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

CrossMark

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

A B S T R A C T

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.

© 2017 Elsevier Inc. All rights reserved.

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).

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