Soc. Netw. Anal. Min. (2016) 6:105
DOI 10.1007/s13278-016-0415-0

ORIGINAL ARTICLE

Resilience of the World Wide Web: a longitudinal two-mode network analysis

George A. Barnett[1] **· Ke Jiang**[1]

Received: 21 June 2016 / Revised: 19 October 2016 / Accepted: 28 October 2016 / Published online: 9 November 2016
© Springer-Verlag Wien 2016

Abstract Daily data on the use of the world's 425 most frequently visited websites by 118 countries were mined over the 4-month period September 1 to December 31, 2015, to create a longitudinal two-mode network (countries and websites). This paper describes the changes in the international World Wide Web (WWW), as a network, to determine the effects of unanticipated shocks (the terrorist attacks on Paris and San Bernardino) and predictable events [national holidays (Golden Week, Christmas, New Years) and shopping days, Black Friday and 11/11]. The results indicate that while there are changes in the use of individual websites, due to weekly cycles in viewing specific websites, the shocks and other social and cultural events, the overall network is remarkably stable. This resilience is due to the constraints that network ties place on the relationships among the websites, which limits their potential behavior.

Keywords Worldwide Web · Two-mode network analysis · Shocks · Resilience · Network dynamics

1 Introduction

As a concept, resilience is related to the process of adapting to or maintaining equilibrium in the face of significant stress. From the perspective of social networks, components (nodes) do not act independently of one another, and as such their behavior is to a certain extent linked to one another. Thus, resilience can be understood as how the relations among the nodes place constraints among the range of each other's possible behavior limiting the degree of freedom of the individual nodes to act independently. In other words, the functional stability of a social network is achieved through interconnectivity. When one component in a network fails to provide the necessary resources (material or information) because it is disabled or removed from the network, another will perform the necessary function for the system to return to and maintain a state of equilibrium. Thus, social networks' resilience is the ability of the networks to adjust their activities to retain basic functionality in times of stress, such as their response to unanticipated shocks, such as terrorist attacks, and predictable events, like holidays. So while individual nodes behavior may change in response to perturbations—shocks resulting from unplanned events, or planned events, the overall network remains remarkably resilient. For example, when one airport is closed due to the weather, other airports in the transportation network with routes to same destinations takes over its role routing passengers through it to their final destinations. Resilience is maintained through redundant interconnectivity. This paper examines the resilience of the World Wide Web (WWW) by analyzing a combination of factors contributing to network resilience including the power law distribution of links, the existence of network hubs, network density, heterogeneity and symmetry. Specifically, this paper uses daily data on the use of the world's most frequently visited websites by individual countries over the 4-month period September 1 to December 31, 2015, to describe the changes in the international WWW to determine the effects of unanticipated shocks and predictable events on two networks. The

✉ George A. Barnett
gabarnett@ucdavis.edu

Ke Jiang
kejiang@ucdavis.edu

[1] Department of Communication, University of California, Davis, Davis, CA 95616, USA

networks are composed of the world's countries and the most frequently visited websites.

2 Theory

Albert-László Barabási (2014) argued that generally, the distribution of links in social networks, and specifically, the Internet with its distributed hypertext system, the WWW, follows the power law. Unlike the bell curve, the power law distribution does not peak at the mean with an equal number of cases greater and less than the average. Rather, the power law is a continuously decreasing curve with very few large values and many small values. Its mean has few values greater and more less than the average. A network whose degree distribution follows a power law is also called a scale-free network. Barabási (2014, p. 113) wrote, "… (A)n significant fraction of nodes can be randomly removed from any scale-free network without its breaking apart. … As the Internet, the World Wide Web, the cell, and social networks are known to be scale-free, the results indicate that their well-known resilience to errors is an inherent property of their topology…"

Faloutsos et al. (1999) had shown that the Internet is a scale-free network, where a majority of the information passes through a limited number of routers. Cohen et al. (2000) indicated that while the Internet becomes more diluted as more nodes become inoperable, it remains a connected network even if nearly all the routers breakdown. Essentially, it is resilient to random breakdown because its connectivity is distributed by the power law. Barnett and Park (2005) and Park et al. (2011) had reported that the international WWW is also distributed as a power law; a few websites (i.e., Google and Facebook) have the greatest number of visitors, and the majority of sites see few visitors. Thus, it is expected that the international WWW would be resilient to shocks and other perturbations because the links connecting the nodes are distributed as a power law, creating a scale-free network.

Barabási (2014) explained that networks' topological robustness is due to the existence of hubs, a few well-connected nodes that hold a network together. If shocks to the network have an equal chance of impacting all the nodes, then the small nodes are more likely to be affected because there are more of them. However, if a few of the large nodes and the hubs were impacted, the network would be more vulnerable to failure. For example, the airline network might breakdown if the major hubs, London-Heathrow or the airports in New York, Chicago or Tokyo were to close. This was the case when in 2010, Iceland's Eyjafjallajökull volcano erupted and shut down London-Heathrow, which paralyzed flights across Europe, although some international flights were rerouted and the network eventually returned to equilibrium.

There are other properties of networks that facilitate their resilience and the ability to adjust their activities to retain basic functionality in times of stress. Gao et al. (2016) suggested three properties: network density, heterogeneity and symmetry. Density is the proportion of possible ties linking the network's nodes. The greater a network's density, the greater is the redundancy in the relations among its components. Thus, if some of the nodes or links were removed due to a shock to the system, others would take their place to ensure the functioning of the system. Heterogeneity refers to the variance in link strength or weighted degrees (Gao et al. 2016). As is the case with the power law, networks with high heterogeneity tend to be more resilient. Symmetry is the correlation between in-degree and out-degree. An undirected network ($x_{ij} = x_{ji}$) is perfectly symmetrical. In asymmetric networks ($x_{ij} \neq x_{ji}$), where one or more of the nodes have a larger in-degree than out-degree, the system tends to be less resilient. In sum, "…dense, symmetric and heterogeneous networks are most resilient, and sparse, antisymmetric and heterogeneous networks are least resilient" (Gao et al. 2016, p. 311). Here antisymmetric and heterogeneous refer to a distribution of links based on a combination of both properties.

Past research indicates that the international WWW has a center-periphery structure (Barnett et al. 2001; Barnett and Park 2005, 2014), where the more central websites, such as Google and Yahoo, act as hubs (Barnett and Park 2014; Ruiz and Barnett 2014). Also, there are a number of linguistic and cultural subgroups, such as clusters of Chinese, Russian, Spanish and Arabic sites (Barnett and Sung 2005; Barnett et al. 2015). However, their examination has been limited to the 200 most frequently visited websites. Further, research indicates that this network has grown tremendously over the last two decades. Park et al. (2011) reported that the international hyperlink network grew from 91 million links in 1996 to 356 million in 2003 and to over 14 billion by 2010. However, this research was limited to hyperlink data between top-level domains (country labels such as .cn and.jp) with annual data (Park et al. 2011; Rosen et al. 2011; Barnett et al. 2013). Barnett et al. (2013) did find that while the international hyperlink network did expand, the structure of the network was stable over time and that it co-evolved with the international telephone network. This is one indicator of the resilience of the WWW, which was due to the fact that the large central websites (e.g., Google and Facebook) located in the USA act as hubs, in a dense scale-free network.

We expect that an examination of the two networks will result in center-periphery structures with the USA at center of the international network, and Google and Facebook at center of WWW network. This is due to the fact that US-based companies such as Google, Facebook and Twitter together accounted for more than 70% of the international network traffic (Barnett et al. 2015). Additionally, there will be separate linguistic and cultural communities. Also, due to the network properties identified by Barabási and colleagues, it is expected that while there will be changes in the use of individual websites, due to shocks and planned social and cultural events, the network structures will remain relatively stable. That is, due to the ties among the websites and countries, when perturbations occur the network will absorb the shocks and become resilient in the sense that the system would soon return to equilibrium. Further, while there are changes in the use of individual websites, due to various perturbations, the overall network will be stable. This resilience is due to the constraints that network ties place on the relationships among the websites, which limits their potential behavior allowing the network to maintain a state of equilibrium.

3 Methods

Daily data on the use of the world's 500 most visited websites by 118 countries were mined from Alexa.com (http://www.alexa.com/topsites) using a Python script written for that purpose. Alexa's traffic estimates are based on the browsing behavior of a global panel, which is a worldwide sample of millions Internet users. The data are a rolling three-month average updated daily. The number of individual panel members who visit a site on a given day determines visitors. Page views are the total number of user URL requests for a site. Multiple requests for the same URL on the same day by the same user are counted as a single view. Alexa's traffic data are for top-level domains only. They do not provide separate rankings for subpages within a domain. Data on sites with relatively low traffic may not be reliable. However, the more frequently a site is visited, the more reliable the data.

This collecting process began September 1, 2015. This paper discusses the data for the period September 1 to December 31, 2015. The data are the percentage of each site's users that come from a particular country. The data are two mode, such that two countries are considered linked to the extent that they both visit a common website, and two websites are related to the degree that they are both visited by the same country. This is based on the aggregate online behavior of people within countries as they access web-based information. The one-mode country (country $\times$ country) and website (website $\times$ website) data have been extracted from the two-mode data. The analysis was performed with the 425 sites from among the 500 that were visited at least 90% of the time during this 4-month period. Websites that exited or entered the network during this time were not analyzed. Missing data were added as the mean of the value for the date prior to the missing value and the value of the date that followed, $x_t = (x_{t-1} + x_{t+1})/2$. Further, each percentage was multiplied by the proportion of the total number of Internet users who visited the website that day. According to Alexa.com, there are 26 news sites and 35 shopping sites among the 425 that compose the network data.

In order to determine the resilience of the WWW, change in the country and website networks was examined through the use of network analysis. Network analysis is a set of research methods for identifying structures in systems based on the patterns of relations among system components (Wasserman and Faust 1994). In this case, the system is the WWW and the relations are based on the use of the various websites by the countries of the world. Two countries are related to the extent that they both visit the same website, and two websites are considered linked to the extent that their visitors come from the same country.

There are a number of indicators of network structure, including density, the proportion of links among the system components divided by the total possible number of links, which in this case is the number of countries (or websites) times the number of countries (or websites) minus 1 divided by two ($(n \times n - 1)/2$) (Monge and Contractor 2003). Another measure of the interconnectivity of a network is transitivity, which refers to the extent to which the relation that links two nodes in a network that are connected by an edge is transitive. As such, transitivity may be considered an indicator of density. It demonstrates the density of transitive triples in a network. Specifically, three nodes (i, j, k) are considered transitive if whenever i is linked to j and j is linked to k, then i is also linked to k (Wasserman and Faust 1994). Transitivity leads to the notion of triads, three nodes that are linked together in transitive relations. It is the ratio of transitive triads to potentially transitive triads. Triads have been studied extensively in order to predict such phenomenon as coalition formation and attitude change (Holland and Leinhardt 1970). For calculating transitivity and the number of triads, the networks were first transformed to the binary networks.

Centrality is the extent to which a node occupies a prominent or critical position in a network (Valente et al. 2008). The more central nodes act as hubs and are vital for network resilience. The measures of the individual nodes' centralities should be distributed as a power law, indicating that the network is scale-free and therefore more resilient. There are a number of different measures of centrality; among them are degree, betweenness, distance and eigenvector centrality (Freeman 1979). This study examined degree and eigenvector centrality. Degree is the total number of direct links or the sum of these link strengths (Freeman 1979). Eigenvector centrality is an indicator of a node's overall centrality in a network (Bonacich 1972). The measure requires valued data and is ideal for dense networks.

Further, it takes into account the positions of a node's contacts such that its centrality increases relatively if it is tied to more central nodes. Density, transitivity and both centrality measures were examined over time to determine the impact of shocks on the structure of the network.

Also, to study the overall changes in these networks, the corresponding cells of each sociomatrix at time t were correlated with the same network at $t+1$ for the entire series, creating a $n-1$ vector of correlations, which may be examined for differences at specific points in time to determine the impact of perturbations on the network (Barnett et al. 2015). The networks were analyzed using R. *Gephi* (Bastian et al. 2009) was used to visualize the international and the WWW networks.

4 Results

4.1 Patterns of individual website usage

We begin by describing the over-time changes in individual websites to examine the impact of critical events on the two networks. We found the critical events had impacts on the use of individual websites. Perhaps the great shock during the fall of 2015 was the Paris terrorist attacks, which took place in the evening of November 13. As seen in Fig. 1a, lemonde.fr, a French news site's worldwide usage rose dramatically from about .2 to over .6% overnight. After the event, the number of users declined reaching its former level by the end of the year. Prior to the attacks, LeMonde was the 22nd most frequently visited website in France. Throughout November, it was more popular, as the eighteenth or seventeenth-ranked site.

Other news sites also responded to the Paris attacks, as well as the incident in San Bernardino California that took place on December 2. Figure 1b shows that nytimes.com's pattern of usage was cyclic with peak viewership on Thursdays. However, at the time of the Paris attacks, its proportion of web users worldwide rose from about 1.12 to over 1.6%. Afterward, it dropped back to previous levels until the attacks in San Bernardino, where it rose again to about 1.35%. Twitter.com and cnn.com showed similar patterns.

Individual shopping sites usage patterns responded to planned shocks and cultural events. Figure 3a shows that amazon.com visits were cyclic, ranging between 9 and

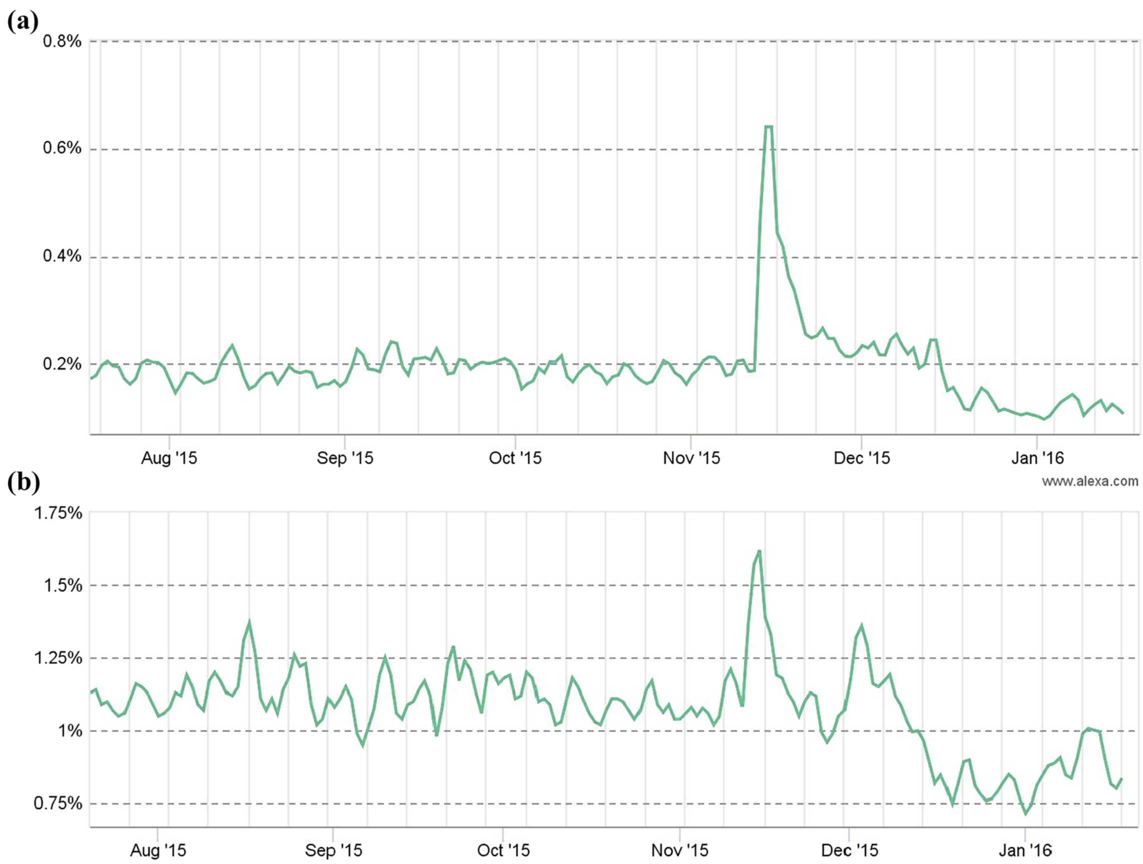


Fig. 1 Global reach (% viewers worldwide) of news websites, July 2015–February 2016. **a** lemonde.fr.. **b** nytimes.com. The New York Times website has a weekly cycle peaking on Thursday. It reached its greatest number of viewers following the Paris terrorist attacks (November 15) and the San Bernardino attacks (December 3)

10%, peaking on Monday and Tuesday. It hit its highest point, 14% of users worldwide, on the Friday after the Thanksgiving holiday, November 27, which is known as "Black Friday" and the beginning of the holiday shopping season in the USA. Its use declined thereafter as Christmas and New Years approached. EBay.com pattern of visits was also cyclic, peaked on Sunday at about 3.25% of users worldwide and bottomed out on Thursday at around 3.0%. Like Amazon, it reached its highest point on Friday November 27, with over 3.75%. Its use then declines toward Christmas and the end of the year.

Tmall.com, a Chinese shopping site, had its greatest number of visitors on November 11 (11/11) and to a lesser extent on December 12 (12/12). The 11/11 shopping day began in 2009, and the 12/12 shopping day started in 2015. They are all newly formed cultural concepts for sales and business in China. For November 11, its percentage of users worldwide rose from about 3.5 to 5.5%, and on December 12, from about 3.0 to 3.5% (see Fig. 2b).

The most frequently visited websites, google.com and facebook.com, also had weekly cycles in their number of visitors. Google (see Fig. 3a) has a weekly cycle that peaks on Tuesday (worldwide, Monday in the USA). Between 43 and 48% of web users worldwide use the site daily. However, toward the holiday season at the end the year, it had fewer visitors. Facebook (see Fig. 3b) also has a weekly cycle, ranging from 36.5 to 39%, with its greatest use on Sunday. Its use at the end of the year is more volatile, declining for Christmas and increasing for the New Year.

Baidu (Fig. 3c), the Chinese search engine, also has a weekly cycle ranging between about 13.25 and 14.25% of users worldwide with a peak on Saturday. Like the Western websites, it bottoms out just before New Year.

While the individual websites change in response to perturbations—shocks resulting from unplanned events, such as the Paris and San Bernardino terrorist attacks or planned events, like special shopping days, holidays and other cultural celebrations, how do these patterns affect the overall structure of the network over time? The shocks to the global system were unplanned and took place at different times. Further, the websites had cycles with peaks

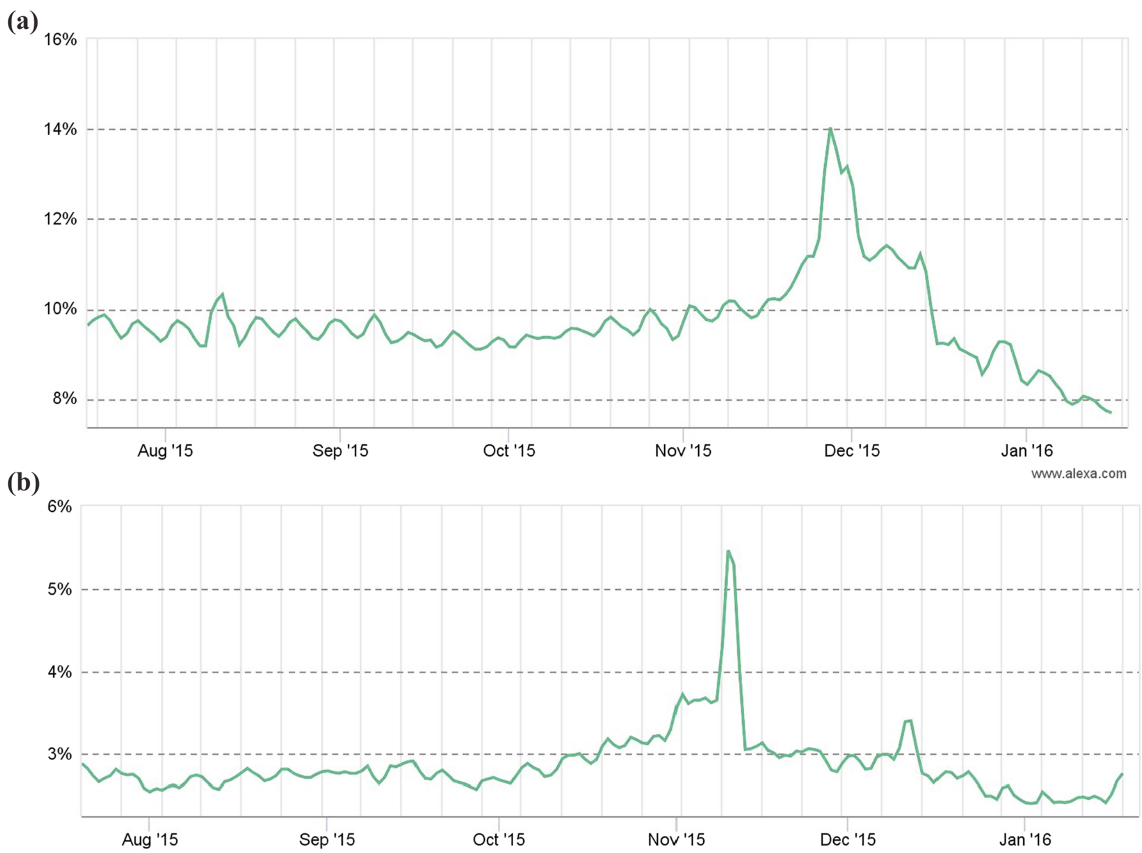


Fig. 2 Global reach (% viewers worldwide) of shopping websites, July 2015–February 2016. **a** amazon.com. It has a weekly cycle peaking on Monday/Tuesday and reaching its lowest point on Friday. The website's greatest number of visitors was November 27, the Friday after Thanksgiving in the USA, the most important shopping days in the USA. **b** tmall.com, a Chinese shopping website. It has its greatest number of visitors on November 11 (11/11), and there is increased viewership on December 12 (12/12), important shopping days in China

Fig. 3 Global reach (% viewers worldwide) of major websites, July 2015–February 2016. **a** google.com, which has a weekly cycle with the greatest number of viewers on Tuesday (worldwide). Its viewership was the lowest on Christmas and New Years. **b** facebook.com. It has a weekly cycle with the greatest number of viewers on Sunday. Its viewership declined on Christmas and rose around New Years. **c** baidu.com, the Chinese search engine. This website has a weekly cycle with the greatest number of viewers on Saturday

and valleys at different times that could result in interference among the patterns negating the weekly cycles in web use. Also, the sites had variable connectivity to different countries and websites. We will now examine the changes in the international network and the WWW network to determine the impact of the shocks, planned and unplanned, the weekly cycles and variable connectivity on overall system.

4.2 The two-mode network

Figure 4 shows the average two-mode network with 425 websites and 118 countries. The size of the website equals the percentage of users worldwide on January 16, 2016. As expected, the USA is at the center of the network with links to the other countries though the various websites. Also, there are separate clusters around USA, China and Russia.

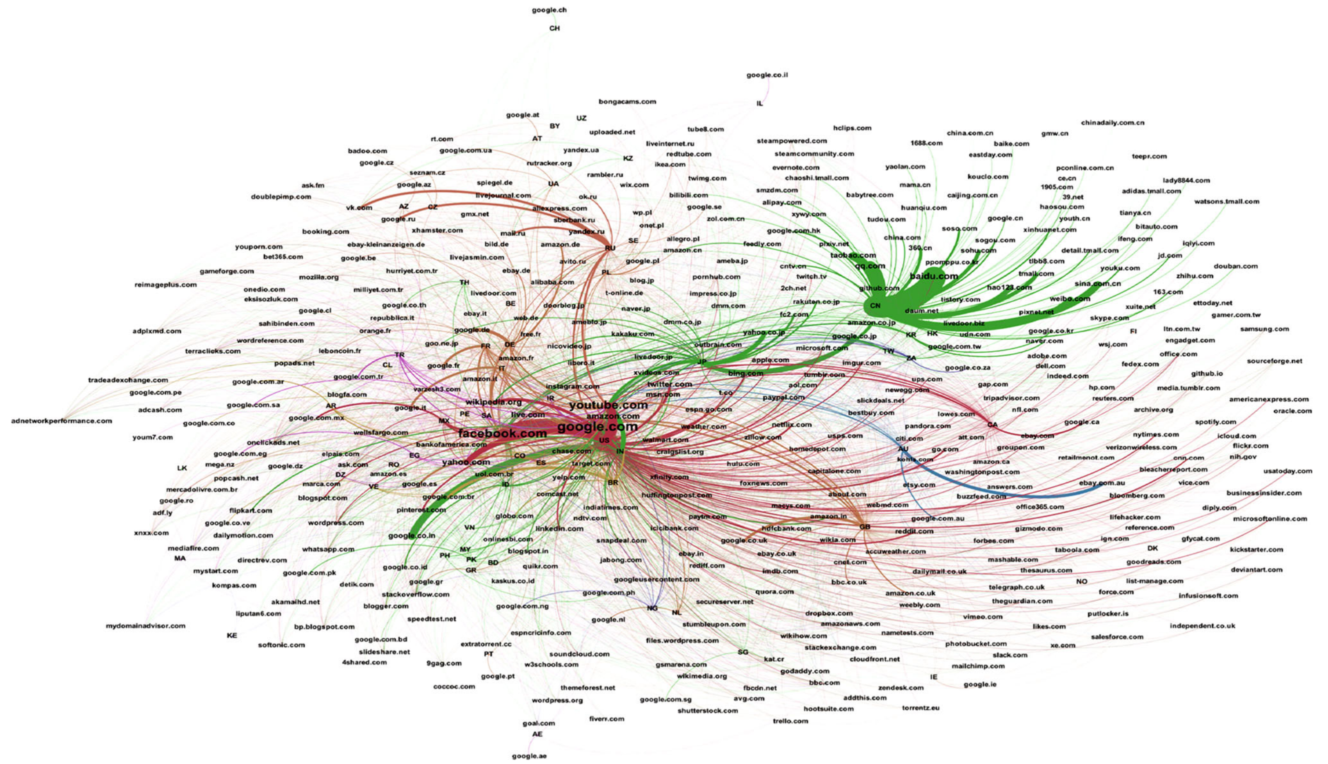


Fig. 4 Two-mode network of the average daily relations from September 1 to December 31, 2015, among countries and websites. Website size is the percentage of users worldwide on January 16, 2016. *Color* represents continent, *red* is North America, *green* Asia, *brown* Europe, *blue* Australia, *violet* Africa, and *yellow* South America

4.3 The country network

The average country network is shown in Fig. 5. It is a center to periphery network with the USA at the center. Again, the colors represent the links between continents. All countries have at least one tie to another based on their shared use of common websites. As indicated above, the average number of websites used by an individual country was about 80. The Bahamas, Montenegro and Palestine used only one site, while the USA makes use of 364. The average density for the period September 1 to December 31, 2015, was .882. In other words, 88.2% of possible links among the countries were present. This is a dense network. As a result, there are no apparent cultural or linguistic clusters, although the European countries tend to be toward the left of the graph. The average transitivity was .889. This means that, on average, the chance that people from three countries used a common website was almost 90%. The average number of triads is 1,423,239.

The eigenvector centralities for the countries are distributed by the power law ($R^2 = .697$, $F = 259.39$, $p < .000$). The most central country was the USA, followed by the UK, India and Canada, all English-speaking nations. Germany and France, other wealthy countries, are the next most central. They are followed by Mexico, which is more central than expected due to its relationship with the more central USA. Table 1 shows the eigenvector centralities for the 20 most central countries, with the USA at 1.0 and UK at .95. The mean centrality was .28.

How did the network of countries based on the shared website use change over time? Its density remained stable. On September 1, it was .871, and by December 31, it reached its maximum at .905, a change of less than .02% per day ($Y = .871 + .00017t$; $r^2 = .680$). The over-time changes in density are displayed in Fig. 6.

Transitivity remained stable. On September 1, it was .871, and by December 31, it was .889, and it reached its maximum at .902 on October 13. For the number of triads, on September 1, it was 1,394,644, and by December 31, it was 1,423,258. It also reached its maximum at 1,444,256 on October 13. The overall time changes in transitivity are displayed in Fig. 7. Besides stability, we found that both the Paris and San Bernardino attacks may have had slight impacts on the transitivity of the country networks. From November 11 to November 13 (Paris attacks), transitivity dropped from .896 to .892. And from November 17 (4 days after Paris attacks) to November 18, the transitivity dropped suddenly from .893 to .884. From December 1 to December 2 (San Bernardino attack), transitivity decreased from .888 to .885 and continued to drop to .881 two days after the attack.

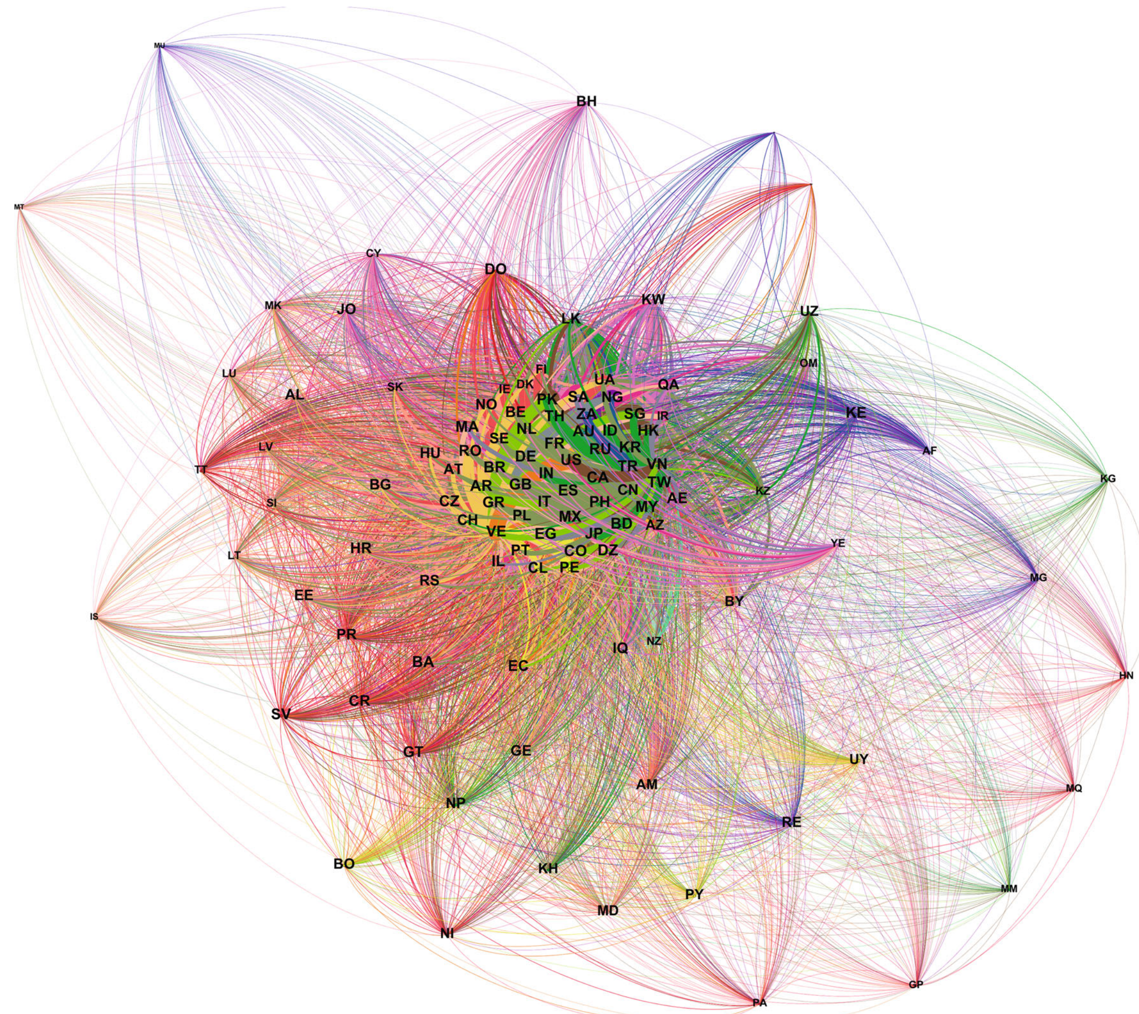


Fig. 5 Country network of the average daily relation to websites from September 1 to December 31, 2015. *Colors* represent the links between continents

The relative positions of the countries were also generally stable. Figure 8 presents the normalized eigenvector centralities of some key countries over time. They include the six most central nations (USA, UK, India, Canada, Germany and France), as well as Brazil, Russia and China who were among the top 20 countries. The USA remained the most central. Its normalized value stayed at 1.0. The UK was stable at .95, and Canada and India at .925, Germany at .90 and Brazil at .875. Russia was slightly more volatile ranging from .84 to .86. The exceptions were France and China. France experienced a slight increase in centrality (.88–.90) beginning with the Paris attacks and remained at the higher level before declining.

China's position in the network was relatively more dynamic. Its centrality increased during September and then declined during the first week in October—"Golden Week," a national holiday and a time for travel. It was stable through early November (the Paris attacks) and rose at the end of the month, before declining until the end of December when it again became relatively more central. Its instability may be explained by the fact that it is less tightly connected to rest of the network than other countries due to the "Great Firewall." The Chinese use their own domestic websites, such as Baidu, QQ, Taobao, Weibo and Tmall. The nationalities of the visitors to the ten most frequently visited Chinese websites averaged 96.9% Chinese. Since it was among the top 20 countries in terms of eigenvector centralities and located in the large group at the center of the international network (see Fig. 9), the dynamics of China's position may impact the overall patterns of use and the resilience of the use of the network.

The analysis of the changes in the relative positions of the nodes in the international network focused primarily on the 20 most central countries. Had the analysis also examined the more peripheral nations more volatility may

Table 1 Eigenvector centrality of the 20 most central countries

Rank	Country	Eigenvector centrality
1	USA	.215
2	UK	.204
3	India	.198
4	Canada	.198
5	Germany	.192
6	France	.191
7	Mexico	.189
8	Brazil	.189
9	Australia	.188
10	Spain	.188
11	Netherlands	.187
12	Italy	.186
13	Russia	.182
14	China	.176
15	South Korea	.172
16	Indonesia	.161
17	Saudi Arabia	.160
18	South Africa	.157
19	Japan	.155
20	Poland	.154
Mean		.280

Mean centrality for the period September 1, 2015–December 31, 2015, based on shared website visits

have been revealed. Therefore, this finding should be viewed somewhat skeptically.

Much of the research on network evolution has focused on the change in community or group structure and the shift in individual nodes' group membership (Backstrom et al. 2006; Gliwa et al. 2013; Singhal et al. 2014). In this case, the number of communities changed from three to five based on the countries' centrality, with additional groups being identified, as the countries became less connected to the core. These changes took place after the Paris attacks as the holiday season approached at the end of November.

4.4 The World Wide Web network

Figure 9 represents the average worldwide web network. Because of the large number of nodes, only those with links strengths greater than the mean plus two standard deviations are presented. This network's links strengths are also distributed by the power law ($R^2 = .794$, $F = 2252.25$, $p < .000$). While google.com made contact with 45% and facebook.com reached 37% of web users world wide, only 92 sites were visited by at least 1.0% of users and 183 by .5%. The average density for the period September 1 to December 31, 2015, was .905, and 90.5% of possible links among the websites were present. This was also a dense center-periphery network. The average transitivity was .937. This means that, on average, the chance that three websites used by people from the same country is almost 94%. The average number of triads is 63,380,282.

How did the website network based on how they were shared by countries change over time? Its density remained stable. On September 1, it was .882, and by December 31, it fell to .830, its minimum, a change of less than .04% per day ($Y = .882 - .000365t$, $r^2 = .656$). As can be seen in Fig. 10, most of this change took place during the holiday season, after December 15.

Transitivity remained relatively stable, but also demonstrated a slightly decreasing trend ($Y = .8652 - .0005t$; $r^2 = .531$), dropping .05% per day. On September 1, it was .842, and by December 31, it reached to the minimum at .767. The number of triads, on September 1, was 63,723,124, reaching its minimum at 58,048,394 on December 31. The over-time changes in transitivity are displayed in Fig. 11. From November 11 to November 13 (Paris attacks), transitivity dropped from .843 to .841 and

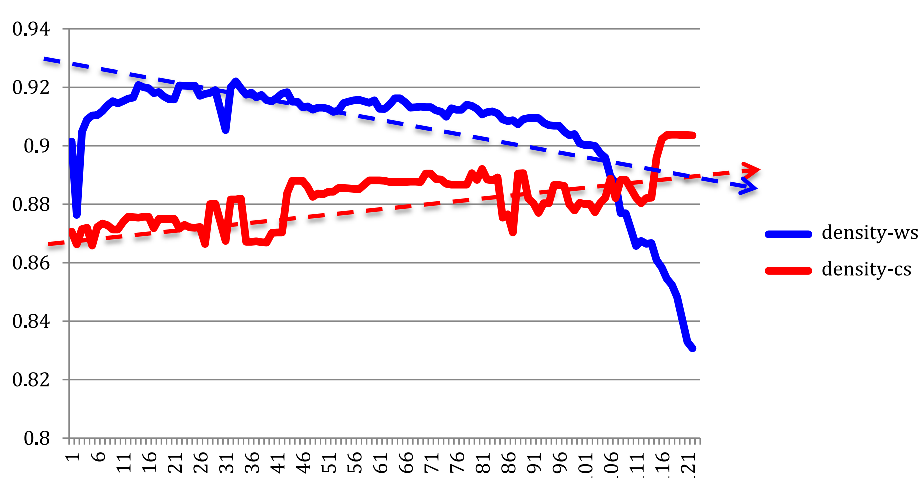


Fig. 6 Network density for the country and website networks over time. Large changes took place around December 12, the beginning of the holiday season

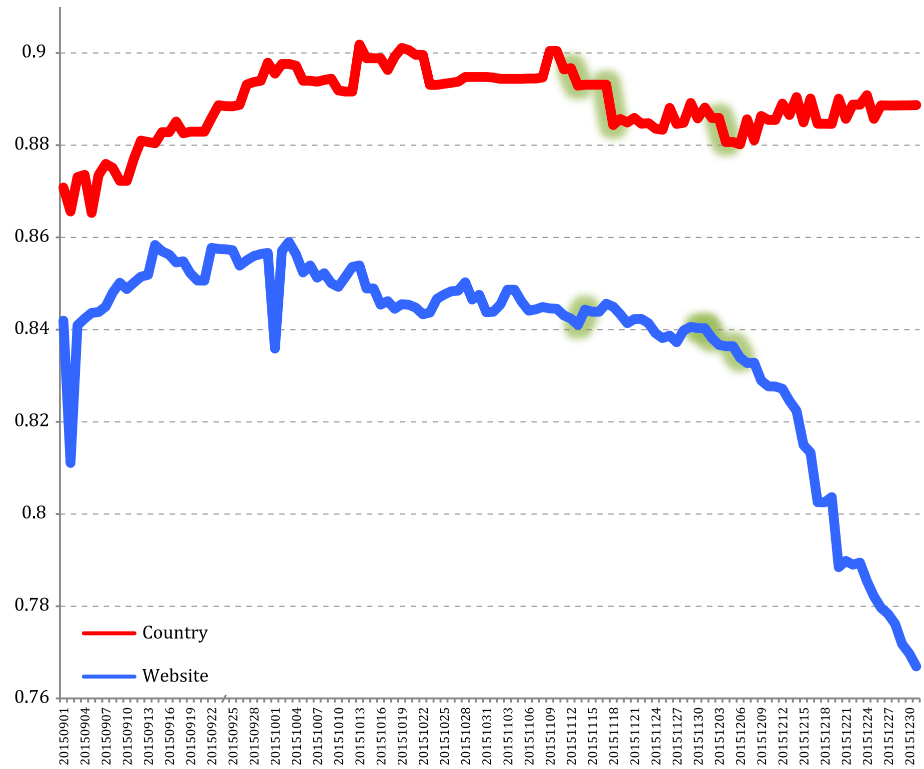


Fig. 7 Network transitivity for the country and website networks over time. Large changes occurred in the website network around December 12, the beginning of the holiday season

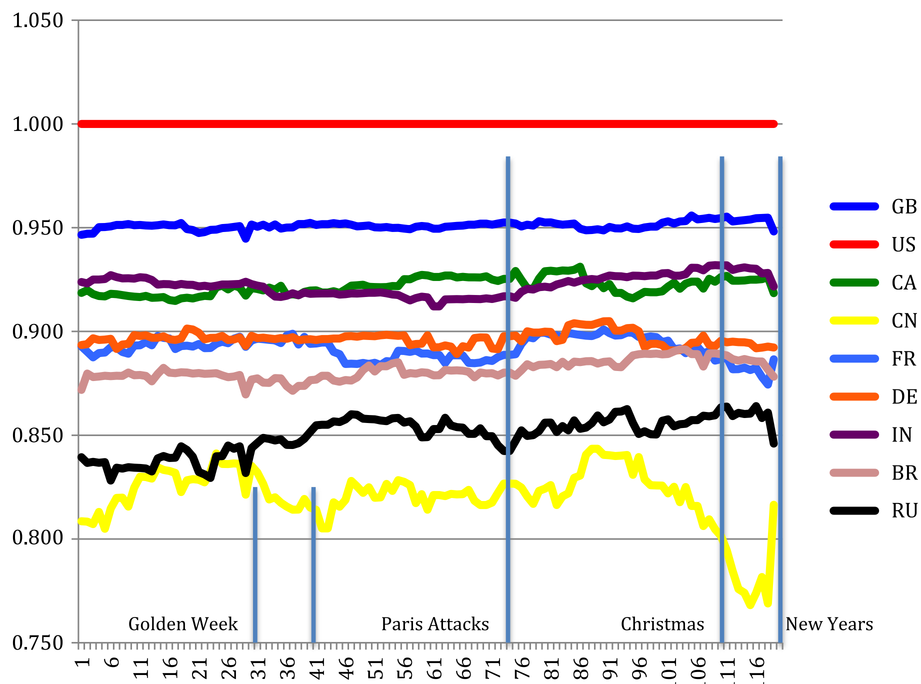


Fig. 8 Eigenvector centrality of key countries over time

then increased to .844 on November 14 (one day after the Paris attacks). From December 1 to December 2 (San Bernardino attack), transitivity decreased from .840 to .838 and then continued to drop to .833 on December 7 (5 days after San Bernardino). These small changes indicate that both the Paris and the San Bernardino attacks may have had slight impacts on the transitivity of the website network. It is also interesting to find the significant correlation ($r = .717, p < .001$) between the transitivity of the website networks and the eigenvector centralities of China in the international networks. This indicates that the unstable positions of China in international networks might have had

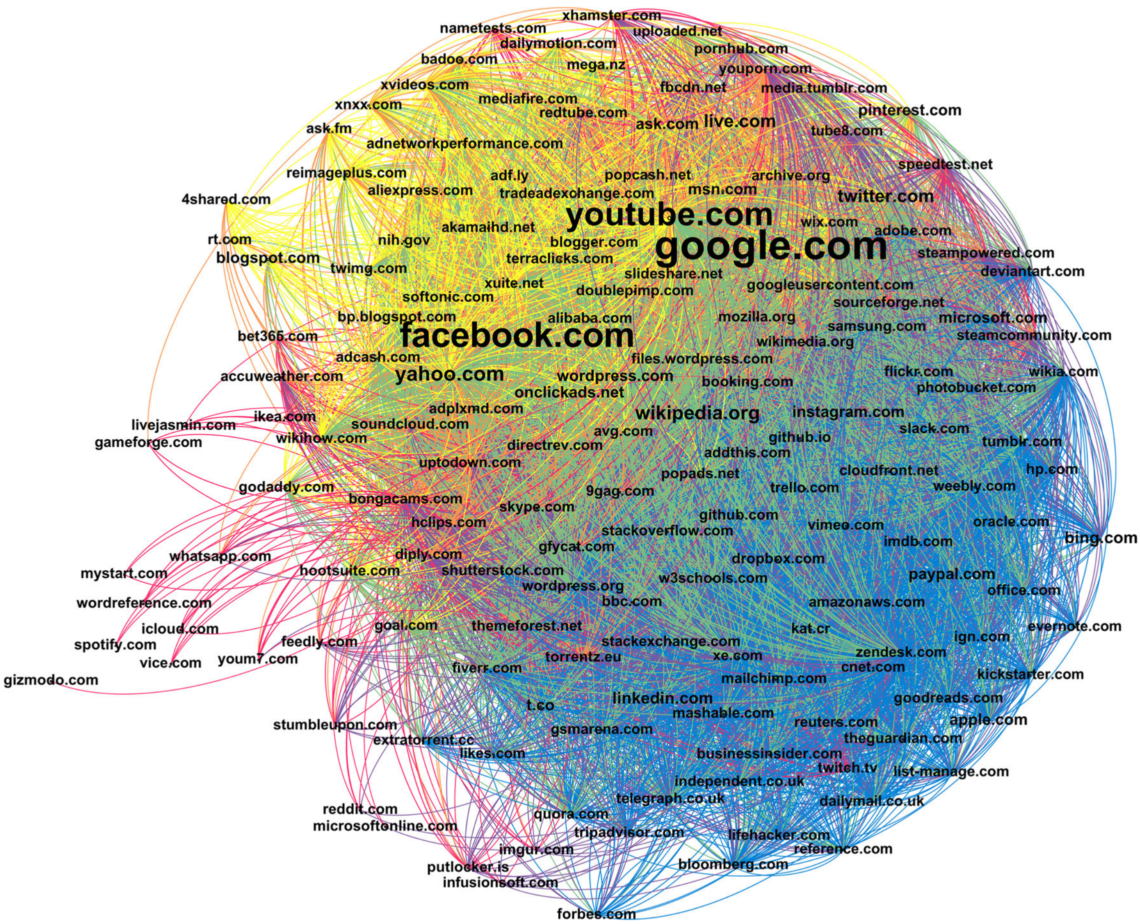
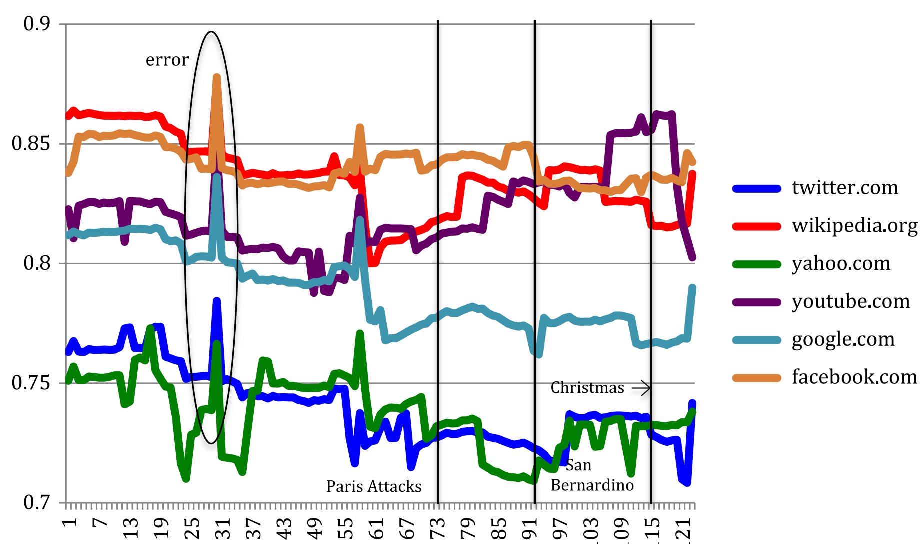


Fig. 9 Website network of the average daily relation to countries from September 1 to December 31, 2015. The size of a site's name indicates its centrality. *Colors* represent the links between continents

Fig. 10 Eigenvector centrality of most frequently visited websites over time. Note the error at times 30 and 58



an impact on the patterns of the use of websites by countries.

Figure 10 shows the normalized eigenvector centrality of the most frequently visited websites over time. While this network seems to be more volatile than the international one, the Paris attacks did not seem to impact these sites, with the exception of Wikipedia. People visited this site to get the most up-to-date news and background

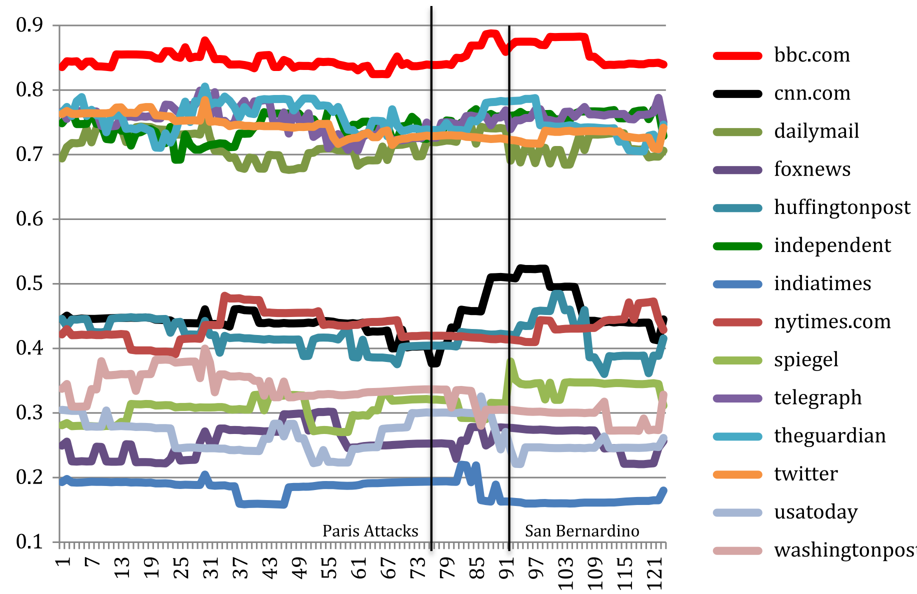


Fig. 11 Eigenvector centrality of news websites over time

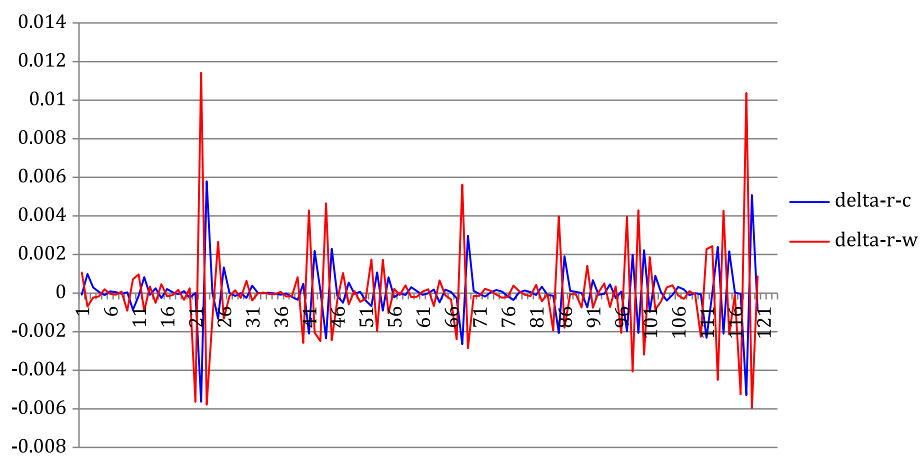


Fig. 12 Change in network correlation over time

information on the events in France (Keegan et al. 2013). The San Bernardino attack seems to impact all sites (Twitter, Yahoo and Google), except YouTube and Facebook. The impact on Wikipedia was especially apparent. These are all US-based sites, and Christmas and New Years appear to have had a big impact. Missing from the figure is Baidu, the fourth most visited website in the world. Its average centrality was only .112 and remained at the periphery of the network. Like the more central nodes, it showed some volatility at the end of the year. Also, note the discrepant values at times 30 and 58. These have been treated as measurement errors in the data and were treated the same as missing data.

The over-time normalized eigenvector centralities of the most frequently visited news sites are displayed in Fig. 11. With the exception of CNN.com, the positions of none of the websites seem to have changed a great deal. CNN centrality rose from less than .4 to over .5 immediately after the Paris attacks. Further, it received an additional boost after San Bernardino. With a delay, the BBC and Huffington Post centrality rose after the Paris attack, and the New York Times and the Guardian rose after San Bernardino. As was the case with Wikipedia, CNN's change is probably due to people seeking immediate information. The other news outlets lag may be attributed to individuals seeking in-depth follow-up analysis. The centralities of the major shopping sites remained stable throughout this time period.

The website network was composed of three communities at the beginning of September, Asian, US and a set of those from other parts of the world. There were five at end of the year. They were: (1) the core Asian sites; (2) the peripheral Asian, mostly Chinese; (3) the most central websites, mostly US based; (4) a community made up of mostly news sites; and (5) a set of others. Because the greatest change in the web network occurred at the end of the year, as it became less dense, it is hard to tell whether the addition of communities was temporary, due to the holiday season or a more permanent development. Additional longitudinal data are required to answer this question.

4.5 Overall changes in country and website networks

To examine the overall changes in these networks, the corresponding cells of each point in time (t) were correlated with the same network at the next point in time ($t + 1$). The over-time correlations indicate that the networks are quite stable despite planned and unplanned shock to the network. For the country network, the mean correlation was .9995 (SD = .0009), and for the web network, it was .9982 (SD = .0041). The differences in the correlations at specific times reveal the impact of perturbations on the networks. Figure 12 displays these differences. Overall, the differences are very small. For the country network, the average difference was only 1.6799E – 06, and for the website network, it was only .00017. A visual examination of the figure suggests that every 22–25 days, there is a spike on the amount of change and country network lags web network by one day. Fourier analysis of density and eigenvector centrality for both the web and country networks confirmed this cycle (Barnett et al. 2015). However, it was not associated with any specific events. The overall network is resilient, absorbing shocks and keeping the network at a relatively stable equilibrium. Further, consistent with other analysis, there appears to be greater change at the end of the year, during the holiday season.

The Fourier analysis also identified a seven-day cycle for the networks' degrees. The average weekly change in link strengths for the countries was 1.75% and 1.40% for the website network. Web use was at a minimum on Sunday and peaked during the workweek. The Fourier analysis also identified a seven-day cycle for the country networks' transitivity. The average weekly range in transitivity for the countries was 7%. However, in spite of these periodic changes, the relationships among the countries and websites return to a state of equilibrium. This suggests that the overall networks are resilient.

5 Discussion

The results of the examination of the two-mode network consisting of the 425 most frequently visited websites by 118 countries over a 4-month period of time indicated that both the network composed of countries linked to the extent that they visited common websites and the international website network, where websites were linked to the extent that visitors came from the same countries, were remarkably stable. This was in spite of planned and unplanned shock during this time period that had an impact on the behavior of individual websites and individual countries. For example, the terrorist attacks in Paris and San Bernardino have a major impact on lemonde.com, wikipedia.com and cnn.com. Golden Week altered the behavior of web use in China and its major shopping site, tmall.com. Black Friday changed the number of visitors to amazon.com and eBay. The holiday season at the end of the year had an impacted a number of websites, as well as having the greatest effect on the overall system as fewer people spent time using the WWW and more time engaging in other activities. The shocks also have slight impacts on the formation of triads of the country and website networks.

Barabási and colleagues (Barabási 2014; Gao et al. 2016) suggested a number of reasons for the network's resilience. One is that the strength of ties and their centralities are distributed according to the power law. These are scale-free networks. Second, resilience is a result of the networks' density. The average densities were .887 and .905 for the countries and websites, respectively. Another indicator of the redundancies in network ties was transitivity. Eighty-nine percentage of country ties and 94% of web links were transitive. Third, the networks are symmetrical. This was due to the data collection methods, which produced non-directional networks. Fourth, the link strengths were heterogeneous. They were distributed according to the power law.

There are a number of shortcomings with this research. First, we only examined exogenous shocks and not endogenous system shocks. For example, what would happen if a major trans-Atlantic or trans-Pacific fiber link were severed that would impede international Internet traffic, or if Google, the most central website, went down for an extended period of time. Barabási (2014) suggested that since it is a hub in the network, the WWW's resilience would be vulnerable to failure.

Second, we did not examine the entry and exit of websites (or countries) into the network. Daily data were collected on the 500 most visited websites, resulting in a total of 619 different sites by March. Only the 425 sites that remained among the most frequented 90% of the time were examined as part of the network. By March, there were only 383 sites from that pool. The decision to analyze the network with only the nodes that met this criterion may have biased the findings and produced a higher degree of stability. Future research is planned to examine how shocks to the system impact the exit existing of nodes and the entry of new nodes into the network.

Third, the Fourier analysis suggests that both networks expand and contract between one and two percentage each week. These oscillations may be a source of instability that could potentially lead to the failure of the WWW, particularly when coupled with an extreme event. Future research is planned to examine the dilation and shrinking of the network and the impact on the system's resilience.

Fourth, four months is too short of a period of time upon which to make knowledge claims about the resilience of the international web. For example, there could be shock far more significant than those that took place over the examined time period. Countries could go to war, or a major earthquake could strike Northern California home of many of the websites. Further, there are other cultural events that were not captured in our data that may have as significant impact as the holidays at the end of the year, such as Chinese New Year, and the summer season when students and the French take vacations. There are plans to continue gathering these data over a much longer time period, for at least one year.

There are practical implications for policy planners wishing to create resilient social systems. First, create systems with networks of redundant ties. Create dense networks. Thus, when shocks, such as natural catastrophes or man-made disasters, occur, the impacted actors will have backup relationships upon which they can rely. For example, if one road is closed due to a toxic spill, there should be alternative routes connecting potential destinations. Second, if the possibility of a shock is equally probable across the system, create large hubs with the surplus resources that can be distributed to outlying actors. In other words, develop scale-free networks. If possible, build these hubs in the safest places and reinforce them to withstand potential perturbations. Third, develop symmetrical (equal) relations among the actors. Make the ties equally reciprocal. That is, have as much traffic (people, information and material) into a node as it sends out. Do not make the larger nodes information, capital or material sinks. They should be transitive with the ability to pass on the resources to the other nodes in the network.

Alternatively, planners may wish to make certain networks more vulnerable to failure, as in the case of dark networks (terrorist or criminal) or social networks that spread diseases during times of epidemics. In these cases, activities that make the network sparse would impede the epidemic and the illegal activity. Quarantines make a network sparser by limiting the social interaction among an at-risk population. Removing the central actor (the hub) would also lead the network to functional failure. For example, arresting a drug cartel leader would make this illegal network more vulnerable to failure and thus less resilient.

In summary, through an examination of the use of the WWW by 118 countries, this article demonstrates that while there are changes in the use of individual websites, due to the planned and unplanned shocks, and other social and cultural events, both the overall website and country networks remained remarkably stable. This resilience is due to the constraints that network ties place on the relationships among its nodes, which limits their potential behavior.

Acknowledgements This research sponsored by Army Research Office/Minerva (Award # w911nf-15-1-0502-0).

References

Backstrom L, Huttenlocher D, Kleinberg J, Lan X (2006) Group formation in large social networks: membership, growth, and evolution. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

Barabási A-L (2014) Linked: how everything is connected to everything else and what it means for business, science, and everyday life. Basic Books, New York

Barnett GA, Park HW (2005) The structure of international Internet hyperlinks and bilateral bandwidth. Ann Telecommun 60:1115–1132

Barnett GA, Park HW (2014) Examining the international Internet using multiple measures: new methods for measuring the communication base of globalized cyberspace. Qual Quant 48:563–575

Barnett GA, Sung EJ (2005) Culture and the structure of the international hyperlink network. J Comput Mediat Commun 11(1):217–238

Barnett GA, Chon BS, Rosen D (2001) The structure of international Internet flows in cyberspace. NETCOM (Netw Commun Stud) 15(1–2):61–80

Barnett GA, Ruiz JB, Hammond JR, Xin Z (2013) An examination of the relationship between international telecommunication networks, terrorism and global news coverage. Soc Network Anal Min 3:721–747. doi:10.1007/s13278-013-0117-9

Barnett GA, Ruiz JB, Park HW (2015) Globalization or decentralization of hyperlinked content among websites: an examination of website co-citations. In: Proceedings of Hawaii International Conferences on System Science (HICSS-48), pp 1779–1787

Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. ICWSM 8:361–362

Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. J Math l Soc 2:113–120

Cohen R, Erez D, ben-AvrahamS. D, Havlin D (2000) Resilience of the Internet to random breakdowns. Phys Rev Lett 85(21):4626–4628

Faloutsos M, Faloutsos P, Faloutsos C (1999) On power-law relationships of the Internet topology. Comput Commun Rev 29:251

Freeman LC (1979) Centrality in social networks: conceptual clarification. Soc Networks 1:215–239

Gao J, Barzel B, Barabási A-L (2016) Universal resilience patterns in complex networks. Nature 530:307–312

Gliwa B, Bródka P, Zygmunt A, Saganowski S, Kazienko P, Koźlak J (2013) Different approaches to community evolution prediction in blogosphere. SNAA 2013 at ASONAM 2013, IEEE Computer Society

Holland PW, Leinhardt S (1970) A method for detecting structure in sociometric data. Am J Sociol 70:492–513

Keegan B, Gergle D, Contractor NS (2013) Hot off the wiki: structures and dynamics of Wikipedia's coverage of breaking news events. Am Behav Sci 57(5):595–622

Monge PR, Contractor NS (2003) Theories of communication networks. Oxford University Press, Oxford

Park HW, Barnett GA, Chung C (2011) Structural changes in the global hyperlink network 2003–2009. Glob Netw 11(4):522–544

Rosen D, Barnett GA, Kim JH (2011) Social networks and online environments: when science and practice co-evolve. Soc Network Anal Min 1:27–42

Ruiz JB, Barnett GA (2014) Who owns the international Internet? Int J Commun 20:1–20

Singhal A, Roy A, Srivastava J (2014) Understanding co-evolution in large multi-relational social networks. IEEE IRI 2014, San Francisco

Valente TW, Coronges K, Lakon C, Costenbader E (2008) How correlated are network centrality measures? Connections 28(1):16–26

Wasserman S, Faust K (1994) Social network analysis. Cambridge University Press, Cambridge