# Investigating biomedical research literature in the blogosphere: a case study of diabetes and glycated hemoglobin (HbA1c)*†‡

**Anatoliy Gruzd, PhD; Fiona A. Black, PhD; Thi Ngoc Yen Le, MEC; Kathleen Amos, MLIS, AHIP**

See end of article for authors' affiliations.

DOI: http://dx.doi.org/10.3163/1536-5050.100.1.007

**Objective:** The research investigated the relationship between biomedical literature and blogosphere discussions about diabetes in order to explore the role of Web 2.0 technologies in disseminating health information. Are blogs that cite biomedical literature perceived as more trustworthy in the blogosphere, as measured by their popularity and interconnections with other blogs?

**Methods:** Web mining, social network analysis, and content analysis were used to analyze a large sample of blogs to determine how often biomedical literature is referenced in blogs on diabetes and how these blogs interconnect with others in the health blogosphere.

**Results:** Approximately 10% of the 3,005 blogs analyzed cite at least 1 article from the dataset of 2,246 articles. The most influential blogs, as measured by in-links, are written by diabetes patients and tend not to cite biomedical literature. In general, blogs that do not cite biomedical literature tend not to link to blogs that do.

**Conclusions:** There is a large communication gap between health professional and personal diabetes blogs. Personal blogs do not tend to link to blogs by health professionals. Diabetes patients may be turning to the blogosphere for reasons other than authoritative information. They may be seeking emotional support and exchange of personal stories.

## INTRODUCTION

The widespread use of the Internet by lay individuals who require health information is well known [1]. It has been estimated that almost 80% of Internet users have searched for health information online [2]. The development of Web 2.0 applications, especially those relating to social networking, provides additional ways for individuals to access health information. Patients, along with their families and friends, increasingly seek and share health information in a multitude of public and semipublic online venues. Among Web 2.0 applications, blogs have emerged as a powerful medium through which computer-literate individuals can express themselves. As more people begin to blog, many more people are also starting to rely on information provided by the blogosphere. Patients, caregivers, and doctors are increasingly turning to the blogosphere to search for information, discuss treatment options, and share their stories and experiences [2, 3]. Blogs and other Web 2.0 technologies enable Internet users to actively contribute to the abundance and diversity of online content. Technorati estimates that, between 2002 and 2008, more than 133 million blogs were launched and almost 1 million blog posts were published every day [4].

### Highlights

- Networks of blogs relating to diabetes care tend to be in distinct realms: those that cite authoritative biomedical literature and those that do not cite such literature.
- Bloggers who are not health professionals may be unaware of the availability of biomedical literature recrafted for a lay audience, such as via MedlinePlus.

### Implications

- Biomedical literature is important for health decisions. Discussions in the blogosphere that may influence decisions by patients may not be drawing on the literature.
- Diabetes patients and their caregivers' blogs connect with each other in the blogosphere. This may be more for emotional than for informational support.
- Health communicators and information professionals might work more closely together to aid diabetes patients by enabling blogging, some of which might help interpret recent biomedical literature or link directly to articles in MedlinePlus.

Recent work by Neal and McKenzie indicates that librarians' criteria for evaluating sources should move toward more user-centered criteria. Sources traditionally considered authoritative may not necessarily be what the lay public desires [5]. The blogosphere represents a large source of wide-ranging opinions and attitudes. As with the Internet generally, not all information published in the blogosphere is credible or reliable [6–8]. Patients could be misled by erroneous or outdated information. Bloggers interested in

* This study was supported by a Research Development Fund grant from Dalhousie University awarded to Anatoliy Gruzd in 2010.

† An earlier stage of this study was presented as: Black FA, Amos K, Gruzd A. Disruptive technologies in health information landscapes: the case of diabetes and HbA1c. The American Society for Information Science & Technology (ASIS&T) Annual Conference; Vancouver, BC, Canada; 9–11 Nov 2009.

‡ This research draws on data provided by the University Research Program for Google Search, a service provided by Google to promote a greater common understanding of the web.

health information can increase the quality and reliability of their blogs by following and citing biomedical literature. The increasing availability of scientific literature through open and public access methods, or at least free availability of abstracts, suggests that it might now be easier for laypeople to access such literature. Davis and Walters emphasize the need to explore differences between communities that produce and communities that consume scholarly literature [9]. However, little research has been done to date regarding whether, and how, biomedical literature is cited in the blogosphere. This study explores this area, using the example of diabetes disease management, through the following research questions:

- How many blogs cite biomedical literature and why? Who are the authors of those blogs?
- What are the most influential blogs on diabetes, as measured by the number of links to them? Are blogs that cite biomedical literature more influential in the blogosphere than those that do not?
- To what extent do blogs that do not cite biomedical literature link to blogs that do?

To address the research questions, the study focuses specifically on blogs and biomedical literature related to diabetes. It is estimated that the total number of people worldwide with diabetes will rise from 171 million in 2000 to 366 million by 2030 [10]. Diabetes mellitus, in particular type II diabetes, is the subject of increasingly intense clinical research, particularly related to ongoing disease management strategies, as well as increases in activity by the public and health promotion research communities. This study examines the dissemination of biomedical knowledge related to the relevance and importance of a particular measurement in the management of diabetes, glycated hemoglobin (HbA1c) [11]. Professional opinion on HbA1c testing, as reflected in published biomedical literature indexed in the MEDLINE database, is compared with lay perspectives, measured using quantitative and qualitative analyses of the blogosphere, to assess if and how biomedical information is discussed by the public in the blogosphere.

## BACKGROUND

Health and medical blogs have attracted research attention, in particular from scholars of social media and health communication. A strong focus of prior research has been the characteristics of blog readers. A relatively early study by Kim and Chung explored characteristics of cancer blog users—including demographics, use, and perceptions of blogs—and concluded that blog readers were attracted to blogs for their ability to share emotional support and personal stories rather than for their medical knowledge [12]. A similar conclusion stems from research by Walther et al., who examined several sociotechnical attributes in a cancer-related online discussion system, including interactivity, homophily, social distance, and anonymity [13]. They concluded that a strong benefit of online support groups was ''the way they bring out common experience, or homophily, among participants.'' This finding tied in closely with that of a ground-breaking multidimensional study of diabetes patients, which examined social support, self-management, and quality of life. Glasgow et al. surveyed participants using the Diabetes Network (D-Net), a website designed to support patients with diabetes. Part of their work involved the development of an ''internet-related diabetes support scale,'' which embraced the areas of emotional support, advice, and information [14]. The results of their research suggested an ongoing need to take a ''multidimensional approach to measuring social support and computer-mediated health outcomes.'' Social network analysis (SNA) of health blogs is one such multidimensional approach that may contribute new information relevant to studies of diabetes health outcomes.

In addition to studies of health-related blog readership, scholars have studied blog authorship. Buis and Carpenter explored blog content in relation to credentials of individual bloggers and blog host sites, based on the premise that blogs have the potential to be crucial tools for health consumers, not only due to the links they provide to external websites, but also due to their capacity for supporting discourse through the comment function [15]. The authors pointed out the need for investigations of who reads health blogs and how often they do so. The current project contributes to the first part of this gap in knowledge by exploring links between blogs. Finally, Buis and Carpenter found that the majority of their sampled blogs, due mainly to their chosen population, were written by those with medical or related academic credentials [15]. Similarly, when Miller and Pole analyzed the content and characteristics of influential bloggers on health care, they concluded that health bloggers in their sample were highly educated and that half of them worked in the health field and blogged from a professional perspective [16]. The current study contributes new knowledge due in part to its broader population of blogs, thus capturing health blogs authored by nonspecialists.

A variety of methods can be employed to research themes in blogs and blog use. Kovic, Lulic, and Brumini used an online survey of a select group of medical bloggers to explore the characteristics of the authors and their blogs [17]. Their study investigated motivations for blogging as well as the type of attention bloggers gained from other bloggers. Their analysis was necessarily based on survey respondents' subjective input, with that method's advantages and limitations. Objective analysis, coupled with visualization, is now feasible based on automated SNA. SNA facilitates investigations of human interactions or information flows to explain social phenomena [18] and is often used in studies on the blogosphere. For instance, Pikas used SNA to understand the structure of the science blogosphere and study how scientists influence and help each other in

searching for new knowledge [19]. Furthermore, Kovic, Lulic, and Brumini suggested that future research should include blog content analysis [17]. In contrast to SNA, which usually facilitates a macro view, content analysis is especially useful at the micro level, examining the data from an individual blogger's perspective. Content analysis has been widely used to identify and study characteristics of virtual communities amongst bloggers [20, 21]. The current project makes use of both methods.

Despite the growing body of literature on blogs in general and health blogs in particular, key questions remain: Is there evidence that bloggers are aware of relevant biomedical literature, and to what extent are journal articles discussed in the blogosphere?

## METHODS

The current study used structured searches and web crawling, content analysis, and SNA to investigate bloggers' awareness of relevant academic research and to determine if influential diabetes-related blogs cite biomedical literature. Web crawling techniques were used to automatically collect information about blogs. Next, manual content analysis was applied to a sample of blogs to explore the nature of literature citations in blog posts and types of bloggers who cite them. Finally, SNA was conducted to analyze the network properties of blogs on diabetes.

### Step 1: structured searches and web crawling

Structured searches were conducted in PubMed to construct the first of three primary datasets: published articles about diabetes and HbA1c (D1). Web crawling was used to construct the second and third datasets: blogs that mentioned at least one article from D1 (D2) and all blogs included in D2 plus all blogs on diabetes that did not mention any articles from D1 (D3).

To ensure a focus on recent research and scholarship, two searches were performed in PubMed to retrieve articles on diabetes and HbA1c published in 2008 or 2009 for D1. Both searches used the same strategy, ''Diabetes Mellitus''[Mesh] AND ''Hemoglobin A, Glycosylated''[Mesh], with one search limited to items published in 2008 and the other to items published in 2009. Google Blog Search was selected to locate material for D2, because it has a massive collection of blogs and advanced search functions. It was used to find blogs that mentioned at least one article from D1. The main Google Blog Search query consisted of the article's title and last name of the first author. This search query was satisfactory in most instances. However, when article titles were short or general in meaning, Google Blog Search could return noisy results. For example, a search in blogs for the article, ''Tight Blood Glucose Control,'' by Frank, returned numerous blog posts that included these words, but many did not reference the article in question. In such situations, the results were manually checked and incorrect items were removed. Also, not all of the websites returned by Google Blog Search were blogs. Google considers any website that publishes a really simply syndication (RSS) feed to be a blog, but other websites, such as news sites and journals, also publish RSS feeds. Non-blog websites were automatically excluded using information in uniform resource locators (URLs) and hypertext markup language (HTML) meta tags. The result of this filtering procedure was cross-checked manually to ensure data quality.

Google Blog Search has proved to be a valuable tool for this line of research but does have limitations, which may have resulted in missing a small number of blogs that might have cited articles. Searches sometimes returned inaccurate results if search keywords contained special characters. This could happen when an author's name contained a non-English character, such as in ''Müller.'' The number of affected articles was estimated as less than 5%. Despite these limitations, a large sample of blogs that cited recently published articles on diabetes and HbA1c was collected.

D3 consisted of the blogs in D2 plus other blogs on diabetes that did not cite articles from D1. To find blogs not captured in D2, two additional blog search engines were used: Icerocket and Technorati. Google Blog Search was found to be inappropriate for this task due to a technical limitation that limits how many items could be retrieved automatically. (This was not an issue while collecting D2, because search queries were more specific and usually returned results in the single digits.) As the functionality of the additional search engines differs, the researchers employed two different strategies to identify blogs. Using Icerocket, an automated script retrieved English-language blog posts with titles containing the word ''diabetes'' published between January 1, 2008, and July 1, 2010 (date of data collection). As Technorati does not have a date/time filter, all of the blogs listed in Technorati's category on diabetes were retrieved. Blogs from D2 and those identified by Icerocket and Technorati were merged, with duplicates removed, to create D3.

### Step 2: content analysis: blogs that cite biomedical literature

Once the three datasets were collected and cleaned, manual content analysis was used to discover why blogs cite biomedical literature. The results of content analysis also helped to iteratively improve the web crawling procedure in step 1.

This stage of the study explored in what context biomedical literature tends to be mentioned in blogs, as part of a general discussion or as a simple announcement of a new study, and whether citation of biomedical literature might indicate a type of blogger. To address these questions, 3 independent reviewers conducted manual content analysis of the first 100 English-language blog posts from D2 that were identified as having cited at least 1 of the articles published in 2009 from D1. The sample of 100 was deemed to be a manageable size for human judges to evaluate manually and, at the same time, large

enough to start observing some general trends among blogs and bloggers who cite biomedical literature.

Out of 100 posts, 4 were manually removed as they were not from blogs, having been erroneously included by an initial version of the web crawler in step 1. The remaining 96 blog posts were used for the analysis. Blog posts were separated into 2 broad categories based on article citation patterns: ''announcements'' and ''analyses.'' Announcement posts provided information about an article, without commenting on its significance. Examples included the posting of an abstract, a bibliography, a journal's table of contents, or a news release about an article. Analysis posts went beyond a simple posting of bibliographic information to offer some discussion of the article, such as reviewing the article or using it as a reference in the post. Blog posts that could not be placed in either category remained unclassified. Cohen's kappa was used to measure the level of agreement between the reviewers.

An additional in-depth review of the blog posts by one of the reviewers explored the types of bloggers who cite biomedical literature. Blog authors were categorized as being health professionals, journalists, librarians, or laypeople. When the reviewer could not make determinations about authors, the blogs remained uncategorized. Data for this analysis were gathered directly from each blog, usually from an ''About Me'' page.

### Step 3: social network analysis (SNA)

In the third step, SNA was used to analyze the network of diabetes blogs (D3) to better understand how health information is disseminated in the blogosphere. A network representation of links between blogs was built before applying SNA. The Google search engine was used to determine how blogs in D3 are linked to each other. Because Google does not directly support finding a link between two blogs, the researchers designed an automated script to search each blog in D3 for all of the blog posts that linked to it. Only the links between blogs in D3 were kept. Self-references (when a blog links to itself) were removed, as were blogs that did not link to any other blogs in D3 and smaller isolated groups of two to five blogs that were not connected to the main group of blogs. These ''isolated'' blogs did not participate in direct information exchange in this network; thus, they were not germane to the network analysis.

ORA SNA software [22] was used to visualize and analyze D3 from a network perspective. This method relies on a network representation, where nodes represent individuals or other entities such as websites and lines between nodes represent connections, or ''ties,'' between entities. SNA in the current study included two analytical foci: influence, as measured by in-degree centrality, and homophily. The analysis of influential blogs on diabetes examined general network patterns and identified influential blogs to discover factors that might be affecting information exchange among blogs on diabetes. The analysis began with a network visualization and an examination of its general structure. Then, an SNA measure called ''reciprocity'' was used to determine the level of awareness among bloggers in this network. Reciprocity takes into account the directionality of each link and counts the proportion of mutual links in the network. Low values of reciprocity indicate the prevalence of unidirectional ties between nodes and usually suggest a generally low sense of community among members [23]. Finally, ''in-degree centrality'' was used to find the most influential blogs in this network. In-degree centrality counts the number of incoming links for each blog. In addition to identifying and describing the most influential blogs, this analysis addressed the question of whether blogs that cite biomedical literature were more influential in the network, based on in-degree centrality. To test this, the researchers divided the blogs into two groups: blogs that cite articles from D1 (group 1) and blogs that do not (group 2). A *t*-test for network data available in UCINET [24], an SNA package, was performed to compare and test the significance of the difference between the normalized means of in-degree centrality for the two groups of blogs.

The analysis of homophily in the blogosphere examined whether blogs citing articles on the same topic (HbA1c) were more likely to link to each other than to other blogs. To test this, the blogs were again divided into the two groups described above. Krackhardt and Stern's external-internal index (E-I index) [25] was calculated using UCINET to estimate to what extent blogs tend to link to blogs within their own group versus to blogs from the other group. The formula for how the E-I index is measured is:

$$\text{E-I index} = \frac{\#\text{of external links} - \#\text{of internal links}}{\#\text{of all links}}$$

Values for the E-I index range from $-1$ to 1, where $-1$ indicates that all links are within the group and 1 indicates that all links are with blogs from the opposite group.

## RESULTS

### Step 1: structured searches and web crawling

Searches for articles on diabetes and HbA1c were conducted in PubMed in February 2010, returning 2,246 articles published in either 2008 (1,188 articles) or 2009 (1,058 articles). These articles made up the D1 dataset. A Google Blog Search (conducted in July 2010) found 1,450 websites that mentioned at least 1 article from D1. After non-blogs were removed, 558 actual blogs remained. Of these, 308 were unique and formed the D2 dataset. Some blogs contained multiple posts citing articles from D1, leading to the smaller number of unique blogs. The third dataset (D3) included all of the blogs in D2 plus 1,662 blogs that mentioned diabetes from Icerocket and 1,057 blogs from Technorati. In total, after removing a few

duplicates, D3 consisted of 3,005 unique blogs on diabetes, 308 of which (D2) cited at least 1 article from D1.

### Step 2: content analysis: blogs that cite biomedical literature

Eleven percent of the articles in D1 were directly mentioned in the blogosphere. To investigate who cited them and why, the researchers classified 96 blog posts into 2 broad categories based on article citation patterns: ''announcements'' and ''analyses.'' Inter-rater agreement among reviewers was high, with 2 or more reviewers agreeing on the categorization in 97% (93) of 96 posts. The 3 posts that did not garner agreement were not used in the subsequent analysis. Also removed were 5 posts in which the article citation appeared in a comment to the blog post and 9 posts incorrectly identified as citing articles. Of the remaining 79 posts, 44% (35) were announcements and 57% (44) were analyses. For these 79 posts, there was substantial agreement amongst all 3 reviewers (Cohen's kappa ranged from 0.77 to 0.89 in the pair-wise comparison between the reviewers).

The majority of posts in this sample appeared in blogs written by health professionals, who authored 52% (41) of the 79 posts, while laypeople authored only 10% (8). Librarians were responsible for 11% (9) of the posts and journalists for 3% (2). Health professionals were the most frequent authors for both types of posts but were responsible for proportionately more analyses than announcements. In announcement posts, health professionals authored 34% (12) of the 35 posts, librarians authored 17% (6), laypeople 6% (2), and journalists 6% (2). In analytical posts, health professionals were responsible for 66% (29) of the 44 posts, laypeople wrote 14% (6), and librarians 7% (3). None of the analytical posts were written by journalists.

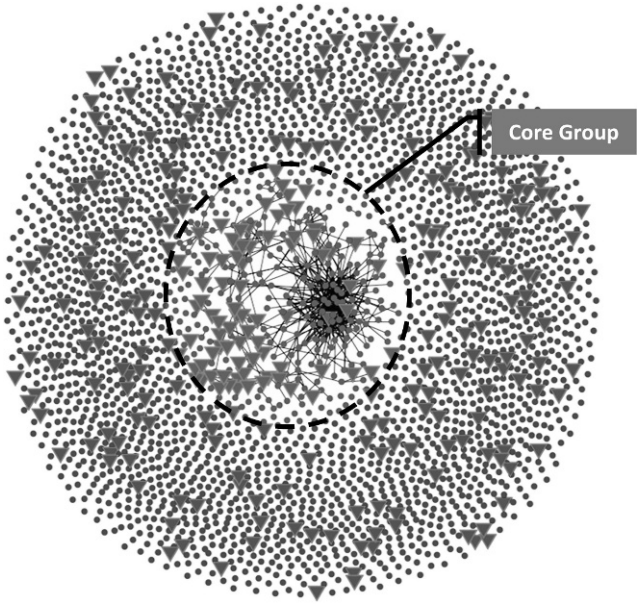
### Step 3: SNA

In building a network representation of D3, the researchers found 20,955 links between blogs in D3 and other blogs in the blogosphere. Only the 607 links connecting any 2 blogs from D3 were kept for further analysis.

**Influential blogs on diabetes.** Figure 1 shows a network view of the full D3 dataset of 3,005 blogs. Each dot or node represents a blog. The triangular nodes represent the 308 blogs that cited at least 1 article in D1, and the circular nodes represent the 2,697 blogs that did not cite any. A line connecting 2 nodes represents a link between 2 blogs. The 238 blogs with links to other blogs are shown in the center of the network. This *core group* is surrounded by numerous isolated nodes and small groups of 2–5 blogs that do not link to other blogs.

The core group contains approximately 1% (563) of all possible links within the network (Figure 2). Only 10% of the links are reciprocal. Most of the links

**Figure 1**
Network of blogs on diabetes (n=3,005)

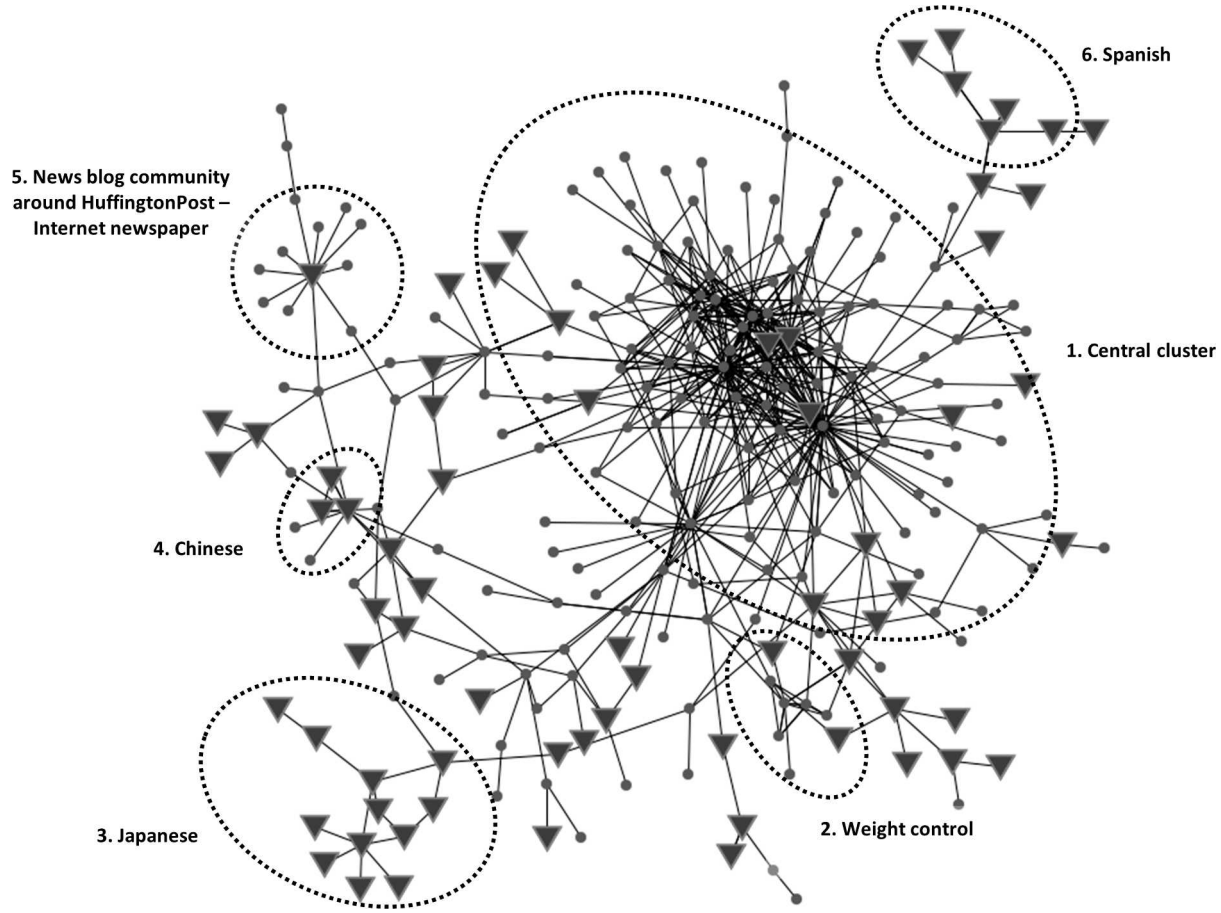



Triangular nodes represent blogs with citations to biomedical diabetes literature; circular nodes represent blogs that did not cite the literature. A line connecting two nodes represents a link between two blogs

within the core group are concentrated within the boundaries of different clusters. The largest cluster of nodes (marked as #1) contains some of the most highly connected blogs on diabetes from the English-speaking blogosphere. This cluster contains a large number of circular nodes, blogs that did not cite biomedical literature from D1. There are also smaller clusters on the periphery of this core network, and, upon closer examination, it appears that many of these blogs are from other countries. For example, the cluster marked as #3 contains Japanese blogs, blogs in cluster #4 are Chinese, and blogs in cluster #6 are from Spain. Most of these blogs from non-English countries cited English-language articles from D1, indicating that English biomedical literature has an international reach. Finally, some clusters of blogs share similar interests, such as cluster #2, a group on weight control and healthy eating, and cluster #5, a community of blogs around the Huffington Post, an Internet newspaper and content aggregating blog.

After reviewing the top 12 rankings of blogs based on in-degree centrality (Table 1), it appears that some of the most influential blogs in the network are personal blogs written by diabetes patients and advocates including Six Until Me, Diabetes Mine, and The Butter Compartment. Only 3 sites on this list (ranked 7 and 10) cited biomedical literature from D1. This suggests that blogs that cite biomedical literature are not necessarily more central or influential in the network. A *t*-test performed to check this for all of the blogs in the core group found no statistically significant difference in in-degree centrality between

**Figure 2**
Analysis of core group of blogs on diabetes (n=238)




Triangular nodes represent blogs with citations to biomedical diabetes literature; circular nodes represent blogs that did not cite the literature. A line connecting two nodes represents a link between two blogs

blogs that do and do not cite biomedical literature (difference=0.002, $P$=0.065). There is a tendency for blogs that do not cite biomedical literature to be more popular (though this finding is not statistically significant).

**Homophily in the blogosphere.** Approximately 29% (70) of the 238 blogs in the core group cited at least 1 article from the D1 dataset. This ratio is about 3 times larger than the same ratio for the whole dataset of 3,005 blogs, where only about 10% (308) of blogs cited biomedical literature from D1. This may indicate that blogs citing articles on the same topics are more likely to link to each other than to other blogs. The result of an E-I index calculation to test this is presented in Table 2. The E-I index is −0.55, which is statistically significant, being about 3 times lower than would be expected by chance alone. This indicates that there are more internal and fewer external links than would be expected by chance alone, confirming that blogs that do not cite biomedical literature do not tend to link to blogs that do cite biomedical literature as a way to connect to credible information on diabetes.

## DISCUSSION

Based on a large sample of 3,005 blogs on diabetes and 2,246 articles about diabetes and HbA1c, it was found that only 10% of the blogs cited about 11% of the articles. This suggests that bloggers interested in diabetes may have a relatively low interest in or awareness of recent literature that might be important to their readers. This low interest and awareness may not be surprising given that biomedical literature is not typically aimed at imparting complex information to a lay audience [26]. Nevertheless, a public dataset from the Public Library of Science (PLoS) [27] was used to check that this result was not due to the narrowness of the topic of HbA1c as a comparison. PLoS tracks how often each of its articles is referenced in the blogosphere. Of 10,565 articles published between 2007 and 2009, 12% received at least 1 link in the blogosphere. This supports the observation that only a small portion of published articles is cited in the blogosphere. This finding may reflect the fact that bloggers and blog readers are not using blogs to discuss or look for health information, but for other purposes such as sharing personal stories and

**Table 1**
Top 12 rankings of blogs based on in-degree centrality

| Rank | Site | In-degree centrality (normalized) | # of incoming links | Blog authors in D3 (information gleaned from each blog) |
|---|---|---|---|---|
| 1 | sixuntilme.com | 0.135 | 32 | Diabetes patient, freelance writer, and social media consultant |
| 2 | diabetesmine.com | 0.127 | 30 | Founder/editor is a diabetes patient, freelance writer, and consultant in health and diabetes industries |
| 3 | thebuttercompartment.com | 0.114 | 27 | Diabetes patient and art therapist |
| 4* | diabeteslinks.blogspot.com | 0.101 | 24 | Layperson and author of books on low-carbohydrate diet |
| 5 | thecornerboothcc.blogspot.com | 0.084 | 20 | Diabetes patient-advocate and journalist, blogging for personal rather than professional reasons |
| 6 | diabetes24-7.com | 0.076 | 18 | Diabetes patient, develops programs for physicians and patients with chronic illnesses, occasionally contributes health and lifestyle items to newspaper |
| 7† | **diabetesdaily.com** | 0.072 | 17 | Community/support network for diabetes patients founded by a patient and her husband; blogger is a diabetes patient |
| 8 | talesofmy30s.wordpress.com | 0.068 | 16 | Diabetes patient, freelance writer |
| 9 | countrygirldiabetic.blogspot.com | 0.059 | 14 | Diabetes patient |
| 10 | jnpedersen2.blogspot.com | 0.042 | 10 | Diabetes patient |
| 10† | **huffingtonpost.com** | 0.042 | 10 | News and content aggregating blog; one blogger is an author of books on healthy living and diabetes advocate; another blogger holds an MD and is a medical editor of the *Huffington Post* |
| 10† | **blog.sina.com.cn** | 0.042 | 10 | Blog service of the largest infotainment portal in China; one blogger is a medical student; another blog is maintained by the Cardiovascular Chinese group |
| 11 | 25unitstogo.wordpress.com | 0.038 | 9 | Diabetes patient |
| 12 | lemonlemonade.wordpress.com | 0.034 | 8 | Patient-advocate, works in social media for a health care public relations agency |
| 12 | threeyearsfree.blogspot.com | 0.034 | 8 | Nurse and diabetes advocate, blogging for personal rather than professional reasons |
| 12 | diabetesaliciousness.blogspot.com | 0.034 | 8 | Diabetes patient-advocate, consultant, motivational speaker |
| 12 | d-mom.com | 0.034 | 8 | Community/support network for parents of children with diabetes; founder and principal author is a diabetes advocate |

* As of August 1, 2011, this blog is no longer available. The blogger remains active on another blog called The Livin' La Vida Low-Carb Show.
† A blog that cited at least one article from D1.
Sites in bold cite biomedical literature from D1.

emotional support, as reported by Kim and Chung [12]. This is reinforced by the fact that only about 6% (168) of the located diabetes blogs (D3) linked to MedlinePlus, a resource specifically created to provide health information in nonspecialist language [28].

Based on manual content analysis of 96 sample blog posts with citations to biomedical literature, blogs that cite biomedical literature are mostly authored by health professionals or librarians and more likely to be analytical than announcement-related in nature. This confirms the initial expectation that laypeople such as diabetes patients do not tend to cite recent biomedical literature in their blogs. There are a number of possible reasons for this. Bloggers who are not health professionals may not know about or have access to these publications. Patient bloggers or their readers may not have the medical expertise to interpret biomedical literature. Additionally, biomedical literature frequently focuses on new developments in the field that may not yet have direct practical implications or be readily accessible to patients. As a result, laypeople may be less interested in discussing this work among themselves. Regardless of possible reasons, this discovery, coupled with the observation from SNA that blogs that do not cite biomedical literature are unlikely to link to blogs that do, suggests that laypeople are somewhat isolated from information about recent developments in the medical field. This is a good example of the social phenomenon ''homophily,'' summarized in the well-known proverb ''birds of a feather flock together.'' This is important as homophily has been shown to limit people's access to information and resources [29].

Furthermore, a low level of reciprocity in links between blogs in this network suggests a generally low sense of community and lack of explicit awareness among bloggers in the network. This is unfortunate especially because personal blogs written by

**Table 2**
External-internal (E-I) index for the blogs in the core group

| | Observed | Average (based on 5,000 random permutations) | Standard deviation | *P*<=observed |
|---|---|---|---|---|
| % internal links | 77% | 58% | 4 | 1 |
| % external links | 23% | 42% | 4 | 0.00 |
| E-I index | **−0.55** | −0.17 | 0.08 | 0.00 |

Bold indicates statistical significance.

laypeople tend to be influential among blogs on diabetes; they are the blogs most often linked to from other blogs. In addition, patients looking for discussions about recent authoritative literature on diabetes may benefit from blogs that cite biomedical literature, especially blogs by librarians. Because many librarians are not health professionals by training, the language they use on their blogs to discuss biomedical literature may be more accessible to laypeople.

These findings point to a possible but potentially dangerous scenario where a patient could be misled by erroneous or outdated information found on popular blogs. This possibility is of concern given that a recent Pew Internet report indicates that 41% of e-patients seek ''just-in-time someone-like-me'' information through personal commentaries in blogs and related sites [30]. To combat and overcome some of these potential risks, the authors believe that greater effort should be made to promote blogs by health professionals and especially those maintained by librarians that contain useful information for diabetes patients. These blogs need to be made more visible and approachable in the blogosphere. Successful examples of well-connected blogs by patients that do cite recent biomedical literature include a community of blogs on ASweetLife, an online magazine on diabetes run by a couple with type 1 diabetes, or Scott's Web Log, a blog by a writer and diabetes patient who believes that ''[p]atients need to read the medical and scientific literature themselves before drawing conclusions.'' From these success stories and from the network perspective, blogs that follow and discuss recent research in the field would be wise to connect to and participate in discussions on highly connected personal or community blogs such as Six Until Me, Diabetes Mine, and Diabetes Daily in order to expose their content to a greater audience. Such exposure could be achieved, for example, by linking to and commenting on popular blogs in this community or by inviting diabetes patients or health organizations to contribute blog posts. By doing so, a blog is likely to appear closer to the community core and become more visible and findable to patients looking for information on diagnosis, treatment, and disease management options.

## CONCLUSION

This study indicates that there is a large communication gap between health professional blogs and personal blogs on diabetes and that active diabetes bloggers do not generally cite recent biomedical literature in their posts. Blogs that do not cite the literature (mostly personal blogs) do not tend to link to blogs that cite the literature (mostly blogs by health professionals). This may be because health professionals' blogs are not well known outside health professional circles. Furthermore, diabetes patients who turn to the blogosphere may be looking more for emotional support and exchange of personal stories than for information about recent research. While further research is needed to pinpoint exact reasons, this finding aligns with the ''community engagement'' component of the framework suggested by Seeman for patient-centered health blogs [31].

The social web should not be overlooked as an important player in e-health, broadly defined. Health communication studies make clear the myriad benefits and challenges of technology applications [32]. Large numbers of health care organizations and individual physicians now have a web presence. In spite of this, we are just beginning to understand the power of social media and online communities in helping patients to manage chronic conditions such as diabetes. There is much scope for further work on health information in the blogosphere, building on this current study by probing the conceptual model of Widén-Wulff et al., which bridges the concepts of information-seeking behavior and social capital [33].

The study's findings also suggest the need to explore whether librarians cite biomedical literature when they author or contribute to health blogs. The analytical strategies used by information professionals to assess the effectiveness of information resources and their dissemination might be enhanced by greater network and content analysis of the ever-expanding blogosphere.

## REFERENCES

1. Koch-Weser S, Bradshaw YS, Gualtieri L, Gallagher SS. The Internet as a health information source: findings from the 2007 Health Information National Trends Survey and implications for health communication. J Health Commun. 2010;15(suppl 3):279–93. DOI: http://dx.doi.org/10.1080/10810730.2010.522700.

2. Fox S, Pew Internet & American Life Project. The engaged e-patient population: people turn to the Internet for health information when the stakes are high and the connection fast. Washington, DC: The Project; 2008.

3. Adams SA. Blog-based applications and health information: two case studies that illustrate important questions for consumer health informatics (CHI) research. Int J Med Informatics. 2010 Jun;79(6):e89–e96. DOI: http://dx.doi.org/10.1016/j.ijmedinf.2008.06.009.

4. Winn P. State of the blogosphere: introduction. Technorati state of the blogosphere 2008 [Internet]. San Francisco, CA: Technorati; 21 Aug 2009 [cited 7 Sep 2010]. <http://www.technorati.com/blogging/article/state-of-the-blogosphere-introduction/>.

5. Neal DM, McKenzie PJ. Putting the pieces together: endometriosis blogs, cognitive authority, and collaborative information behavior. J Med Lib Assoc. 2011 Apr;99(2):127–34. DOI: http://dx.doi.org/10.3163/1536-5050.99.2.004.

6. Craigie M, Loader B, Burrows R, Muncer S. Reliability of health information on the Internet: an examination of experts' ratings. J Med Internet Res. 2002;4(1):e2. DOI: http://dx.doi.org/10.2196/jmir.4.1.e2.

7. Ashlin A, Richard JL. Environmental science adrift in the blogosphere. Science. 2006 Apr 14;312(5771):201. DOI: http://dx.doi.org/10.1126/science.1124197.

8. Gunter B, Campbell V, Touri M, Gibson R. Blogs, news and credibility. ASLIB Proceedings. 2009;61(2):185–204. DOI: http://dx.doi.org/10.1108/00012530910946929.

9. Davis PM, Walters WH. The impact of free access to the scientific literature: a review of recent research. J Med Lib

Assoc. 2011 Jul;99(3):208–17. DOI: http://dx.doi.org/10.3163/1536-5050.99.3.008.
10. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for 2000 and projections for 2030. Diabetes Care. 2004 May;27(5):1047–53. DOI: http://dx.doi.org/10.2337/diacare.27.5.1047.
11. National Library of Medicine (US). MedlinePlus: HbA1c [Internet]. Bethesda, MD: The Library [updated 6 Oct 2010; cited 20 Apr 2011]. <http://www.nlm.nih.gov/medlineplus/ency/article/003640.htm>.
12. Kim S, Chung DS. Characteristics of cancer blog users. J Med Lib Assoc. 2007 Oct;95(4):445–50. DOI: http://dx.doi.org/10.3163/1536-5050.95.4.445.
13. Walther JB, Pingree S, Hawkins RP, Buller DB. Attributes of interactive online health information systems. J Med Internet Res [Internet]. 2005;7(3) [cited 11 Apr 2011]. <http://www.jmir.org/2005/3>.
14. Glasgow RE, Barrera M, McKay HG, Boles SM. Social support, self-management, and quality of life among participants in an Internet-based diabetes support program: a multidimensional investigation. Cyberpsychol Behav. 1999 Jan 29;2(4):271–81. DOI: http://dx.doi.org/10.1089/cpb.1999.2.271.
15. Buis LR, Carpenter S. Health and medical blog content and its relationships with blogger credentials and blog host. Health Commun. 2009 Dec;24(8):703–10. DOI: http://dx.doi.org/10.1080/10410230903264014.
16. Miller E, Pole A. Diagnosis blog: checking up on health blogs in the blogosphere. Am J Pub Health. 2010 Aug;100(8):1514–9.
17. Kovic I, Lulic I, Brumini G. Examining the medical blogosphere: an online survey of medical bloggers. J Med Internet Res. 2008;10(3):e28. DOI: http://dx.doi.org/10.2196/jmir.1118.
18. Mitrović M, Tadić B. Bloggers behavior and emergent communities in blog space. Eur Phys J B. 2010;73(2):293–301.
19. Pikas CK. Detecting communities in science blogs. Proceedings of the 2008 4th IEEE International Conference on e-Science; [Indianapolis, IN]; 7–12 Dec 2008. IEEE Computer Society; 2008. p. 95–102.
20. Blanchard A. Blogs as virtual communities: identifying a sense of community in the Julie/Julia project. In: Gurak LJ, Antonijevic S, Johnson L, Ratliff C, Reyman J, eds. Into the blogosphere: rhetoric, community and culture of weblogs [Internet]. St. Paul, MN: University of Minnesota; 2004 [cited 20 Apr 2011]. <http://blog.lib.umn.edu/blogosphere/blogs_as_virtual.html>.
21. Nardi BA, Schiano DJ, Gumbrecht M, Swartz L. The blogosphere: why we blog. Commun ACM. 2004;47(12):41.
22. Carley KM, Reminga J. ORA: organization risk analyzer [Internet]. Pittsburgh, PA: Institute for Software Research International, School of Computer Science, Carnegie Mellon University; 2004 [cited 19 Apr 2011]. <http://www.casos.cs.cmu.edu/projects/ora/>.
23. Gruzd A, Wellman B, Takhteyev Y. Imagining Twitter as an imagined community. Am Behav Scientist [special issue on imagined communities]. 2011;55(10):1294–318. DOI: http://dx.doi.org/10.1177/0002764211409378.
24. Borgatti SP, Everett M, Freeman L. UCINET 6 for Windows: social network analysis [Internet]. Lexington, KY: Analytic Technologies; 2002 [cited 13 Sep 2011]. <http://www.analytictech.com/ucinet/>.
25. Krackhardt D, Stern RN. Informal networks and organizational crises: an experimental simulation. Soc Psychol Q. 1988;51(2):123–40.
26. Priest SH, ed. Encyclopedia of science and technology communication. Thousand Oaks, CA: SAGE Publications; 2010.
27. Public Library of Science [Internet]. San Francisco, CA: Public Library of Science [cited 20 Apr 2011]. <http://www.plos.org>.
28. National Library of Medicine (US). MedlinePlus [Internet]. Bethesda, MD: The Library [cited 20 Apr 2011]. <http://www.nlm.nih.gov/medlineplus/>.
29. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Ann Rev Sociol. 2001;27(1):415–44.
30. Fox S, Jones S, Pew Internet & American Life Project. The social life of health information. Washington, DC: The Project; 2009.
31. Seeman N. Inside the health blogosphere: quality, governance and the new innovation leaders. Healthc Q. 2009;7(3):101–8.
32. Suggs LS. A 10-year retrospective of research in new technologies for health communication. J Health Commun. 2006;11(1):61–74. DOI: http://dx.doi.org/10.1080/10810730500461083.
33. Widén-Wulff G, Ek S, Ginman M, Perttilä R, Södergård P, Tötterman AK. Information behaviour meets social capital: a conceptual model. J Inf Sci. 2008 Feb 5;34(3):346–55. DOI: http://dx.doi.org/10.1177/0165551507084679.

## AUTHORS' AFFILIATIONS

**Anatoliy Gruzd, PhD,** Anatoliy.Gruzd@dal.ca, Assistant Professor; **Fiona A. Black, PhD,** Fiona.Black@dal.ca, Associate Professor; School of Information Management, Dalhousie University, Halifax, NS B3H 4R2, Canada; **Thi Ngoc Yen Le, MEC,** yle@thirdwaveconsulting.ca, Salesforce Developer, Thirdwave Consulting, 1791 Barrington Street, Suite 300, Halifax, NS B3J 3K9, Canada; **Kathleen Amos, MLIS, AHIP,** kamos@phf.org, Project Manager, Council on Linkages Between Academia and Public Health Practice, Public Health Foundation, 1300 L Street Northwest, Suite 800, Washington, DC 20005, USA

*Received May 2011; accepted August 2011*