World Wide Web (2013) 16:621–644
DOI 10.1007/s11280-012-0167-3

Who blogs what: understanding the publishing behavior of bloggers

Kang Zhao · Akhil Kumar

Received: 30 June 2011 / Revised: 10 April 2012 /
Accepted: 2 May 2012 / Published online: 17 May 2012

Abstract Are bloggers' topical coverages related to their contributions, impacts, and publishing styles in the blogosphere? We investigated this question by grouping bloggers on the basis of their topical coverages and comparing their publishing behaviors. From a blog website with more than 370,000 posts, we first identified two types of bloggers: specialists and generalists. Then we studied and compared their respective publishing behaviors in the blogosphere. Our analysis suggested that bloggers with different topical coverages do behave in different ways. Specialists generally make more contributions than generalists. Specialists also tend to publish more on weekdays, during business hours, and on a more regular basis. We also revealed that specialists also have different publishing behaviors, with only a small fraction creating a large "buzz" or producing a voluminous output. As blogs start to gain more business value, an extensive analysis like ours can help various stakeholders in the blogosphere maximize their share of the value chain.

Keywords blogosphere · topical profile · specialist · generalist · publishing behavior · blogger clustering

K. Zhao (✉)
Department of Management Sciences, Tippie College of Business,
The University of Iowa, S210 John Pappajohn Bus Bldg,
Iowa City, IA 52242, USA
e-mail: kangzhao7@gmail.com

A. Kumar
Smeal College of Business,
The Pennsylvania State University,
University Park, PA 16802, USA
e-mail: akhilkumar@psu.edu

1 Introduction

With the emergence of Web 2.0 and social media applications, Internet users are able to create and disseminate online content, as well as interact and collaborate with each other. Examples of Web 2.0 and social media applications include blogs/microblogs, social networking sites, social tagging, and so on [30]. A blog (weblog) is a special type of website that allows its owner(s) to publish their entries in a timely and easy way. Entries in a blog are often organized in the reverse-chronological order. Many blogs also allow readers to leave comments to posts, so that readers can interact with the blogger.

The Blogosphere, the world-wide community of blogs, is becoming more popular and assuming greater importance. According to a study by the Pew Research Center, about 10% of all online adults in the U.S. maintain a personal online journal or blog [25]. BlogPulse, an online archive of blogs, has indexed more than 157 million public blogs, with more than 1 million new posts published every day [9]. According to the Alexa ranking of Web traffic, popular blogging services Blogger.com and WordPress.com rank 5th and 21st respectively among all sites on the Web [4]. Meanwhile, the blogosphere has become an important online media to publish, disseminate, and discuss information and opinions. For example, some blogs (e.g. Politico.com) have had a major impact on government and corporate policies, and have become a must reading for officials [39].

Along with their popularity and importance, the business value of blog websites is also on the rise. In September 2010, AOL acquired the technology-oriented blog website TechCrunch for more than $25 million [22]. Two months later, The Daily Beast, a news reporting and opinion blog website, merged with the Newsweek magazine and formed a 50–50 joint-venture [36]. In February 2011, AOL announced that it would pay $315 million to acquire popular news blog–The Huffington Post [33]. Consequently, the blogosphere has attracted a lot of research interests from multiple disciplines [32].

With a focus on bloggers' topical publishing patterns, this study tries to address the following research question: *Are bloggers' topical coverage in their posts related to their publishing behaviors, such as contributions, impacts, and temporal publishing styles, in the blogosphere?* We believe answers to this research question will not only deepen our understanding of bloggers' behaviors, but also improve the effectiveness of online advertising and inform the design of blogging applications to generate more economic value from them. The remainder of the paper is organized as follows. Section 2 reviews related work. Then we introduce our dataset and perform preliminary analysis in Section 3. The next section describes how we grouped bloggers using the topical distribution of their posts and compared contributions and behaviors of different types of bloggers. The different publishing behaviors of bloggers with the same topical patterns are also illustrated. Section 5 discusses the implications of this research at length in the context of a proposed framework for blogosphere analysis. We conclude the paper by offering directions for future work in Section 6.

2 Related work

Previous research on blogosphere analysis has studied a wide range of topics, from community-level topics, such as the evolution of blogosphere [24], to individual-

level topics, such as reading behaviors of blog users [16]. Despite such variety in blogosphere analysis, the implications of the analysis are mainly in three areas: advertising strategies, market intelligence, and blog system design. Next, we describe these areas and briefly introduce how previous blogosphere analysis helps in each one.

First, blogosphere analysis can improve the effectiveness of online advertising, one of the most important revenue sources for personal, professional, and corporate blogs [40]. A major goal of deploying online advertisements, such as Web page banners, is to get high click-through rates from readers of the page. A popular targeted advertising strategy is contextual advertising: placing advertisements that are related to the topic of a Web page (e.g., show an iPod retailer's advertisement on a Web page about portable music players), as readers of the page are more likely to be interested in the advertisement. Through content analysis of a blog post, one can extract keywords from the post [3, 5] so that the advertising network can learn about the topic of the post and thus match it with related advertisements. In some studies, content analysis of a blogger's post can help to identify what type of products the blogger would like to purchase [28]. In addition, the identification of influential bloggers [2] can help an advertising network to differentiate the rates for placing advertisements on different bloggers' posts.

Second, market intelligence derived from blogosphere analysis can help corporations to design, improve, and market their products [18]. To obtain such intelligence, a corporation can maintain a corporate blog, such as Microsoft's MSDN, and also encourage its employees to contribute to various public blogs. On one hand, a corporate blog can help a company and its employees to interact directly with existing and potential customers and learn about customers' feedback [34, 38]. On the other hand, analyzing public blogs adds more breadth. For example, sentiment analysis of posts related to a product [11] can help a company to evaluate whether this product is well received by customers. Some companies went a step further and used customer reviews or feedback gathered from the blogosphere to design and improve their products accordingly [41]. In addition, the general understanding of how information spreads [20, 26] and the identification of key members in online communities [2, 43] will be helpful for viral marketing campaigns in the virtual world [15].

Last, blogosphere analysis can help a blog site attract more users by improving the information system. Topical clustering can group posts, such that posts with similar content or about similar topics are placed in the same group [10]. Such clustering often takes advantage of keyword co-occurrences or inter-post links [6, 35]. Further research has augmented these approaches with tags that the readers assigned to posts [10, 27]. Such grouping can enable still better post recommendation based on what a blog reader is reading. It could also improve search results when users try to find posts by querying keywords through a blog website's search engine. In addition to topical clustering, revealing the temporal trend of buzz-words in the blogosphere [12, 19] can help readers to keep up with the ever-changing hot topics in the virtual world. Asking readers what topics they are interested in and providing this information to bloggers can help to build an active and interactive blog community [17].

Topical analysis can contribute to all of the three ares. However, two areas in topical analysis need still further investigation. First, most previous research investigated topical patterns for posts, but studied it only little from the perspective of bloggers. Schmidt [37] provides an analytical framework of blogging practices

based on bloggers' procedures and routines, social and virtual relations, and blogging software. The work of [23] clustered bloggers based on personal interests that are listed in their online personal profile. However, in an online setting, a user's online profile can often be incomplete or inaccurate. Also, personal interests in such profiles do not necessarily correspond to a blogger's topical coverage in her/his blog.

Second, previous topical analyses focused mostly on microscopic topics, which are mainly represented by a few buzz words that appear in the posts. While such microscopic topics can help readers find posts about popular products or events, they have limitations as well. On one hand, such buzz words may not reveal what a post is really about. For instance, if a post reviews a car whose stereo system has an iPod dock and happens to contain a link to another post about iPod, then this post could be labeled with the topic "iPod", even though it is actually about a car. On the other hand, those buzz-word-based topics are often ad-hoc and lack breadth. For example, if one reads a post about iPods, she may also be interested in other digital gadgets, such as navigation systems and smartphones. However, focusing only on buzz words may not be able to capture the relationship between buzz words that are semantically related but may not co-occur frequently in the same post. While the research described in [1] groups political bloggers at the ideological level (e.g. liberals versus conservatives) based on the political issues covered in their posts, there is little research that studies bloggers' publishing patterns at a broader topic level (e.g., as sports, technology, etc.) like we did in this research.

3 Dataset

At the outset, we first show some basic information and analysis about the dataset we used. In fact, the analysis on the citation network revealed interesting phenomena that inspired our topical analysis on bloggers. Our research is based on the data from an Italian blog website, which covers a broad range of topics. The dataset contains more than 370,000 posts, published between September 2009 and June 2010, from 2,275 blogs. Although bloggers did not specify the topical category of their blogs, the website extracted the content of their posts and conducted text analysis using proprietary natural language semantic analysis tools to classify the topic of each post into eight macroscopic topic categories: news, cars, culture, entertainment, dining, sports, technology, and others. The classification results were validated by manual inspections of each topical category for a large number of posts. The assignment of such macroscopic topical categories enables us to study the topical publishing patterns of bloggers.

We first constructed a post citation network, where posts represent nodes and edges denote the citation relationship among posts. If post A cites post B, it means that post A contains a hyperlink pointing at B. This network is similar to the network of pages on the World Wide Web as both netowkrs are built upon hyperlinks. However, we call it a citation network as we only include hyperlinks in the body of a post, not those in places like sidebars. Thus the in-degree of a node is the number of other posts that cite this post, while the out-degree is the number of other posts that this post cites. This network is sparse because most posts (about 88% of all the posts in the dataset) do not cite other posts or get cited in this website. Only 43,047 posts from 906 blogs have non-zero in- or out-degrees, and are connected by 50,434

edges in the network. The distributions of in- and out-degrees generally follow Power Laws $P(k) = k^{-r}$ [7] (Figure 1), meaning that most posts cite few posts or get cited by few, while some cite many posts or get cited by many. However, there is no giant component among these non-zero-degree posts (a giant component is a connected sub-network that contains a majority of all the nodes). Instead, these 43,047 nodes are isolated into 9,754 components or sub-networks, which are disconnected from each other. The largest sub-network has only 4,228 nodes (posts). The low density of the citation network also makes it difficult for us to make use of citation information of posts or bloggers in later analysis.

We also examined assortative patterns [29, 42] of the citation network. It was found that citations among posts is topically assortative as a post tends to cite another post with the same topic. We represent the topical assortativity with the cross-topic citation density. The citation density between a pair of topics (X, Y) measures how likely a citation link exists between a post on topic X and another on topic Y, and is defined as:

$$D_{XY} = C_{XY}/P_X P_Y \tag{1}$$

where, C_{XY} is the number of citation links between a post on topic X and another one on topic Y; P_X and P_Y are the total number of posts on topic X and Y, respectively. Take the citation network in Figure 2 as an example. Between three sports posts and two technology posts, a maximum of six links, which point from sports posts to technology posts, could exist. The network only has three citation links. Thus the citation density $D_{\text{sports,technology}} = 3/(3 \times 2) = 0.5$.

We illustrate the topical citation densities with a density map in Figure 3, each cell representing the citation density between two topics. As the "others" category does not really correspond to a specific topic, we only include seven topic categories in the map. The density values along the diagonal range from 8.7×10^{-5} to 66×10^{-5}.

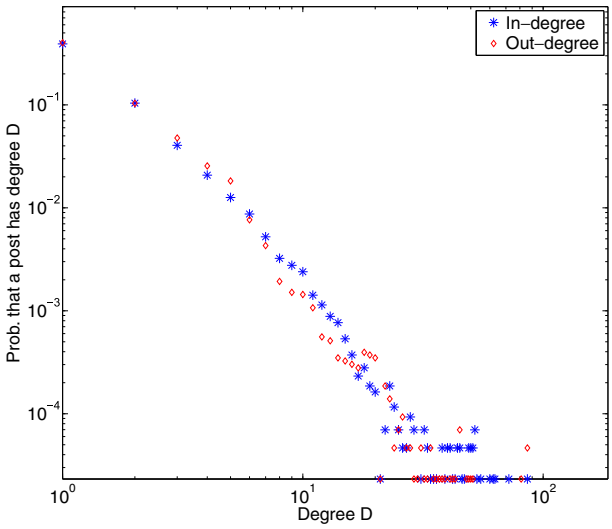

Figure 1 The log-log degree distribution for the post citation network.

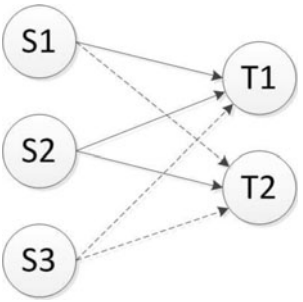

Figure 2 An example citation network between sports posts (denoted with *S*) and technology posts (denoted with *T*). *Solid lines* are existing citations; *dashed lines* represent non-existing but possible citations.

Meanwhile, the density values off the diagonal are much lower, with a maximum of 4.98×10^{-5} and a mean of 5.97×10^{-6}. Posts on culture, for instance, are 30 times more likely to cite other posts within culture than posts on other topics; however, this is less so for news posts, where the same-topic citation is only three times more likely than cross-topic citation. We conjecture that this is because the coverage of news posts is generally broader than of other topics, and so there is less of an "incestuous" tendency among them.

On the basis of the post citation network, we then look at citations from the perspective of bloggers, i.e., how a blogger cites other bloggers' posts. Similar to the sparse post citation network, the blogger citation network, in which nodes denote bloggers and edges represent citation relationship between bloggers, also have low density: it has only 2,038 edges between 740 bloggers. Other bloggers (about 2/3 of all bloggers) have zero degrees in this network, meaning that they do not cite or get cited by others' posts. Interestingly, unlike the post citation network, this network has a giant component that includes 702 out of the 740 non-zero-degree bloggers. Recall that the post citation network has more than 50,000 inter-post citations, almost 24 times larger than the number of inter-blogger citations. Thus we examine the inter-post citations and find the reason for the huge difference: the data reveals a rather narcissistic trend of self-citation. Among all the 50,434 inter-post citations, 86% of the time the bloggers cite their own blog posts! Thus there are much more inter-post citations than inter-blogger citations.

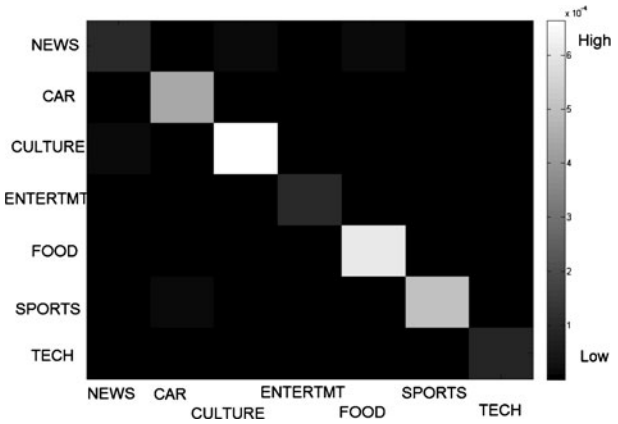

Figure 3 Topical citation densities in the post citation network.

Now we know that a blogger tends to cite her own posts and cross-post citation is often between posts with the same topic. Does that mean a blogger's posts are likely to focus on one specific topic only? Next we try to answer this question through a more in-depth analysis of bloggers' topical coverages and behaviors.

4 Analysis

4.1 Identifying generalists and specialists

To study the relationship between bloggers' topical coverages and their behaviors, we first need to reveal what topic(s) bloggers cover. Thus we used a topic vector T_i to represent the topical profile of blogger i.

$$T_i =< t_{i,1}, t_{i,2}, ..., t_{i,3} >, \text{ where } \sum_{j=1}^{8} t_{i,j} = 1. \quad (2)$$

where $t_{i,j}$ represents the ratio of blogger i's posts on topic j. For example, if a blogger has published a total of ten posts, with two on news and eight on technology, her topic vector will be $< 0.2, 0, 0, 0, 0, 0, 0.8, 0 >$. Figure 4 shows the distribution of bloggers' topical vectors. The existence of many peaks on various topics mean that a great number of bloggers' topic coverages are not evenly distributed. Then can we group bloggers on the basis of which topic(s) their posts are more likely to cover, so that bloggers in the same group publish posts on similar topic(s)? Such a grouping will make it easier to compare the behaviors of bloggers with different topical coverages.

In this research, we tried two approaches to group bloggers on the basis of their topical vectors. The first one is an intuitive threshold-based approach. Basically, a threshold values S ($S \geq 0.5$) is picked. If $t_{i,j} \geq S$, we call blogger i a specialist on topic j. Otherwise, blogger i is considered a generalist who covers a broader range of topics. The second approach is to use clustering algorithms, such as the classic k-means algorithm. This iterative algorithm partitions all bloggers into k clusters so

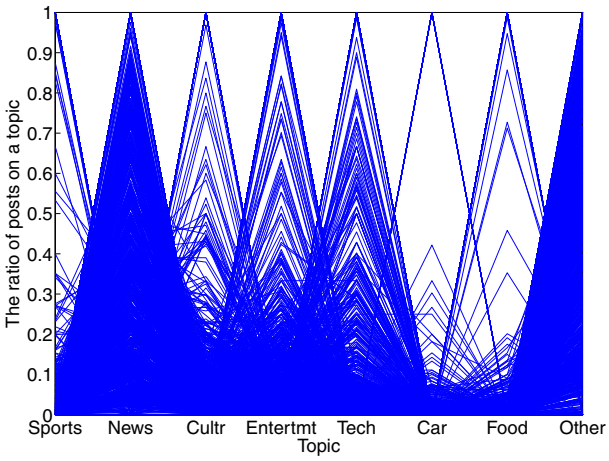

Figure 4 The topical distribution for all bloggers (values of $t_{i,j}$ in blogger i's topical vector are connected by a line. Note that some lines overlap with each other.).

Table 1 The DBIs for the threshold-based grouping with various S values.

S	DBI
0.5	0.1099
0.6	0.1053
0.7	0.1026
0.8	0.1020
0.9	0.0970

that bloggers with similar topical vectors belong to the same cluster. To run k-means, one has to specify the value of k, i.e., how many clusters will be generated.

To find the best grouping of bloggers for this dataset, we tried various S values for the threshold-based approach and different k values (from 2 to 19) for k-means. We evaluated the quality of the grouping outcome with the Davies–Bouldin Index (DBI) [13]. DBI is defined in (3), where $D_{\text{intra}}(C_i)$ is the average distance from all members of cluster C_i to the center of C_i, and $D_{\text{inter}}(C_i, C_j)$ is the distance between the centers of clusters C_i and C_j. In this research, we use Euclidean distance. Briefly speaking, DBI is based on a compactness measure of clusters divided by an inter-cluster distance measure. On one hand, DBI favors smaller clusters because the intra-cluster distance tends to be lower in a smaller cluster. On the other hand, it also penalizes short inter-cluster distances so that partitioning the data into a large number of small clusters that are very close to each other is also discouraged. The solution with the lowest DBI gives a balanced clustering.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{j:i\neq j}\left\{\frac{D_{\text{intra}}(C_i) + D_{\text{intra}}(C_j)}{D_{\text{inter}}(C_i, C_j)}\right\} \tag{3}$$

We summarized the DBIs for both approaches in Tables 1 and 2. The results suggest that k-means with $k = 9$ yields the smallest DBI and thus is used for the clustering of bloggers' topical publishing profiles. While alternative clustering algorithms exist, the k-means converges fast on our dataset and generates reasonable, stable, and compact clusters of bloggers.

Among the nine clusters of bloggers we discovered (see Table 3), seven are topic-specific clusters, as there is a one-to-one mapping between each of the seven topics and a cluster. On average, bloggers in a topic-specific cluster were found to publish more than 90% of their posts on a single topic. For example, one cluster of 278 bloggers focuses heavily on technology, because, an average of 95.4% of their posts are about technology. For sports bloggers in a 147-blogger cluster, the average

Table 2 The DBIs for the k-means clustering with various k values.

k	DBI	k	DBI
2	0.0335	11	0.0518
3	0.0357	12	0.0601
4	0.0400	13	0.0747
5	0.0445	14	0.0812
6	0.0466	15	0.0877
7	0.0276	16	0.0969
8	0.0326	17	0.1066
9	0.0270	18	0.1171
10	0.0354	19	0.1254

Table 3 Topical clusters of bloggers.

Cluster	Number of bloggers	Top topic(s) in blogs (avg. percentage of posts) on the topic(s)
Specialists-1	441	News (93%)
Specialists-2	312	Entertainment (98%)
Specialists-3	278	Technology (95%)
Specialists-4	147	Sports (98%)
Specialists-5	129	Dining (99%)
Specialists-6	158	Culture (98%)
Specialists-7	21	Car (100%)
Generalists-1	423	News (36%), Culture (10%), Technology (9%), Entertainment (12%), Others (28%)
Generalists-2	366	Other (70%), News (10%)

percentage of sports posts is 98%. Similarly, we also find clusters for entertainment, dining, news, cars and culture. Because the 1,486 bloggers (about 65% of all bloggers in the website) in the seven topic-specific clusters publish posts mainly on a single topic, we call them specialists of that topic.

In contrast to the seven topic-specific clusters, the other two clusters do not exhibit such a strong focus on one specific topic. For example, a cluster of 423 bloggers published 36% of their posts in news, 12% on entertainment and 28% on other topics. This means bloggers in the two clusters cover a broader range of topics in their posts than specialists do. Thus, we combine the two clusters and classify the 789 bloggers (about 35% of all bloggers) in the two clusters as generalists.

4.2 Publishing behaviors of generalists and specialists

4.2.1 Contributions to the blogosphere

In the previous subsection, we grouped bloggers into specialists and generalists. Now we investigate how specialists and generalists contribute to the blogosphere. We first need to develop metrics to measure a blogger's contribution. While many factors could reflect the contribution of a blogger, no single one can solely represent such contribution. Thus a blogger's contribution in the blogosphere is often approximated by combining multiple factors, such as the number of posts, the length of posts, the number of citations, etc. [2].

In this paper, we use various metrics to examine the quantity, quality, and temporal patterns of generalists' and specialists' behaviors. It is also worth noting that some bloggers may abuse the impact metric to boost their contribution ranking by publishing spam posts and comments. However, the consideration of this type of behavior is beyond the scope of this research.

The first metric we use is *productivity*. It is based on the total number of posts a blogger publishes in a given period, 10 months in our case. The assumption here is that, in the virtual community of blogosphere, publishing posts is one of the most important and tangible ways for a blogger to contribute. The number of posts one publishes is a surrogate measure of productivity.

Figure 5 compares the distribution of productivity for generalists and specialists. The approximate Power Law curve for specialists ($r \approx 1$) lies above the one for generalists ($r \approx 1.2$), and its slower decay suggests that specialists are generally more

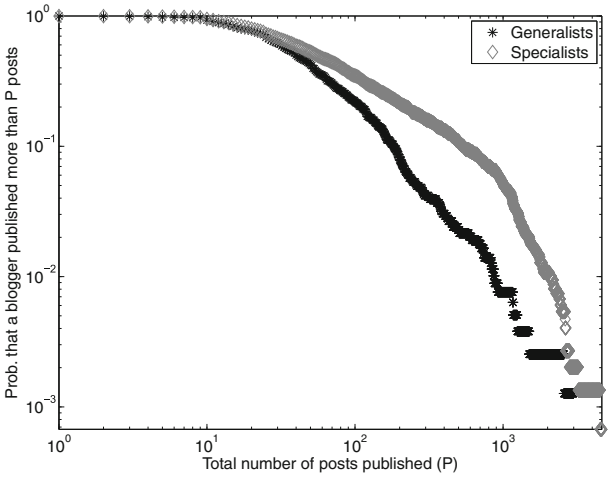

Figure 5 Complementary cumulative distributions for the total number of published posts.

productive than generalists. Our statistical analysis also confirms that, on average, a generalist published 89.8 posts, with a 95% confidence interval of [76.2, 103.4]. By contrast, a specialist had an average of 201.5 posts, with a 95% confidence interval of [180.8, 222.1].

Admittedly, productivity does not reflect the quality of one's posts. Thus we introduce a second metric: buzz-factor (BF), which is a measure of how much "buzz" a post generates. Similar to evaluating scientific researchers, one possible way to measure a blogger's buzz would be to look at a blogger's in-degree in the blogger citation network. As Figure 6 shows, specialists have slightly higher in-degrees than generalists do. While intuitive, this metric has obvious limitations too. As we mentioned in Section 3, the blogger citation network is very sparse, with only 1/4 of all bloggers getting cited by other bloggers. More importantly, this network just reflects a small part of the whole picture of post citations as it does not include citations coming from outside this blog website. For example, if a blogger's post is

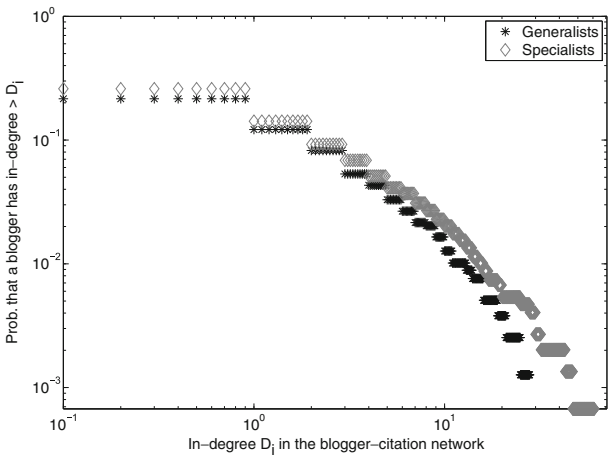

Figure 6 Complementary cumulative in-degree distributions for the blogger citation network.

cited by a journalist's article published on CNN.com, such citation will not show up in the blogger citation network but will certainly help the blogger to get attention from around the world. Although one can hypothesize that a node's in-degree in this small-scale blogger citation network inside the blog website may be correlated with its in-degree in the actual full-scale citation network across the World Wide Web, testing such a hypothesis requires collecting and analyzing all the Web pages of or a carefully selected sample of the World Wide Web.

Thus we approximate a blogger's buzz-factor with the number of comments for her/his post. Compared to citations, comments of a post, from this blog website or not, are much easier to track. As suggested in previous research [2], a post that can attract readership and generate discussion among readers will likely receive many comments. However, the number of comments a post receives follows a near power-law distribution with $r \approx 2.1$ (see Figure 7), meaning that most receive none or few comments, while a select few draw many comments. If we calculate buzz-factor as the average comments per post across all the posts of a blogger has, we might penalize bloggers who published a lot of posts. Therefore, we consider only the Top-N most commented posts (MCPs) of a blogger (e.g. "Top-1", "Top-5", "Top-10"), and average across them to determine the blogger's BF. The BF of blogger i is defined as BF_i in (4) below, where $N_{i,j}$ is the number of comments received by blogger i's j-th most commented post.

$$BF_i = \frac{1}{N} \sum_{j=1}^{N} N_{i,j} \tag{4}$$

Figure 8 compares the buzz-factor metric based on the average number of comments received by a blogger's Top 5 MCPs. The two Power Law curves in this figure suggest that specialists ($r \approx 0.9$) tend to attract more comments than generalists ($r \approx 1.3$). However, the difference is not statistically significant: generalists have a mean of 7.12, with a 95% confidence interval of [5.00, 9.23]; specialists have a higher average of 14.73 posts but a wider 95% confidence interval of [8.58, 20.88].

While productivity and buzz-factor measure a blogger's contribution from two aspects, it will be easier to measure which group has made a higher contribution

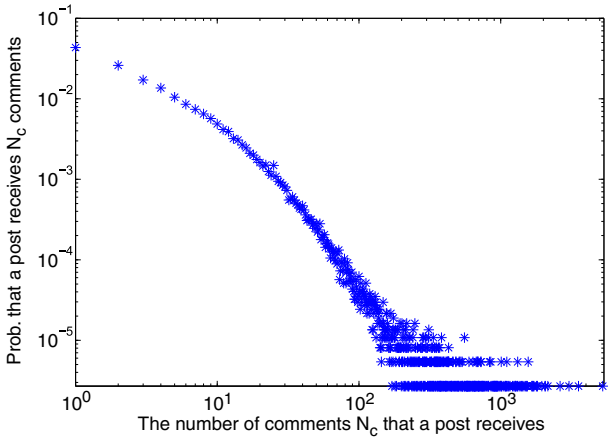

Figure 7 The log-log distribution for the number of comments a post received.

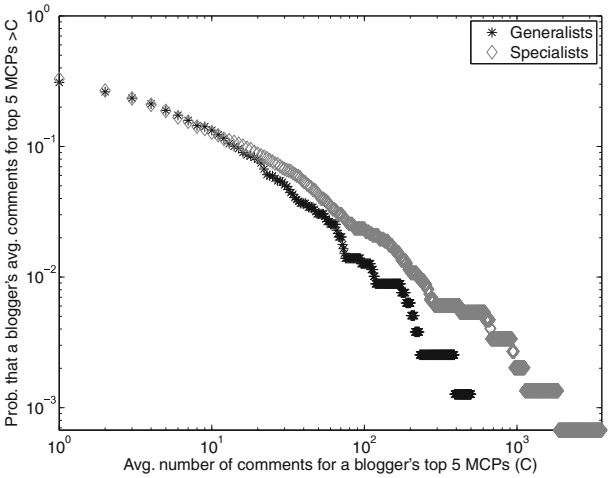

Figure 8 Complementary cumulative distributions for the average number of comments for the top 5 MCPs.

if we can use a metric that combines both. Therefore, we develop a metric called Blogger-index (B-index). This metric is inspired by the well-known H-index which measures both the productivity and impact of a scholar [21]. A blogger is said to have a blogger-index of b, if b of her posts have attracted at least b comments each. In Table 5, we compare the B-index of generalists and specialists (also see Figure 9), and find that the latter outperform the former with a higher B-index: the mean B-index for specialists is 3.26 (95% confidence interval [2.72, 3.80]) and the mean B-index for generalists is 2.29 (95% confidence interval [1.93, 2.64]).

The comparison of contribution reveals that specialists by far make larger contributions to the blogosphere than generalists. One might hypothesize that the difference in contribution reflects the difference between *professional* and *amateur bloggers*. Compared with generalists, specialists tend to be more professional bloggers who treat blogging as an occupation. Thus they contribute more posts. They also

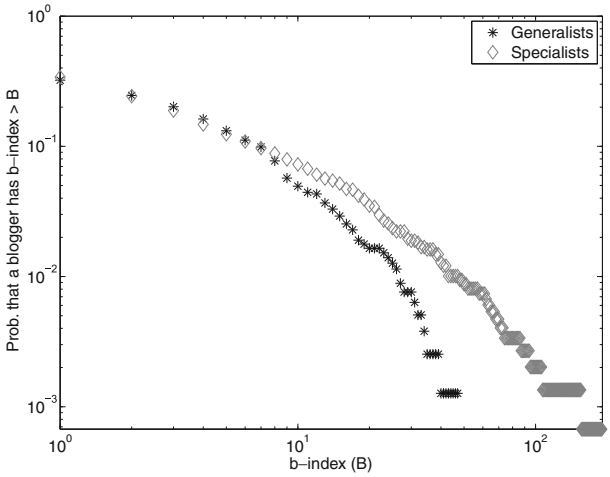

Figure 9 Complementary cumulative distributions for bloggers' b-index.

Table 4 Summary of the metrics used to evaluate bloggers' publishing behaviors.

Metric	Description
Productivity (PD)	The total number of posts published by a blogger.
Buzz-factor (BF)	The average number of comments for the Top-N most commented posts.
B-index (BI)	A blogger has a B-index of b, if b of her/his posts has attracted at least b comments each.
Percentage of weekday posts (PW)	The percentage of a blogger's posts that are published on weekdays.
Daily publishing fluctuation (DPF)	The normalized average daily variation in the number of a blogger's posts during the blogger's temporal span of activity.

bring more expertise and dedication to their topic, making their posts more appealing to readers. Generalist blogs, on the other hand, tend to come from more amateur bloggers who publish on more than one topic of general interest to them. Hence, their posts may tend to lack depth and thus draw less attention. Next, we will try to gain more insights into this hypothesis through a temporal pattern analysis. For reading convenience, we summarize all the metrics we use in this research Table 4.

4.2.2 Temporal publishing patterns

The purpose of this analysis is to better understand whether bloggers' topical patterns are related to their temporal publishing behaviors. To achieve this goal, we compared the temporal publishing patterns of specialists and generalists.

The first temporal pattern examines whether the posts of a blogger tend to appear more on weekdays or weekends. We measure the percentage of posts published by bloggers on weekdays (Monday through Friday) as a fraction of the total number of posts on all days of the week. The cumulative distributions for the percentage are shown in Figure 10. Recall that if one's temporal publishing pattern is random, then a blogger would publish 5/7 (approximately 71.43%) of her posts on the five weekdays, and the rest 2/7 on the weekend. We find that specialists are more active from Monday through Friday, with an average of 73.88% of their posts being

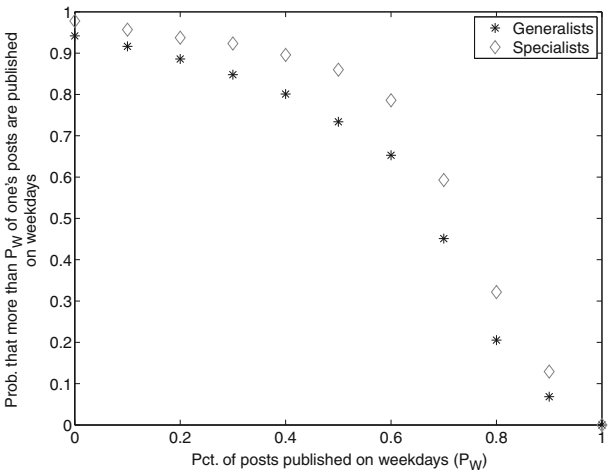

Figure 10 Complementary cumulative distributions for the percentage of posts published on weekdays.

Table 5 The average values on 5 metrics for generalists and specialists.

Metric	Generalists' average (95% CI)	Specialists' average (95% CI)
Productivity	89.80 (76.15–103.44)	201.51 (180.81–222.21)
Buzz-factor	7.12 (5.00–9.23)	14.74 (8.58–20.89)
B-index	2.29 (1.93–2.64)	3.26 (2.72–3.80)
Percentage of weekday posts	65.1% (63.3–67.0%)	73.9% (72.8–75.0%)
Daily publishing fluctuation	1.34 (1.31–1.38)	1.21 (1.18–1.23)

published on weekdays. This is significantly higher than the generalists' average of 65.12% (confidence intervals are summarized in Table 5). We also looked the time of publishing to see whether generalists and specialists tend to publish posts in different times within a day. As Figure 11 suggested, specialists are more productive during business hours in the morning (8:00–12:00), while generalists work on more posts during off-hours (20:00–00:00). The differences between the two groups are not significant in other time slots. These temporal patterns strengthen our hypothesis that specialists are more likely to be professional bloggers who devote more time to blogging during business hours on working days, while generalists tend to include more amateur bloggers who get more time to blog when they are off work in evenings and during weekends.

Another interesting question to pose is "How regularly or sporadically does a blogger publish?" Does a blogger constantly publish a number of posts almost every day? Or does she blog sporadically (for instance, no posts for three days, followed by ten posts in one day, etc.)? One intuitive way to measure temporal regularity is to use auto-correlation. Higher auto-correlation coefficients mean higher periodicity in time series. Thus we chose to examine the number of posts published per day by bloggers during a 20-week period, which is also the most active period in this blog website, and calculated auto-correlation coefficients for time lags varying from 1 day to 14 days. As shown in Figure 12, the coefficients for specialists are significantly

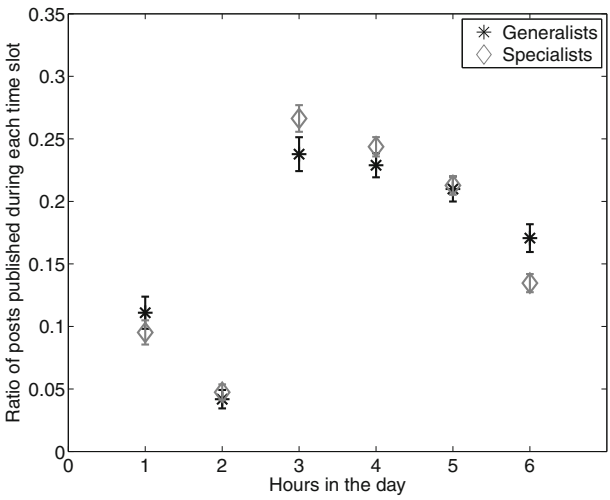

Figure 11 The ratio of posts published in 6 time slots of a day (vertical bars indicate 95% CIs).

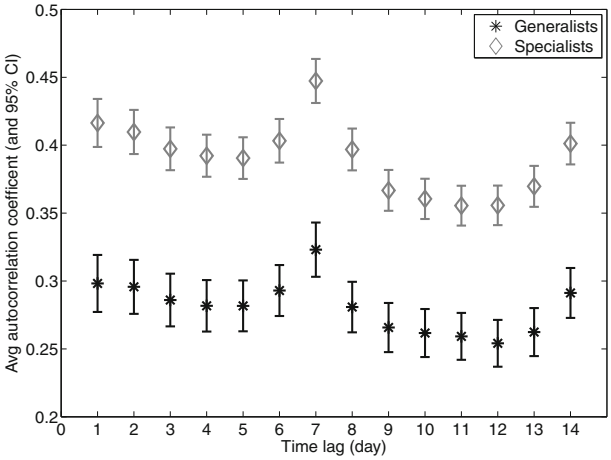

Figure 12 Compare auto-correlation coefficients on the number of posts published per day (*vertical bars* indicated 95% CIs).

higher than those for generalists for all the time lags we examined. This suggests that the temporal publishing patterns for specialists are more periodic or regular than those for generalists. Another interesting observation is that the coefficients reach maximum values on the 7-day time lag. In other words, bloggers' temporal publishing behaviors seem to follow a weekly routine.

However, the auto-correlation coefficient for a blogger's number of posts each day is a function of time-lags. Can we find a more straightforward approach to describe the temporal regularity or sporadicity? Is it possible to use one numeric value to denote how the number of posts a blogger publishes varies from day to day? Take a 7-day period as an example. Bloggers A and B published 21 posts each. Assume that the numbers of posts blogger A published each day are 0, 6, 1, 6, 2, 6, 0, while blogger B published exactly three posts per day. As shown in Figure 13, A's temporal publishing sequence is more sporadic or irregular than B's. What metric could reflect the difference in temporal sporadicity? One may wish to use variance, which measures how the numbers spread out from the mean. Variance will suffice in this scenario with bloggers A and B, where A has a variance of 8.33, and B of zero. However, variance is not suitable in other scenarios. For instance, say blogger C also published 21 posts but in the order 0, 0, 1, 2, 6, 6, 6 over a 7-day period. Although C's temporal sequence does not fluctuate as much as A's (see Figure 13), the variance in their daily posts is the same. This is because variance ignores the temporal ordering of data.

Hence, we propose a new metric, daily publishing fluctuation (DPF), to measure the temporal sporadicity. The DPF_i of blogger i is defined in (5) below, where $P_{i,j}$ is the number of posts that blogger i published on day j; S_i and E_i are the starting and ending days of i's blogging activity. Briefly speaking, a blogger's DPF is the average daily variations in productivity between successive days during her temporal interval of activity $[S_i, E_i]$ normalized by the average number of posts the blogger publishes per day in this interval. A higher DPF value indicates more fluctuation or sporadicity in daily publishing behaviors. We use the average daily variations because using the

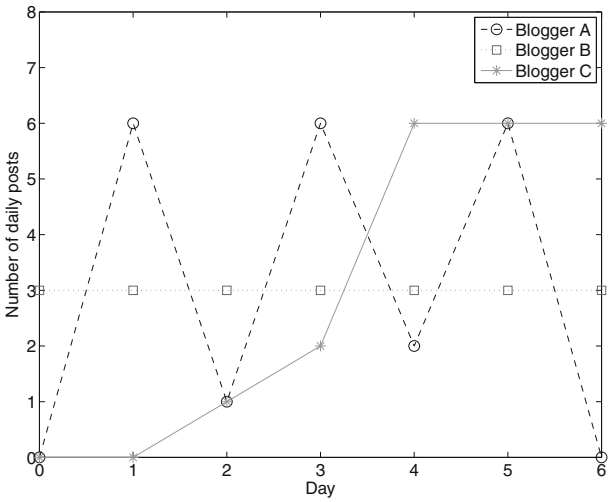

Figure 13 A sample temporal publishing sequence for three bloggers.

sum of daily variation may unfairly lead to higher DPF for bloggers who have been active for a longer time.

$$DPF_i = \frac{\frac{1}{E_i - S_i} \sum_{j=S_i}^{E_i - 1} |P_{i,j} - P_{i,j+1}|}{\frac{1}{E_i - S_i + 1} \sum_{j=S_i}^{E_i} P_{i,j}} \tag{5}$$

Figure 14 compares the DPF of generalists and specialists and suggests that generalists have a higher DPF. Statistical analysis also shows that generalists' average DPF of 1.34 is significantly higher than specialists' 1.21 (see confidence intervals in Table 5). Similar to the outcome from our auto-correlation analysis, DPF also pointed out that a specialist generally publishes more regularly than a generalist blogger. These patterns support our hypothesis that specialists tend to include more professional bloggers who blog regularly and during business hours to make a living; while generalists generalists' publishing behavior is more sporadic and publish more

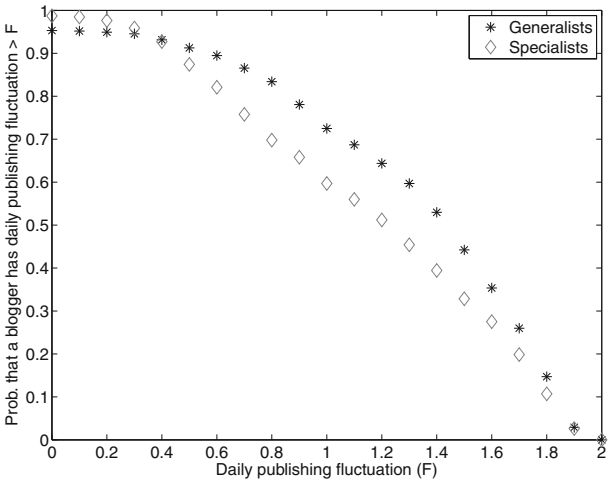

Figure 14 Complementary cumulative distributions for daily publishing fluctuation.

during weekends and off-hours. Although this analysis is by no means a substitute for a rigorous empirical study, it provides additional insights on the publishing behaviors of specialists and generalists, and helps to denominate them.

4.3 A drill-down into specialists

Our dataset has more specialists than generalists. As discussed above, bloggers' topical publishing patterns are related to their publishing behaviors. Specialists, whose topical coverages are more focused, make greater contributions than generalists do when they are evaluated with our metrics. Specialists also blog more regularly than generalists do. A next logical step is to understand whether specialists with similar topical patterns behave differently. Are there different types or subgroups of specialists? As an example of a specialists group, we decided to focus on the 441 "news" specialists, who constitute the largest topic-specific cluster in our topical clustering of bloggers.

For each news blogger, we consider four of the metrics that were mentioned earlier (summarized in Table 4): each blogger i is represented by a 4-tuple contribution vector $B_i =< PD_i, BF_i, PW_i, DPF_i >$, where

PD_i Blogger i's productivity;
BF_i The buzz-factor (measured by the average number of comments on the blogger's top 5 MCPs);
PW_i the percentage of weekday posts out of all weekly posts;
DPF_i The daily post fluctuation.

Figure 15 shows scatter plots and correlation coefficients (and p values) between the six possible pairs among the four metrics (all metrics have been normalized

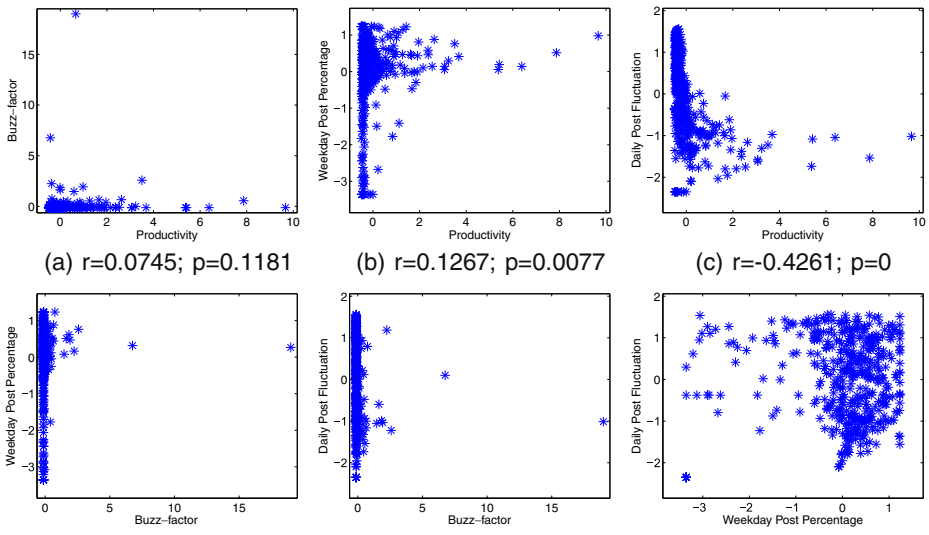

Figure 15 Scatter plots and correlation coefficients for the four metrics used to group news specialists (all metrics have been normalized).

to their z-scores). The results suggest that news bloggers do blog differently from others as the correlation is not strong among the four metrics. For example, a highly-productive blogger does not necessarily attract many comments; and a blogger who can generate high buzz-factor may not publish regularly. Also, productivity and buzz-factor have more outliers and thus wider range than weekday posts percentage and daily post fluctuation do.

To identify subgroups within news bloggers, we again use a clustering algorithm. While k-means algorithm works well for topical clustering, the heterogeneous ranges on different dimensions and the existence of outliers deteriorate its performance on this contribution vector dataset. For example, the algorithm often creates a one-member cluster for some of the outliers with very high productivity or buzz-factors. Thus we used Gaussian Mixture Models (GMM) to cluster news bloggers. The basic idea is to find a mixture of k (k is the number of clusters) Gaussian models that fit news bloggers' contribution vectors. Instead of determining cluster membership using only the center of a cluster as in k-means, GMM takes a probabilistic approach and tries to find a group of Gaussian models that fit the data. Parameters of Gaussian models were estimated using the Expectation-Maximization (EM) algorithm [14]. The outcome of GMM clustering depends on two factors: the value of k (the number of clusters) and the randomly chosen initial setting. One can evaluate the quality of the clustering results by calculating the expected likelihood of the data given the statistical model using cross-validation.

We applied GMM-EM on the dataset with different k values (from 2 to 20). For each k, we ran the algorithm 50 times. Each run started with a random initial setting, and thus may lead to a different result. Then we evaluated each result using 10-fold cross-validation for 30 repetitions, and calculated the average log-likelihood of the testing data. Finally, among all k and initial setting combinations, we chose the one that gets the highest log-likelihood as the clustering configuration on this dataset.

Table 6 lists the 14 clusters of "news" bloggers identified by our GMM clustering and the mean of their corresponding Gaussian model. The results echo the outcome

Table 6 Clusters of "news" specialists.

		Mean of the Gaussian model for the cluster			
Cluster	Number of bloggers	Productivity	Buzz-factor	Weekday posts pct.	Daily post fluctuation
1	19	0.5159	−0.1068	0.7892	−0.877
2	28	−0.3913	−0.0795	0.3098	0.5739
3	6	0.4623	0.0149	−1.4916	−0.8732
4	15	−0.4157	−0.1289	−0.4078	0.7542
5	10	−0.4196	−0.1006	−2.7031	1.0611
6	50	0.1411	−0.0971	0.1171	−1.0108
7	5	−0.3173	0.356	0.2227	0.4234
8	56	−0.4226	−0.1185	−0.4261	1.1909
9	4	0.1943	6.222	0.4033	−0.6907
10	7	5.5479	−0.0189	0.3517	−1.3009
11	16	−0.369	−0.128	−3.3086	−1.7608
12	133	−0.2259	−0.1259	0.4298	−0.0093
13	36	1.2631	0.2711	0.2369	−1.1321
14	56	−0.4007	−0.1256	0.7126	0.9512

of our correlation analysis that "news" specialists have different publishing behaviors. For example, the 7 "busy beavers" in Cluster 10 published many posts, and yet generated below-average buzz. By contrast, in Cluster 9, there are 4 "star" bloggers who did not have to blog much but gained high buzz-factors. The results of a previous study on the activity and influence of bloggers in a technology blog website [2] were similar, but we did not find a category that combines high productivity and high buzz-factor in our dataset. In addition, Cluster 11 represents 16 "regular leisure blogger", who blogged regularly and tended to publish on weekends, but had below-average productivity and buzz-factor. The 66 bloggers in Clusters 5 and 8 are "occasional contributors" who published sporadically. Of course, the majority are "average Joes" (such as those in Clusters 1, 2, 4, 7, 12, and 14), whose values on all metrics are within one standard deviation of the mean values.

5 Discussion

By analyzing of bloggers' topical publishing patterns and comparing the behaviors of generalists and specialists, our research found that bloggers with different topical patterns and coverages have different contributions, impacts, and temporal publishing styles. Although some results of this study may not apply to all blog websites, such as the ratio of generalists and specialists, and the number of bloggers in each topical group, the study adds further to the general understanding and analysis of the blogosphere and how value is realized in it.

The major contribution of this study can be better appreciated when we look at the ecosystem of blogosphere. To illustrate various stakeholders and their relationships, we developed a framework for blogosphere analysis (in Figure 16). The framework describes a network among five major stakeholders in the blogosphere ecosystem: bloggers, publishers (e.g., blog websites), readers, corporations, and advertising networks. Bloggers create content that is published by blog publishers. Readers can subscribe to posts about certain topics or those written by certain bloggers. Publishers can also recommend posts of likely interest to readers or notify them through RSS feeds or emails. Based on their own marketing goals, corporations hire online advertising networks, such as Google AdSense, which will pay a blog website to place advertisements on the website's pages. A blog website may offer bloggers monetary or non-monetary rewards for attracting high readership or achieving high advertisement click-through rates.

Our analysis of bloggers' topical profiles and publishing behaviors has implications for all the five stakeholders in the blogosphere ecosystem by improving advertising strategies and post recommendation. Blog readers can receive better pointers to content that is of interest and relevant to them. This will increase the frequency of their visits to the blog. Bloggers will be able to attract more relevant readership to their blogs and this will generate more comments and in turn make their posts more attractive to advertisers. Consequently, advertising networks will be willing to pay more for placing their advertisements with the knowledge that they will attract more suitable viewers who are likely to click on the advertisements more often. Moreover, corporations will get superior market intelligence in addition to better value for their advertisement dollars.

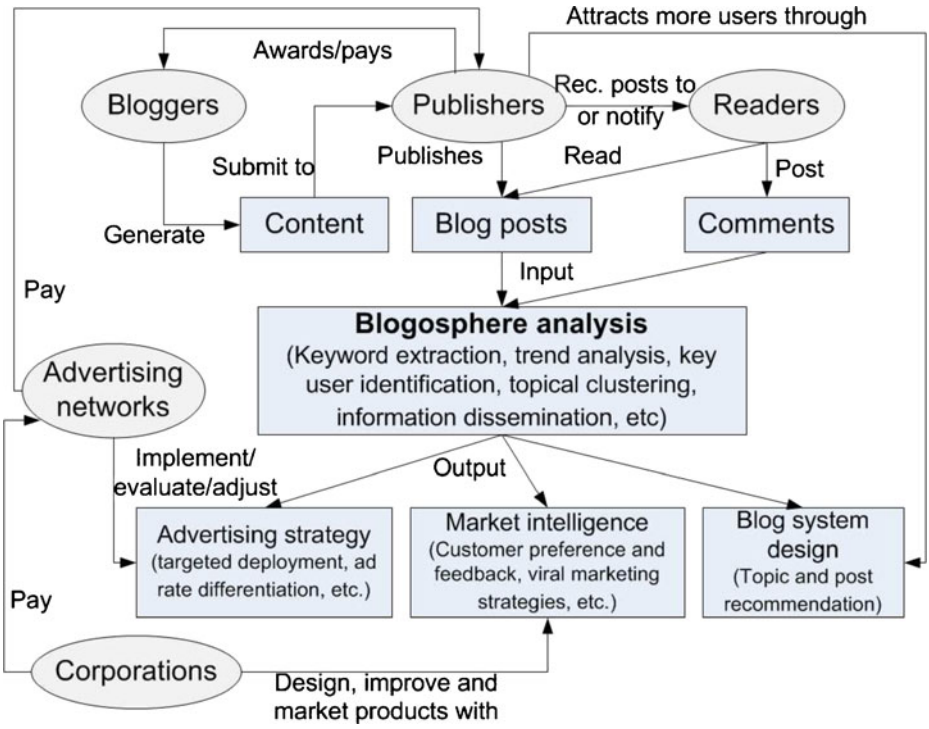

Figure 16 A framework for blogosphere analysis.

First, our analysis may help the ecosystem to generate greater economic value because a successful advertising strategy can directly help corporations, advertising networks and publishers to increase their income. More revenue from advertising will motivate the publisher to retain existing high-impact bloggers, as well as to recruit new ones, through rewards and better blog system design. This will in turn lead to a more active blogosphere and more high-quality posts to blog readers.

Specifically, we believe incorporating a blogger's topical profile can help to improve contextual advertising. As mentioned earlier, current deployment of online contextual advertisements is mainly based on keywords which may not accurately reflect the content of a webpage and sometimes lead to irrelevant or entirely inappropriate advertisements [31]. While many text mining techniques are available, real-time analysis of all content in a webpage is still challenging within the short time period of loading the Web page. Discovered using more in-depth text analysis and clustering algorithms offline, bloggers' topical profiles can help an advertising network to deploy contextual advertisements in real time and in a more targeted way. The example of a car review post that mentions an iPod dock in the car stereo system is helpful. If the blogger's topical profile suggests that she is a car specialist, then the advertising network should focus on advertisements that are related to cars (such as from car manufacturers or dealerships), instead of iPod-related ones, even though keywords from both topics are present.

In addition to contextual advertising, a blogger's topical profile and publishing behavior may help an advertising network to differentiate the advertisement rates that they charge. Empirical data from the Italian blog website suggests that the click-through rate varies across posts and topics. For instance, the click-through rate is generally higher for technology posts than for news posts. Thus an advertising network can charge more for advertisements deployed on the posts of technology specialists. Similarly, the charge may also be raised on posts from specialists with higher Blogger-index as these bloggers have a track record of attracting more eyeballs, or bloggers who contribute on a more regular basis because their blogs may attract more regular visitors and even subscribers. In fact, the advertising rates can be determined by a formula based on the various metrics related to our study. For instance, the cost per thousand views of an advertisement (CPM) can be represented as

$$CPM_{i,j} = f(BI_i, DPF_i, TA_j, BTP_i, NC_i...) \tag{6}$$

Equation (6) basically shows that, among many others, five factors we studied in this paper could be used to find the CPM of an advertisement deployed on blogger i's post j: the blogger-index of the blogger (BI_i), the blogger's temporal blogging sporadicness (DPF_i), the topical category of the advertised product/service (TA_j), the topical profile of the blogger (BTP_i), and the centrality (e.g., degree centrality, PageRank, etc) of a blogger in the blogger citation network (NC_i). While there are other factors that could determine CPM (e.g., keyword matches) and finding the exact way to calculate CPM needs further investigations, we show in (7) an example instantiation, where $\alpha, \beta, \delta, \gamma, \lambda > 0$:

$$CPM_{i,j} = \alpha * BI_i + \beta / DPF_i + \delta * BTP_i + \gamma * match(TA_j, BTP_i) + \lambda * NC_i \tag{7}$$

In this example, the CPM is greater for a high-impact blogger with higher BI_i and NC_i, a regular contributor with low DPF_i, a specialist on certain topics (such as technology), and when the topical category of the advertisement matches the topical specialty of the blogger.

The second important implication of a study like ours is that it can help to develop a better post recommender system within the blog. For blog readers, a good recommender should assist them to find related posts easily. More importantly, if readers can find more posts that match their interests, they will spend more time on the blog site. Clearly, this helps to bring more page visits to the site and in turn generates more revenue. How will our analysis help to improve a post recommender? For one, the bloggers' topical publishing profiles and different publishing behaviors may help the publisher to decide what posts to highlight on the front page of the website by integrating popular specialist posts about different topics to create variety. In addition, as noted earlier, bloggers' topical publishing profiles and different contributions can also help to improve the breadth of post recommendations to individual readers. For example, when one is reading a post about "iPods," the recommender may want to prefer iPod-related posts from specialists with high blogger-indexes, as well as pick non-iPod-related posts from other high-impact technology specialists to complement current keyword-based posts recommendation.

6 Conclusions and future work

In this research, we studied publishing behaviors of bloggers by focusing on their topical publishing patterns. Using a large dataset from a blog website, we were able to partition bloggers into groups, and distinguish between general and special interest bloggers based on the topical coverages of their posts. To reveal how bloggers' topical publishing patterns relate to their publishing behaviors, we compared specialists and generalists in terms of contribution, impacts, and temporal publishing styles. The outcome of our research suggests that specialists made more contributions than generalists, perhaps from their deeper subject matter expertise. Specialists also tend to blog much more on weekdays and during business hours, and publish posts more regularly. Further analysis of a group of specialists revealed that their publishing behavior styles are also different: some are very productive but generate average buzz ("busy beavers"); some are able to create a lot of buzz from only a few posts ("stars"); some blog regularly but do not have high productivity or generate high buzz ("regular leisure bloggers").

There are still many unanswered questions for the future. With additional data, we plan to study the impact of bloggers using networks based on various relationships such as trackbacks and blogrolls [16]. This will allow us to devise more comprehensive metrics for bloggers' contributions and behaviors. We do recognize that one limitation of this research is that the dataset from only one blog website was used. In the future, we expect to apply our approach to other blog websites, so that we can make comparisons and generalize our findings still further. For blog websites that do not have topic labels readily available, we plan to use topic models, such as Latent Dirichlet Allocation [8], to discover topics from posts automatically. Such an approach will also allow us to generate the topical distribution for each blogger, so that we can apply similar methods to find bloggers with similar topical coverage and analyze their publishing behavior. Another possible topic for future research is to conduct fuzzy clustering on bloggers and allow a blogger to belong to a cluster with a given probabilities. Such "soft" clustering of bloggers will help us better understand the behaviors and topical coverages of bloggers, especially generalists, whose contributions span several topics. Another fruitful area of research would be to perform an economic analysis to determine a pricing model for online advertising based on the popularity of a post and characteristics of a blogger.

Acknowledgements The authors would like to thank Mr. Massimiliano Spaziani from Telecom Italia for providing the dataset. Kang Zhao also thanks Santa Fe Institute for the opportunity to start this research during the 2010 Complex Systems Summer School.

References

1. Adamic, L.A., Glance, N.: The political blogosphere and the 2004 U.S. election: divided they blog. In: Proceedings of the 3rd International Workshop on Link Discovery, pp. 36–43. ACM (2005)
2. 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. ACM (2008)
3. Agarwal, D., Gabrilovich, E., Hall, R., Josifovski, V., Khanna, R.: Translating relevance scores to probabilities for contextual advertising. In: Proceeding of the 18th ACM Conference on Information and Knowledge Management, pp. 1899–1902. ACM (2009)

4. Alexa.com: Top sites. http://www.alexa.com/topsites (2011). Accessed 18 March 2011
5. Anagnostopoulos, A., Broder, A.Z., Gabrilovich, E., Josifovski, V., Riedel, L.: Just-in-time contextual advertising. In: Proceedings of the 16th ACM Conference on Conference on Information and Knowledge Management, pp. 331–340. ACM (2007)
6. Bansal, N., Chiang, F., Koudas, N., Tompa, F.W.: Seeking stable clusters in the blogosphere. In: Proceedings of the 33rd International Conference on Very Large Databases, pp. 806–817. VLDB Endowment (2007)
7. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
8. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
9. BlogPulse.com: Blogpulse stats. www.blogpulse.com (2011). Accessed 9 March 2011
10. Brooks, C.H., Montanez, N.: Improved annotation of the blogosphere via autotagging and hierarchical clustering. In: Proceedings of the 15th International Conference on World Wide Web, pp. 625–632. ACM (2006)
11. Chesley, P., Vincent, B., Xu, L., Srihari, R.: Using verbs and adjectives to automatically classify blog sentiment. In: AAAI symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW), pp. 27–29 (2006)
12. Chi, Y., Tseng, B.L., Tatemura, J.: Eigen-trend: trend analysis in the blogosphere based on singular value decompositions. In: Proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM), pp. 68–77. ACM (2006)
13. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-1**(2), 224–227 (1979)
14. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the em algorithm. J. R. Stat. Soc. Ser. B (Methodological) **39**(1), 1–38 (1977)
15. Domingos, P.: Mining social networks for viral marketing. IEEE Intell. Syst. **20**(1), 80–82 (2005)
16. Furukawa, T., Ishizuka, M., Matsuo, Y., Ohmukai, I., Uchiyama, K.: Analyzing reading behavior by blog mining. In: Proceedings of the 22nd National Conference on Artificial Intelligence, vol. 2, pp. 1353–1358. AAAI Press (2007)
17. Geyer, W., Dugan, C.: Inspired by the audience: a topic suggestion system for blog writers and readers. In: Proceedings of the 2010 ACM conference on Computer Supported Cooperative Work (CSCW), pp. 237–240. ACM (2010)
18. Glance, N., Hurst, M., Nigam, K., Siegler, M., Stockton, R., Tomokiyo, T.: Deriving marketing intelligence from online discussion. In: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, pp. 419–428. ACM (2005)
19. Glance, N., Hurst, M., Tomokiyo, T.: Blogpulse: automated trend discovery for weblogs. In: WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics (2004)
20. Gruhl, D., Guha, R., Liben-Nowell, D., Tomkins, A.: Information diffusion through blogspace. In: Proceedings of the 13th International Conference on World Wide Web (WWW), pp. 491–501. ACM (2004)
21. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. **102**(46), 16,569–16,572 (2005)
22. Kopytoff, V.G.: Aol to acquire techcrunch to bolster its news coverage. http://www.nytimes.com/2010/09/29/technology/29aol.html (2010). Accessed 1 May 2011
23. Kumar, R., Novak, J., Raghavan, P., Tomkins, A.: Structure and evolution of blogspace. Commun. ACM **47**(12), 35–39 (2004). 1035162
24. Kumar, R., Novak, J., Raghavan, P., Tomkins, A.: On the bursty evolution of blogspace. World Wide Web **8**(2), 159–178 (2005)
25. Lenhart, A., Purcell, K., Smith, A., Zickuhr, K.: Social media and mobile internet use among teens and young adults. Tech. rep., Pew Research Center (2010)
26. Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., Hurst, M.: Patterns of cascading behavior in large blog graphs. In: SIAM International Conference on Data Mining, pp. 551–556 (2007)
27. Li, X., Guo, L., Zhao, Y.E.: Tag-based social interest discovery. In: Proceeding of the 17th International Conference on World Wide Web, pp. 675–684. ACM (2008)
28. Mishne, G., Rijke, M.d.: 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 (WWW), pp. 925–926. ACM, 1135947 (2006)
29. Newman, M.E.J.: Mixing patterns in networks. Phys. Rev. E **67**(2), 13 (2003)

30. Oreilly, T.: What is Web 2.0: Design patterns and business models for the next generation of software. Communications Strategies **65**(1st quarter 2007), 17–31 (2007)
31. Pak, A., Chung, C.W.: A Wikipedia matching approach to contextual advertising. World Wide Web **13**(3), 251–274 (2010)
32. Parameswaran, M., Whinston, A.: Research issues in social computing. J. Assoc. Inf. Syst. **8**(6), 336–350 (2007)
33. Peters, J.W., Kopytoff, V.G.: Betting on news, aol is buying the Huffington post. http://www.nytimes.com/2011/02/07/business/media/07aol.html (2011). Accessed 1 May 2011
34. Porter, L.V., Sweetser Trammell, K.D., Chung, D., Kim, E.: Blog power: examining the effects of practitioner blog use on power in public relations. Publ. Relat. Rev. **33**(1), 92–95 (2007)
35. Qamra, A., Tseng, B., Chang, E.Y.: Mining blog stories using community-based and temporal clustering. In: Proceedings of the 15th ACM International Conference on Information and Knowledge Management, pp. 58–67. ACM (2006)
36. Reuters.com: Newsweek and daily beast agree merger. http://www.reuters.com/article/2010/11/12/us-newsweek-dailybeast-idUSTRE6AB0JI20101112 (2010). Accessed 1 May 2011
37. Schmidt, J.: Blogging practices: an analytical framework. J. Comput-Mediat. Comm. **12**(4), 1409–1427 (2007)
38. Scoble, R., Israel, S.: Naked Conversations: How Blogs are Changing the Way Businesses Talk with Customers. John Wiley and Sons Ltd (2006)
39. Wallsten, K.: Many sources, one message: political blog links to online videos during the 2008 campaign. J. Polit. Market. **10**(1), 88–114 (2011)
40. White, D.: State of the Blogosphere 2008. Tech. rep., Technorati Inc. (2009)
41. Wright, J.: Blog Marketing: The Revolutionary New Way to Increase Sales, Build Your Brand, and Get Exceptional Results. McGraw-Hill (2005)
42. Zhao, K., Ngamassi, L.M., Yen, J., Maitland, C., Tapia, A.: Assortativity patterns in multi-dimensional inter-organizational networks: a case study of the humanitarian relief sector. In: S.K. Chai, J.J. Salerno, P.L. Mabry (eds.) Advances in Social Computing—Proceedings of the 2010 International Conference on Social Computing, Behavioral Modeling, and Prediction (SBP10). LNCS, vol. 6007/2010, pp. 265–272. Springer (2010)
43. Zhao, K., Qiu, B., Caragea, C., Wu, D., Mitra, P., Yen, J., Greer, G.E., Portier, K.: Identifying leaders in an online cancer survivor community. In: Proceedings of the 21st Annual Workshop on Information Technologies and Systems (WITS'11), pp. 115–120 (2011)