Journal of Communication ISSN 0021–9916

ORIGINAL ARTICLE

Mapping User-Centric Internet Geographies: How Similar are Countries in Their Web Use Patterns?

Yee Man Margaret Ng [1] & Harsh Taneja [2]

1 Department of Journalism and Institute for Communication Research, University of Illinois at Urbana Champaign, Urbana, IL, 61801, USA
2 Department of Advertising and Institute for Communication Research, University of Illinois at Urbana Champaign, Urbana, IL, 61801, USA

With half the world now online, a handful of websites dominate globally. Yet little is known about the homogeneity or geographical distinctness of global web use patterns. Focusing beyond popular sites, we inquired into how and why countries are similar in their web use patterns, developing a framework drawing on the literatures on media globalization, as well as Internet geographies. To compute similarities in web use between countries, we utilized an algorithm that considered both ranking positions and overlap counts on ranked lists of the 100 most popular websites for 174 countries, totaling 6,252 unique websites. Findings from a network analysis and from regressions suggest that countries with similar languages and shared borders, as well as those vastly different in their Internet market sizes, tend to have similar web use patterns. Neither are countries particularly similar to the US in web use nor does the prevalence of English speakers have an influence.

Keywords: Global Web Use, Global Media Flows, Cultural Proximity, World Systems Theory, Rank Biased Overlap, Cluster Analysis, Quadratic Assignment Procedure, Network Analysis

doi:10.1093/joc/jqz030

The World Wide Web presumably creates the foundation for a more united world, facilitating ideas across once-prohibitive national boundaries and geographical distances. At the same time, the availability of nationally focused web content in one's preferred language might leave little incentive for users to explore what is foreign or unfamiliar. Given both these possibilities, what people around the world actually do online remains an open question.

Corresponding author: Harsh Taneja; e-mail: harsh.taneja@gmail.com
The authors contributed equally to the work and authorship has been determined by alphabetical order.

Studies, based on analyses of web traffic across a large number of websites, often show that while web use, on average, is driven by language and geography, the usage of immensely popular websites with user-generated content, such as YouTube, tends to be more homogenous across countries. Little is known, however, about how different or similar countries are in their web use patterns. Altering the level of analysis, therefore, from websites to countries, our study intervenes in this discussion with an aim to examine the extent to which countries are similar in their web use patterns. Instead of sampling the 1,000 most popular websites in the world or focusing on specific domains, such as YouTube, Twitter, or Wikipedia, as previous studies have done, we sampled the most popular websites from each country. This design captures usages from both large and smaller countries.

To predict how similar or different countries are from one another in their web use, we turned to the canonical literature on media globalization, as well as the relatively recent literature on geographies of the global Internet. Distinct from a focus on technical infrastructures, we adopted a user-centric perspective, assuming that both institutional and cultural factors shape global cultural consumption. We thus posited that languages spoken in countries, their geographical locations, and the relative sizes and influences of their Internet markets all play a role in shaping their web use patterns. Consequently, similarities and differences between countries along these traits could explain how similar or dissimilar they are from one another in their web use.

We utilized web traffic data from 174 countries to create a network of countries, based on the same websites appearing in their ranked lists of the 100 most popular websites. Our sample included 6,252 distinct web domains. We analyzed this network to detect communities of countries with similar web use and also predict these ties using regressions based on quadratic assignment procedures. We replicated these analyses for another time point, and also after removing several websites by tech giants. We also analyzed the website traffic data as a network of websites, tied on the basis of their usage in the same countries. All analyses suggest that countries with high language similarities or those with shared borders had similar web use patterns. Surprisingly neither were countries particularly similar to the United States, and nor did the prevalence of English language speakers explain similarities between countries' web use.

We now briefly describe how this manuscript is organized. We first review the literature on media globalization and Internet geographies to motivate a set of hypotheses and questions for comparing web use patterns between countries. We then describe our data sources and elaborate on the method we used to compute a country-by-country matrix by comparing website traffic ranks. The third section reports our findings, based on network and cluster analyses, to detect communities of countries with similar web use, and also reports on the models identifying factors that predict these similarities. What follows are reports of additional analyses and the replication we conducted to ensure our findings are robust. Finally, we discuss how the study contributes to understanding the factors shaping usage of the World Wide Web on a global scale.

Theorizing global web use

The global circulation of media has been studied for decades. The debate has primarily been between theories of cultural imperialism and cultural proximity. Early scholarship feared cultural imperialism (Galtung, 1971; Schiller, 1969), predicting a global dominance of content originating in the United States and other western countries (Jayakar & Waterman, 2000). Later studies proposed the idea of cultural proximity, demonstrating that, with the growth of domestic media industries, people generally preferred content that focused on their region and was in their own language (Straubhaar, 1991). Yet, some content across media does have cross-cultural appeal, and there is evidence, at least in movie watching, of a global convergence toward American tastes (Fu & Govindaraju, 2010). Likewise, recent literature, which we review next, has pointed to both homogenization and diversification in patterns of global web use.

Mapping global Internet structure

The homogenization thesis stems from the World Systems Theory (Wallerstein, 2010), which posits that the world is organized according to a hierarchy of core nations and dependent, peripheral ones. For instance, many (peripheral) postcolonial countries still depend on their former colonizers (core). George Barnett's two decade–long, extensive research program showed that the global Internet reflects this structural hierarchy (e.g., Barnett & Park, 2014). Mapping interconnections between countries on the basis of hyperlinks between websites, bandwidth, and other supply-side infrastructures revealed the core-periphery organization of nations, as the World Systems Theory predicted. Local cultural factors do little to moderate these hierarchies. A recent study mapping global news flows across 70 countries also reiterated this core-periphery structure (Guo & Vargo, 2017).

We agree with recent work suggesting that technical connectivity should not be equated with online user activity, as each represents distinct geographies of the Internet (see Graham, 2013; Graham, De Sabbata, & Zook, 2015). For instance, even if news organizations in the global South link disproportionately to stories of *The New York Times*, as the World Systems Theory would predict, it does not necessarily mean that users of those news websites visit *The New York Times*. Relatedly, hyperlink analyses do not reveal the motivations of link providers (De Maeyer, 2012). An audience-centric perspective and mapping usage instead revealed that, even though the Internet has a centralized technical infrastructure, cultural factors largely drive people's web use (Taneja, 2017; Wu & Taneja, 2016).

Mapping global web use

Focusing on usage instead of technical infrastructures, recent studies have conceptualized the media as audience networks, with connections between outlets based on shared usage (Ksiazek, 2011; Webster & Ksiazek, 2012). Analyzing shared traffic between the world's 1,000 most popular websites, Taneja and Webster (2016) found that language and geography are the most important factors in explaining shared website audiences. They also found that among websites with high language similarities, audience overlap was higher when websites focused on the same country. Using the same approach on traffic data from 2009 to 2013, Wu and Taneja (2016) found that, with the broadening and deepening of Internet access to include more countries and users in the global South, stronger regional cultures of web use emerged. Relatedly, Bail, Brown, and Wimmer (2019) analyzed longitudinal data on Google search trends, and found that "global cultural diffusion" is rather rare. In effect, evidence points to a gradual de-Americanization and decentralization in global web use.

Since these studies were conducted, the global Internet user base has doubled to about four billion (Internet World Stats, 2018). This explosive growth in Internet users, particularly in parts of Asia, Africa, and the Middle East, may have contributed to even stronger online regional cultures, each with distinct use patterns shaped by local online landscapes (Burrell, 2012). Other studies focusing on specific domains also echo the importance of language, geography, and overall cultural similarity in web use. For instance, Internet users tend to email other culturally similar users (State, Park, Weber, & Macy, 2015) and form Twitter ties based on geographical proximity (Takhteyev, Gruzd, & Wellman, 2012); Wikipedia editions also vary based on language (Graham, Hogan, Straumann, & Medhat, 2012; Hecht & Gergle, 2010).

Convergence in global web use

Even though web use is largely culturally aligned, content does diffuse across linguistic and national boundaries. Video-sharing websites that afford user-generated content especially play a significant role in this spread. For example, YouTube has facilitated the popularity of Korean pop music ("K-pop," especially Gangnam Style), not just in East Asia but also in South America and across the world (Baek, 2014). Sites such as YouTube play a role in bridging otherwise dissimilar patterns of cultural consumption (Platt, Bhargava, & Zuckerman, 2015). Cultural proximity, based on individual cultural values rather than linguistic or geographic proximity, explains these patterns (Park, Park, Baek, & Macy, 2017). Content with cross-border appeal also gets discussed on social media, such as Twitter, beyond their countries of origin.

In recent decades, English has grown to become the primary global bridging language. Its influence in both offline and online settings has eclipsed all other languages. Besides being a preferred language for book translations, English is the main second language for Twitter users, Wikipedia editors, and bloggers who contribute in more than one language (Hale, 2012; Ronen et al., 2014). Thus, the influential role of English, along with the disproportionate popularity of a handful of social networking sites, might have contributed to homogenizing patterns of global web use.

Mapping user-centric Internet geographies

There appear to be contradictory claims about the patterns of global web use. One set of studies, focused on usage at the level of websites, suggested that language and geographic factors potentially differentiate web use across countries. Other studies, focused on specific web domains, such as Twitter and YouTube, point to a growing trend of convergence in global cultural consumption. However, we argue that neither of these studies can fully capture the extent to which countries are (dis)similar in their web use patterns. First, the study of any single domain, even if the domain is among the most popular online entities, with a huge global user base, is not representative of overall web use patterns. Second, although the top 1,000 websites capture 99% of all web traffic (Wu & Ackland, 2014), such a sample is now insufficient to capture the web use of smaller countries, as most global Internet traffic tends to come from countries with large online populations. Take a small country, Singapore, where English predominates, as an example. The website of *The Straits Times,* its nationally focused newspaper, may not make it to a list of the 1,000 or even 2,000 most popular websites in the world. Having excluded these local websites, one may assess web use patterns in Singapore, based on Singaporeans' usage of globally popular web domains, to be more similar than they actually are to other, larger nations where English dominates. Therefore, sampling a list of the globally most popular websites could suppress overall heterogeneity in web use, especially if small countries rely on their own local websites.

Given these considerations, we suggest that, to investigate global web use similarity, one needs to sample a comparable number of popular websites from as many countries as possible for which web traffic is separately available (see also Barnett & Park, 2014). Following this approach, the first goal of this study was to empirically assess the extent to which global web use is similar between countries. This translated into a rather broad research question:

RQ1: How similar are countries in their web use patterns?

What factors might explain the similarities observed in RQ1? First, despite the global availability of most websites, users prefer to consume content in their own languages (de Sola Pool, 1977) and content that resonates with the specificities of their own regions (La Pastina & Straubhaar, 2005; Taneja & Webster, 2016). This leads to the following hypothesis:

H1: Language similarity between countries positively correlates with similarity in their web use patterns.

However, since English is a global bridging language, a pair of countries with different primary languages, but both with a high proportion of English speakers, could be quite similar in their web use patterns. This led to the following research question:

RQ2: Does similarity in the proportion of English speakers between countries explain similarity in their web use patterns?

Given the correlations between online and offline networks that prior studies found, geographic proximity could play a role in determining web use similarity between countries. However, physical distances are not accurate indicators of proximity between countries. For instance, the distance between the United States and the United Kingdom (3,461 miles) is similar to that between India and Austria (3,462 miles) or Hong Kong and Oman (3,513 miles). These three country pairs are otherwise quite different from one another. Therefore, the role of distance, operationalized as such, in explaining web use similarity is questionable. In contrast, the effect of geographic proximity would be particularly strong if a pair of countries share a border, as interactions between people in adjacent countries are believed to be more frequent. Thus, we posited a second hypothesis:

> H2: A shared border between countries explains similarity in their web use patterns.

Third, the relative size of the Internet market could influence web use similarities between countries (Park et al., 2017). For nations with smaller domestic Internet markets, the local Internet industry may not be well developed; hence, the use of the Internet would be more likely to resemble patterns in countries with larger markets. Therefore, two countries with large and developed Internet markets will likely show more dissimilar web usage, as compared to a pair of countries with a significant difference in their Internet market sizes. We thus put forward a third question:

> RQ3: How do the relative sizes and Internet penetration rates of countries explain web use similarity between those countries?

When investigating this factor, we paid special attention to two markets: the United States and China. With its superior economic prowess, the United States has enjoyed a long-standing advantage in exporting cultural products globally. Coupled with the global influence of the English language (Ronen et al., 2014) and the dominance of U.S.-based digital conglomerates in the Internet market (Statista, 2018), we theorized that web use patterns of many countries would be disproportionately similar to those of the United States. This would be especially true if theories such as the World Systems Theory and cultural imperialism still explain patterns of global web use. As already noted, recent studies have shown lower purchase for these theories. In favor of ideas of cultural proximity, the extent to which global web usage remains similar to the United States remains a matter worthy of empirical inquiry.

For China, we speculated an effect opposite to that of the United States. China has more Internet users than any other country. With its distinct language, unique historical culture, and a thriving domestic market for tech companies, China has developed a distinct web use ecosystem, where Chinese users mainly rely on domestic websites. This is facilitated in part by the Chinese state's access blockage of several foreign websites through the "Great Firewall" (Taneja & Wu, 2014). For these reasons, we anticipated that Internet use in China would be quite dissimilar from most other countries. However, we also wanted to discern whether the low similarities we expected to find for China could be explained by factors beyond China's large Internet penetration, language, and shared borders. Our expectations about the United States and China raised the following question.

RQ4: How does one of the countries being the United States or China impact web use similarity between countries?

Method

Our empirical strategy was to calculate similarities based on shared web use between all possible pairs of countries. We computed a country-by-country (174 x 174) similarity matrix based on Alexa traffic data, which served as the dependent variable. We first descriptively analyzed this matrix to detect clusters of countries with similar web use. Subsequently, we used regressions based on quadratic assignment procedures to identify the significant factors that explained these pairwise similarities.

Web use data

We obtained the ranked lists of the 100 most-visited websites for 174 different countries, which were available from Alexa, a web analytics company, for July 2018. Alexa tracks the online activity of millions of users who have installed Alexa's toolbar browser extensions (Alexa, 2018). Based on its panel's web browsing activity, Alexa estimates the top 100 websites for each country, according to the relative number of monthly unique users. A country's list is based on user browsing behavior, but not a site's country of origin. For example, a Ghana-based website can figure in the list for the United States, provided the site is among the most popular websites in the States. Additionally, Alexa reports traffic separately for each regional domain of websites such as Google and eBay (e.g., www.google.com vs www.google.de or ebay.fr vs ebay.ca). To check reliability, we compared Alexa's U.S. list with ComScore's 100 most popular sites in the United States in July 2018, and found a high correlation ($\rho = .69$; see Supporting Information Table S2).

Web use similarity via rank biased overlap

Our focal variable was web use similarity between countries, which, to the best of our knowledge, only one prior study (Barnett & Park, 2014) has focused on. However, Barnett and Park (2014) quantified website similarity between countries merely on the basis of the presence or absence of the same sites in each country's list of top 100 sites. Accordingly, two countries with one common website had the same similarity score even if that website ranked first for both countries or ranked first for one country and 100th for the other country. Such a dichotomization overlooked the rank order and significantly decreased the variance accounted for in the resulting similarity scores.

Website traffic generally follows a power law distribution (Huberman, 2001; Webster & Lin, 2002). A pair of countries that has multiple common websites at higher-ranked positions ought to be given a higher similarity score than a pair of countries that only shares common sites at lower ranks. Similarity scores based on traditional correlational measures, such as Kendall's τ and Spearman's ρ, also do not capture variations in ranking positions. Further, these measures require identically sized, finite lists, and weigh all items equivalently, regardless of their ranks.

To improve the similarity measure based on ranks, we used a relatively novel measure: the rank biased overlap (RBO), developed by Webber, Moffat, and Zobel (2010). RBO handles non-conjoint lists, counts the average fraction of common items between them, and weights similarities observed at higher ranks more heavily than those observed at lower ones. We used the following formula to compute web use similarities between countries:

$$\mathrm{RBO} = (1-p)\sum_{k=1}^{\infty} p^{k-1}\frac{|A_{1:k}\cap B_{1;k}|}{k}$$

$A_{1:k}$ and $B_{1:k}$ denote the top k visited websites in Country A and Country B, respectively. Here, p is a weighting parameter, which can take a value between 0 and 1, and the influence of the weight decreases geometrically along the ranked list (consistent with positively skewed website traffic distributions), but never reaches zero. This minimizes the impact of choosing an arbitrary threshold: our choice of the top 100 sites was of an arbitrary threshold dictated by data availability.

Since Alexa does not provide unique web user counts, to determine an appropriate value of p, we examined the distribution of unique user reaches of the top 100 sites for the United States from ComScore. Based on its power law–like distribution, we set p to .973, which allocated a share of 50% of all traffic to the first 10 sites. RBO similarity scores range from 0 (disjointed) to 1 (identical). We then used these pairwise similarity data to construct a symmetric country-by-country (174 x 174) similarity matrix.

Language similarity

Recent studies (e.g., Akaliyski, 2017; Kohl & Brouwer, 2014) have used a binary classification to denote whether a pair of countries shared a common native/official language or whether their languages were under the same language family, as proxies for language similarity. However, binary measures are unable to capture linguistic heterogeneity and underestimate the impact of language on web use. Thus, we collected the estimated number of speakers of 460 different languages spoken among the 174 countries, primarily from Ethnologue, cross-checked with other updated sources, such as Central Intelligence Agency (CIA) World Factbook. Ethnologue is a comprehensive catalog of the known spoken languages globally (Lewis, 2009) and provides estimated numbers of total language speakers for all languages spoken by more than 1% of a country's population (see also Ronen et al., 2014, for a study using this source).

We calculated the proportion of speakers for each language in each country. For each language, we included people who spoke it as either a first or second language. We quantified the language similarity of each pair of countries using cosine similarity:

$$\cos(a, b) = \frac{\sum_{i=1}^{m} a_i b_i}{\sqrt{\sum_{i=1}^{m} a_i}\sqrt{\sum_{i=1}^{m} b_i}}$$

with a indicating Country A, b indicating Country B, and m representing the 460 languages. Each country (a, b, etc.) is a vector, with languages $\{a_1, a_2, a_3 \ldots\}$, $\{b_1, b_2, b_3 \ldots\}$ as components, weighted by the proportion of speakers.

English prevalence

To account for both the proportion and the actual number of English speakers in each country, we calculated the relative English prevalence between a pair of countries as the absolute difference between the multiple of a country's proportion of the English-speaking population and the number of English speakers. We used the formula: $PE_A E_A - PE_B E_B$, where PE_A and PE_B represent the proportions of English speakers in Country A and Country B, respectively, and E_A and E_B represent the numbers of English speakers in Country A and Country B, respectively.

Relative size of the Internet market

To account for the effect sizes for both penetration rate and the actual population of Internet users, we calculated, using data from Internet World Stats (2018), the relative size of the Internet market between each pair of countries as the absolute difference between the multiple of each country's Internet penetration rate and Internet population. We used the formula: $PI_A I_A - PI_B I_B$, where PI_A and PI_B represent the Internet penetration rates in Country A and Country B, respectively, and I_A and I_B represent the numbers of Internet users in Country A and Country B, respectively.

Geographic distance

We obtained the pairwise distance between countries from the French Research Centre in International Economics' distances and geographical database, using great-circle distance: the distance between latitudes and longitudes of countries' most populous cities.

We created a symmetric country-by-country matrix for all of these four variables, where each cell represented the language similarity, relative prevalence of English speakers, Internet market, and relative distance between the countries in rows and columns. To compare coefficients, we standardized the values in all matrices to their z-scores.

Shared borders

We created a fifth matrix, in which each cell with a value of 1 indicated a shared land border between a pair of countries, and 0 indicated no shared border.

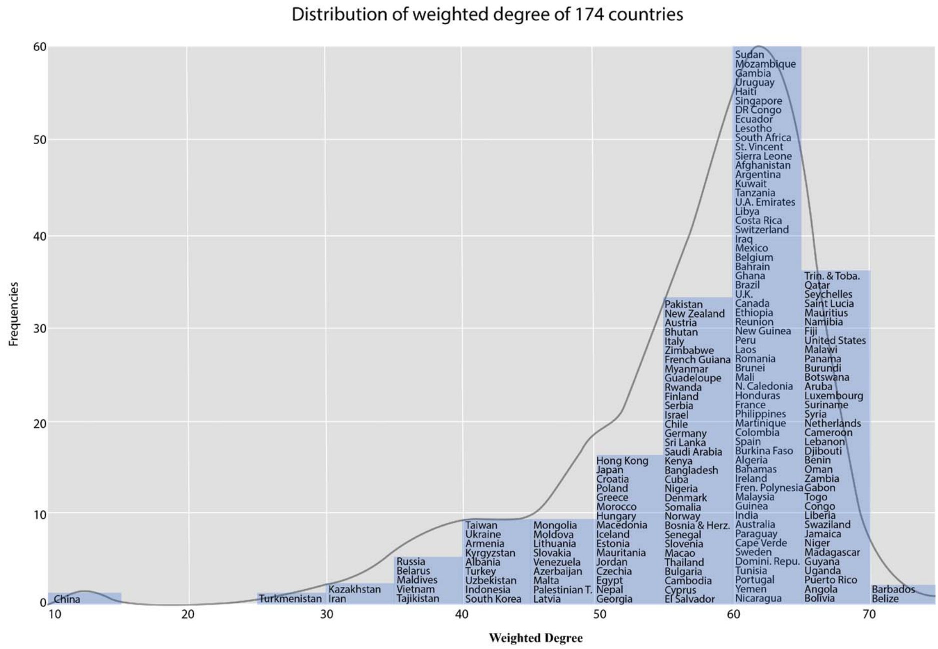

Figure 1 Weighted degree for each country ($N = 174$, Gini coefficient = .08, and the coefficient of variance = .15).

The effect of the United States and China

Likewise, we created two additional matrices of countries, in which a value of 1 indicated that one of the countries was the United States in the first matrix and that one was China in the second matrix. Henceforth, we term these matrices as "U.S. effect" and "China effect."

Analysis and findings

Analyzing the rank-based overlap matrix

To answer RQ1, we analyzed the country-by-country similarities. Since this is a completely connected network, the higher a country's weighted degrees, the more similar it is to the rest of the countries. Figure 1 presents the weighted degree distributions ($M = 58.08$, $SD = 9.04$, $Median = 60.83$). The Gini coefficient was 0.08 and the coefficient of variation was 0.15, suggesting a low centralization. In other words, no particular country stood out as exceedingly similar to most others in terms of website usage. Countries in the Caribbean had the highest weighted degrees, beginning with Barbados (72.2) and followed by Belize (70.76) and by Trinidad and Tobago (68.98). The United States ranked tenth (66.97). Conversely, Iran (30.29) and Turkmenistan (29.25) were among the lower rank, along with China (13.82), scoring the lowest in weighted degree among all 174 countries.

Country clusters based on web use similarity

Next, we performed an agglomerative hierarchical cluster analysis on the similarity matrix and generated a dendrogram (Supporting Information Figure S1) to identify clusters of countries with similar web use patterns. Hierarchical clustering first gathers the most similar observation pairs, then progressively aggregates other observations according to their similarities until all observations unite into one group. However, as is often the case with a cluster analysis, setting a cut-off point to separate cohesive subgroups required qualitative judgment. Thus, we further created 29 choropleth world maps (from 2 to 30 clusters) to interpret the relationship between clusters and spatial patterns. Here, we describe the visual representations of solutions with 5 clusters and 18 clusters to illustrate the strong spatial associations we found (Figure 2).

When global web use manifested as 5 clusters, the groupings were rather broad. The largest cluster (green in color) was composed by major countries from South and Southeast Asia (e.g., India, Myanmar, Malaysia, Laos, and Singapore), the Middle East, Africa, and Western Europe (e.g., Belgium and Luxembourg). It also included the majority of Caribbean and Latin American countries (e.g., Mexico and Brazil), as well as the United States. For the 18-cluster solution, this large cluster was split into seven smaller groups. The Caribbean countries and South Africa became their own clusters. Latin American countries remained as a cluster. The United States, Singapore, and a few Western European countries (e.g., Belgium and Luxembourg) clustered into one group. Likewise, regions of France (i.e., France, Réunion, French Guiana, Guadeloupe, and Martinique) formed their own cluster. There was one cluster among South (e.g., India and Bangladesh) and South East Asian countries (e.g., Malaysia). The sixth cluster consisted of Central and Southern African countries (e.g., Angola and Zambia), and the seventh cluster was formed among countries in West Africa. Similarly, the other four large clusters were split into smaller groups of geographically contiguous and/or linguistically similar regions. Figure 2 lists each of the large clusters in the 5-cluster solution and their corresponding, smaller clusters in the 18-cluster solution.

In summary, the clustering hinted at some role of language, as well as of shared borders, in global web use similarity. However, to isolate the effects of each of these factors, we conducted further confirmatory analyses.

Quadratic assignment procedure

To further the exploratory inferences presented so far, we used quadratic assignment procedure (QAP) to evaluate the power of various predictors in explaining pairwise similarities between countries' web use. Since our dependent and independent variables manifested as relational matrices, QAP regression was more appropriate than Original Least Squares (OLS). Unlike OLS, QAP generates a non-parametric, null distribution using row-column permutations of input matrices, retaining their dimensionality and degree distributions.

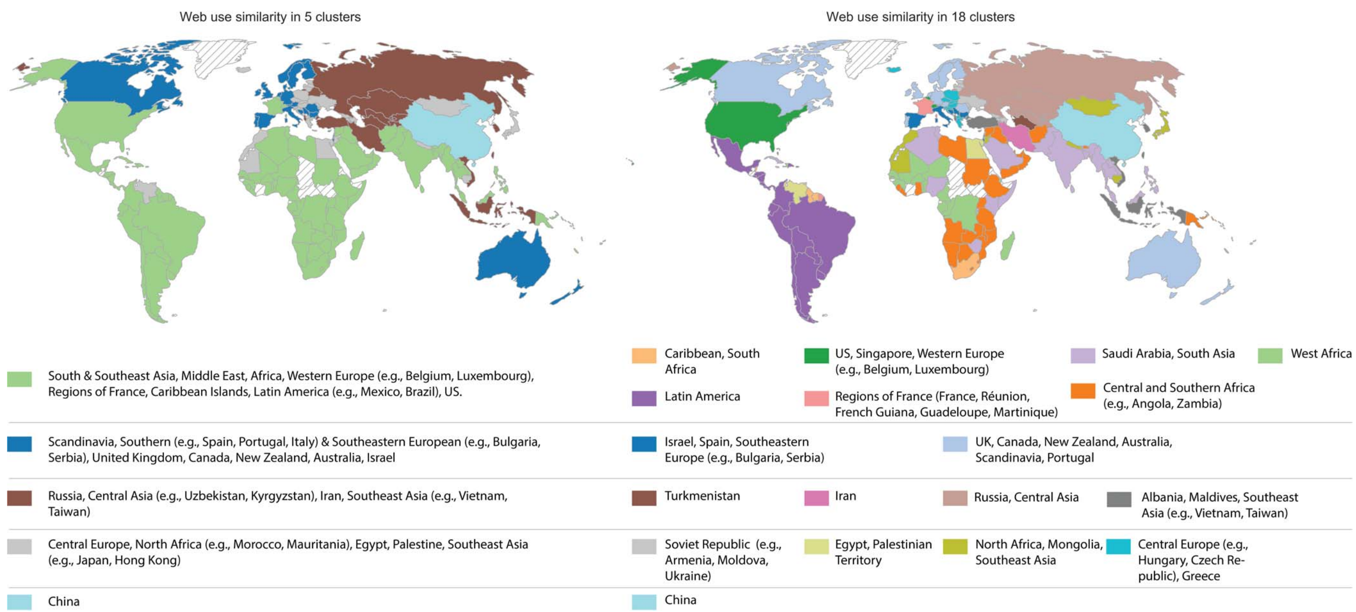

Figure 2 Choropleth world maps: web use similarity in 5 clusters and 18 clusters. We did not deem these two cluster solutions as the most acceptable or the most appropriate. The aim of showing these two solutions is to illustrate the hierarchy of clusters we found, which followed a geo-linguistic pattern. These two solutions are chosen as illustrations of this pattern.

Table 1 Quadratic Assignment Procedure Correlations Across Variable Matrices

	1	2	3	4	5	6
1. Web use similarity	...	...	...	...	...	...
2. Language composition	.38***	...	...	...	...	...
3. Sharing border	.11***	.14***	...	...	...	...
4. Internet market size	−.03**	−.02	−.05***	...	...	...
5. US effect	.06	.08**	−.005	.42**	...	...
6. China effect	−.31**	−.05	.06**	.65**	−.006	...
7. English prevalence	.06	.10**	−.001	.45**	.57**	−.01

Note: $^{**}p < .01$, $^{***}p < .001$. $N = 174$.

We conducted QAP using the *statnet* family of packages in the R statistics language. Table 1 shows that most variables were only slightly correlated with each other, allaying serious concerns about multicollinearity. However, some moderate correlations required more attention. For instance, English prevalence was moderately correlated ($r = .57$) with the U.S. effect, as the United States had a high overlap in English language speakers with all other countries. Likewise, Internet market size was moderately correlated with both the U.S. effect ($r = .42$) and English prevalence ($r = .45$). We addressed these correlations by building a series of models in blocks, first including only one of the two variables from these correlated pairs.

Table 2 reports the results of the QAP regressions. In the first model, similarity in language composition significantly predicted similarity in global web use ($b = .03$; $p < .001$), as did shared borders ($b = .05$; $p < .001$) and difference in Internet market size ($b = -.02$; $p = .001$). We additionally included geographical distance (Supporting Information Table S3), which was not statistically significant ($b = .002$; $p = .48$) and did not add to the predictive power. Specifically, the results suggested that, on average, a country pair which either had a similar language composition or a shared border would have similar web use patterns, supporting H1 and H2. In contrast, the negative coefficient of Internet penetration indicated that an increase in the gap between the Internet penetration rates of a country pair was associated with less similar web use patterns, answering RQ3.

We added the indicator matrices for the U.S. and China effects in the second model. None of these were significant (United States: $b = .10$, $p = .09$; China: $b = -.11$, $p = .17$), answering RQ4. In order to test RQ2, we built a third model, where we added English prevalence. Its effect was insignificant ($b = .005$; $p = .75$), and it impacted neither the coefficients nor the standard errors of the U.S. effect matrix, alleviating potentially concerning effects of multicollinearity. More notably, adding indicator matrices for either China or the United States as one of the country pairs did not change the magnitude or significance of the language and border effects, and nor did

Table 2 Quadratic Assignment Procedure Regressions for Web Use Similarity Across Countries

Variables	b		
	Block 1	Block 2	Block 3
Intercept	.34***	.34***	.34***
Language composition	.03***	.03***	.03***
Shared border	.05***	.05***	.05***
Internet market size	−.02**	−.02*	−.02*
U.S. effect	...	.10	.07
China effect	...	−.11	−.11
English prevalence	...	...	.005
R^2	.22***	.25***	.25***
Adjusted R^2	.22***	.25***	.25***

Notes: We used 1000 permutations for estimating standard errors.
The coefficients presented for the continuous variables—language composition, Internet market size, and English prevalence—are standardized, which allows for the comparison of their magnitudes to infer effect sizes. Other variables are indicator variables (0/1), the coefficients of which should be compared with one another. $N = 174$.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

adding the English prevalence matrix as a predictor. The QAP regression explained 25% of the total variance.

Robustness check

We conducted four additional analyses to assess the robustness of the findings. First, we repeated the QAP regressions without China. Second, we repeated the analysis, removing the three globally most popular websites. Third, we replicated the analysis using Alexa traffic data from a more recent month. Fourth, we projected Alexa traffic data to its other network projection—a website-by-website similarity matrix—to assess whether our country-by-country projection resulted in a loss of valuable information.

China as an outlier

Since the network analysis and the clustering both pointed to China as a dissimilar outlier, we replicated the QAP regressions after removing China altogether from the data set. As expected, it reduced the predictive power of the model (*R*-square dropped to 16.5%; full model in Supporting Information Table S4), but the effects of language, border, and Internet market size remained significant and their directions did not change.

Removing websites by tech giants

Since the literature suggested a convergence in global web use because of the worldwide popularity of a handful of sites, we replicated our analysis after removing Google.com, YouTube.com, and Facebook.com. As expected, the overall web use similarity, measured in weighted degrees, went down ($M = 33.31$, $SD = 8.28$, $Median = 35.72$) and there was more variability in the data, suggesting the somewhat homogenizing role of these websites. The QAP regressions retained all previously significant factors. Supporting Information S5 and S6 have more details.

Replication

We collected the Alexa traffic data for a second month (February 2019) and constructed a country-by-country RBO similarity matrix for 171 countries. We had to exclude Aruba and Saint Lucia, as well as Saint Vincent and The Grenadines, as the Alexa data for February 2019 did not report on these three countries. The two matrices had a very high correlation ($r = .94$). Keeping the same predictor matrices (assuming these factors remained relatively stable over a short span of 6 months), we replicated the QAP regressions. The effects remained largely the same, with language similarity explaining the most variation, and shared borders and Internet market size being the other significant predictors (see Supporting Information Table S7). This additional analysis helped establish the conceptual validity of our model, allaying concerns that these effects were artifacts of a specific month of data collection.

Website similarity matrix

The Alexa web traffic data can be treated as a two-mode (affiliation) network, with the 174 countries and 6,252 unique websites as the two sets of nodes. Guided by the research objective, our analysis focused on the country-to-country one-mode projection. However, projecting to one mode might lead to a loss in valuable structural information. Yet, due to their interpretive simplicity, one-mode projections are more attractive than the direct analysis of the two bipartite matrices (Everett & Borgatti, 2013; Melamed, 2014). To assess whether we lost information in projecting our data to a one-mode network of countries, we followed Everett & Borgatti's (2013) dual projection approach and also analyzed the website-by-website projection.

To construct the website-by-website network, we followed the intuition that the more frequently two websites appeared together in any two countries' lists, the more similar they were. However, instead of simple overlap counts, we factored both websites' co-occurrences and their relative ranks in each country's top websites list. For each pair of websites, we thus divided the number of countries in which both websites appeared by the sum of the geometric means of their rankings in each of these countries, as per the following equation, where A and B are websites and i refers to the country in which both appear:

$$similarity = \frac{no\ of\ countries\ in\ which\ both\ sites\ appear\ (n)}{\sum_n \sqrt{RankA_i\ RankB_i}}$$

The numerator of this expression captures the co-occurrence, whereas the denominator adjusts for the similarity in rankings. For example, Google.com and Youtube.com, which appeared together in lists of all 174 countries, had a similarity of 88.7. In contrast, Google.com and Instagram.com, which also coexisted in 174 countries' lists, had a lower similarity, of 31.9, as Instgram.com had a lower rank in most countries compared to Youtube.com.

Just as we did for the country-by-country matrix, we also conducted a cluster analysis on the website-by-website matrix, using a community detection algorithm that is appropriate for large weighted but undirected networks (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008). This revealed 17 clusters (modularity = 0.256). In general, websites with the same language, especially when their content focused on countries that share a border, clustered together. We profiled these clusters in Table 3, which provides a sampling of constituent websites and our interpretations. To be clear, this analysis did not formally test our hypotheses, but helped corroborate our analysis of web use similarity between countries. Both country-by-country and website-by-website projections led to similar inferences. Supporting Information Figure S8 visualizes the bipartite graph, showing both sets of nodes: the websites and the countries.

Discussion

The number of Internet users has increased 10-fold in the last 15 years, with almost every country's Internet penetration increasing. Relatedly, studies have shown that web use tends to be along the lines of languages and geographies. At the same time, with the global dominance of platforms which capitalize on user-generated content, some web use does transcend national and linguistic boundaries. Given both these trends, how does geography shape global web use? Our study analyzed web use patterns from 174 countries, utilizing global web traffic data from Alexa, and found that having a similar language composition and shared borders best explained similarities in web use between countries. The similarity to the United States and the prevalence of the English language had surprisingly little influence. Finally, China, with its large market and blockage to foreign content, isn't particularly isolated. In the remainder of this section, we reflect more on these results, with an eye on the study's theoretical and methodological contributions.

Overall, our study shows that, driven by cultural proximity, global web use manifests as a mosaic of regional cultures, composed of geographically adjacent and linguistically similar countries. On the contrary, we do not find any significant effect for one of the countries being the United States, allaying fears of cultural imperialism or homogenization of global web use patterns. Investigating web use beyond a handful of popular websites reveals much more heterogeneous patterns along regional lines than studying only popular domains would have suggested. These findings are consistent with a recent analysis of Google trends, which concluded that cultural diffusion online is seldom led by the United States (Bail et al., 2019). Our

Table 3 Community Detection: 17 Website Clusters

Clusters	Number of sites, $N = 6{,}562$		Google sites (n), $n = 159$	Example Countries
0	84	1.28%	Google.co.ve (1 Google site)	Venezuela
1	498	7.59%	Google.com.eg; Google.ae; Google.co.ma . . . (9)	Egypt, Iran, United Arab Emirates, Morocco, Lebanon, Jordan, Palestine, Seychelles, and Cuba
2	2,056	31.33%	Google.es; Google.com.mx; Google.com.ng, etc. (68)	Spain, Ethiopia, Kenya, Angola, Senga, Madagascar, Saudi Arabia, South Africa, Mexico, Nigeria, etc.
3	2,319	35.34%	Google.ru; Google.co.uk; Google.com.ua, etc. (53)	Russia, Ukraine, United Kingdom, Germany, Iceland, Finland, Netherlands, Australia, United States, Thailand, Singapore, etc.
4	205	3.12%	Google.fr; Google.gp (2)	France, French Guiana, *Réunion*
5	361	5.50%	Google.com.hk; Google.com.tw; Google.cn, etc. (6)	Taiwan, Hong Kong, China, Korea, Japan, and Malaysia
6	125	1.90%	Google.com.br; Google.pt; Google.cv (3)	Brazil, Portugal, and Cape Verde
7	52	0.79%	Google.al (1)	Albania
8	47	0.72%	Google.so (1)	Somalia
9	73	1.11%	Google.co.id (1)	Indonesia
10	75	1.14%	Google.com.tr (1)	Turkey
11	119	1.81%	Google.cz; Google.sk (2)	Czech Republic and Slovakia
12	69	1.05%	Google.hu (1)	Hungary
13	58	0.88%	Google.bg (1)	Bulgaria
14	124	1.89%	Google.com.vn; Google.si (2)	Vietnam and Slovenia
15	81	1.23%	Google.it; Google.com.mt (2)	Italy and Malta
16	216	3.29%	Google.co.in; Google.com.bd; Google.lk, etc. (5)	India, Sri Lanka, Maldives, Bhutan, and Bangladesh

Notes: Modularity = 0.256.

findings question the theoretical purchase of frameworks such as World Systems Theory, which emphasize technical connectivity to theorize the global Internet structure, and which recent studies continue to employ to study global media flows (e.g., Guo & Vargo, 2017).

Instead of analyzing shared traffic between websites, as recent studies have done, we focused on shared usage between countries. Had we sampled the world's 1,000 or even 2,000 most popular sites, as prior studies did, we would have most definitely excluded usage from small countries, especially those with distinct use patterns. We took care of this potential issue by sampling the 100 most popular sites from each country. This further allowed us to factor how smaller countries contribute to major online geo-linguistic formations. For example, we found that Singapore's web usage, presumably due to its use of English as an official language, is quite similar to that of the United States. We also found unique patterns of similarity in web use between many small, French-speaking Caribbean nations and Belgium and Luxembourg, two Francophone Western European countries.

Language has long been a crucial factor in determining media use. Studies have shown that websites having content in the same language and a similar geographic focus have high audience overlaps. Our study raised the level of analysis from websites to countries and examined what makes two countries' web use similar. We found that countries cluster on the basis of shared usage of websites with other adjacent countries, as well as with those that speak the same languages. For example, web use in Singapore, where English is a majority language, is more similar to the United States than to other East Asian countries. Thus, consistent with prior studies, we found language to be a significant factor, but this level of analysis allowed us to understand how language similarity shapes web use beyond national territories.

We speculated, based on the recent Ronen et al. (2014) study, that the growth of English as a highly influential bridging language could weaken the overall role of linguistic similarity in explaining web use similarities between countries. However, we found that English has a minimal impact on countries' web use similarities. Divergent research designs might explain the differences in findings. Ronen et al. (2014) constructed two online global language networks, based on narrow slices of multilingual online usage that arguably skew elite: Tweets by Twitter users and edits on Wikipedia. Our study focused on a much larger sample—the 100 most-frequently visited websites for each of the 174 countries—arguably capturing the usage of average Internet users in different countries, for whom English may not be their preferred language. Future studies may wish to reconsider relying on singular domains, such as Twitter or YouTube.

Shared physical borders generally signal frequent interactions between countries: greater movement of people and goods. Regardless of overall linguistic similarities, people in border areas tend to know the languages of both countries. Taneja and Webster (2016) found that focusing on the same geography further enhanced audience overlaps between same-language websites. Extending their work, our study found that, even after controlling for language similarities, adjacent countries had significant overlaps in their web use patterns. Thus, belonging to the same region (Griswold, 2008; Wu & Taneja, 2016) is an independently significant factor in shaping similar patterns of web use, and perhaps other cultural consumption.

We found that the difference between Internet market sizes was negatively related to web use similarities between countries. Consider three countries: the United States, Kenya, and Myanmar. The United States has an Internet penetration rate of 89% and 287 million Internet users; Kenya has an 85% penetration rate and 43 million users; and Myanmar has a 33% penetration rate and 28 million users. Our model indicates that owing to this difference in Internet penetration, *ceteris paribus*, web use in Myanmar would be more similar to the United States than web use in Kenya. The latter's higher Internet penetration would foster a self-sufficient, domestic online ecosystem, but users in Myanmar would more likely rely on foreign websites. Our findings empirically support the argument that increasing Internet penetration, especially in large countries, fosters more local content, in effect creating thick online regional cultures (Wu & Taneja, 2016).

Finally, we found China to be quite dissimilar from all other countries in its web use patterns. At first glance, it appeared that the comprehensive Great Firewall could be responsible. However, somewhat surprisingly, when we controlled for language, shared borders, and differences in Internet market sizes, one of the countries being China—a reasonable proxy for capturing the effects of its comprehensive blockage—did not significantly explain (dis)similarities between countries' web use patterns. In other words, all other factors in which China is also unique explain its low similarities with most other countries. For example, China is the largest Internet market in terms of user numbers and has a language that is only spoken in a few other neighboring countries. Thus, our findings corroborate another recent study that showed China's access blockage has indirectly habituated its Internet user population to largely visit local websites, which theories of cultural proximity would predict anyway (Taneja & Wu, 2014).

As observational data become more common in communication research, at least two techniques we used might be useful for future studies. First, ranked lists, such as the ones we used for website traffic, are generally accessible, observational data, often valuable for international and comparative work. Instead of simply counting overlaps in different ranked lists, we utilized RBO. Even though we did not have the unique user numbers for websites on which these rankings were based, both prior knowledge and other related data allowed us to assume that web traffic follows a power law–like distribution. RBO thus allowed us to draw on this prior knowledge and assign higher similarity scores to common objects ranked at higher-ranked positions. Such a situation, where researchers have reliable rank order data and some knowledge of the distribution, but not actual numbers, is not uncommon in social science research. We demonstrated that methods such as RBO can work well in such situations.

Second, we used cosine similarity to calculate language similarities between countries. This allowed us to factor the varying proportions of speakers of all major languages spoken in each country. It is a significant improvement, compared to measures based on the binary presence or absence of languages, such as the Jaccard coefficient (see Taneja & Webster, 2016).

No design, including ours, is perfect. Alexa reports the rankings of top sites by countries and does not provide raw user numbers. This is most likely because unique user numbers for websites are hard to estimate precisely from Alexa's traffic data. Although the data are based on millions of Internet users from around the world, the panel is self-selected. Rankings deduced from such data, although less granular, are more likely to be reliable. In the absence of user numbers, we approximated their distributions based on prior knowledge. Yet such approximations do not factor any uneven distances between ranks and assume that the traffic in each country follows a similar distribution.

We had web traffic rankings available at the level of web domains. Most companies, such as Google, Amazon, eBay, Yahoo, and Microsoft, have country specific domains, but some, such as Wikipedia.org, are available worldwide as one domain, with different language subdomains. Thus, a user in Germany accesses de.wikipedia.org and a user in the United Kingdom accesses en.wikipedia.org. For a handful of such websites, our data set did not allow us to discern such differences in usage. Thus, we might have overestimated web use similarities between some countries, which, in turn, would mean we underestimated the role of language similarity.

Alexa web traffic data does not capture usage via mobile apps, which has now surpassed web use via personal computers (StatCounter, 2016). How might have this affected our findings? On one hand, mobile devices are responsible for deepening Internet penetration beyond educated elites in much of the global South. Based on this, one would expect mobile-based consumption to be even more regionally specific, as it is driven by language and geography. On the other hand, globally popular apps, such as YouTube and Google search (often preinstalled on phones), would make consumption patterns across countries even more similar.

Arguably, we have ignored other factors that also potentially contribute to web use similarity between countries. One of these is the role of inter-country migration (e.g., Rauch & Trindade, 2002). Besides shared borders and geographic distances, the connections between global transportation networks and web use similarities might stand out (Takhteyev et al., 2012). Future studies could also include connections between countries based on myriad physical and technical layers of infrastructure, including locations of data centers or undersea cables, in formally modeling patterns of global web use.

The World Wide Web turned 30 years old in 2019, with half the world online. Yet it is far from being a global platform with a universal language, as early visionaries had imagined. Both political and Silicon Valley elites continue to suggest that the growth of the Internet is making distances, languages, and geographies somewhat irrelevant. Studies like ours offer an empirical reality check against such normatively optimistic prescriptions. The technical architecture of the web does render it as a frictionless platform for global conversations. In reality, it appears that, on average, web use is blossoming along traditional lines, defined by languages and geographies.

Supporting Information

Additional Supporting Information may be found in the online version of this article. Please note: Oxford University Press is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Acknowledgments

Y. M. M. N. led the data collection and preparation. H. T. led the conceptualization and research design.

References

Akaliyski, P. (2017). Sources of societal value similarities across Europe: Evidence from dyadic models. *Comparative Sociology,* 16, 447–470. doi:10.1163/15691330-12341432

Alexa. (2018). *What exactly is the Alexa traffic panel*? Retrieved from https://support.alexa.com/hc/en-us/sections/200077030-Site-Overview

Baek, Y. M. (2014). Relationship between cultural distance and cross-cultural music video consumption on YouTube. *Social Science Computer Review,* 33(6), 730–748. doi:10.1177/0894439314562184

Bail, C. A., Brown, T. W., Wimmer, A. (2019). Prestige, proximity, and prejudice: How Google search terms diffuse across the world. *American Journal of Sociology,* 124(5), 1496–1548. doi:10.1086/702007

Barnett, G. A., & Park, H. W. (2014). Examining the international Internet using multiple measures: New methods for measuring the communication base of globalized cyberspace. *Quality and Quantity,* 48, 563–575. doi:10.1007/s11135-012-9787-z

Blondel, V. D., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment,* 2008(10), P10008. doi:10.1088/1742-5468/2008/10/p10008

Burrell, J. (2012). Invisible users: Youth in the Internet cafes of urban Ghana. Cambridge, Mass: MIT Press

De Maeyer, J. (2012). The journalistic hyperlink: Prescriptive discourses about linking in online news. *Journalism Practice,* 6, 692–701. doi:10.1080/17512786.2012.667273

Everett, M., & Borgatti, S. (2013). The dual-projection approach for two-mode networks. *Social Networks,* 35(2), 204–210. doi:10.1016/j.socnet.2012.05.004

Fu, W. F., & Govindaraju, A. (2010). Explaining global box-office tastes in Hollywood films: Homogenization of national audiences' movie selections. *Communication Research,* 37, 215–238. doi:10.1177/0093650209356396

Galtung, J. (1971). A structural theory of imperialism. *Journal of Peace Research,* 8(2), 81–117. doi:10.1177/002234337100800201

Graham, M. (2013). Geography/Internet: Ethereal alternate dimensions of cyberspace or grounded augmented realities? *The Geographical Journal,* 179(2), 177–182. doi:10.111/geoj.12009

Graham, M., De Sabbata, S., & Zook, M. A. (2015). Towards a study of information geographies: (Im)mutable augmentations and a mapping of the geographies of information. *Geo: Geography and Environment,* 2, 88–105. doi:10.1002/geo2.8Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014). Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Annals of the Association of American Geographers,* 104(4), 746–764. doi:10.1080/00045608.2014.910087

Griswold, W. (2008). In University of Chicago (Ed.), *Regionalism and the reading class*. Chicago, IL.

Guo, L., & Vargo, C. J. (2017). Global intermedia agenda setting: A big data analysis of international news flow. *Journal of Communication,* 67, 499–520. doi:10.1111/jcom.12311

Hale, S. A. (2012). Net increase? Cross-lingual linking in the blogosphere. *Journal of Computer-Mediated Communication,* 17(2), 135–151. doi:10.1111/j.1083-6101.2011.01568.x

Hecht, B., & Gergle, D. (2010). The tower of Babel meets web 2.0: User-generated content and its applications in a multilingual context. In *Proceedings of the 28th Annual CHI Conference on Human Factors in Computing Systems* (pp. 291–300). ACM. Atlanta, GA. doi:10.1145/1753326.1753370

Huberman, B. A. (2001). *The laws of the web* (). Cambridge, MA: MIT Press.

Stats, I. W. (2018). *Internet user in the world by region in June*, 30, 2018. Retrieved from. www.Internetworldstats.com/stats.htm.

Jayakar, K. P., & Waterman, D. (2000). The economics of American theatrical movie exports: An empirical analysis. *The Journal of Media Economics,* 13, 153–169. doi:10.1207/s15327736me1303_1

Kohl, T., & Brouwer, A. E. (2014). The development of trade blocs in an era of globalisation. *Environment and Planning A,* 46, 1535–1553. doi:10.1068/a46261

Ksiazek, T. B. (2011). A network analytic approach to understanding cross-platform audience behavior. *Journal of Media Economics,* 24, 237–251. doi:10.1080/08997764.2011.626985

La Pastina, A. C., & Straubhaar, J. D. (2005). Multiple proximities between television genres and audiences: The schism between telenovelas' global distribution and local consumption. *International Communication Gazette (*Leiden, Netherlands*),* 67, 271–288. doi:10.1177/0016549205052231

Lewis, M. (2009). *Ethnologue: Languages of the world* (16th ed.). Dallas, Retrieved from: www.ethnologue.com/.

Melamed, D. (2014). Community structures in bipartite networks: A dual-projection approach. *PLOS ONE,* 9(5), e97823. doi:10.1371/journal.pone.0097823

Park, M., Park, J., Baek, Y. M., & Macy, M. (2017). Cultural values and cross-cultural video consumption on YouTube. *PLOS One,* 12(5), 1–13. doi:10.1371/journal.pone.0177865

Platt, E. L., Bhargava, R., & Zuckerman, E. (2015). The international affiliation network of YouTube trends. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media* (pp. 318–326). ICWSM, Oxford, UK.

Rauch, J. E., & Trindade, V. (2002). Ethnic Chinese networks in international trade. *Review of Economics and Statistics,* 84, 116–130. doi:10.1162/003465302317331955

Ronen, S., Gonçalves, B., Hu, K. Z., Vespignani, A., Pinker, S., & Hidalgo, C. A. (2014). Links that speak: The global language network and its association with global fame. *Proceedings of the National Academy of Sciences,* 111, E5616–E5622. doi:10.1073/pnas.1410931111

de Sola Pool, I. (1977). The changing flow of television. *Journal of Communication,* 27, 139–149. doi:10.1111/j.1460-2466.1977.tb01839.x

StatCounter. (2016, November). *Mobile and tablet Internet usage exceeds desktop for first time worldwide*. Retrieved from http://gs.statcounter.com/press/mobile-and-tablet-internet-usage-exceeds-desktop-for-first-time-worldwide/

State, B., Park, P., Weber, I., & Macy, M. (2015). The mesh of civilizations in the global network of digital communication. *PLOS One,* 10, e0122543. doi:10.1371/journal.pone.0122543

Schiller, H. I. (1969). *Mass communication and American empire*. Boston, MA: Beacon.

Statista. (2018). *Top Internet companies: Global market value 2018*. Retrieved from https://www.statista.com/statistics/277483/market-value-of-the-largest-Internet-companies-worldwide/

Straubhaar, J. D. (1991). Beyond media imperialism: Assymetrical interdependence and cultural proximity. *Critical Studies in Media Communication,* 8, 39–59. doi:10.1080/15295039109366779

Takhteyev, Y., Gruzd, A., & Wellman, B. (2012). Geography of twitter networks. *Social Networks,* 34, 73–81. doi:10.1016/j.socnet.2011.05.006

Taneja, H. (2017). Mapping an audience-centric world wide web: A departure from hyperlink analysis. *New Media and Society,* 19, 1331–1348. doi:10.1177/1461444816642172

Taneja, H., & Webster, J. G. (2016). How do global audiences take shape? The role of institutions and culture in patterns of web use. *Journal of Communication,* 66, 161–182. doi:10.1111/jcom.12200

Taneja, H., & Wu, A. X. (2014). Does the great firewall really isolate the Chinese? Integrating access blockage with cultural factors to explain web user behavior. *The Information Society,* 30, 297–309. doi:10.1080/01972243.2014.944728

Wallerstein, I. (2010). A world-system perspective on the social sciences. *The British Journal of Sociology,* 61, 167–176. doi:10.1111/j.1468-4446.2009.01244.x

Webber, W., Moffat, A., & Zobel, J. (2010). A similarity measure for indefinite rankings. *ACM Transactions on Information Systems,* 28, 1–38. doi:10.1145/1852102.1852106

Webster, J. G., & Ksiazek, T. B. (2012). The dynamics of audience fragmentation: Public attention in an age of digital media. *Journal of Communication,* 62, 39–56. doi:10.1111/j.1460-2466.2011.01616.x

Webster, J., & Lin, S.-F. (2002). The Internet audience: Web use as mass behavior. *Journal of Broadcasting & Electronic Media,* 46, 1–12. doi:10.1207/s15506878jobem4601_1

Wu, A. X., & Taneja, H. (2016). Reimagining Internet geographies: A user-centric ethnological mapping of the world wide web. *Journal of Computer-Mediated Communication,* 21, 230–246. doi:10.1111/jcc4.12157

Wu, L., & Ackland, R. (2014). How web 1.0 fails: The mismatch between hyperlinks and clickstreams. *Social Network Analysis and Mining,* 4, 1–7. doi:10.1007/s13278-014-0202-8