Technological Forecasting & Social Change xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

The world is not flat: Evaluating the inequality in global information gatekeeping through website co-mentions

George A. Barnett [a], Jeanette B. Ruiz [a], Weiai Wayne Xu [b], Ji-Young Park [c], Han Woo Park [d,*]

[a] University of California, Davis, United States
[b] Northeastern University, United States
[c] YeungNam University, Gyungsan 712-749, South Korea
[d] YeungNam University, South Korea

ARTICLE INFO

Article history:
Received 11 November 2015
Received in revised form 7 December 2016
Accepted 7 January 2017
Available online xxxx

Keywords:
Co-mention analysis
Webometrics
Globalization
Inequality

ABSTRACT

This study explores inequality in global internet by looking at structure of co-mentions across global top website domains. Findings show that websites of the U.S. were central and dominant in the global content flow. The network based on the level of corporate ownership was even more centralized, in which the top 10 sites producing at least 1% of all Internet citations were from U.S.-based companies such as Google, Facebook, and Twitter, which together accounted for >70% of the network ties. In particular, Google was at the center of the network and serves as the Internet "gatekeeper". Additionally, the global web is divided into two clusters of websites, one represented by websites owned by American firms and the other by Chinese companies. The study discusses how such divide might be the outcome of geopolitics, internet governance and media conglomeration.

1. Introduction

In his much acclaimed book *The World is Flat*, Thomas L. Friedman (Friedman, 2005) argued that globalization has created a more level playing field for global competition. Yet, while concerns for inequality linger, mostly in the realm of economics (Wade, 2004), it is becoming clear that information breeds inequality as well. One cannot consider information inequality without taking into account the World Wide Web (WWW).

The global web is a collection of networks, consisting of interconnected entities, ranging from individual bloggers at the micro level, to giant multinational corporations, and nation-states at the macro level (Castells, 2004, 2011; Chang et al., 2012). Among them, the hyperlink network is a type of network established on mutual acknowledgement of relevancy and information flow. Much like academic citations, hyperlinks point one source document to another (Thelwall, 2009), forming the web structure for content diffusion. Also, hyperlinking is a conscious and sometimes strategic behavior. Its patterns also reveal the politics of association (Rogers and Ben-David, 2008). Given that the global web has become a marketplace of ideas and the public sphere for the discussion of issues and social movements, how websites are interlinked affect the size and shape of the public sphere (Dahlgren, 2005; Turow, 2008).

There are two ways to look at hyperlinking patterns. The traditional approach looks at inter-linkages, that is, direct citations between a pair of websites (Barnett and Park, 2014). It shows relationships between site authors—how the site authors acknowledge one another. But the approach does not necessarily show contextual connectedness between site content, that is, how the content on one site is perceived as important to the content on another site, through the judgment of a third-party user. Additionally, this perspective has been criticized for including erroneous links, irrelevant information and inconsequential relationships (Weber and Monge, 2011; Ackland, 2013). The examination of website co-mentions represents an improved alternative.

Co-mention analysis is useful for identifying the contextual connectedness between two sites from a third-party perspective (He and Hui, 2002; Barnett et al., 2017). Co-mentions occur when two different sites are mentioned by a third site. Co-mention analysis is similar to bibliometrics, the study of the structure of literature and author collaboration (White and Griffith, 1981). Underlying co-mention ties is the mutual recognition of relevance and worthiness in terms of subject and content (He and Hui, 2002; Kenekayoro et al., 2015; Kim et al., 2016). Co-mention networks produce a concise road map to navigate users through content. The location of content in the co-mention

* Corresponding author.
E-mail addresses: gabarnett@ucdavis.edu (G.A. Barnett), jbruiz@ucdavis.edu (J.B. Ruiz), weiai.wayne.xu@gmail.com (W.W. Xu), heytreee@gmail.com (J.-Y. Park), hanpark@ynu.ac.kr (H.W. Park).

http://dx.doi.org/10.1016/j.techfore.2017.01.011
0040-1625/

network and the overall network structure dictate how much attention the content garners, consequently revealing the influence of the content providers (He and Hui, 2002).

2. Hyperlinking and information inequality

Hyperlinking's socioeconomic importance lies in its gatekeeping power. Traditionally, gatekeeping refers to news editors and journalists selectively choosing what content to make public (Shoemaker, 1991). This content selection affects what issues get public attention. In the digital age, gatekeeping takes on multiple forms. For example, Internet users can participate in the collective selection of sources, as in the case of social movements and breaking news (Meraz and Papacharissi, 2013; Xu et al., 2016). For this study, the focus is on hyperlinking - exercising the power to decide what content should get audience attention (Van Dijck, 2009) as hyperlinking dictates web traffic while its structure affects search engine indexing (Barzilai-Nahon, 2008). Highly linked websites tend to appear first in search results, making them more likely to be found and their information accessed (Page et al., 1999).

Hyperlinking networks are self-organizing web ecosystems without central planning (Barnett and Houston, 2005), and no central international governing body exists that dictates flows of information (Ruiz and Barnett, 2015). However, even assuming that hyperlinks grow organically, there are forces that naturally lead to inequality. For example, hyperlinking follows the scale-free power-law pattern as do most digital phenomenon—a small set of actors control the majority of hyperlinking traffic (Barabási and Albert, 1999). The web likely shows a bow-tie-like structure in which the frequency distribution of numbers of incoming and outgoing ties decays, and as a network grows, the probability of a given node receiving a tie is proportional to that node's current connectivity. Also, hyperlinking follows preferential attachment (Barabási and Albert, 1999) which posits that websites prefer to form links with more connected and thus more influential websites over time as a way to obtain popularity and influence, leading to the principle of "the rich get richer" (Pennock et al., 2002). The first two research questions examine the role of the organic forces in shaping unequal co-mention network.

RQ1: What is the network structure of the global web based on website co-mentions?

RQ2: Do structural characteristics of web-based co-mentions reflect preferential attachment and power law?

Along with the organic sociological forces, cultural differences and geopolitics play a role in the structure of the worldwide web. Communication networks show economic disparity as well as cultural differences (Barnett and Sung, 2005). For example, cultural and linguistic factors can lead certain countries and regions to be more closely connected through hyperlinks (Barnett and Sung, 2005). But more convincingly, the divisions reflect an uneven distribution of world power. Previous studies in this context have used world-system theory to reveal a core-peripheral structure in which Western countries play a central and prominent role in influencing online communication (Chung et al., 2014). Economically and politically powerful countries (e.g., the U.S. and Western European countries) tend to form the core, frequently linked by websites in countries with less influence (Park et al., 2011). This is evident in hyperlinks between global nongovernment organizations (Shumate and Dewitt, 2008; Yang, 2013), between firms (Nam et al., 2014), between international media outlets (Barnett et al., 2013), and between academic institutions (Park and Thelwall, 2006; Barnett et al., 2014). Hence, the following question addresses the role of geopolitics in shaping the global inequality imbedded in co-mention hyperlinking network.

RQ3: Do structural characteristics of web-based co-mentions reflect the core-peripheral structure in geopolitics?

Global information inequality can be also driven by the ideological differences involving whether and how the Internet should be governed to accommodate national interests. There is a debate within the international community about whether the Internet should be regulated by national governments in the service of their national interest (van Eeten and Mueller, 2013; Shackelford and Craig, 2014). This issue has led to heated debates after increasingly sophisticated and coordinated cyber-attacks targeting certain countries (DeNardis, 2014). The tension has pushed certain countries, mainly authoritarian ones, to propose the notion of "cyber sovereignty," that a national government can control its digital realm and activities within the country, much like a nation controlling its border (Gasser et al., 2013). China's Great Firewall aptly exemplifies such efforts to control domestic cyberspace. This firewall blocks access to global social networking sites such as Google, Facebook and Twitter and monitors Internet traffic through keyword filtering (Freedom House, 2013). Similarly, Iran has implemented the "halal internet" to purge Western influence (Shirazi, 2014). What is at stake here is not just the restriction of free speech but also the creation of a sealed-off web ecosystem disconnected from the rest of the world (van Eeten and Mueller, 2013). In China's case, the firewall has led to a separate and closed monopoly in which domestic internet service providers have grown rapidly to control a majority of Internet traffic (Zhong, 2012). In contrast to China's vibrant domestic Internet economy and cyber-culture, the Chinese Internet as a whole is less connected to the global Internet (Xu and Feng, 2015; Zhong, 2012). This disconnect supports the argument that Internet governance can create separate clusters on the web (Shackelford and Craig, 2014). However, it is less clear, whether such separation is reflected in patterns of co-mentions.

RQ4: Do structural characteristics of co-mentions reflect separate clusters of websites from certain countries?

Geopolitics aside, global information inequality can be shaped by media conglomeration. The ownership consolidation of the media industry has been an ongoing phenomenon since the last century (Arsenault and Castells, 2008). In the digital realm, a few dominant players, including Google, Facebook, and Amazon own web content and services (Haucap and Heimeshoff, 2014). This type of concentration has raised concerns over a possible monopoly of discourse and opinions. Recent empirical and critical studies have examined biases in search engines' content (Jiang, 2014; Mager, 2012). However, few studies have examined how the consolidated ownership of Internet firms is reflected in patterns of global information flow observed through website co-mentions.

RQ5: Do structural characteristics of website co-mentions reflect the ownership consolidation across Internet websites?

3. Methods

3.1. Co-mention network analysis

Network analysis is a research method for identifying the structure of information and social relationships (Carrington et al., 2005; Jung and Park, 2015, 2016). In contrast to traditional social science research methods such as the survey, content analysis and interview, which largely examine individual attributes of a person or content, the focus of network analysis is on how various attributes are interconnected to form a system (Knoke and Yang, 2008). Thus, network analysis is an optimal approach to examine web ecosystem formed on co-mentions. A network system is comprised of *nodes* and *ties*. Nodes, often referring to a person in social networks, and in the current context, individual websites, are connected to one and another through ties. In social

Please cite this article as: Barnett, G.A., et al., The world is not flat: Evaluating the inequality in global information gatekeeping through website co-mentions, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.01.011

networks, ties are formed based on friendship, interaction and proximity (Carrington et al., 2005). In content networks (e.g. hyperlink co-mention networks), ties are formed based on information flow, topical similarity or affiliation (Park, 2003). Ties can be analyzed for their presence or absence. In addition, they can be considered for the variation in its intensity and strength.

With node and tie being the two basic network elements, a network can be analyzed for network positions and its overall structure. Specifically, a basic network data set is an $n \times n$ matrix **S**, where n is the number of nodes in the network and s_{ij} is the quantified relationship between nodes i and j. In the current context, s_{ij} refers to the measured relationship between two websites, which is the number of co-mentions.

Network positions can be revealed through centrality. A higher score of centrality is associated with importance, prominence, or persuasive power in information dissemination (Knoke and Yang, 2008). Degree centrality is a commonly used centrality measure. Degree centrality refers to the number of ties or the total tie strength possessed by a node in a network. In the current context, a high degree centrality means that a website shares much topical and content similarity with many other sites. Overall, high degree centrality puts a node in the central position in a network. Another centrality measure, betweenness, reveals the control or brokerage of information *control* among different nodes. It is measured by the frequency with which a node lies in the shortest path connecting everyone else in the network (Freeman, 1977). Websites with higher betweenness centrality are in a better position to create new exchange ties with other websites and in dominating the general discourse.

The overall network structure can be revealed through density, which refers to how extensive or complete ties in a network are. The density of a network ranges from *sparse*, that is a network has few ties, to *dense*, whereas nodes in a network are highly connected. Density is calculated as the actual number of ties divided by the number of possible ties $[n(n-1)/2]$. To tap into the inequality of distribution of power across a network, the Gini coefficient measures is used, which scores from 0, indicating perfect equality to 1, showing maximum inequality.

A network can breakdown into sub-groups, often referred to as *clusters* in network terms. Cluster analysis involves the identification of such sub-groups. Specifically, a complete link hierarchical clustering using UCINET (Borgatti et al., 2002) was used in this study to group nodes into subsets with similar or structurally equivalent characteristics.

UCINET (Borgatti et al., 2002) was used to calculate network measures, and diagrams were visualized using NetDraw, a component program in UCINET (Borgatti, 2002). Visual representations are important for facilitating a better understanding of networks and illustrating analysis results.

3.2. Data

The 200 most visited websites in the world as of March 31, 2014, were determined using Alexa.com (http://www.alexa.com/topsites). Redundant websites were not merged because Alexa.com treated them as separate websites. Within the top 200 there were only three sets of redundant websites, t.co and twitter.com, bbc.co.uk and bbc.com and livedoor.jp and livedoor.com. However, amazon, ebay, google, naver, and yahoo had multiple website within this sample differentiated by language. 200 websites were chosen because the proportion of website users is distributed according to the power law (Barabási and Albert, 1999; Barnett and Park, 2014) and represent the vast majority of site visits in the world. These websites were based in 39 different countries. In terms of the place of origin, the U.S. had 93 sites; China, 24; the U.K., 9; Japan and Russia, 7; India and Canada, 6; and Germany, 5. Co-mentions of a pair of the top 200 sites or co-mentions of these sites' domains based on their symmetric relationships were measured using Boolean operators for data collection purposes. More specifically, double quotation marks were used to increase accuracy, for example, "Google.com" (AND) "Facebook.com."

Hit counts were obtained from Google.com (through paid access keys using the Google Custom Search Engine) because Yahoo.com no longer provided hit numbers through its API service. Numbers of hits for each of the 19,990 search queries were gathered between May 7 and May 9, 2014, by using proprietary web-mining software based on Google API. Because only 10,000 search queries per day were allowed, queries were evenly split across the three days. Search parameters included search scope keywords or links and all search regions. Ownership information was obtained for each of the 200 top websites. Websites were aggregated to the corporate level or to the level of corporate ownership, which reduced the number of nodes from 200 to 127. Ownership data were collected through an online search of each website.

4. Results

The first research question addressed the network structure of the Web based on website co-mentions. The results indicate a completely connected network. Its density was 1.0. Every website shared or co-cited at least one other website with another website. When the mean value of co-mentions was required for a link, the density was only 0.027, indicating a very sparse network. In addition, the results for a valued network (number of co-mentions) tell a different story. The composite Gini coefficient for the distribution of cited websites was 0.161, and there were a few sites cited disproportionately. Table 1 shows the normalized centrality of the top 20 websites. Betweenness centrality was calculated by dichotomizing data at the mean number of co-mentions (354,663). The most central website was YouTube (U.S.), which accounted for 20.0% of all website co-mentions, followed by Facebook (U.S.; 18.9%), Twitter (U.S.; 16.1%), Google.com (U.S.; 4.4%), Instagram (U.S.; 4.2%), BlogSpot (U.S.; 3.7%), and akamaihd.net (U.S.; 3.6%). In terms of betweenness centrality, Google.com was the most central website, followed by Blogspot.com, and Baidu.com (China).

Concerning the second question, the Spearman rank-order correlation between degree centrality and the website rank for the 200 most visited websites was -0.269 ($p < 0.001$), and that between betweenness centrality and the rank was -0.359 ($p < 0.001$). This indicates that less popular websites cited well-connected websites to increase their status and connections. The power law led to a significant fit for the distribution of centrality measures ($p < 0.001$ for all measures),

Table 1
Co-mention centrality for Top 20 websites.

Website	Degree	Share	Betweenness centrality
youtube.com	1.446	0.200	3.428
facebook.com	1.368	0.189	2.770
twitter.com	1.164	0.161	1.416
google.com	0.318	0.044	7.955
instagram.com	0.300	0.042	0.085
blogspot.com	0.268	0.037	5.880
akamaihd.net	0.260	0.036	0.064
wordpress.com	0.208	0.029	0.986
tumblr.com	0.185	0.026	0.475
Bp.blogspot.com	0.157	0.022	0.588
t.co	0.153	0.021	0.074
files.wordpress.com	0.151	0.021	0.037
pinterest.com	0.100	0.014	0.277
soundcloud.com	0.082	0.011	0.071
amazon.com	0.081	0.011	1.305
googleusercontent.com	0.068	0.009	0.064
yahoo.com	0.064	0.009	2.206
wikipedia.org	0.056	0.008	1.720
flickr.com	0.047	0.007	0.058
wikimedia.org	0.038	0.005	0.085
Mean	0.036	0.005	0.029
Standard Deviation	0.167	0.023	0.877

Note: Degree: Normalized degree; Betweenness: Normalized betweenness.

Please cite this article as: Barnett, G.A., et al., The world is not flat: Evaluating the inequality in global information gatekeeping through website co-mentions, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.01.011

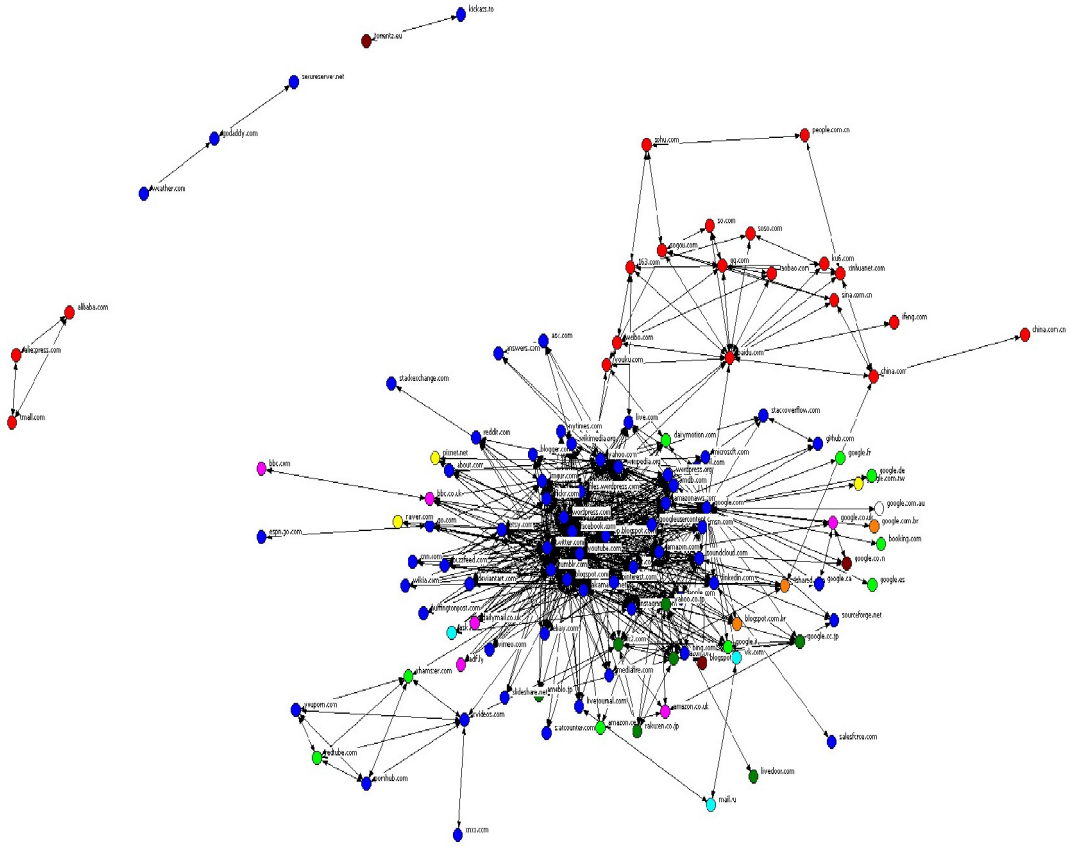
but not to the best goodness-of-fit (Park et al., 2011). The power law accounted for <50% (49.4%) of the total variance in degree centrality and 73.3% of that in betweenness centrality. In addition, the inverse function was significant ($p < 0.001$) but accounted for 86.7% of the total variance in degree centrality, and 92.3% of that in betweenness centrality because the network was more concentrated around the most central websites.

Research question 3 through 5 addressed structural differences possibly shaped by geopolitical and policy factors. Four different cluster analyses were conducted: single-link (nearest neighbor or minimum method), complete-link (diameter or maximum method), average, and weighted average methods. A visual analysis produced three distinct clusters. There was a highly centralized group around U.S. websites such as YouTube, Twitter, Facebook, and Google.com. The second group was a cluster composed of pornographic websites based in the U.S. and Europe, including sites such as XVideos.com (U.S.) and Pornhub.com (U.S.). Finally, there was another cluster around a group of Chinese websites such as Baidu.com, China.com, and Weibo.com. A visual examination revealed three clusters, and the results of a statistical analysis of the adequacy of the clustering solution indicate the best description to be a network composed of only two clusters: a core cluster of American websites and another composed of Chinese sites. The E-I index for the two-cluster solution was 0.908, and that for the three-cluster solution was 0.855. The E-I index is calculated by the number of external ties to a group minus that of internal ties divided by the total number of ties (Krackhardt, 1988).

Fig. 1 shows a website co-mention network with prominent website connections (based on the mean as the minimum for a link) and the country of origin. The figure shows the most central U.S. websites (blue) at the center, Chinese websites (red) near the top and slightly to the right, and pornographic websites toward the lower left. Fig. 2 shows only the most central nodes in the website co-mention network (based on the mean plus one standard deviation). These websites were based almost exclusively in the U.S., with U.K. and Chinese websites at the periphery.

To address RQ4 specifically—whether structural characteristics of co-mentions would reflect separate clusters of websites from different countries. For China, the most central website was Baidu, but Baidu accounted for only 0.1% of all co-mentions. This indicates a much lower level of ownership concentration in comparison to the U.S. The two main clusters were based in China and the U.S., but the U.S.-based cluster held the larger share and the higher centrality. A hierarchical cluster analysis of the website co-mention ownership network revealed two main clusters: a China-based cluster and a U.S.-based cluster. Initially, there were three clusters (including a cluster of pornographic websites), but an additional analysis clearly revealed the cluster to be part of the larger U.S.-based cluster. The E-I index for the two-cluster solution was 0.961, and that for the three-cluster solution was 0.886.

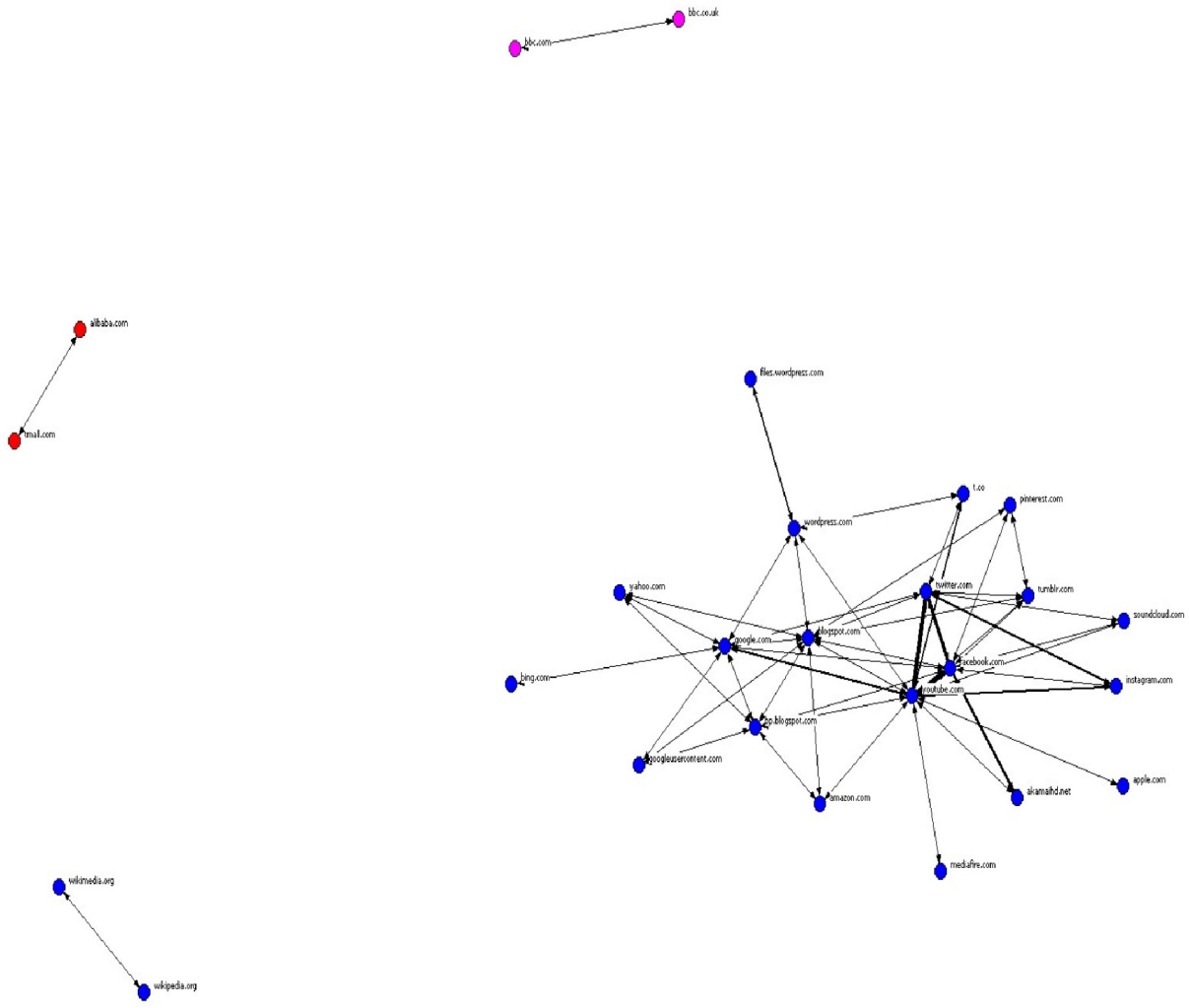
To address RQ 5, whether structural characteristic of co-mentions would reflect concentration of ownership among corporate owners of websites, the results indicate a network with a density value of 1.0.



Note: (the mean is the minimum for a link). U.S.: Blue; China: Red; Japan: Yellow; U.K.: Violet; Russia: Light blue; E.U.: Light green; India: Brown; Latin America: Orange; Australia: White.

Fig. 1. A website co-mention network showing prominent website connections. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: Barnett, G.A., et al., The world is not flat: Evaluating the inequality in global information gatekeeping through website co-mentions, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.01.011



Note: U.S.: Blue; China: Red; U.K.: Violet.

Fig. 2. A website co-mention network based on the mean plus one standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

However, when the mean value was required for a link, the density was 0.30. The composite Gini coefficient for the distribution of website ownership was 0.939, indicating a highly concentrated network (more concentrated than the network of individual website URL co-mentions). Table 2 shows the normalized centrality, share, and betweenness centrality of websites aggregated by ownership and a list of firms and their websites. The network was clearly centered on U.S. firms. The share was concentrated in three firms: Google (28% of website co-mentions), Facebook (24%), and Twitter (20%). These three firms are based in the U.S. All of the top 10 websites with at least 1% of all Internet co-mentions were owned by U.S. firms (except for Wikimedia, a site not commercially owned). In terms of betweenness centrality, all the websites analyzed were connected through a Google-owned website. So overall, Google dominated access regardless of the language or country of origin of websites.

Fig. 3 visualizes a website co-mention ownership network showing prominent website connections (based on the mean as the minimum for a link) and the country of origin in terms of website ownership. The most central U.S. websites (blue) were at the center, and Chinese websites (yellow) were near the top and to the right. Here the size of a node indicates its degree centrality in the network.

5. Discussion

This study shows that co-mention network analysis, a derivative of hyperlink analysis, is useful in discerning global inequality in term of information gatekeeping. As argued in the beginning of the study, such inequality can reflect or reinforce dominance or marginalization in the attention economy. In particular, inequality could be a product of power law and preferential attachment—two organic forces shaping the web ecosystem. But inequality may be also shaped by domestic internet governance that emphasizes on sovereignty and control as well as conglomeration of media outlets. The findings first show that the global co-mention network follows power law and the rule of preferential attachment in that less cited websites tend to be connected to the most widely cited sites. The most popular websites tend to be more central in the co-mention network.

Second, the study shows that the global web has a highly connected network structure in which the world's top 200 websites cite one and another. Yet, the power in influencing web traffic is disproportional, predominately controlled in the hand of U.S.-based websites. Specifically, YouTube, Facebook, and Twitter, followed by Google, Instagram, and Blogspot are dominant in the network. The

Please cite this article as: Barnett, G.A., et al., The world is not flat: Evaluating the inequality in global information gatekeeping through website co-mentions, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.01.011

Table 2
Centrality by website ownership for top 20 entities.

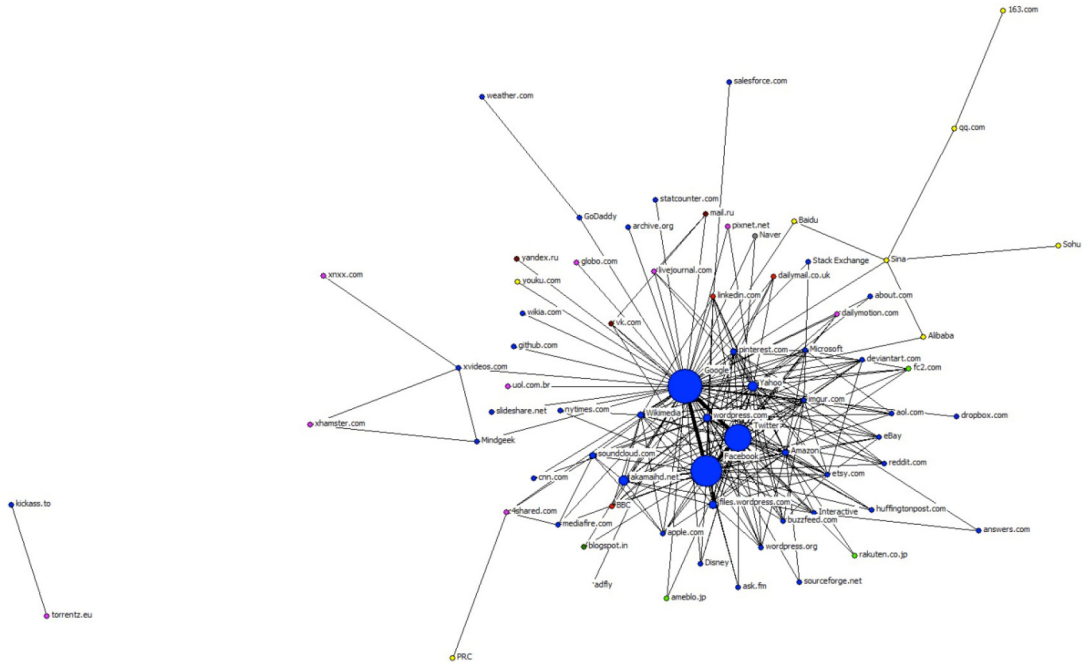
Website owner	Degree	Share	Eigenvector	Betweenness	Sites owned
Google	2.051	0.275	84.480	22.646	36
Facebook	1.804	0.242	80.151	1.290	2
Twitter	1.525	0.204	77.495	1.090	2
Yahoo	0.339	0.045	11.733	1.740	4
akamaihd.net	0.306	0.041	14.893	0.036	1
wordpress.com	0.244	0.033	3.672	0.636	1
files.wordpress.com	0.178	0.024	1.508	0.010	1
Amazon	0.131	0.018	4.166	0.350	6
pinterest.com	0.117	0.016	2.945	0.184	1
soundcloud.com	0.096	0.013	4.835	0.106	1
Wikimedia	0.066	0.009	1.499	0.068	3
Microsoft	0.043	0.006	1.253	0.025	5
etsy.com	0.040	0.005	0.805	0.023	1
imgur.com	0.034	0.005	0.739	0.126	1
eBay	0.020	0.003	0.376	0.000	4
Interactive	0.025	0.003	0.488	0.265	3
apple.com	0.023	0.003	0.908	0.004	1
deviantart.com	0.021	0.003	0.500	0.000	1
Mindgeek	0.018	0.002	0.037	0.425	3
blogspot.in	0.012	0.002	0.637	0.000	1

Note: Degree: Normalized degree; Betweenness: Normalized betweenness.

web also exhibits an interesting bio-polar mode, with the dominant U.S. presence and rising emergence of China. China, as the world's second largest economy and the host of the largest Internet population in the world has two sites listed as the top. This is consistent with the findings of previous network structure studies suggesting U.S. dominance. Such dominance can be attributed to historical and economic factors: historically, the Internet was created and developed in the U.S. and has only recently been subjected to any ownership rules, as in the case of other media communication channels. Therefore, it is not surprising to find a U.S.-based firm at the center of the network. Theoretically, the finding supports the world-systems theory.

China's domestic Internet governance may explain the salience of Chinese sites. Chinese websites were low in centrality in comparison to U.S.-based websites. Among China-based websites, Baidu, a Chinese-language search engine similar to Google, was the most central website. However, Baidu accounted for only 0.1% of the network share. But Chinese sites as a whole form a unique cluster. This is probably due to Chinese government's internet policies that favor domestic service providers and restrict foreign operations through content censorship. Reinforced fight over Internet governance can create clusters of divides on the web (Shackelford and Craig, 2014). The firewall has led to a separate and closed monopoly board in which domestic internet service providers have grown rapidly to control a majority of internet traffic (Zhong, 2012). In contrast to China's vibrant domestic Internet economy and cyber-culture, the Chinese Internet as whole is less connected to the global Internet (Xu and Feng, 2015; Zhong, 2012). The findings from this study is consistent with the argument. Domestic Internet governance aside, the result of regional difference (between the U.S. and China) might be explained by language, content and cultural differences. Previous studies show that hyperlinks differ based on their topics and language (Kenekayoro et al., 2015). An analysis of the top 50 Korean websites found web portals and search engines occupying the top seven positions in the network (Park et al., 2002a). Another hyperlink-affiliation network analysis found a subgroup centered about financial websites (content/type) (Park et al., 2002b). A review of international dotcom web communities by link found the dominant use of English by central websites. Top website content types included business, the Internet and computers, and recreation and entertainment (Chung et al., 2014).

Yet, at large, regardless of the language or country of corporate ownership, Google served as the global Internet gatekeeper. For the website co-mention ownership network, the level of ownership concentration



Note: U.S.: Blue; China: Yellow; U.K.: Red, Korea: Gray; Russia: Brown; Japan: Light green; Canada: Light blue; India: Green; Other: Magenta. Node size indicates relative co-mentions.

Fig. 3. A website co-mention network by corporate ownership based on the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: Barnett, G.A., et al., The world is not flat: Evaluating the inequality in global information gatekeeping through website co-mentions, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.01.011

indicates the dominance of three firms based in the U.S.: Google, Facebook, and Twitter. These firms accounted for a majority (>70%) of co-mentions in the network.

Taken together, co-mention analyses reveal the division in cyberspace between the United States and China, and between dominant Internet players and smaller regional players. Such divides can be attributed to concentration of corporate ownership and domestic Internet governance.

In addition to the insights, the study also makes a unique methodological contribution. As mentioned, ample studies looking into globalization adopt qualitative approach. The current study provides a structural and webometric approach. In particular, its focus on co-mention network is unique in itself, among hyperlink studies.

6. Policy implications

In a study of internet domain names, Ruiz and Barnett (2015) addressed the issue of internet ownership. While ownership of domains and content reflects worldwide inequality, the inequality in content gatekeeping, as seen in the current study, is perhaps more concerning. The pattern explained above shows a strong favoritism towards bigger players to widen the divide between haves and have-nots. When internet first came into being, there was optimism that internet can level the playing fields to provide alternative voices and opportunities. Yet, much of the developments of the past decades have shown strong policy influence and de facto monopoly. It is debatable whether internet is a private business that should be free of regulation, or an electronic public sphere that needs to be governed in a way to promote public interest. Future internet governance may focus on creating an international governing body, much like the United Nations for geopolitics, to balance various national and corporate interests in the online space. While commercial operations should be free from regulations, preferential domestic internet policies should be subject to the scrutiny of the international body. Meanwhile, technology developers can explore the possibility of developing algorithm that is less *biased* towards existing influential sources.

7. Future directions

Future research should examine the roles of various types of content and the website language in patterns of co-mentions on the Web. Previous studies have classified a large number of content types, including science and research, the Internet and computers, politics, recreation and entertainment, personal interests, business, education, arts, social issues and religion, news, sports, travel, health, regional issues, and government (Chung et al., 2014). This classification system should be refined to take into account variations in the type of content in the sample. For example, the Internet and computers should be further differentiated into portals (Yahoo.com), search engines (Google.com), social media (Facebook.com), software (Microsoft.com), and computer hardware firms (Apple.com).

More generally, future research should extend the investigation of the network structure of the Internet by including various measures and approaches (Barnett and Park, 2014), as done in the present study, as well as by examining how the network changes over time. Here a longitudinal analysis should provide valuable insights into any developments or changes in the structure as a result of the implementation of net neutrality in the U.S. and the rapid adoption of the Internet in China, India, and other countries. This should provide a more detailed picture of the structure of the Internet as well as the roles of content, culture, and technology in determining global society as it becomes fully embedded in a digital world.

Acknowledgement

The earlier draft of this paper was presented in the HICSS 2015 (48th Annual Hawaii International Conference on System Sciences). However, current version is significantly different from conference presentation because the current study added the section on the aggregation of the websites by country of ownership in terms of the inequality in the attention economy.

References

Ackland, R., 2013. Web Social Science: Concepts, Data and Tools for Social Scientists in the Digital Age. Sage.
Arsenault, A., Castells, M., 2008. The structure and dynamics of global mutli-media business networks. Int. J. Commun. 2, 707–748.
Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512.
Barnett, G.A., Houston, R., 2005. Advances in Self-organizing Systems. Hampton Press.
Barnett, G.A., Park, H.W., 2014. Examining the international internet using multiple measures: new methods for measuring the communication base of globalized cyberspace. Qual. Quant. 48 (1), 563–575.
Barnett, G.A., Sung, E., 2005. Culture and the structure of the international hyperlink network. J. Comput.-Mediat. Commun. 11 (1), 217–238.
Barnett, G.A., Park, H.W., Jiang, K., Tang, C., Aguillo, I.F., 2014. A multi-level network analysis of web-citations among the world's universities. Scientometrics 9 (1), 5–26.
Barnett, G.A., Ruiz, J.B., Hammond, J.R., Xin, Z., 2013. An examination of the relationship between international telecommunication networks, terrorism and global news coverage. Soc. Netw. Anal. Min. 3 (3), 721–747.
Barnett, G.A., Xu, W.W., Chu, J., Jiang, K., Huh, C., Park, J.Y., Park, H.W., 2017. Measuring international relations in social media conversation Gov. Inf. Q (in press).
Barzilai-Nahon, K., 2008. Toward a theory of network gatekeeping: a framework for exploring information control. J. Am. Soc. Inf. Sci. Technol. 59 (9), 1493–1512.
Borgatti, S.P., 2002. Netdraw Network Visualization. Analytic Technologies, Harvard, MA.
Borgatti, S.P., Everett, M.G., Freeman, L.C., 2002. Ucinet for Windows: Software for Social Network Analysis. Analytic Technologies, Harvard, MA.
Carrington, P.J., Scott, J., Wasserman, S., 2005. Models and Methods in Social Network Analysis. Cambridge University Press.
Castells, M. (Ed.), 2004. The Network Society: A Cross-cultural Perspective. Edward Elgar, Cheltenham.
Castells, M., 2011. The Rise of the Network Society: The Information Age: Economy, Society, and Culture. vol. 1. Wiley-Blackwell, Malden, MA.
Chang, T.K., Southwell, B.G., Lee, H.M., Hong, Y., 2012. Jurisdictional protectionism in online news: American journalists and their perceptions of hyperlinks. New Media Soc. 14 (4), 684–700.
Chung, C.J., Barnett, G.A., Park, H.W., 2014. Inferring international dotcom web communities by link and content analysis. Qual. Quant. 48 (2), 1117–1133.
Dahlgren, P., 2005. The Internet, public spheres, and political communication: dispersion and deliberation. Polit. Commun. 22, 147–162.
DeNardis, L., 2014. The Global War for Internet Governance. Yale University Press.
Freeman, L.C., 1977. A set of measures of centrality. Sociometry 40 (1), 35–41.
Friedman, T.L., 2005. The World is Flat: A Brief History of the Twenty-first Century. Farrar, Straus and Giroux, New York.
Gasser, U., Faris, R., Heacock, R., 2013. Internet Monitor 2013: Reflections on the Digital World. Berkman Center Research Publication (27).
Haucap, J., Heimeshoff, U., 2014. Google, Facebook, Amazon, eBay: is the internet driving competition or market monopolization? IEEP 11 (1-2), 49–61.
He, Y., Hui, S.C., 2002. Mining a web citation database for author co-citation analysis. Inf. Process. Manag. 38 (4), 491–508.
House, F., 2013. Throttling dissent: China's new leaders refine internet control. http://www.freedomhouse.org/sites/default/files/resources/Throttling%20Dissent_FOTN%202013_China_0.pdf.
Jiang, M., 2014. Search concentration, bias, and parochialism: a comparative study of Google, Baidu, and Jike's search results from China. J. Commun. 64 (6), 1088–1110.
Jung, K., Park, H.W., 2015. A Semantic (TRIZ) Network Analysis of South Korea's "Open Public Data" Policy. Gov. Inf. Q. 32 (3), 353–358.
Jung, K., Park, H.W., 2016. Tracing interorganizational information networks during the emergency response period: A webometric approach to the 2012 gumi chemical spill in South Korea. Gov. Inf. Q. 33 (1), 133–141.
Kenekayoro, P., Buckley, K., Thelwall, M., 2015. Clustering research group website homepages. Scientometrics 102 (3), 2023–2039.
Kim, J.Y., Lee, Y.O., Park, H.W., 2016. Delineating the complex use of a political podcast in South Korea by hybrid web indicators: the case of the Nakkomsu Twitter network. Tech. Forecasting Soc. Chang. 110, 42–50.
Knoke, D., Yang, S., 2008. Social network analysis. Sage.
Krackhardt, D., 1988. Informal networks and organizational crises: an experimental simulation. Soc. Psychol. Q. 51 (2), 123–140.
Mager, A., 2012. Algorithmic ideology: how capitalist society shapes search engines. Info. Commun. Soc. 15 (5), 769–787.
Meraz, S., Papacharissi, Z., 2013. Networked gatekeeping and networked framing on# Egypt. Int. J. Press Polit. 18 (2), 138–166.
Nam, Y., Barnett, G.A., Kim, D., 2014. Corporate hyperlink network relationships in global corporate social responsibility system. Qual. Quant. 48 (3), 1225–1242.

Please cite this article as: Barnett, G.A., et al., The world is not flat: Evaluating the inequality in global information gatekeeping through website co-mentions, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.01.011

Page, L., Brin, S., Motwani, R., Winograd, T., 1999. The pagerank citation ranking: bringing order to the web. Technical Report. Stanford InfoLab (Retrieved from http://ilpubs.stanford.edu:8090/422/).
Park, H.W., 2003. Hyperlink network analysis: a new method for the study of social structure on the web. Connections 25 (1), 49–61.
Park, H.W., Thelwall, M., 2006. Web-science communication in the age of globalization. New Media Soc. 8 (4), 629–650.
Park, H.W., Barnett, G.A., Chung, C.J., 2011. Structural changes in the global hyperlink network: centralization or diversification. Global Netw. 11 (4), 522–542.
Park, H.W., Barnett, G.A., Nam, I.Y., 2002a. Interorganizational hyperlink networks among websites in South Korea, NETCOM: Networks and Communication Studies. Special issue on the Internet development in Asia 16 (3/4), 155–173.
Park, H.W., Barnett, G.A., Nam, I.Y., 2002b. Hyperlink-affiliation network structure of top websites: Examining affiliates with hyperlink in Korea. J. Am. Soc. Inf. Sci. Technol. 53 (7), 602–611.
Pennock, D.M., Flake, G.W., Lawrence, S., Glover, E.J., Giles, C.L., 2002. Winners don't take all: characterizing the competition for links on the web. Proc. Natl. Acad. Sci. 99 (8), 5207–5211.
Rogers, R., Ben-David, A., 2008. Winners don't take all: characterizing the competition for the Israeli NGO. New Media Soc. 10 (3) (The Palestinian J., & Giles, C. L. (2002). 497): 28).
Ruiz, J.B., Barnett, G.A., 2015. Who owns the international Internet networks? J. Int. Commun. 21 (1), 38–57.
Shackelford, S., Craig, A., 2014. Beyond the new 'Digital Divide': analyzing the evolving role of national governments in internet governance and enhancing cybersecurity. Stanf. J. Int. Law 50, 119.
Shirazi, F., 2014. Interrogating Iran's restricted public cloud: an Actor Network Theory perspective. Telematics Inform. 31 (2), 228–236.
Shoemaker, P.J., 1991. Gatekeeping. Sage, Newbury Park, CA.
Shumate, M., Dewitt, L., 2008. The north/south divide in NGO hyperlink networks. J. Comput.-Mediat. Commun. 13 (2), 405–428.
Thelwall, M., 2009. Introduction to webometrics: quantitative web research for the social sciences. Synthesis Lect. Inf. Concepts Retrieval Serv. 1 (1), 1–116.
Turow, J., 2008. Introduction: on not taking the hyperlink for granted. In: Turow, J., Tsui, L. (Eds.), The Hyperlinked Society. University of Michigan Press, Ann Arbor, MI, pp. 1–17.
Van Dijck, J., 2009. Users like you? Theorizing agency in user-generated content. Media Cult. Soc. 31 (1), 41–58.
van Eeten, M.J., Mueller, M., 2013. Where is the governance in Internet governance? New Media Soc. 15 (5), 720–736.
Wade, R.H., 2004. Is globalization reducing poverty and inequality? World Dev. 32 (4), 567–589.
Weber, M.S., Monge, P., 2011. The flow of digital news in a network of sources, authorities, and hubs. J. Commun. 61 (6), 1062–1081.
White, H.D., Griffith, B.C., 1981. Author cocitation: a literature measure of intellectual structure. J. Am. Soc. Inf. Sci. 32 (3), 163–171.
Xu, W., Feng, M., 2015. Networked creativity on the censored Web 2.0: Chinese users' Twitter-based activities on the issue of internet censorship. J. Contemp. East. Asia 14 (1), 23–43.
Xu, W.W., Park, J.Y., Kim, J.Y., Park, H.W., 2016. Networked cultural diffusion and creation on YouTube: an analysis of YouTube memes. J. Broadcast. Electron. Media 60 (1), 104–122.
Yang, A., 2013. When transnational civil network meets local context: an exploratory hyperlink network analysis of Northern/Southern NGOs' virtual network in China. J. Int. Intercult. Commun. 6 (1), 40–60.
Zhong, Y., 2012. The Chinese Internet: a separate closed monopoly board. J. Int. Commun. 18 (1), 19–31.

George A. Barnett (PhD, Michigan State University, 1976) is Distinguished Professor of Communication at the University of California, Davis. He has written extensively on organizational, mass, international, intercultural and political communication, as well as the diffusion of innovations. Professor Barnett served as Chair of the Communication and Technology Division of the International Communication Association and President of the International Network for Social Network Analysis (INSNA). He edited the *Handbook of Organizational Communication*, *Organization ↔ Communication: Emerging Perspectives*, *Advances in Communication Sciences*, and *Encyclopedia of Social Networks*. His current research focuses on international information flows and their role on socioeconomic and cultural change, and the process of globalization.
Correspondence: gabarnett@ucdavis.edu

Weiai (Wayne) Xu, Assistant Professor, Department of Communication, University of Massachusetts - Amherst. His research exams internet communities and their impact on people's social and civic life. Email: weiaixu@umass.edu

Jeanette B. Ruiz (PhD, University of California, Davis, 2015) is a Lecturer (LPSOE, Senate Faculty) at the University of California, Davis. Her research focuses on international Internet governance and online public health communication.

Ji-Young Park (MA, Yeungnam University, South Korea) is a doctoral candidate in Interdisciplinary Program of East Asian Cultural Studies at Yeungnam University. Her research areas include social media & big data in Asia, cross-cultural & intercultural communication, and new media & digital technology.

Dr. Han Woo Park is a Full Professor in the Dept. of Media & Communication, Interdisciplinary Programs of East Asian Cultural Studies, and of Digital Convergence Business at YeungNam University, South Korea. His research focuses on the use and role of open (big) data in extending academic, governmental, and business networks in scientific, technical, and innovative activities. He was a pioneer in network science of open big data in the early 2000s (often called *Webometrics*). He is currently the president of the *World Association for Triple Helix and Future Strategy Studies*. Since the establishment of open data law in Korea, he has actively participated in the *Strategic Committee of Public Data* chaired by Prime Minister. He has founded a prestigious conference on big data, government 3.0, and triple network in Asia, called *DISC* (Daegu Gyeongbuk International Social Network Conference). He also sits on the consulting boards of European Union project on open data (http://www.data4policy.eu) in collaboration with *Oxford Internet Institute* where he used to be a visiting scholar. He has been invited to give a keynote speech in a number of conferences including the *International Conference on e-Democracy and Open Government-Asia 2012* (CeDEM-Asia-2012). Further, he is strongly affiliated with several prestigious journals on open government and big data such as *Big Data & Society*. Further, he is a principal investigator of *OpenData500.com* on Korean side (http://www.opendata500.com/kr). Recently, he published the following article: Jung, K., & Park, H.W. (2015). A Semantic (TRIZ) Network Analysis of South Korea's "Open Public Data" Policy. *Government Information Quarterly*. 32 (3), 353-358.

Please cite this article as: Barnett, G.A., et al., The world is not flat: Evaluating the inequality in global information gatekeeping through website co-mentions, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.01.011