite Internet content and tag these hyperlinks with free text. The user can choose these links freely without complying with any form of taxonomy or ontology. These applications, such as Delicious, are known as social tagging systems. We use the word "social" to emphasize the fact that these tags are adopted by a large number of users in the network. This categorization of content by the user via tagging gave rise to the term folksonomy [10,11].

Folksonomy is a phenomenon involving three group types: users, resources and tags, as well as the associations that occur among them [12]. Thus, the structure of social tagging websites can be viewed as a network of three different types of nodes: U users, R resources (URLs) and T tags. Two users $u$ and $u'$ are practically related if and only if there exists at least one resource $r$ that both $u$ and $u'$ share. Likewise, two users $u$ and $u'$ are symbolically related if there exists at least one

**\*Corresponding author:** Juan D. Borrero, Departamento de Dirección de Empresas y Marketing, Universidad de Huelva, Plaza de la Merced, Spain, Tel: +91 9994992532; E-mail: jdiego@uhu.es

tag *t* that both *u* and *u'* have used. These relations then may be used to generate networks of users, on the basis of resources or on the basis of tags. The online network that emerges can be graphically illustrated using the *1* link between the *u* user and the *r* URL that passes through the *t* tag.

Although there are numerous recent empirical studies related to social tagging systems [13-16] our approach is based on both social - a network of social relations between users [14,15] - and semantic - a straight forward representation of user affiliations to concepts – networks [17,18] and involves community users searching out other content that could help to forge a collective identity via the formation of "hidden" links among them.

### Online collective identity

Information Communication Technologies (ICTs) have changed the media's geographical scope and have consequently altered the construction of collective identities [19,20]. Scholars have found evidence of similar processes occurring over the Internet [21,22].

In our case, an individual may access bookmarks made by a set of people who all happen to have accessed each other's bookmarks; such a set of individuals are operating in a somewhat constrained topical space [23], are likely to share similar interests, and are likely to draw information from similar sets of digital resources.

Leveraging the new social movement theory [8], our methodology is centered on the concept of collective identity in social media. This is demonstrated by taking account the significance of the process through which individuals attribute meanings (translated in our methodology as 'concern') to events (or causes) and find association with each other through such processes facilitated by social media networks [24].

Thus, collective identity is an interactive and shared process through which several individuals share certain orientations in common [8].

The concept of frame is central to collective identity [25] and our use of frames draws from "semantic networks". Some researchers use the semantic layer to interpret and analyze links, not only as markers of quality but also of common interests and affiliations [26,27]. A frame component is a word that is part of a frame. An example of a frame component is the word "trade," which is an important component of the anti-Globalization network (frame). In general terms, frame components can be considered as an implicit medium of expression of the collective identity and are highly indicative of the closeness between actors. An online frame network is an undirected network in which nodes represent users and ties represent mutual use of a particular "frame component" (word or term that is part of a frame). For example, if user **u** and user **u'** both use the frame component (tag) "trade" to mark a website, then a (undirected) tie between the two users in the online frame network will exist [28].

Without shared meanings –frame components – and also without practical and symbolic resources via informal networks [29], it is unlikely that individuals will establish a collective identity [30].

We define practical resources as those that can be valued and measured objectively [28]. Thus, a practical exchange network can be a directed network in which hyperlinks are formed between users and URLs, such as opinion networks [31,32] in which users connect to the objects that they gather. For example, Delicious users are connected with the websites they collect.

We define a symbolic exchange network as an undirected network where the links between users reflect a mutual acknowledgement of shared characteristics and objectives. Users build an "identity (implicit) in a network (symbolic exchange network)" with information they share (links) through common interests (frame components).

While we model online frame development as a purely symbolic action, we regard hyperlinks as facilitating the exchange of both symbolic and practical resources [28]. According to these two interpretations, the Globalization of Agriculture in Delicious is a social network propelled by the forces of homophily in which users create links to other similar links. These networks are essential vectors in online collective identity because their main objective is not practical but symbolic [28].

### Hypothesis

Our theoretical framework is based on the idea that collective identity on the Web reveals itself in two processes: in the frame network (u-u) and in the opinion network (user$\rightarrow$URL). This theoretical framework leads us to propose two hypotheses concerning the detection of online collective identity.

### H1. The online frame networks will exhibit structural homophily

The literature on homophily in social networks specifies various classification criteria [33-35]. In our case, we consider that structural homophily [36] can appear when ties are restricted by characteristics, contexts and external situations such as sharing knowledge using social tagging on the Web and is induced because the individuals relate to each other implicitly via the context they occupy (social tagging systems).

Our view is that the exchanges are responsible for making individuals more or less similar, resulting in the generation of the structure and the attributes.

Following from this, Hypothesis 1 seeks only to determine whether the users are aware of different interests when establishing links in terms of the concept of social cohesion. We consider that a comparison can be made between the concepts of homophily and social cohesion. Therefore, we start from the basic idea that the configuration of social groups and, consequently, the attributes and social categories have their origins in interactive phenomena. The results of these interactions are distributed across networks that are more or less cohesive between users, giving birth to groups in other words, a social structure.

Thus, inside a collective, the internal ties or bonding enable its agents to generate status, or categories, according to the different positions within the network structure and, therefore, different configurations of social cohesion [37]. This interactive dynamic enables the formation of complementarity and linking as a single phenomenon of social cohesion and homophily.

Likewise, in the way we have defined the relational structure that is constitutive of the network, we can to relate it to assortative mixing concept. It refers to a positive correlation in the attributes of nodes that are adjacent in a network. A measure of assortative mixing might tell us that nodes that share a particular characteristic have a higher probability of being connected, but it gives no indication about the exact processes that have led to the formation of a particular network [38]. Hypothesis 1 simply states that if the users comprise distinct clusters then there should be statistical evidence of homophily on the basis of cluster affiliation.

### H2

The opinion and online frame networks will show different network

"effects", reflecting the differing degrees of explicitness that underlie their construction.

Online frame development does not bring practical resources into play; it is a merely symbolic action and it expresses implicit behaviour. On the contrary, hyperlinking facilitate the exchange of a practical resource and for this reason can be referred to as an explicit activity. Statistically significant network effects can be regarded as structural signatures, or indicators of the particular social forces underlying the network. We expect that these differences will be reflected in the structural signatures [39-41] of these networks.

## Method

We built the network of globalization of agriculture using a combination of search techniques proposed for researching "issue networks" [42]. The process of data retrieval and the representation of the *Delicious* community as a network follows a procedure that we present in Figure 1.

First (A), following links from an authoritative news source, we identified the search attributes on the basis of an original sample of a set of 26 webpages according to the Wikipedia definition of "critics of globalization". The important factor in this phase was to have an authoritative news source as a baseline to find keywords connected to globalization and then to narrow that idea to the globalization of agriculture as the main issue for the present work.

Through associative reasoning [42], we made educated guesses about relevant issues and found nine keywords commonly linked to all seed websites – globalization, agriculture and seven more words: trade, poverty, activism, development, food, organic, and GMO. The key concepts were extracted manually from the website homepages and from tag clouds or topics that appeared on the homepage (B). In a third step (C), we gathered the raw data sample of all the users' records, websites and tags available for the eight tag pairs around the main tag of globalization: globalization + trade, globalization + poverty, globalization + activism, globalization + development, globalization+ agriculture, globalization + food, globalization+ organic and globalization + GMO. These tags were identified by crawling through the social bookmarking website Delicious using a web crawler developed in Perl.

Finally, we developed a program in Haskell to reduce the amount of data (D) by cutting the URLs and using keywords, including the identification of synonyms and the elimination of words with capital letters and derivatives, such as words in the plural form. The data-gathering process covered one full month (April 22, 2011- May 21, 2011) and constituted 3,668 users, 2,148 websites and 4,776 tags.

Using Pajek software [43], SNA was used for the analysis, and algorithms and layout techniques networks [44] were used for the visualizations created with Gephi [43].

## Results

In our opinion, the concept of social cohesion must also be linked to the form and intensity in which the ties are distributed, those that occur between members of the collective; more specifically, it has to do with bonding or ties within the collective. The identity and intensity of social cohesion come from the exchanges-relations between members of a collective. Table 1 shows all the nodes and links for the different opinion and online frame networks represented. Table 2 shows the density and other measures of cohesion (mean k-core and modularity), and the centralization measure.

To test the first hypothesis related to homophily, we examine the cohesion of network (i.e., density, k-core and modularity) [45].

Density is a measure of the level of connectivity within the network and reflects the actual number of links as a proportion of total possible links. The frame network (frame component = tag) has a density of 33.61% (the number of links expressed as a proportion of the total possible number of ties) and the density of the frame network (frame component = url) is 1.95%. As in large-scale networks, this indicates a sparse network, but it is greater than what has been found in other research [28,46,47].

Cohesion is the degree to which actors are connected directly to each other by cohesive bonds. A k-core of a network is a maximal connected sub-network in which all nodes have a degree of at least k. The concept of a *k*-core was introduced to study the clustering structure of social networks [48]. In the frame network, we found two clusters, one with 1,541 users and another with 2,127 users. But, if we extract from the network, those nodes with one or no link, we can see that the new netwok is completely cohesive, with a density of 99.95%, consisting of 2,127 users, all from the second cluster above.

Figures 2 and 3 show the online frame network maps, drawn using Fruchterman and Reingold [49] force-directed graphing layout, where a tie between two users reflects mutual use of at least one of the tags or
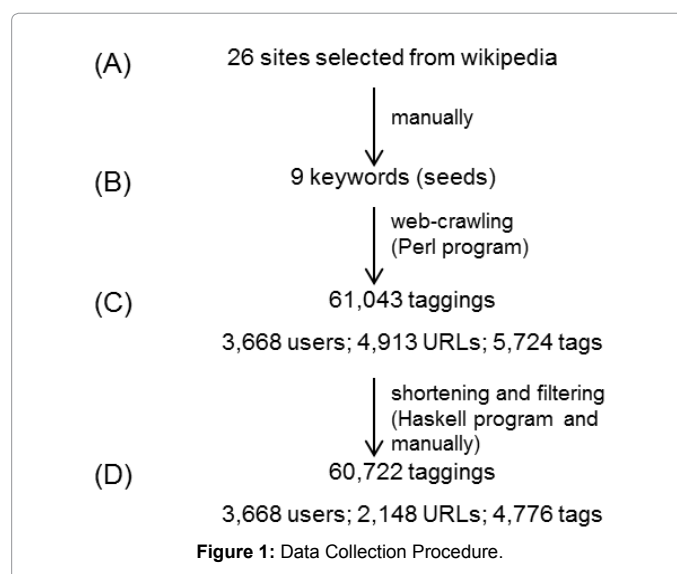


| | |
|---|---|
| (A) | 26 sites selected from wikipedia |
| | ↓ manually |
| (B) | 9 keywords (seeds) |
| | ↓ web-crawling (Perl program) |
| (C) | 61,043 taggings |
| | 3,668 users; 4,913 URLs; 5,724 tags |
| | ↓ shortening and filtering (Haskell program and manually) |
| (D) | 60,722 taggings |
| | 3,668 users; 2,148 URLs; 4,776 tags |

**Figure 1:** Data Collection Procedure.

| Network | | Relation | No. of links | No. of nodes |
|---|---|---|---|---|
| user→url | Opinion | Directed | 5,816 | 7,200 |
| user-user (tag) | Frame1 | Undirected | 3,668 | 2,261,001 |
| user-user (url) | Frame2 | Undirected | 3,668 | 131,165 |

**Table 1:** Social Network Descriptions from *Delicious* Data Set.

| Network | Density | Centralization (Av. Degree) | Communities | Modularity[a] | *K*-core |
|---|---|---|---|---|---|
| user→url | 0.09% | 2.4759 | 473 | 0.7541 | 5 |
| user-user (tag) | 33.61% | 1,232.8250 | 1,544 | 0.0751 | 2,126[b] |
| user-user (url) | 1.95% | 71.5185 | 450 | 0.5881 | 258 |

[a] Louvain Method (Resol. = 1, multy-level coarsening+simple refinement, calculated by Pajek).
[b] Two clusters: 0-core with 1,541 nodes (42.012%) and 2,126-core with 2,127 nodes (57.988%).

**Table 2:** Social Network Statistics from Delicious Data Set.

URLs used in bookmarking websites. Node color in both figures reveal a strong community structure or clustering [50] evidence of network homophily [3,51].

## Centralization

For the second hypothesis regarding different degrees of explicitness, we also examine the measure of centrality. The degree of centralization at least partially provides information on a type of social stratification within the community. How networks of relationships in online communities are structured has important implications for how social capital may be generated, which is critical to both attract and govern the necessary user base to sustain the site [52].

The average degree of a network is computed over the degree of all nodes, i.e., the average number of neighbors of nodes. As expected, the online frame networks are highly centralized than the opinion network (Table 2).

## Discussion and Conclusions

This article has empirically examined the existence of collective identity in social bookmarking sites.

Our model introduces a boundary between practical and symbolic resources, which are transferred by via online networks by differentiating between expressive behaviour that is totally explicit (such as the creation of opinion networks) and expressive behaviour that is implicit (as in online frame networks).
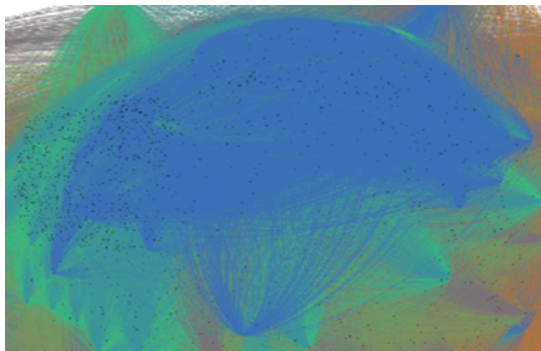


**Figure 2:** Online frame_1 network Energy-Fruchterman Map. Node color reflecting modularity clusters.
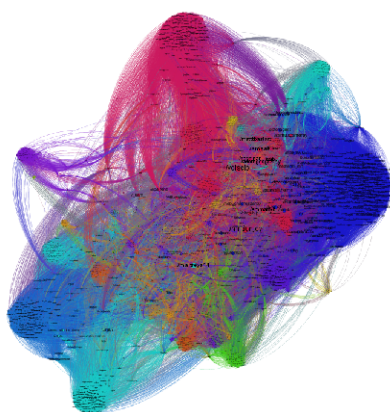


**Figure 3:** Online frame_2 network Energy-Fruchterman Map. Node color reflecting modularity clusters.

This leads us to formulate hypotheses on the existence of homophily in the online frame networks, on the informal structure of the opinion network and the behavioural differences between the opinion and online frame networks. An empirical application based on digital data compiled automatically from Delicious has been the support for these hypotheses.

Findings on the density, components and force-directed visualizations in the previous section indicate there is homophily in the online frame networks.

Social cohesion is part of a relational or reticulated perspective based on bonding. Social cohesion emerges from a collective that is not necessarily defined or delimited by categories or predetermined social attributes; it originates in those agents with the ability to interact and relate. As previously mentioned, we can relate to the concept of assortative mixing.

We can also identify three significant differences between both network types which demonstrate the various degrees of explicitness in the expressive behaviour that underlies the Delicious online community.

Firstly, it is important to note that the density of the network of users is greater than the density in the user→url network, which shows that the community of users is inherently more cohesive and it exists an implicit expressive behaviour. Secondly, the density of the frame network (frame component = tag) is greater than that of the frame network (frame component = url), which shows that on the basis of implicit links, the users are more closely linked to common terms than to common websites. Thirdly, the evidence of greater centralization in the online frame networks compared to the centralization apparent in the opinion network also coincides with our theoretical model, which that places the creation of online frames towards the less degree of explicitness. Because actors in Delicious place such a premium on informality and horizontality, the comparatively more explicit quality of the act of hyperlinking results in the opinion network being less centralized than the corresponding online frame networks.

By defining and empirically testing for the structural signatures of online collective identity, our approach enables the accurate and effective mapping of the contours of online collective identity, enabling large-scale comparative work across other social media.

Finally, our approach also represents an important first step towards the development of empirical techniques capable of automatically discovering the existence of online collective identity and the formulation of strategies on the knowledge base of collective interests.

## References

1. Boyd DM, Ellison NB (2007) Social Network Sites: Definition, History and Scholarship. Journal of Computer-Mediated Communication 13: 210-230.

2. Kane GC, Fichman RG, Gallaugher J, Glaser J (2009) Community Relations 2.0. Harvard Business Review 87: 45-50.

3. Kwon O, Wen Y (2010) An Empirical Study of the Factors Affecting Social Network Service Use. Computers in Human Behavior 26: 254-263.

4. Ames M, Naaman M (2007) Why we tag: motivations for annotation in mobile and online media. In: Proceedings of the SIGCHI conference on Human factors in computing systems. San Jose, California, USA.

5. Golder SA, Macy MW (2013) Social Media as a Research Environment Cyberpsychology, Behavior and Social Networking 16: 627-628.

6. Diani M (1992) The Concept of Social Movement. Sociological Review 40: 1-25.

7. Melucci A (1989) Nomads of the Present: Social Movements and Individual Needs in Contemporary Society. Temple University Press, Philadelphia.

8. Melucci A (1995) The Process of Collective Identity. In Johnston H, Klandemans B eds. Social Movements and Culture. University of Minnesota Press, Minneapolis.

9. McPherson M, Smith-Lovin M, Cook JM (2001) Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology 27: 415-444.

10. Golder SA, Huberman BA (2006) Usage Patterns of Collaborative Tagging Systems. Journal of Information Science 32: 198-208.

11. Noruzi A (2006) Folksonomies: (Un) Controlled Vocabulary? Knowledge Organization 33: 199-203.

12. Smith G (2008) Tagging: People-powered metadata for the social web. New Riders Press, Berkeley.

13. Derntl M, Hampel T, Motschnig-Pitrik R, Pitner T (2011) Inclusive social tagging and its support in Web 2.0 services. Computers in Human Behavior 27: 1460-1466.

14. Gray P, Parise S, Iyer B (2011) Innovation Impacts of Using Social Bookmarking Systems. MIS Quarterly 35: 629-643.

15. Huang SL, Lin SC, Chan YC (2012) Investigating effectiveness and user acceptance of semantic social tagging for knowledge sharing. Information Processing & Management 48: 599-617.

16. Panke S, Gaiser B (2009) "With My Head Up in the Clouds" Using Social Tagging to Organize Knowledge. Journal of Business and Technical Communication 23: 318-349.

17. Monge P, Eisenberg E (1987) Emergent Communication Networks. In: Jablin F, Putnam L, Roberts K, Porter L eds. Handbook of Organizational Communication. Newbury Park: Sage.

18. Stohl C (1993) European Managers' Interpretations of Participation: A Semantic Network Analysis. Human Communication Research 20: 97-117.

19. Arquilla J, Ronfeldt D (2001) The Advent of Netwar (revisited) . In: Networks and netwars: the future of terror, crime, and militancy. Santa Monica, CA: Randpp 1-25.

20. Myers DJ (2000) The Diffusion of Collective Violence: Infectiousness, Susceptibility and Mass Media Networks. American Journal of Sociology 106: 173-208.

21. Ayers MD (2002) Comparing Collective Identity in Online and Offline Feminist Activists. In: Cyberactivism: Online Activism in Theory and Practice.New York: Routledgepp 145-164.

22. Wall MA (2007) Social Movements and Email: Expressions of Online Identity in the Globalization Protests. New Media & Society 9: 258-277.

23. Marlow C, Naaman, Boyd D, Davis (2006) Ht06, Tagging Paper, Taxonomy, Flickr, Academic Article, to Read. In: Proceedings of the 17th Conference on Hypertext and Hypermedia. New York pp. 31-40.

24. Agarwal N, Lim M, Wigand R (2012) Raising and Rising Voices in Social Media. A Novel Methodological Approach in Studying Cyber-Collective Movements. Business & Information Systems Engineering 4: 113-126.

25. Hunt SA, Benford RD, Snow DA (1994) Identity Fields: Framing Processes and the Social Construction of Movement Identities. In: Larana E, Johnston H, Gusfield JR eds. New Social Movements: From Ideology to Identity,Temple University Press, Philadelphia.

26. Adamic LA, Adar E (2003) Friends and Neighbors on the Web. Social Networks 25: 211-230.

27. Huberman BA (2001) The Laws of the Web: Patterns in the Ecology of Information. Cambridge, MA: MIT Press.

28. Ackland R, O'Neil M (2011) Online Collective Identity: The Case of the Environmental Movement. Social Networks 33: 177-190.

29. Diani M (2003) Networks and Social Movements: A Research Programme. In Diani M, McAdam D eds. Social Movements and Networks: Relational Approaches to Collective Action. Oxford: Oxford University Press.

30. McAdam D, McCarthy JD, Zald MN (1996) Comparative Perspectives on Social Movements: Political Opportunities, Mobilizing Structures, and Cultural Framings. New York: Cambridge University Press.

31. Blattner M, Zhang Y-C, Maslov S (2007) Exploring an opinion network for taste prediction: An empirical study. Physica A 373: 753-758.

32. Maslov S, Zhang YC (2001) Extracting Hidden Information from Knowledge Networks. Physical Review Letter 87: 248701-248705.

33. Lane PJ, Salk JE, Lyles MA (2001) Absorptive capacity, learning, and performance in international joint ventures. Strategic Management Journal 22: 1139-1161.

34. Mowery DC, Oxley JE, Silverman BS (1996) Strategic alliances and interfirm knowledge transfer. Strategic Management Journal 17: 77-91.

35. Pfeffer J, Nowak P (1976) Joint ventures and interorganizational interdependence. Administrative Science Quarterly 21: 398-418.

36. Ahuja G, Polidoro F, Mitchell W (2009) Structural Homophily or Social Asymmetry? TheFormation of Alliances by Poorly Embedded Firms. Strategic Management Journal 30: 941-958.

37. Shaw ME (1981) Group Dynamics: The Psychology of Small Group Behavior. NewYork, NY: McGraw-Hill.

38. Wimmer A, Lewis K (2010) Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. American Journal of Sociology 116: 583-642.

39. Faust K, Skvoretz J (2002) Comparing networks across space and time, size and species. Sociological Methodology 32: 267-299.

40. Skvoretz J, Faust K (2002) Relations, species, and network structure. Journal of Social Structure Volume3.

41. Welser H, Gleave E, Fisher D, Smith M (2007) Visualizing the signatures of social roles in online discussion groups. Journal of Social Structure, Volume 8.

42. Rogers R, Zelman A (2002) Surfing for knowledge in the information society. In: Elmer G, Rowman L eds. Critical Perspectives on the Internet. Lanham, MD.

43. De Nooy W, Mvrar A, Batagel JV (2005) Exploratory Social Network Analysis with Pajek. Cambridge University Press.

44. Fortunato S (2010) Community Detection in Graphs. Physics Reports 486: 75-174.

45. Bastian M, Heymann S, Jacomy M (2009) Gephi: An open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media Paris, France.

46. Lusher D, Ackland R (2011) A relational hyperlink analysis of an online social movement. Journal of Social Structure, Volume 12.

47. Shumate M, Dewitt L (2008) The North/South Divide in NGO Hyperlink Networks. Journal of Computer Mediated Communication 13: 405-428.

48. Seidman, SB (1983) Network structure and minimum degree. Social Networks 5: 269-287.

49. Fruchterman TMJ, Reingold EM (1991) Graph Drawing by Force-Directed Placement. Software: Practice and Experience 21: 1129-1164.

50. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of community hierarchies in large networks. J Stat Mech 10: P10008.

51. Noack A (2009) Modularity clustering is force-directed layout. Physical Review E 79, 026102.

52. Ganley D, Lampe C (2009) The Ties that Bind: Social Network Principles in Online Communities. Decision Support Systems 47: 266-274.