University of Pennsylvania
ScholarlyCommons

Internet Policy Observatory | Center for Global Communication Studies (CGCS)

8-2014

The Rise of the Global South on the World Wide Web: Bridging Internet Policies and Web User Behavior

Harsh Taneja

Angela Xiao Wu

Follow this and additional works at: http://repository.upenn.edu/internetpolicyobservatory

 Part of the Communication Technology and New Media Commons

Recommended Citation

Taneja, Harsh and Wu, Angela Xiao. (2014). The Rise of the Global South on the World Wide Web: Bridging Internet Policies and Web User Behavior. *Internet Policy Observatory.*
Retrieved from http://repository.upenn.edu/internetpolicyobservatory/4

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/internetpolicyobservatory/4
For more information, please contact libraryrepository@pobox.upenn.edu.

The Rise of the Global South on the World Wide Web: Bridging Internet Policies and Web User Behavior

Abstract
Scholars of internet governance have traditionally focused on how institutions such as sovereign nation states and multilateral organizations establish public policy. In doing so, experts and policy makers often presume the impact of Internet policies on Internet usage, but rarely do they examine usage aggregated from the behavior of individual web users.

In this study, authors Harsh Taneja and Angela Xiao Wu examine the relationship between internet governance and internet user behavior, empirically investigating web user behavior on a global scale. The authors utilize web use data from ComScore to construct a network for the 1,000 most visited websites globally in September 2009, 2011 and 2013. Analysis of these networks revealed a number of "clusters" of websites, whereby sites within the cluster had more users in common than they did with sites outside the cluster. In each of the three years, the most salient means upon which websites clustered together were both language and geography (and not content type). Thus, the authors interpret such clusters as online expressions of place-based cultures, or "regional cultures", with data suggesting a de-Americanization and rise of the Global South on the WWW since 2009.

Disciplines
Communication | Communication Technology and New Media

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

This report is available at ScholarlyCommons: http://repository.upenn.edu/internetpolicyobservatory/4

The Rise of the Global South on the World Wide Web:

Bridging Internet Policies and Web User Behavior

July 25, 2014

Harsh Taneja, University of Missouri

Angela Xiao Wu, Chinese University of Hong Kong

Table of Contents

The Formation and Flourishing of Regional Cultures on the WWW 3
Data, Methods, and Measures 4
Audience Duplication 5
Data 6
Network Measures 6
Preliminary Results: Network Analysis of Audience Duplication among the Top 1000 sites (2009, 2011 and 2013) 8
Table 1 Overall Network Measures - 2009 to 2013 8
Figure 1 Degree Distribution 10
Figure 2 Network Visualization of Global Web Usage 11
Table 2 Clusters of WWW Usage 15
Table 3 Isolation of Clusters 16
WWW Usage over the Years 17
Internet Policies and User Behavior: Initial Observations 19
Toward an Explanatory Model: Additional Contributing Factors 22
Table 4 Factors Affecting Internet Usage 23
Rethinking Government Intervention with a Broader Concept of Internet Freedom 24
Works Cited 26

The field of Internet governance has traditionally focused on how institutions such as sovereign nation states and multilateral organizations establish public policy. Recent research has documented the power struggles over the creation of the Internet's technical infrastructure such as transmission protocols and domain name systems between government institutions and private players (DeNardis, 2012). Despite the hypothesized relationship between Internet policies and user behavior, the study of Internet governance and that of Internet usage have remained disparate fields (cf. DeNardis, 2014).

Our study is an attempt to bridge this gap by linking Internet policy and governance research with Internet user behavior. It builds on our recently concluded study, which reveals that global web usage manifested as many clusters based on linguistic and geographic proximity regardless of the existence of access blockage such as China's Great Firewall (Taneja & Wu, in press). Furthermore, by empirically investigating web user behavior on a global scale, we aim to move beyond the existing concentration of policy debates on information control and censorship practiced by various states. We highlight that various other state policies, including those fostering the expansion of telecommunication and broadband infrastructure and domestic ICT cultural industries, may affect web use patterns in terms of access and preference.

Our emphasis on the connection between these Internet policies and web user behavior provides a fresh perspective to reconsider the recent development of web use in the global South. In line with the dominant Internet policy discourse originating in the US, much existing scholarship on Internet governance pertaining to these developing countries focuses on censorship as a form of political constraint, which curbs Internet freedom. The prevalent imagery of the World Wide Web (WWW) is thus of one being partitioned by access blockage implemented by evil states. In contrast, we foreground that Internet freedom is also hindered by constraints of economic nature, an example of which is digital inequality. Specifically, our research program intend to examine the WWW as a constellation of worldwide user behavior in light of various state policies promoting domestic access and web content of relevance. We foster a new imagery of the WWW as interconnected regional cultures witnessing the rise of the global South.

In the first section that immediately follows, we elaborate on the idea of regions and why we expect to see regional cultures on the WWW. In the second section, we elaborate on the web traffic data we rely on and our empirical strategy for extracting online regional cultures. In the third section, we report the findings based our analysis of web traffic at three time periods. The fourth section links our findings from analysis of user traffic with state policies. In the concluding section we reflect on how our findings motivate the idea of expanding the scope of Internet freedom beyond censorship and information control to consider a wide spectrum of state policies that affect Internet use.

The Formation and Flourishing of Regional Cultures on the WWW

Our preliminary study indicates that people do not necessarily stay away from certain content because they are denied access to it (Taneja & Wu, in press). Instead, despite the vast universe of websites available to them, they stick to a thin slice due to preference and habit. In particular, we found that this slice of websites tends to be “culturally proximate” content anchored by language and geographic focus. In this project, we describe such content as “regional” and people’s tendency to use regional content “regional orientation.”

We chose "region" because the term has the conceptual richness required for dissecting the complex phenomena involved in global web usage patterns. The term is also malleable enough to be appropriated in our network-based approach without causing confusion. Basically, regional cultures are place-based cultures. Associated with places, regional cultures are maintained in specific language traditions, and shared climates, natural resources, and historical experiences give rise to some cultural commonalities. Notably, regional boundaries are not always congruent with political boundaries, and a region may be above, below, or intersecting an individual country. The designated patch of geography has fuzzy boundaries dependent upon the viewer’s perspective (Griswold, 2008). For example, to introduce Hindustani Music sung by a singer who hails from the holy Indian city of Varanasi, we could describe the “region” of this music as the city of Varanasi, North Indian Indo-Gangetic Plains, the Hindi Speaking states of India, India, or even the Hindustani speaking parts of South Asia comprising North India, Pakistan and Afghanistan.

In recent years, the fate of regional cultures has become a major concern as globalization deepens and new communication technologies spread worldwide. Many assume the "no sense of place" thesis, which maintains that "as place and information access become disconnected, place-specific behaviors and activities begin to fade" (Meyrowitz, 1985, p. 148).[1] The Internet has been viewed as the borderless utopia upholding unrestricted cultural flows, making place increasingly irrelevant. This vision is articulated in the following quote from Google's chairman Eric Schmidt (2013) recent book. "Through the power of technology," he writes, "age-old obstacles to human interaction, like geography, language and limited information, are falling and a new wave of human creativity and potential is rising" (p. 4).

In this project, we engage with this larger conversation by exploring how regional cultures have evolved on the WWW. Our preliminary research on global web use patterns indicates that the link between cyberspace and physical place is far from being severed. Rather, these usage patterns are better considered as place-based social formations on the WWW. Viewed historically, we see the evolution of online regional cultures as a complex dynamic whereby situated users "mak[e] space into place, 'mapping the invisible landscape' through words and symbols, [...] a cultural process that, while by no means innocent of politics and economics, fulfills a human need for meaningful spatial orientation" (Griswold, 2008, p. 17).

Data, Methods, and Measures

To examine online regional cultures expressed through shared user traffic, we conducted network analysis on the usage data of the world's top 1000 web domains at three time points (September of 2009, 2011 and 2013). We then attempt to contextualize changes in various regional cultures through some preliminary archival research. Our goal is to develop a scheme of contextual factors that explain web use patterns on a global scale.

[1] Also see Vincent Mosco (2004) *The Digital Sublime*.

Audience Duplication

Despite having at their disposal a large number of media options, most media users do not consume all of these options and instead focus only on small subsets commonly termed "media repertoires" (Taneja, Webster, Malthouse, and Ksiazek 2012). These repertoires could be based on a number of factors such as users' preferences and the platforms and content they are able to access. When analyzed in the aggregate, the composition of these repertoires indicates what media and/or content are consumed by the same audiences. Hence, in order to identify these repertoires we not only need to know how many people consume each media outlet, but how they move between different outlets. This information is often provided by a measure called audience duplication.

Audience duplication is the extent to which two media outlets are consumed by the same audiences. For instance, on any given day, if out of 100 people in a population, 20 people watched both FOX News and CNN, the audience duplication between them would be 20% for that day. Likewise, duplication can be calculated for all possible pairs of media outlets for a given audience. This results in a symmetric audience duplication matrix, where the elements $A_{i,j}$ represent the extent to which media outlets i and j have audiences in common. Such a matrix can be analyzed further to identify clusters of media outlets that have audiences in common. The earliest application of audience duplication was in identifying user defined program types, the subsets of television programs that were watched by the same set of users (Webster 1985). Following our preliminary study (Taneja and Wu, in press), we use such an approach in this research.

We obtained audience duplication figures between all possible pairs of most popular websites. In doing so, we define duplication between any two websites as the % of unique users that visited both websites across any possible pair. This reveals the extent to which audiences move between all pairs of websites. The resulting audience duplication matrix can help investigate the expression of regional cultures on the WWW. Before detailing specific network measures we apply, we describe our source for audience duplication data.

Data

In this study, we analyze web usage data from comScore, a panel based service that provides Internet audience measurement data once a month. It is currently the largest continuously measured audience panel of its kind. With approximately 2 million consumers worldwide in 170 countries under continuous measurement, the comScore panel utilizes a meter that captures behavioral information through a panelist's computer. Data are collected from both work and home computers of the panel members. Complementing the panel is a census-level data collection method, which allows for the integration of the aggregate level Internet behavior obtained through servers with audience information gained through the comScore panel.

ComScore organizes websites by web domains and subdomains. We decided our sample to be the top 1000 Web domains (ranked by monthly unique users) in the world, as this number not only captures most of the domains that 99% of Web users visit, but also ensures an adequate representation of sites in different languages and different geographies. For many large websites such as Google, the different geo-linguistic variants are classified into separate domains (e.g., www.google.es, www.google.de, etc.). For certain large domains such as Wikipedia, language versions are sub-domains of the main domain (e.g., es.wikipedia.org). In such cases, these sub-domains have been considered in the final sample instead of restricting to top level domains. These data reflect traffic during September 2009 (1018 websites included in the final sample), September 2011 (1022 websites included), and September 2013 (1031 websites included). For each website, we obtained its audience duplication with all other websites from the same annual sample. Thus the final dataset, we have 517,653 ((1018 *1017)/2) pairs of audience duplication in 2009, 521,731 ((1022*1021)/2) pairs in 2011, and 529,935 ((1030*1029)/2) pairs in 2013.

A coder visited each website to note all the languages in which the website offers content.

Network Measures

To begin, we conducted a descriptive network analysis of three resultant audience duplication networks for 2009, 2011 and 2013. The various measures described as follows address

our first research objective regarding the evolution of online regional cultures. These include network analytic measures at the node level (i.e., properties based on connections of individual nodes, in this case websites) as well as at the global level (i.e., properties of network as a whole). Both sets of measures inform this analysis.

At the global level, we computed various network measures such as clustering coefficient, network centralization and density. These indicate the overall shape of the network as well as the distribution of network ties. *Network centralization* indicates the extent to which the ties are more concentrated on a small set of nodes or more uniformly distributed. A high centralization score would indicate that a few nodes receive the lion share of ties and these would be highly central in the network. In a network with low centralization score, however, the ties are more evenly distributed and it is hard to identify a set of few central nodes that receive most of the links. Network centralization for a media audience network such as this one can be considered analogous to Herfindhal-Hirschman index, a well-established measure of market concentration (Webster & Ksiazek, 2012).

Likewise, *clustering coefficient* is a measure that indicates to what extent nodes in a graph tend to cluster together. This is calculated by considering sets of three nodes or triplets that occur in the network. First the clustering coefficient is obtained for each node as the ratio of the number of closed triangles that exist in the network to the total number of triangles (both closed and open) theoretically possible (Watts & Strogatz, 1998). The average of the clustering coefficient of each node is the global clustering coefficient. It varies between 0 and 1, with 1 indicating a fully clustered network and 0 indicating no clustering.

In addition to the two measures above, we report three other measures. First is the network *density,* which simply is the proportion of all ties present to the total number of ties possible. Second is the average *geodesic distance,* which is the average length of all possible shortest paths between all nodes in a network. Third, we computed the network *diameter*, which is the length of the longest geodesic, in other words the longest shortest path between any two nodes.

At the node level, we computed various measures of centrality, which help identify the central nodes in the network. The simplest of these measures, *degree centrality*, counts the number of nodes that each node is tied to. Further, we report the distribution of degree centrality of each node, which also indicates the extent of network centralization.

Preliminary Results: Network Analysis of Audience Duplication among the Top 1000 sites (2009, 2011 and 2013)

Table 1 Overall Network Measures - 2009 to 2013

Measure	2009	2011	2013
Centralization	0.448	0.515	0.572
Clustering Coefficient	0.855	0.851	0.857
Density	0.425	0.399	0.350
Average Geodesic distance	1.578	1.609	1.651
Diameter	3	3	3
Average Degrees *(SD)*	432.3 *(221.8)*	407.5 *(209.7)*	359.7 *(144.1)*
N (number of nodes)	*1018*	*1022*	*1031*

In Table 1 we list a summary of all global level descriptive network measures for all three years. We find that network centralization is moderately high, indicating that positional advantages are distributed rather unequally in this network, which in other words, suggests that a few nodes are linked to a disproportionately large number of nodes (Haneman & Riddle, 2005). In terms of global patterns of web use, this suggests that a small number of sites have duplicated audiences with most other sites. Further, the central position of these sites increases even more over the years, as indicated by the steady increase of the centralization score from 2009 (44.8%) to 2013 (57.2%).

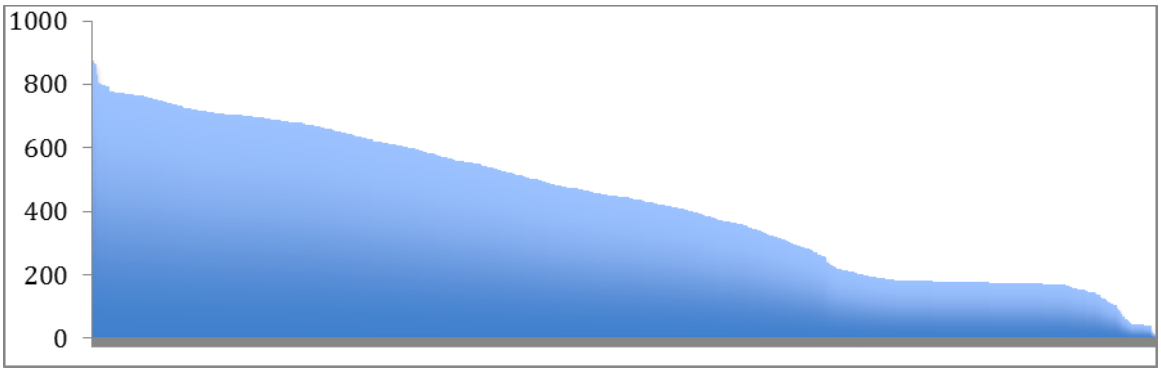
To better interpret centralization, we obtained the degree distributions of the three graphs. First, we find that the average number of degrees of nodes in the graph has decreased from 2009 to 2013 (Table 2). Thus, an average website shared audiences with a smaller number of sites in 2013 (360 sites) than in 2009 (432), a measure also corroborated by the decreasing network density. However, the charts in Figures 1(a), 1(b) and 1(c), which show the degree distributions of the graph in 2009, 2011 and 2013 respectively reveal that this reduction in degrees is not uniform for all sites in the network. On the contrary, first, the number of sites with the highest degrees decreased. In 2009 the graph had 10 sites with more than 800 degrees, whereas in 2011 and 2013 only 8 sites and 6 sites made the cut. Second, sites with the highest degrees in each network gained more degrees over the years. For instance, the degrees of the most central sites increased from 2009 to 2013.

The clustering coefficient of the graph remains high in all three years (Table 1), suggesting that despite the high concentration of audience duplication towards a few central nodes, there is strong evidence of clustering. This high clustering coefficient suggests that websites cluster into groups in a manner where all sites belonging to the same group have high audience duplication among themselves and have relatively low duplication between sites that belong to different groups.

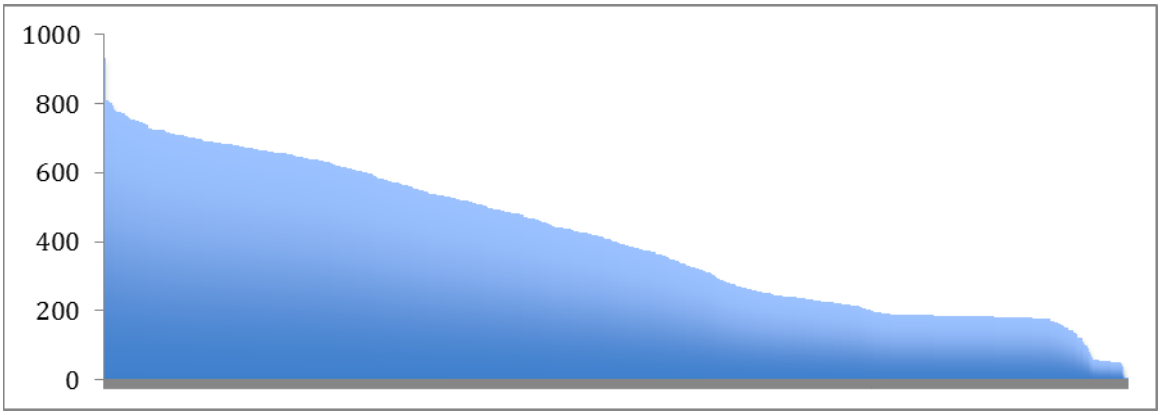
Usually a dense network is expected to have a high clustering coefficient due to the presence of a large number of links. However, in our case the fact that graph continues to display a high level of clustering despite its decreasing density suggests an overall increase in its tendency to cluster into subgroups. This finding corresponds to the growing width and depth of Internet usage worldwide and the accompanying increase in linguistic diversity in our sample. We detail these aspects in the visualizations we describe next.

Figure 1 Degree Distribution

(a) 2009



(b) 2011



(c) 2013

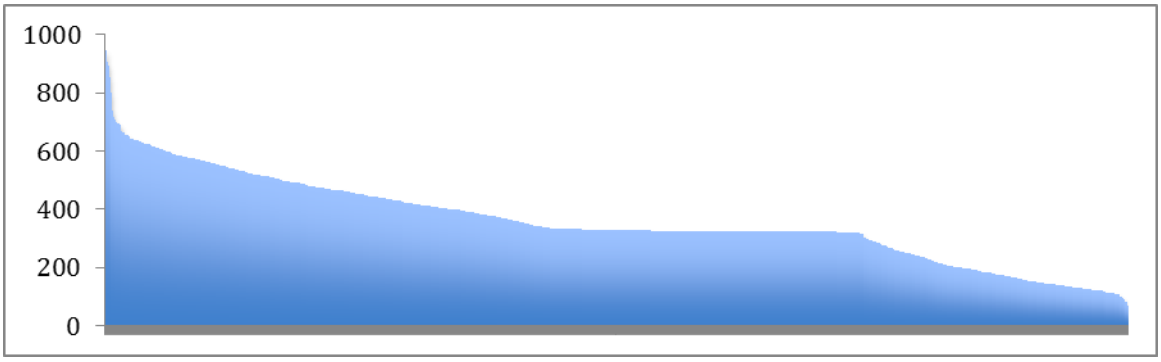
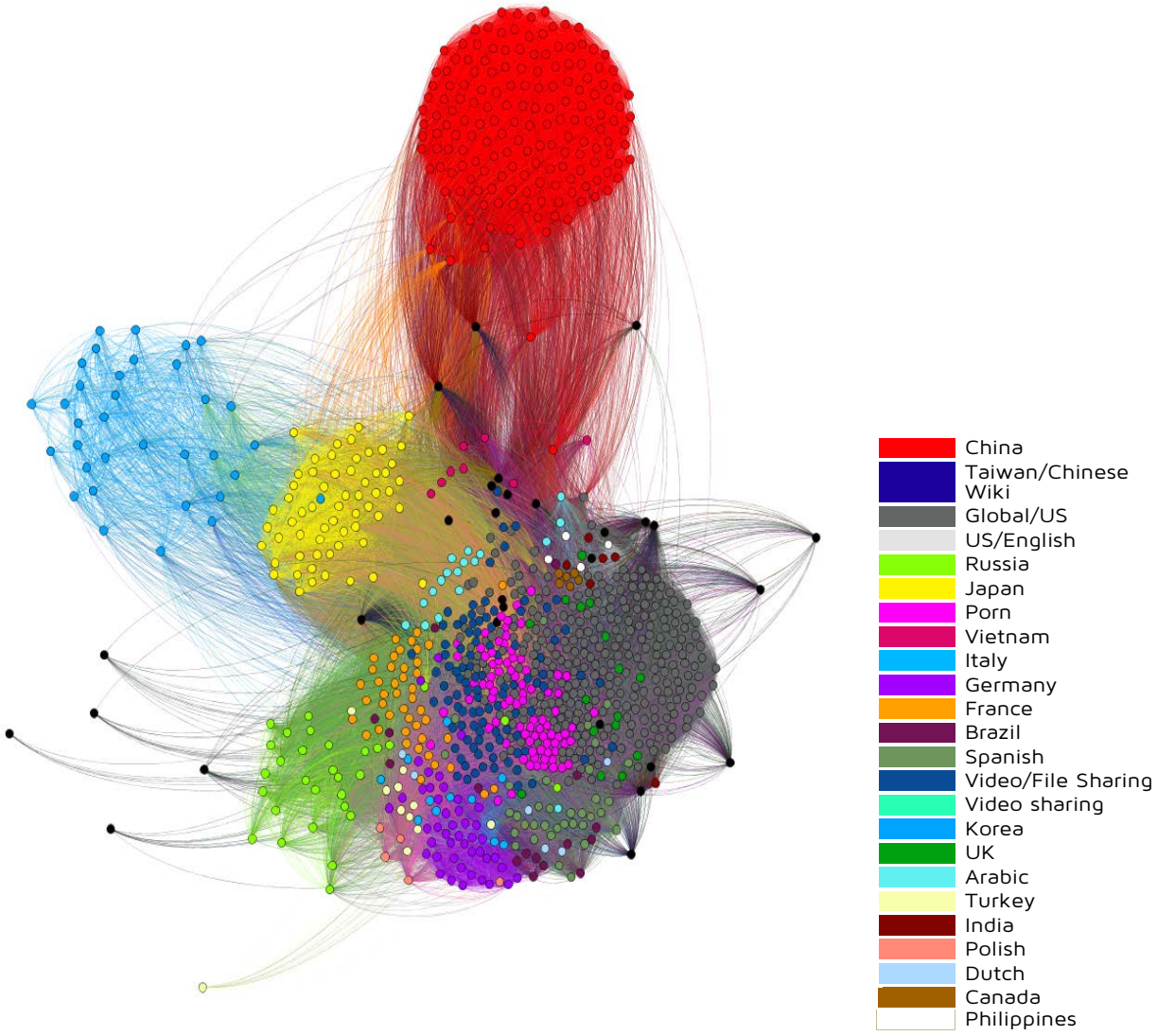
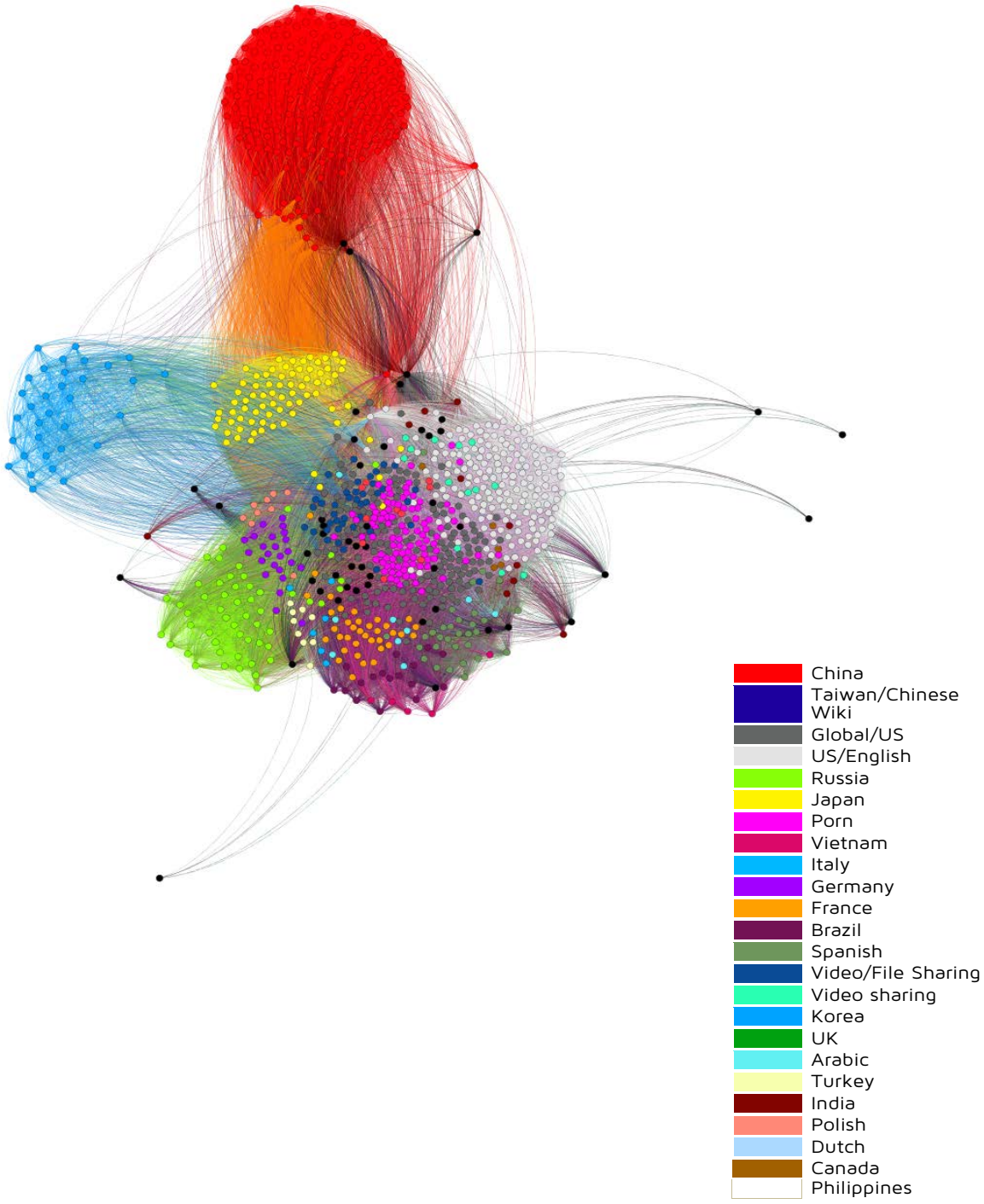


Figure 2 Network Visualization of Global Web Usage

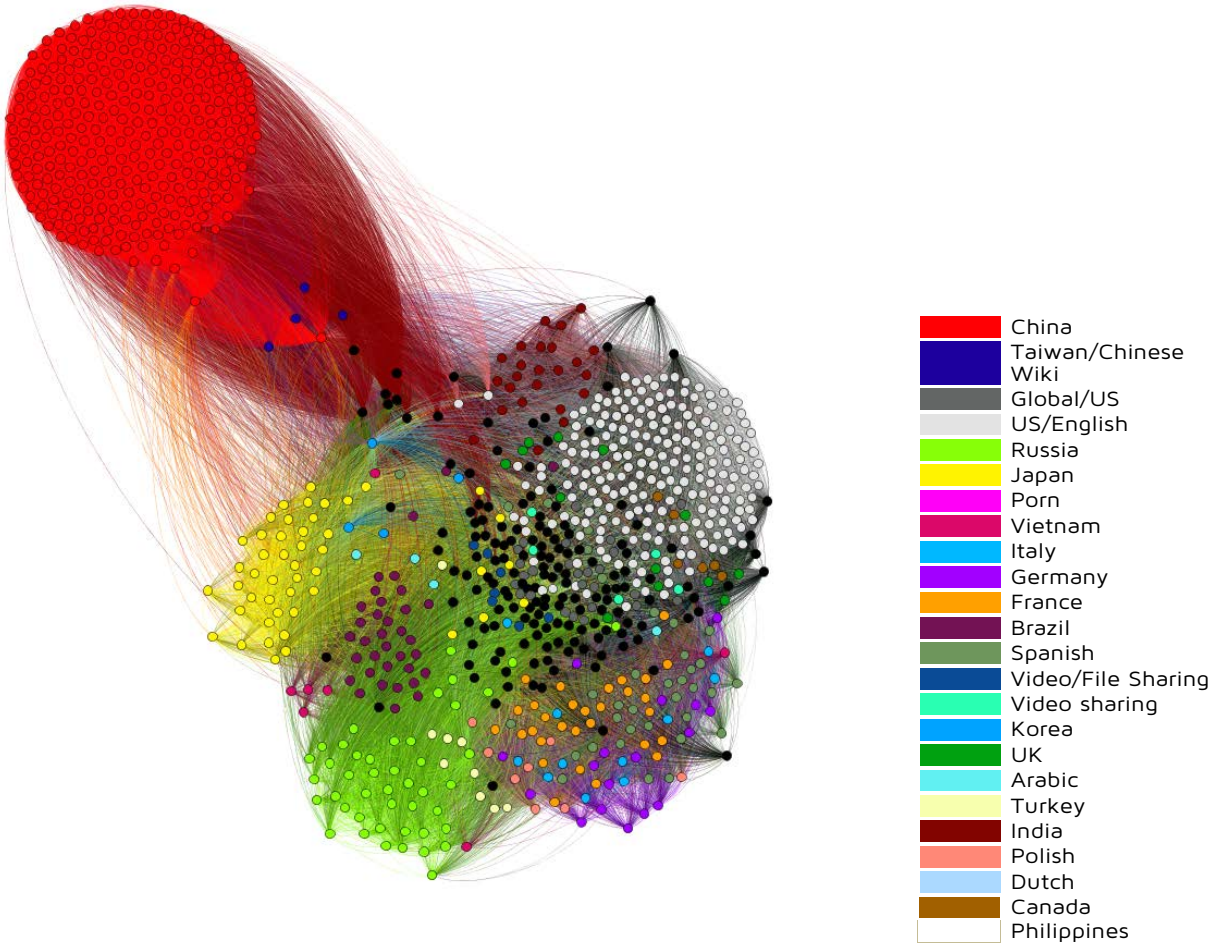
(a) 2009



(b) 2011


China
Taiwan/Chinese Wiki
Global/US
US/English
Russia
Japan
Porn
Vietnam
Italy
Germany
France
Brazil
Spanish
Video/File Sharing
Video sharing
Korea
UK
Arabic
Turkey
India
Polish
Dutch
Canada
Philippines

(c) 2013



Figures 2 (a), (b) and (c) are visualizations of the 2009, 2011 and 2013 graphs. In each of the visualizations, the dots are the nodes (websites) and the lines the ties between them based on audience duplication. The size of the nodes corresponds to the website's popularity: the diameter is proportional to the absolute number of monthly unique visitors. We employed the Fruchterman and Reingold (1991) visualizing algorithm, which belongs to a class of visualization techniques known in graph theory as force directed graphs. The basic mechanism is that there are repulsive forces between all nodes; whereas nodes that are topologically adjacent to each other also have attractive forces. Therefore in the final visualization, nodes that tend to have ties with one another are placed

in close vicinity to form tightly knit clusters; such clusters are relatively separated from other similarly formed clusters.

Consistent with our calculation of clustering coefficient for each graph, these visualizations confirm the presence of clusters (or communities) in these networks. We divided these graphs into clusters using a hierarchical clustering approach, using Pearson correlation coefficient between any two pairs of sites as measure of similarity. In other words, a high correlation between any two sites suggests that they tend to share audiences with the same sets of sites. Consequently, groups of sites with high inter-correlations tend to cluster together. Simply put, these clusters are groups of websites that share audiences with one another to a greater extent than with websites outside their groups. Further, we assigned each cluster with a unique color for easy identification, consistent across all visualizations. In Table 2 we report the number of clusters we found in each year, along with their size indicated by the number of sites.

In Table 3, we report how isolated each cluster was from the rest of the WWW. We calculate this isolation score using a basic node-level network measure termed *closeness (or farness) centrality*, an indicator of the average closeness (or distance) of a node from all other nodes in the network. Each cluster in our WWW networks is basically a group of nodes. In order to calculate the isolation of each cluster, we treated the entire cluster as one node, and calculated it's average shortest distance from all other nodes in the network. We calculated this distance for each cluster we observed in 2009, 2011 and 2013.

In the section that follows, we describe the changes in patterns of web use over time as informed by the composition of these clusters and their isolation scores.

Table 2 Clusters of WWW Usage

2009 (*1018 Sites*)		2011 (*1022 sites*)		2013 (*1030 sites*)	
Cluster	**Sites (No.)**	**Cluster**	**Sites (No.)**	**Cluster**	**Sites (No.)**
Global/US	284	China	179	China	316
China	167	US/English	168	English/US	218
File and Video Sharing	99	Global	146	Japan	47
Porn	84	Porn	94	Fr. Soviet	45
Japan	60	Japan	76	Spanish	35
Germany	45	Fr. Soviet	52	Brazil	35
France	44	File Sharing	41	France	30
Spanish	35	France	36	Global	27
Korea	34	Korea	33	India	24
Fr. Soviet	30	Spanish	31	Germany	17
Arabic	18	Germany	24	UK	13
Brazil	16	Brazil	17	Italy	11
UK	16	Global	16	Turkey	9
Turkey	11	UK	13	Poland	8
Italy	11	India	10	File Sharing	7
BBC (Other Languages)	8	Video Sharing	10	Vietnam	7
Vietnamese	8	Arabic	9	Gaming	6
Dutch	6	Poland	9	Canada	6
India	6	BBC (Other Languages)	9	Video Streaming	5
Canada	6	Turkey	9	Torrents	5
Poland	5	Italy	6	Peripheral Search	4
Philippines	4	Vietnam	5	Tech	4
Pakistan	3	Canada	5	Photo	4
Indonesia	3	Video Porn	3	Online Games	4
Other Sites	15	E-comerce	3	Indonesia	4
				Arabic	4
				Mine craft (games)	4
				Taiwan/ Chinese Wiki	4
				Korea	4
				Directories	3
				Sport	3

Table 3 Isolation of Clusters

2009 (*1018 Sites*)		2011 (*1022 sites*)		2013 (*1030 sites*)	
Cluster	**Isolation**	**Cluster**	**Isolation**	**Cluster**	**Isolation**
Global/US	1.005	China	1.183	China	1.521
China	1.343	US/English	1.056	English/US	1.227
File and Video Sharing	1.022	Global	1.015	Japan	1.322
Porn	1.197	Porn	1.218	Fr. Soviet	1.37
Japan	1.515	Japan	1.67	Spanish	1.381
Germany	1.272	Fr. Soviet	1.239	Brazil	1.355
France	1.351	File Sharing	1.13	France	1.482
Spanish	1.262	France	1.449	Global	1.313
Korea	1.751	Korea	1.855	India	1.53
Fr. Soviet	1.236	Spanish	1.364	Germany	1.362
Arabic	1.443	Germany	1.576	UK	1.405
Brazil	1.272	Brazil	1.648	Italy	1.534
UK	1.278	Global	1.21	Turkey	1.698
Turkey	1.458	UK	1.368	Poland	1.791
Italy	1.428	India	1.477	File Sharing	1.331
BBC (Other Languages)	1.465	Video Sharing	1.212	Vietnam	1.844
Vietnamese	1.801	Arabic	1.475	Gaming	1.34
Dutch	1.414	Poland	1.628	Canada	1.526
India	1.465	BBC (Other Languages)	1.556	Video Streaming	1.347
Canada	1.458	Turkey	1.769	Torrents	1.328
Poland	1.718	Italy	1.493	Peripheral Search	1.384
Philippines	1.451	Vietnam	1.829	Tech	1.326
Pakistan	**1.65**	Canada	1.43	Photo	1.363
Indonesia	**1.83**	Video Porn	1.477	Online Games	1.412
Other Sites	_	E-commerce	1.192	Indonesia	1.435
				Arabic	1.659
				Mine craft (games)	1.489
				Taiwan/ Chinese Wiki	1.428
				Korea	1.564
				Directories	1.354
				Sport	1.349

WWW Usage over the Years

In all three years, we find that the WWW manifested in user behavior largely clusters on linguistic and geographical lines, exhibiting distinct "regional orientations." In many cases, these clusters could be linked explicitly to contiguous geographical spaces defined by nation states such as China[2] (which sometimes coalesce with sites on Taiwan and Hong Kong), France, Poland, Korea, Italy and Japan where the principal language is exclusive to the geography. We witness distinct clusters for certain countries (such as United Kingdom, Canada) despite their languages being spoken very widely in many other countries. In some cases, websites focusing on multiple countries clustered together based on a common language. For example, this occurred for Spanish language sites from Spain, Mexico and Argentina. Yet another category of geolinguistic clusters were clusters of global sites (such as Facebook, Youtube and Ikea) that are present in multiple languages and have specific variants focusing on different countries. Finally, we find some evidence of genre based global clusters, with porn sites as well as file and video sharing platforms being two salient examples. Such platforms' reliance on language is minimal and hence they seem to transcend regions easily, and their cross-regional appeal is a function of network externalities—i.e., users benefit by using these platforms because other similar users are already on these platforms and not from content provided by the publisher.

Whereas we find that some major language clusters are largely consistent across years, we notice some striking changes that reflect the global shifts in web use patterns. One of these changes visible in the diagrams is that the Korean cluster, while prominent in 2009 and 2011, is missing from the 2013 visualization. The cluster solutions confirm this, as the number of sites in the Korea cluster declined from 34 in 2009 to 33 in 2011 and only 4 in 2013. Germany, France and Japan are other prominent clusters that shrank in size, but their decline is more gradual than those of Korea (we explain Korea's decline in the next section). Given the similar sample size in each year (approx.1000 sites), the decline of certain clusters is balanced by the growth in others. The most striking growth is witnessed in the Chinese cluster, which nearly doubled in size between 2009 and 2013. Other

[2] It is hard to tell from our data whether the Chinese cluster has users only from mainland China or from HK and Taiwan as well. However in 2013, we see a separate cluster of Taiwanese sites including the Chinese Wikipedia.

clusters with large growth include the Brazil and India clusters, clusters focusing on former USSR countries, Poland, Vietnam as well as the cluster of Spanish language sites. In the latter case, the number of sites focusing on Spain has declined over the years, but the number of sites focusing on Latin American countries has increased.

As stated earlier we computed isolation scores for each of the clusters (reported in Table 3). For a geolinguistic cluster, specifically, a high score suggests that users of sites in that cluster tend to limit their use fewer sites outside that cluster compared to a cluster with a lower score. We find Vietnam, Poland, Korea, Turkey and Japan to be examples of clusters with very high scores in all three years. The cluster which we term as US /Global in 2009 is the least isolated, due to the presence of many "global" sites in this group. However, the isolation of the cluster of US-focused sites increases in 2011 and 2013 (where we accordingly term this cluster, US/English) as in these years, many of the "global" sites (such as Facebook and Twitter) separate from the US cluster to form a cluster of their own. That the US-focused sites (in 2011 and 2013 data) appear as a regional cluster just like many other online regional cultures (distinct from the Global cluster) may be indicative of the ongoing de-Americanization of the WWW.

We are unable to see a clear association between a general understanding of access blockage and cyber censorship in certain countries and the isolation scores of corresponding clusters. For instance, China is less isolated than Korea in each of the years. Moreover, China's scores are in the same range as those of Japan. The least isolated clusters are those with sites that provide platforms for user-generated content (such as File and Video Sharing and SNSs). In addition, fast inclusion of users from and websites focusing on formerly underrepresented regions online may cause changes that conflate that of governmental blockage. One noteworthy observation is the general increase in isolation of clusters in 2013, which is consistent with the finding that the network is more clustered in 2013 compared to previous years (the isolation score of least isolated cluster is higher in 2013 compared to both 2009 and 2011). This suggests that as more regions of the world add users and content to the WWW, we would witness seemingly isolated clusters of use. Access blockage may further accentuate this tendency. In sum, at this stage the association between cyber censorship measures and user behavior is very preliminary. To further investigate this association would require

systematic research on contextual variables and a nuanced understanding of the assumptions and agendas underlying each of the available Internet freedom indices.[3]

As already stated, we believe that the online regional cultures witnessed in these clusters result from interactions between many state policies, and cannot be attributed to any one particular policy imperative. In the following section, we attempt to highlight a few examples that link some of these cases to state policies.

Internet Policies and User Behavior: Initial Observations

South Korea is one of the best illustrations where government policy has been instrumental in achieving near ubiquitous broadband access. In fact, according to trade reports, South Korea has the fastest broadband in the world, with connections four times faster and significantly less pricey than those in the United States.[4] This level of penetration has been realized by the government's policy to promote competition among broadband providers as well as subsidies on Internet ready devices. Supplementing this penetration is an innovation culture that fosters a startup community (for instance CYworld, a Korean social networking site/SNS launched 5 years before Facebook and remained Korea's most popular SNS for a decade).[5] Yet, on most Internet freedom indices, South Korea is rated as having partial Internet censorship.

In recent years, although South Korea remains at the top among Asian Pacific countries in terms of Internet penetration, reaching an early saturation means a stagnating Internet user base compared to late comers with large population base. This led to South Korea's decreased prominence on the WWW in terms of web user behavior, as we find in our ongoing study. Korea's contribution to total global web users has declined from 2.53% (of the world's 1.565 billion) at the end of 2008 to 1.68% (of 2.501 billion) at the end of 2012.[6] In other words, while the Internet usage in Korea remained stable, the absolute number of Internet users has risen in many other emerging economies such as India, China, Vietnam, Brazil and Russia. One can deduce from the numbers just

[3] We have proposed this research component in our second proposal for funding to the IPO.
[4] http://edition.cnn.com/2010/TECH/03/31/broadband.south.korea/
[5] http://www.economist.com/blogs/economist-explains/2014/02/economist-explains-3
[6] The 2013 data are yet to be released by the World Bank.

reported that 900 million people (net) joined the Internet in this period, of which only 5 million were Koreans. This rapid decline is also an artifact of ComScore's measurement methodology. ComScore's media metrix data is restricted to measuring PC based web use (desktops and laptops), and does not include consumption via smartphones and tablets. South Korea is one market where consumers have rapidly adopted smartphones, but the total number of Internet users has stayed roughly the same. Consequently, we witness a decline in PC based web consumption. For instance, ComScore reports Naver.com, the largest Korean Internet portal, at about 30 million monthly unique users in 2009 and 2011 but only 12 million users in 2013. (The data report a similar decline in user numbers in 2013 for other Korean Websites in the sample).

In a stark contrast to Korea, Internet user base in **Brazil** and **India** has taken much longer to grow. Brazil had only 24 million Internet users (15% of its population) in 2003 (compared to Korea at more than 60% in the same year), and it surpassed 100 million Internet users (50% of its population) only in 2013, a fact consistent with the growth of the Brazilian cluster in 2013 seen in our data. Much of the growth in the Brazilian Internet user base in the last 3 years seems to be from Internet use on mobile devices. Likewise, India's fast increased online population also comes from its growing mobile Internet. In 2004, India had less than 2% of its population online. In 2013, this number has almost reached 20%, corresponding to about 200 million users. However, presumably due to its linguistic plurality and overall weaker economic indicators, India still lags behind most developing nations in Asia Pacific and elsewhere (eg., Brazil) in Internet adoption. This is despite its position as an international information technology outsourcing hub. Both India and, more so, Brazil have relatively liberal state policies regarding information control, compared to South Korea.

It is also worth discussing **China** here briefly, due to its notoriously comprehensive access blockage measures. As already noted in our first study, China's Great Firewall worked indirectly to help establish a robust domestic Internet industry (Taneja & Wu, in press). For instance, even years after blockage on Wikipedia has been lifted, Chinese users contribute to Baike, a local equivalent owned by the behemoth Baidu. Likewise, Tencent Technologies, valued at a staggering $139 billion, owns a wide portfolio of Internet services including WeChat, which has grown to be the world's

largest mobile messaging service. The success of Chinese local Internet companies is often associated with the blockage of Global competitors (Liu, 2010).

> But Tencent's success can't be pinned on that handicap. The company embraced mobile years before Facebook, and has built a platform, used by 355 million active users, that functionally offers every popular service that Americans are familiar with—including Facebook, Twitter, WhatsApp, and Zynga, all wrapped up in one app. It keeps adding new functions at a fast clip, such as a new Uber-like taxi finder that was used 21 million times in the first few weeks.[7]

Japan with its completely open Internet is a case study in stark contrast to China. Despite being completely free to access any part of the Internet, we find the prevalence of a strong online regional culture in Japan (and similar levels of isolation). It boasts of at least three companies in the top 10 Internet companies in the world based on market capitalization. Only a few companies in United States, China and South Africa have a higher market cap than the Yahoo Japan. More importantly, like Korea, Japan is an early mover in enabling its citizens with Broadband access in early 2000s and major telecom corporations were involved in such initiatives. For instance, Rakuten, a large online portal and e-commerce website in Japan started operations in 1997, much before Google was born. Likewise, Yahoo Japan is not a subsidiary of Yahoo Inc. but a joint venture between Yahoo and Softbank, a Japanese company founded in 1996. Japan has 64.5 million Internet hosts, the second largest in the world.

The search engine market is another curious case in point hinting at the complex factors behind Internet adoption and usage in different geographic areas. Google may be the best known search engine in the Western world, however, in some other countries it is just a distant second or even an insignificant player. For instance, Russia's Yandex corporation is the leading search provider in Russia, Ukraine and Kazakhstan, much ahead of any competitor. Similarly, Baidu leads the search market in China and Korea (more popular than the Korean corporation Naver, which ranks second ahead of Google Korea); Yahoo still leads Internet search in Japan. Our cursory analysis suggests that

[7] http://www.fastcompany.com/3029119/most-innovative-companies/tencent-the-secretive-chinese-tech-giant-that-can-rival-facebook-a#!

Google is more popular in countries where the languages use a roman script. Another possible explanation is that certain regions, such as China and the EU, regulate search engines.

Toward an Explanatory Model: Additional Contributing Factors

Our initial attempt at contextualizing our findings about global web use through a focus on several country cases further points to a number of relatively quantifiable factors that may contribute to the online manifestation of regional cultures.

First, and heavily influenced by policy, is the uneven penetration of the Internet all over the world, which has changed both the geographic distribution and social composition of the global online population. In terms of geographic distribution, the proportion of Internet users from less developed countries have been fast increasing, which may have given rise to a growing number of distinct regional orientations in web use. Regional orientations manifest strongly in global web use patterns as web adoption in a region transcends socio-economic boundaries. In terms of social and economic status, the Internet used to be dominated by the elite segment (especially in developing countries) that has higher cultural capital for navigating and making sense of cultures outside of the region. The concept of hybrid identity is particularly relevant here. "Hybrid" is a more accurate term to describe what is commonly known as "global" or "cosmopolitan" identity (Kraidy, 2002). These elite individuals enact their hybrid cultural identities through using media products from multiple but definite cultures. For example, some well-educated individuals in Korea may inhabit both Korea-based websites and certain English websites, which in effect configures the boundary of the Korean regional culture on the web. Over time, however, the percentage of users from lower strata has grown rapidly, accentuating regional cultures online. Within a region, as the Internet penetrates down the social ladder, it accommodates more users, characterized by emergence of intense regional orientations.

In addition to domestic elites, diasporas also exhibit hybrid identities via their media consumption, as studies have found among Arabs in London (Georgiou, 2011) or Italians in Brazil (La Pastina & Straubhaar, 2005). Regional cultures bear a socially recognized relationship to some place, usually their place of origin (Griswold, 2008). In this light, the web use of a diaspora population may

connect the region of its origin to that of its destination on the WWW. The architecture of the WWW allows the diaspora to enact their identities linking to their region of origin on the same platform as the locals of the region, in turn accentuating the online regional culture.

Languages too contribute to the formation of online regional cultures. People who speak the same language, within a country, geographic region or even in geographically non-contiguous parts of the world can be thought of as a region. Studies have found that regions such as the Spanish speaking world (Spain and Latin America), or the French-speaking world spanning France and Africa, often tend to consume the same content (Straubhaar, 2007). Likewise within multilingual countries such as India or Switzerland, each linguistic region exhibits a distinct regional culture. We argue that Internet use patterns would be influenced by the way regions are formed on the lines of linguistic similarities and differences, both due to shared languages between geographic regions (or countries) as well as linguistic diversity within geographic regions.

Using the variables described above, we made cross-national comparisons to convey the hidden dynamics behind the evolution of the WWW in terms of user behavior globally. Table 4 also serves as a point of departure for future archival research and statistical analyses for this project.

Table 4 Factors Affecting Internet Usage

Country	Local IT Industry	Internet User Base	Internet Adoption	Access Blockage	Diaspora Size	Languages
China	Huge & booming	Big	Expanding	Pervasive	Large	Chinese
Japan	Huge & Stable	Big	Stable	No	Small	Japanese
India	Small & Booming	Small	Expanding	No	Large	Multiple (English as common official language)

Brazil	Small & Booming	Medium	Expanding	Minimal	Small	Portuguese
Korea	Big	Small	Stable	Substantial	Small	Korean

Rethinking Government Intervention with a Broader Concept of Internet Freedom

Our analysis of global web use over time revealed the expanding online presence of the global South. The varied regional orientations involved, clearly distinguishable from the cluster of websites run by US-based multinationals, have grown in both scope and degree. This trend is confirmed by our network visualizations, network measures of centralization, clustering coefficient and density, and finally the growth in the number of network clusters. In attempt to contextualize this finding, our extant research foregrounds that complex local development in Internet infrastructures and industries, in addition to governmental blockage, may correspond with the evolution of online regional cultures as we observed. This motivates our future research agenda to bring cross-country policy and market measures (including cyber censorship indices) into a single dataset, a project that we have proposed to IPO to undertake next. Underlying this agenda is our reconceptualization in connecting the dots in Internet governance debates, user behavior, and global digital connectivity and inequality.

First, our study strives to expand the scope of Internet policy debates beyond a narrow focus on state censorship. This focus is inherent in the dominant view, which posits that the Internet can and should create a global community, and that territory-based government censorship immensely threatens this technological potential. Broadly stated, such censorship includes two types of measures: content censorship over domestic websites, and access blockage targeting websites outside the state's jurisdiction. Access blockage, in particular, has garnered most public attention due to the powerful imagery it invokes of the "balkanization" of the Internet.

Access blockage in the US Internet policy discourse is often projected as the digital Iron Curtain that curbs the "free flow" of online information. Two months after attending a celebration to

commemorate the fall of the Berlin Wall, then US Secretary of State Hillary Clinton announced: “A new information curtain is descending across much of the world.” Western policy and popular discourse tend to presume that nations restricting their citizens’ online access effectively prevent them from joining the world, thus depriving them of necessary means for democratization. Scholars critical of this discourse have pointed out that its “implied causal relationship places stress on the role of the free flow of information in leading to democratization, simplifying a complex historical process into a relatively clear schematic” (McCarthy, 2011, pp. 99-100). How to think of causal relationships turns, first of all, on what we know about human behavior. Our project sets out to gauge the interplay between aggregated personal web use and consequences of governmental blockage.

We empirically deciphered the isolation of the various place-based online regional cultures, many of which were linked to nation states, but we were unable to see a clear role of access blockage in explaining the extent of their isolation. This empirical evidence supports conceptual critiques as just discussed. Also, our findings reinforce recent contention by critical scholars that Internet policy debates are motivated by symbolic understandings surrounding the Internet, which are rooted in larger geopolitics (eg. McCarthy, 2011; Cramer 2013). Specifically, US foreign policy rhetoric serves the interest of large corporations. When it comes to Internet technologies, this rhetoric articulates the invested goal of “opening markets to US capital and liberalizing foreign polities with the values of human rights and democracy” (p. 89). Considered in this context, other states enact censorship policies “in order to counteract the liberalization of their public sphere, or to protect their fractions of information capital from competition within American multinationals” (p. 93).

Furthering this line, we emphasize that extant policy debates have largely ignored a wide range of economically oriented initiatives and regulations because of their one-sided conception of Internet freedom. US policy makers and the public at large tend to understand freedom as the absence of political coercion, leading to strong denouncement of censorship policies. In contrast, the meaning of Internet freedom may also derive from the freedom to act, the other aspect of freedom conceived for centuries in political philosophy (Cramer, 2013). “Internet freedom to act” requires certain technological infrastructures, both hardware and software, to which government intervention

may be indispensable. However, due to the prevalent understanding of Internet freedom as being free from political intervention, even within the US, policies such as those ensuring network neutrality or those ameliorating digital inequality that accompanies economic inequality failed to garner much popular support (Cramer, 2013).

Across the international system, more pronouncedly, it has largely been ignored that numerous developing states exercised government interventions to ensure wide access to web content their citizens prefer, which in effect extend people's Internet freedom to act. This involves formulating various telecommunication policies aiming to facilitate the expansion of Internet infrastructure and the development of domestic Internet industries, some instances of which are mentioned in this report. However, since many of these states have political systems at odds with the US ideal of liberal democracy, in US popular and policy rhetoric, these countries/regions are the problematic trouble-makers at the margin, obstructing the realization of a truly "borderless," "global" internet.

In sum, our research program aims to explore the linkage between the evolution of online regional cultures and various government interventions with a broader conception of Internet freedom, taking into account the economically oriented Internet policies that affect local Internet infrastructures and industries.[8]

Works Cited

Cramer, B. W. 2013. The two Internet Freedoms: Framing victimhood for political gain. *International Journal of Communication* 7: 1074-1092.

DeNardis, L. (2012). Hidden Levers Of Internet Control. *Information, Communication & Society*, *15*(5), 720–738. doi:10.1080/1369118X.2012.659199

DeNardis, L. (2014). The global war for Internet governance. New Haven : Yale University Press

[8] It is worth clarifying that our method relies on aggregated user behavior en masse instead of that of the elites, a usual target of media searchlights. We focus on a behavior level and our project does not engage with philosophical, legal, and ethical issues of agency and free choice, nor does it aim to claim consequences of specific information consumption patterns in terms of political mobilization and institutional reform.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(11), 1129–1164. doi:10.1002/spe.4380211102

Georgiou, M. (2011). Watching soap opera in the diaspora: cultural proximity or critical proximity? *Ethnic and Racial Studies*. doi:10.1080/01419870.2011.628040

Griswold, W. (2008). *Regionalism and the reading class*. Chicago, Ill: University of Chicago Press.

Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside.

Kraidy, M. M. (2002). Hybridity in Cultural Globalization. *Communication Theory*, *12*(3), 316–339. doi:10.1111/j.1468-2885.2002.tb00272.x

La Pastina, a. C., & Straubhaar, J. (2005). Multiple Proximities between Television Genres and Audiences: The Schism between Telenovelas' Global Distribution and Local Consumption. *International Communication Gazette*, *67*(3), 271–288. doi:10.1177/0016549205052231

Liu, C. (2010). Internet Censorship as a Trade Barrier: A Look at the WTO Consistency of the Great Firewall in the Wake of the China-Google Dispute. *Georgetown Journal of International Law, 42*.

McCarthy, D. R. (2011). Open Networks and the Open Door: American Foreign Policy and the Narration of the Internet1. *Foreign Policy Analysis*, *7*(1), 89–111. doi:10.1111/j.1743-8594.2010.00124.x

Meyrowitz, J. (1995). *No Sense of Place: The Impact of Electronic Media on Social Behavior*. New York: Oxford University Press.

Porter, T. M. (1995). *Trust in Numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Schmidt, E., & Cohen, J. (2013). *New digital age*. [S.l.]: Random House.

Straubhaar, J. (2007). *World television : from global to local*. Los Angeles [u.a.]: Sage.

Taneja, H., Webster, J. G., Malthouse, E. C., & Ksiazek, T. B. (2012). Media consumption across platforms: Identifying user-defined repertoires. *New Media & Society*, *14*(6), 951–968. doi:10.1177/1461444811436146

Taneja. H & Wu. A.X. (in press). Does the Great Firewall really isolate the Chinese? Integrating access blockage with cultural factors to explain web user behavior. *The Information Society, 30*(5).

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world"networks. *Nature*, *393*(6684), 440–442.

Webster, J. G. (1985). Program audience duplication: A study of television inheritance effects. *Journal of Broadcasting & Electronic Media*, *29*(2), 121–133.

Webster, J. G., & Ksiazek, T. B. (2012). The Dynamics of Audience Fragmentation: Public Attention in an Age of Digital Media. *Journal of Communication*, *62*(1), 39–56. doi:10.1111/j.1460-2466.2011.01616.x