How Influential Are You: Detecting Influential Bloggers in a Blogging Community

Imrul Kayes[1], Xiaoning Qian[1], John Skvoretz[2], and Adriana Iamnitchi[1]

[1] Computer Science & Engineering, University of South Florida, Tampa, FL, USA
[2] Department of Sociology, University of South Florida, Tampa, FL, USA
imrul@mail.usf.edu, {xqian,anda}@cse.usf.edu, jskvoretz@usf.edu

Abstract. Blogging is a popular activity with high impact on marketing, shaping public opinions, and informing the world about major events from a grassroots point of view. Influential bloggers are recognized by businesses as significant forces for product promotion or demotion, and by oppressive political regimes as serious threats to their power. This paper studies the problem of identifying influential bloggers in a blogging community, BlogCatalog, by using network centrality metrics. Our analysis shows that bloggers are connected in a core-periphery network structure, with the highly influential bloggers well connected with each others forming the core, and the non-influential bloggers at the periphery. The six node centrality metrics we analyzed are highly correlated, showing that an aggregate centrality score as a measure of influence will be stable to variations in centrality metrics.

Keywords: social networks, influence, network centrality, blogosphere.

1 Introduction

The new age of participatory web applications commonly known as Web 2.0 has enabled the transition of the traditional information consumers into information producers in a form of *grassroots journalism* [1]. This kind of web applications include blogs, wikis, social annotation and tagging, and media sharing.

Blogging, in particular, distinguishes itself through both popularity and impact. For example, WordPress alone, a free and open source blogging tool, is used by over 14.7% of Alexa Internet's "top 1 million" websites and as of August 2011 manages 22% of all new websites [2]. Citizen journalism had high impact in major events such as South Asia tsunami, London terrorist bombings, and New Orleans Hurricane Katrina [1]. The *blogosphere*, the virtual universe of the blogs on the web, provides thus a conducive platform for different aspects of virtual and real life, such as viral marketing [3], sales prediction [4, 5], business models [6], and counter terrorism efforts [7].

A blog (also referred to as a "web log") is a personal journal published on the World Wide Web consisting of discrete entries ("posts") typically displayed in reverse chronological order. Blogs are usually the work of a single individual, occasionally of a small group, and are often themed on a focused topic. A conventional blog may combine text, images and links to other blogs and to web

K. Aberer et al. (Eds.): SocInfo 2012, LNCS 7710, pp. 29–42, 2012.

pages. Blogging platforms allow the creation of online profiles in which links to other bloggers are specified. This blogger to blogger ties specify the blogger's interest and endorsement of other bloggers, creating a social network through which blog updates are automatically disseminated.

The influence bloggers have on forming public opinions is significant. First, bloggers influence other bloggers' opinions: in 2011, 68% of bloggers claimed to be influenced by the blogs they read [8]. Second, they can influence the opinions of the masses: 38% of bloggers talk about brands positively and negatively on their blogs. Studies [9] show that 83% of people prefer consulting family, friends or an expert over traditional advertising before trying a new restaurant, 71% of people do the same before buying a prescription drug or visiting a place.

The advantages of identifying influential bloggers are already evident: influential bloggers are often market-movers. Identifying these bloggers can help companies better understand key concerns, identify new trends, and smartly affect the market by targeting influential bloggers with additional information to turn them into unofficial spokespersons [10]. About 64% of the companies are shifting their focus to blogging [11].

This paper investigates the position of influential bloggers in the BlogCatalog blogging community. Based on previous research [12, 13] that correlated a node's position in the network to its influence, we conjecture that the influence of a blogger is represented by its location in the blogging network.

The contributions of this work are the following. First, we propose a method that aggregates different network position measurements into an overall influence score and demonstrate quantitatively that variation in one metric is not likely to significantly affect the aggregate score. Second, we discover that the overall pattern of the BlogCatalog community is that of a core-periphery structure, in which the highly influential bloggers are tightly connected to each others and the non-influential bloggers form the periphery.

The remainder of this paper is organized as follows: Section 2 presents the methodology of our quantitative study. Section 3 presents empirical results and analysis. Section 4 describes related work. We conclude in Section 5 with a discussion of the consequences of our results.

2 Methodology

The influence a node has on other nodes in the network can be represented in social network analysis by different centrality metrics. For example, the larger the number of direct neighbors, the larger an audience the node has for direct communication. Alternatively, the larger the number of paths between other pairs of nodes a node is part of, the more it can control the communication between distant nodes. Based on this intuition, we conjecture that a blogger's influence is determined by and manifests via its centrality in the blogging community.

We propose to aggregate different representative centrality metrics into a final influence score. We define the influence score of a node as the average of the positions of that node in decreasing order of centrality scores over various centrality metrics. Specifically, each centrality metric assigns each node a score that can be used to order nodes in decreasing order of importance (according to that centrality). This allows each blogger to receive a rank according to each centrality metric: the first ranked blogger will be the most central one, the last ranked will be the one with the lowest centrality score. Bloggers having the same centrality score are given the same rank. A blogger's final rank is the average rank over all centrality measures. We selected six representative centrality metrics as the focus of our study: degree, betweenness, closeness, eigenvector, hub, and communicability centrality.

Degree centrality is defined as the number of links that a node has. Although simple, degree centrality intuitively captures an important aspect of blogger's potential influence: bloggers who have connections to many others are read by more people, have access to more information, and certainly have more prestige than those who have fewer connections. High degree centrality bloggers can reach many bloggers directly.

Betweenness centrality, which measures the extent to which a node lies on the shortest paths between other nodes, was introduced as a measure for quantifying the control of a human on the communication between other humans in a social network [14]. Bloggers with high betweenness centrality may have considerable influence within a network by virtue of their control over information passing between others: they can comment, annotate, re-interprete the posts originating from a distant blogger and these altered views can be seen by other remote bloggers. The nodes with highest betweenness are also the ones whose removal from the network will most disrupt communications between other nodes because they lie on the largest number of paths taken by messages [15]. Formally, the betweenness centrality of a node is the sum of the fraction of all-pairs shortest paths that pass through :

$$C(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)} \tag{1}$$

where v is the set of nodes, $\sigma(s,t)$ is the number of shortest (s,t) paths, and $(s,t|v)$ is the number of those paths passing through some nodes v other than s,t. If $s = t$, $\sigma(s,t) = 1$, and if $v \in s,t,\sigma(s,t|v) = 0$. Our implementation of betweenness for this research is based upon the Brandes algorithm [16].

Closeness centrality measures the mean distance from a node to other nodes, assuming that information travels along the shortest paths. Formally, the closeness centrality $(C(x))$ of a node x is defined as follows:

$$C(x) = \frac{n-1}{\sum_{y \in U, y \neq x} d(x,y)} \tag{2}$$

where $d(x,y)$ is the distance between node x and node y; U is the set of all nodes; d is the average distance between x and the other nodes. In our blogging network this centrality measure estimates the amount of information a blogger may have access to compared to other bloggers. Specifically, a blogger with lower mean distance to others can reach others faster.

To account for the fact that not all communications take place along the shortest path, we also consider communicability centrality. This centrality measure is defined as the sum of closed walks of all lengths starting and ending at the node [17].

The centrality of a node does not only depend on the number of its adjacent nodes, but also on their relative importance. Eigenvector centrality allocates relative scores to all nodes in the network such that high-score neighbors contribute more to the score of the node. Formally, Bonacich [18] defines the eigenvector centrality $C(v)$ of a node v as the function of the sum of the eigenvector centralities of the adjacent nodes, i.e.

$$C(v) = 1/\lambda \sum_{(v,t)\in E} c(t) \tag{3}$$

where λ is a constant. This can be rewritten in vector notation, resulting in an eigenvector equation with well-known solutions.

Hubs and authorities are other relevant centralities for the blogging community context. Authorities are nodes that contain useful information on a topic of interest; hubs are nodes that know where the best authorities are to be found [15]. A high authority centrality node is pointed to by many hubs, i.e., by many other nodes with high hub centrality. A high hub centrality node points to many nodes with high authority centrality. These two centralities can play a significant role also in our work of finding influential bloggers. They can infer that the bloggers that have high hub and authority centrality are not only influential but also they are connected with influential bloggers.

3 Quantitative Analysis

We computed the centrality metrics presented before on a real dataset from the BlogCatalog blogging community. We implemented the algorithms in Python 2.7 with the NetworkX library for graph processing, and used `awk` for result processing.

3.1 Dataset

For our experiments we used the declared social network of bloggers on BlogCatalog (`www.blogcatalog.com`) available at [19]. BlogCatalog is a blogging service that allows its members to create online profiles, post their blogs, and automatically receive blogging updates from the BlogCatalog users with whom they have declared "friend" relationships. At the time of data collection, BlogCatalog relationships were symmetrical; at the time of this writing, however, BlogCatalog maintains directed relationships, similar to follower–followed relationships in Twitter. The dataset thus represents an undirected social graph. Where needed

Table 1. Average path length, radius, diameter and clustering coefficient of the BlogCatalog network compared to other networks

Network	Nodes	Avg. path len.	Radius	Diameter	Clustering coefficient
BlogCatalog	10,312	2.38	3	5	0.460
Orkut	3,072,441	4.25	6	9	0.171
LiveJournal	5,284,457	5.88	12	20	0.330
Erdös-Renyi	10,312	2.65	3	3	0.006
Web	200M	16.12	475	905	0.081

for centrality metrics computations, we treated an undirected edge as two directed edges, as it is typically done and supported by the meaning of an edge in our dataset. The network size is 10,312 nodes and 333,983 edges (average degree 64.78 and density 0.00628). The structural properties of the network are presented in Table 1.

For a brief characterization, we compare the BlogCatalog network properties with other networks from diverse domains: the LiveJournal blogging network [20], the Orkut online social network [20], the Web graph [21], and the Erdös-Renyi random graph of the same size as the BlogCatalog dataset ($|V| = 10,312$, $p = 0.00628$). Table 1 shows the average path length, radius, diameter and clustering coefficient of all five networks. A notable characteristic of the blogging communities is the high clustering coefficient compared to other networks. Given a network $G = (V, E)$, the clustering coefficient C_i of a node $i \in V$ is the proportion of all the possible edges between neighbors of the node that actually exist in the network [15]. A high clustering coefficient in both blogging networks implies that a blogger's connections are interconnected and have a greater effect on one another. The small average path length (2.38), comparable with that of the corresponding random graph (2.65), together with the high average clustering coefficient, places the BlogCatalog network in the category of small-world graphs [22]. As in many other real networks [23], BlogCatalog exhibits scale-free properties. Figure 1 shows the complementary cumulative degree distribution of bloggers. The distribution fits a power-law distribution with exponent $\alpha = 2.52$. Most real-world networks with power-law degree distributions have values of α in the range $2 \leq \alpha \leq 3$ [15]. The most notable characteristic of a scale-free network is the occurrence of hubs, which hold a much higher number of links than the average node. As hubs control the "connectedness" of the network, we expect that influential bloggers also will be hubs in BlogCalatog.

3.2 Centralities and Influential Bloggers

As described in Section 2, we use centrality metrics to rank blogger's importance in the network. Figure 2 shows cumulative distributions of various ranks. One of the objectives of plotting these distributions is to show how granular the ranks are, more specifically, how successful these centrality metrics are in assigning distinct scores to different nodes in the network. To this end, analyzing

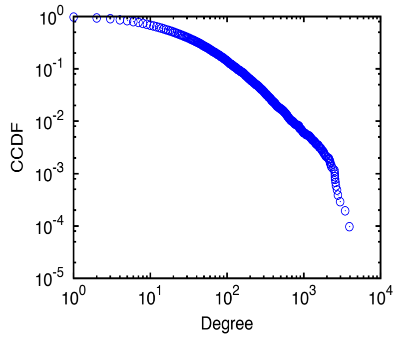

Fig. 1. Degree distribution in BlogCatalog

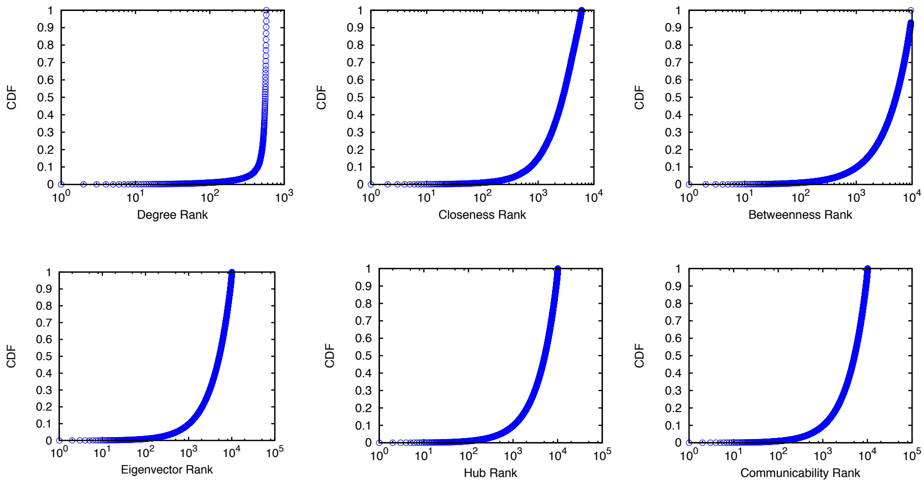

Fig. 2. Distribution of ranks of degree, closeness, betweenness, eigenvector, hub and communicability centralities

the distributions we get these facts: 5% of the bloggers cover the top 64% of the ranks in degree centrality scores, 12% of the bloggers correspond to top 12% ranks in closeness centrality, 10% of the bloggers correspond to top 10.80% ranks in betweenness centrality, 10% of the bloggers rank within 10.20% rank on eigenvector, hub rank distribution and 10% bloggers within top 10.19% rank on communicability rank distribution. So, we observe that all centrality measurements except degree centrality show granular scale of ranking, that is, they are typically capable of assigning a distinct score to each blogger (e.g., 10% bloggers within top 10.20% rank).

Bloggers that appear among the top 15 in multiple centrality metrics are represented in color in Table 2. The average rank of the top 10 most influential bloggers considering all centralities are shown in Table 3.

Table 2. The IDs of the top 15 bloggers according to each centrality measurement, sorted in increasing order by rank from left to right. DC: degree centrality, BC: betweenness centrality, CC: closeness centrality, EC: eigenvector centrality, HC: hub centrality, CoC: communicability centrality. Blogger IDs common to all centralities are colored the same. Black colored IDs represent bloggers who do not appear in the top 15 central bloggers from other centralities.

DC	4839	176	4374	8157	1226	4997	4984	8859	645	446	7098	7806	3198	2521	667
BC	176	4839	4374	8859	8157	645	1226	7806	233	446	3198	1932	4997	4984	7098
CC	4839	176	4374	8157	1226	4984	4997	8859	7098	645	7806	446	3198	2521	233
EC	4839	176	4374	1226	4984	8157	3198	4997	446	645	7098	2521	667	8859	4669
HC	4839	176	4374	1226	4984	8157	3198	4997	446	645	7098	2521	667	8859	4669
CoC	4839	176	4374	1226	4984	8157	3198	4997	446	645	7098	2521	667	8859	4669

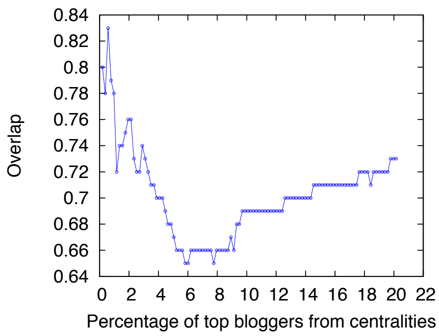

Fig. 3. Fraction of bloggers ranked in top x% by all centrality metrics

Table 3. Average rank of the top ten bloggers

Bloggers' ID	Average Rank
4839	1.17
176	1.83
4374	3.00
1226	4.83
8157	5.17
4984	7.00
4997	8.33
645	9.17
3198	9.17
446	9.83

3.3 Correlation of Centralities

Out of the 15 most influential bloggers listed on each centrality, 12 bloggers (80%) are common in all the centralities. To better understand the correlation between these centrality measures, we run the following experiment.

We incrementally select the top bloggers according to each centrality metric (in increments of 0.2%, from 1 to 20%) and compute the fraction of bloggers who are common. The fraction of common top bloggers according to all centrality measures is shown in Figure 3. This fraction ranges from 0.65 to 0.83, showing that all centrality metrics considered tend to identify about the same individuals. More interestingly, the overlap is higher at the beginning, more specifically for the top 1% most central bloggers.

To observe more closely, we plot the ranks of top 1% of the bloggers assigned by all centralities, showed in Figure 4. As expected, a blogger's assigned ranks from centralities form clusters and together with all the clusters we can visualize a straight line. This show that all the centralities tend to rank the same bloggers in the top.

Table 4. The correlation matrix of six centralities

	Degree	Betweenness	Closeness	Eigenvector	Hub	Communicability
Degree	1.00	0.67	0.65	0.68	0.68	0.67
Betweenness	0.67	1.00	0.85	0.89	0.89	0.88
Closeness	0.65	0.85	1.00	0.98	0.98	0.97
Eigenvector	0.68	0.89	0.98	1.00	1.00	0.98
Hub	0.68	0.89	0.98	1.00	1.00	0.98
Communicability	0.67	0.88	0.97	0.98	0.98	1.00

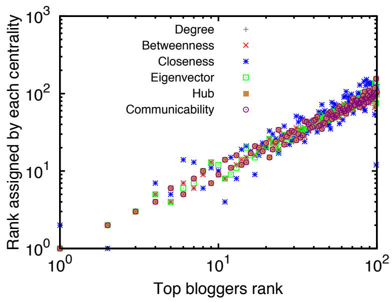

Fig. 4. Assigned rank of top 1% most influential bloggers from all centralities

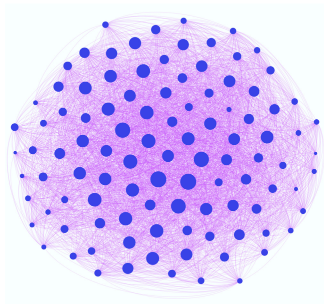

Fig. 5. The subnetwork of the top 1% most influential bloggers, considering only ties among them. Node size is proportional to clustering coefficient.

We consider the blogger ranks as assigned by each centrality and calculate the Pearson correlation coefficient between each pair of centralities, as shown in Table 4. An entry (i, j) in the matrix denotes the correlation coefficient between $Centrality_i$ and $Centrality_j$. The values of the correlation coefficients are high, ranging from 0.65 to 1.00. The high values of correlation coefficients indicate a strong correlation among the centralities in terms of finding influential bloggers. This phenomenon has been observed by other studies: for example, Valente et al. [24] observed high correlation between four centralities: degree, betweenness, closeness, and eigenvector in a network of 58 users. Our study validates their findings using a significantly larger network and a larger set of centrality metrics.

3.4 Interrelations of Influential Bloggers

The average clustering coefficient of influential bloggers is low in BlogCatalog. For 1% of the influential bloggers the average clustering coefficient is 0.07, where the overall network average is 0.46. Figure 6 shows the clustering coefficients of the top 1% influential bloggers. The low clustering coefficients of the influential bloggers imply that they work as network 'hubs' in the BlogCatalog network.

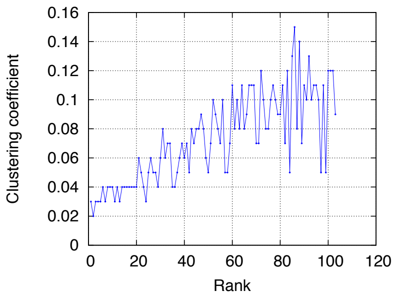

Fig. 6. The clustering coefficient of the top 1% most influential bloggers

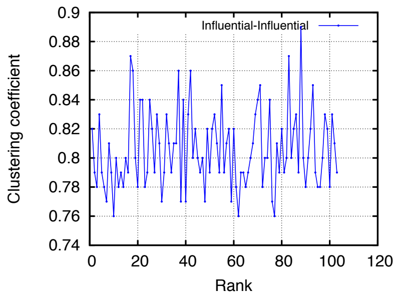

Fig. 7. The clustering coefficient of the top 1% most influential bloggers, considering only the ties among them

However, if we consider only the ties among influential bloggers, then the average clustering coefficient is very high, as shown in Figure 7: above 0.75, for an average of 0.81, thus significantly higher than the average of the entire network. Figure 5 depicts the sub-network of the most influential bloggers by representing the size of a node proportional to its clustering coefficient. Moreover, average path length of these bloggers is 1.22, where the network average is 2.38. This analysis show that influential bloggers in BlogCatalog are highly connected, similar to the way influential users cluster in other communities (such as Facebook [13]). We define the subnetworks of the bloggers and compute assortativity coefficient as shown in Table 5. The assortativity coefficient is a measure of the likelihood for nodes with similar degree to connect to each others, and it ranges between -1 and 1. A positive assortativity coefficient implies that nodes tend to connect to nodes of similar degree, while a negative coefficient implies the opposite. From the negative assortativity -0.25 of the entire network we can infer that nodes likely connect to nodes with very different degree from their own. Furthermore, exclusion of top 1% most influential bloggers from the network increases this trend of likelihood even more, which implicitly implies positive assortative mixing among influential bloggers. This implication is supported by the assortativity coefficient of $+0.07$ of the subnetwork of top 1% most influential bloggers, considering only ties among them. As such, the subnetwork of the top 1% most influential bloggers has negative assortativity (although less than the entire network) as they are connected with non-influential bloggers also. Along with the high clustering coefficient, we conclude that influential bloggers form a tightly-connected “core”, while the non-influential bloggers are located on the fringes of the network. A visualization of this phenomenon can be seen from Figure 8.

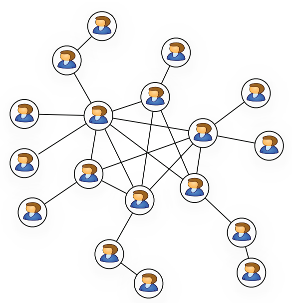

Fig. 8. Summary of interrelations of influential Bloggers

Table 5. Assortativity coefficient of different subnetworks

Subnetwork Definition	Assortativity
Entire network	−0.25
Subnetwork of bloggers excluding top 1% most influential bloggers	−0.67
Subnetwork of top 1% most influential bloggers	−0.16
Subnetwork of top 1% most influential bloggers, considering only ties among them	+0.07

4 Related Work

User influence is empirically elusive in social networks. Manski [25] states that user influence is difficult to identify in social observational data because influence is domain specific, thus domain-specific prior information is required. He argues that even if this information is available, the prospects for inference depend critically on the relationship between the variables defining the population. Inference is difficult to measure if these variables are statistically independent. In a similar vein, Aral et al. [26] observe diversity-bandwidth tradeoffs. The bandwidth of a tie is defined as the information transmission rate. Homophile nodes are connected by strong ties and interact more often, therefore have high bandwidth, but exchange little new information, whereas weak ties interact infrequently but are known to exchange new information. Both diversity of users and diversity of bandwidth are thus important for the diffusion of novel information. Since they are anti-correlated, there has to be a tradeoff to reach an optimal point in the propagation of new information.

Several approaches to identifying influential users have been proposed, including structural models [27], actor-oriented models [28], peer effects models [29], instrumental variable methods based on natural experiments [30], and ad hoc approaches based on specific data characteristics [31]. Our approach fits with structural models, as we used topological position as a measure of influence.

The problem of identifying influential bloggers in a blogging network has been studied empirically in BlogCatalog [32]. The authors compute a blogger's influence score based on four measures: activity, recognition, novelty, and eloquence. The study finds that influential bloggers are not necessarily active bloggers, thus, only considering a blogger's activity (e.g., number of posts or comments generated) may not reflect the blogger's influence in the network. Our approach, instead, considers only the bloggers' position in the network.

Domain-specific information has been used in other studies. Trusov et al. [33] identified influential users in online social networks based on longitudinal records of user log-in activity. They consider a user "influential" if her activity level, as captured by site log-ins over time, has a significant effect on the activity levels of other users. They found that on average, approximately one-fifth of a user's friends actually influence the user's activity level on the site. By using users' real time activity correlated to that of their neighbors (thus, local network topology), this approach captures the local influence and disregards potential distant influences. Aral et al. [13] conducted an experiment to measure influence in the product adoption decisions of a representative sample of 1.3 million Facebook users. The experiment involved the random manipulation of influence-mediating messages sent from a commercial Facebook application. The application lets users share information and opinions about various social contexts. As users adopted and used the product, automated notifications of their activities were sent to randomly selected users of their social contacts. The study shows that influential individuals are less susceptible to influence than non-influential individuals and that they cluster in the network, while susceptible individuals do not. Influence in Twitter has been measured with TwitterRank [34], a variant of PageRank that also considers the topical similarity between users. Tang et al. [35] propose the UserRank algorithm which combines link analysis and content analysis techniques to identify influential users in an online healthcare forum. Han et al. [36] identify influential users in mobile social networks using fixed-length random walks.

New topology-aware centrality metrics have been proposed for measuring influence. Ilyas et al. [37] introduce the principal component centrality metric for identifying influential neighborhoods. The authors take eigenvector centrality as the *de facto* measure of node influence and identify influential nodes in Orkut that are not discovered by eigenvector centrality. This approach takes eigenvector centrality as the sole influence measure, while we consider multiple measures.

Customized ranks and topological similarities have been also studied in identifying influential users in different networks. Subbian el al. [38] propose the supervised Kemeny ranking aggregation method that combines different influence measures to produce a composite ranking mechanism. Ghosh and Lerman's [39] influence model use geodesic-path based distance measures and topological ranking measures. They introduce a normalized α-centrality algorithm that takes as input the score of a node (in this case, number of votes on Digg.com). This centrality measurement is domain dependent and can only be used in networks where voting feature is enabled.

Shetty et al. [40] proposed an entropy model for determining most influential nodes. Their social graph encodes nodes as persons or organizations and edges as the actions they are involved in. Influential nodes are those who affect the graph entropy most when they are removed from the graph (similar to hubs in our case). Zhang et al. [41] use PageRank or HITS link analysis algorithms for expert finding in a closed domain, assuming that the importance of a web page reflects the influence of its author in the social network. Our approach makes similar assumptions in that we also assume bloggers gain influence by virtue of staying structurally important in the network.

5 Summary

The blogging community has established itself as a fast growing and effective social media platform. Understanding influence within a blogging network is a problem with increasing relevance to marketing and information retrieval. We proposed a centrality aggregation method to measure relative influence scores of bloggers in the network. We apply our methodology to the BlogCatalog blogging community and learn the following: (1) some bloggers span significant influence on fellow bloggers due to their strategic location in the network; (2) the six network centrality metrics we studies (degree, betweenness, communicability, closeness, eigenvector and hub) are highly correlated in this community; and (3) influential bloggers form a densely connected core, while non-influential bloggers remain at the periphery of the network, less likely to connect to each other.

The core-periphery structure of the bloggers social network allows us to state the following hypothesis for future research: the structure of any discourse space will tend over time to a core-periphery pattern in which a small subset of contributors to the discourse will exercise hegemonic influence over the remaining vast majority of contributors. This hypothesis could apply, among others, to scientific disciplines viewed as discourse spaces.

Acknowledgements. We thank Nicolas Kourtellis and Xiang Zuo of University of South Florida for their feedback. The US National Science Foundation supported this research under grants CNS 0952420 and CNS 0831785.

References

[1] Gillmor, D.: We the Media: Grassroots Journalism by the People, for the People. O'Reilly (2006)
[2] Rao, L.: Wordpress now powers 22 percent of new active websites in the u.s (2011), http://techcrunch.com/2011/08/19/wordpress-now-powers-22-percent-of-new-active-websites-in-the-us/
[3] Richardson, M., Domingos, P.: Mining knowledge-sharing sites for viral marketing. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 61–70 (2002)
[4] Gruhl, D., Guha, R., Kumar, R., Novak, J., Tomkins, A.: The predictive power of online chatter. In: Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, pp. 78–87 (2005)
[5] Mishne, G., de Rijke, M.: Deriving wishlists from blogs show us your blog, and we'll tell you what books to buy. In: Proceedings of the 15th International Conference on World Wide Web, pp. 925–926 (2006)
[6] Scoble, R., Israel, S.: Naked conversations: how blogs are changing the way businesses talk with customers. John Wiley (2006)
[7] Coffman, T., Marcus, S.: Dynamic classification of groups through social network analysis and hmms. In: Proceedings of IEEE Aerospace Conference, pp. 3197–3205 (2004)

[8] Technorati: State of the blogosphere 2011: Introduction and methodology (2011), http://technorati.com/social-media/article/state-of-the-blogosphere-2011-introduction/
[9] Keller, E., Berry, J.: One American in ten tells the other nine how to vote, where to eat and, what to buy. They are The Influentials. The Free Press (2003)
[10] Aubrey, A.: Mcdonald's courts mom bloggers when changing the menu (2011), http://www.npr.org/blogs/health/2011/07/27/138746335/mcdonalds-courts-mom-bloggers-when-changing-the-menu
[11] Elkin, T.: Just an online minute... online forecast (2005), http://www.mediapost.com/publications/article/29803/just-an-online-minute-online-forecast.html/
[12] Aral, S., Muchnika, L., Sundararajana, A.: Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. PNAS 106, 21544–21549 (2009)
[13] Aral, S., Walker, D.: Identifying influential and susceptible members of social networks. Science 337, 337–341 (2012)
[14] Freeman, L.: A set of measures of centrality based upon betweenness. Sociometry 40, 35–41 (1977)
[15] Newman, M.E.J.: Networks: An Introduction. Oxford University Press (2010)
[16] Brandes, U.: On variants of shortest-path betweenness centrality and their generic computation. Social Networks 30, 136–145 (2008)
[17] Estrada, E., Rodriguez-Velazquez, J.A.: Subgraph centrality in complex networks. Physical Review E 71, 056103 (2005)
[18] Bonacich, P.: Factoring and weighting approaches to status scores and clique identification. Journal of Mathematical Sociology 2, 113–120 (1972)
[19] Zafarani, R., Liu, H.: Social computing data repository at ASU (2009), http://socialcomputing.asu.edu
[20] Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, pp. 29–42 (2007)
[21] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J.: Graph structure in the web. In: Proceedings of the 9th International World Wide Web Conference, pp. 309–320 (2000)
[22] Watts, D.J., Strogatz, S.: Collective dynamics of 'small-world' networks. Nature 393, 440–442 (1998)
[23] Barabasi, A.L.: Linked: How Everything Is Connected to Everything Else and What It Means. Plume (2003)
[24] Valente, T., Coronges, K., Lakon, C., Costenbader, E.: How correlated are network centrality measures? Connections 28, 16–26 (2008)
[25] Manski, C.: Identification of endogenous social effects: The reflection problem. The Review of Economic Studies 60, 531–542 (1993)
[26] Aral, S., Alstyne, M.V.: The diversity-bandwidth trade-off. American Journal of Sociology 117, 90–171 (2011)
[27] Evans, D.: Beyond Influencers: Social Network Properties and Viral Marketing. Psychster Inc. (2009)
[28] Snijders, T., van de Bunt, G., Steglich, C.: Introduction to actor-based models for network dynamics. Social Networks 32, 44–60 (2010)
[29] Bramoull, Y., Djebbari, H., Fortin, B.: Identification of peer effects through social networks. Journal of Econometrics 150, 41–55 (2009)
[30] Sacerdote, B.: Peer effects with random assignment: Results for dartmouth roommates. The Quarterly Journal of Economics 116, 681–704 (2001)

[31] Christakis, N., Fowler, J.: The spread of obesity in a large social network over 32 years. The New England Journal of Medicine 357, 370–379 (2007)

[32] Agarwal, N., Liu, H., Tang, L., Yu, P.S.: Identifying the influential bloggers in a community. In: Proceedings of the International Conference on Web Search and Web Data Mining, pp. 207–218 (2008)

[33] Trusov, M., Bodapati, A., Bucklin, R.: Determining influential users in internet social networks. Journal of Marketing Research 47, 643–658 (2010)

[34] Weng, J., Lim, E.P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 261–270 (2010)

[35] Tang, X., Yang, C.: Identifying influential users in an online healthcare social network. In: Proceedings of the IEEE International Conference on Intelligence and Security Informatics, pp. 43–48 (2010)

[36] Han, B., Srinivasan, A.: Your friends have more friends than you do: identifying influential mobile users through random walks. In: Proceedings of the Thirteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 5–14 (2012)

[37] Ilyas, M.U., Radha, H.: Identifying influential nodes in online social networks using principal component centrality. In: Proceedings of IEEE International Conference on Communications, pp. 1–5 (2011)

[38] Subbian, K., Melville, P.: Supervised rank aggregation for predicting influencers in twitter. In: SocialCom, pp. 661–665 (2011)

[39] Ghosh, R., Lerman, K.: Predicting influential users in online social networks. In: Proceedings of KDD Workshop on Social Network Analysis (2010)

[40] Shetty, J., Adibi, J.: Discovering important nodes through graph entropy the case of enron email database. In: Proceedings of the 3rd International Workshop on Link Discovery, pp. 74–81 (2005)

[41] Zhang, J., Ackerman, M.S., Adamic, L.: Expertise networks in online communities: structure and algorithms. In: Proceedings of the 16th International Conference on World Wide Web, pp. 221–230 (2007)