Computational Social Networks
a SpringerOpen Journal

Open Access

CrossMark

Network approach to internet bandwidth distributions

Hyunjin Seo[1][*] and Stuart Thorson[2]

*Correspondence: hseo@ku.edu
[1]William Allen White School of Journalism and Mass Communications, University of Kansas, Lawrence, USA
Full list of author information is available at the end of the article

Abstract

This study examines the communications networks formed by direct international Internet links, weighted by bandwidth capacity, each year over the 2002–2011 period. Specifically, we analyze changes in bandwidth distributions at country, regional, and continental levels during the period and identify network communities at these different levels. We apply an urn-based model developed with country-level data to bandwidth distributions at regional and continental levels. While the 2011 global Internet network closely resembles that of 2002, the network has become more tightly interconnected over time, and the high international bandwidth regions of Northern Europe, Northern America, and Western Europe have seen a modest decline in their share of total global bandwidth. As a consequence, international bandwidth concentration is showing a slow decline. Relative connectedness as measured by percentage of bandwidth staying within UN geographic regions is decreasing, whereas the percentage remaining within the continent has been fairly constant during the analysis period. All of this must be understood in the context of enormous total international bandwidth growth between 2002 and 2011 at all levels of analysis.

Introduction

Information and communication technologies (ICTs) have resulted in important changes, especially with regard to building and maintaining social relationships and producing and sharing information and knowledge. For example, individuals increasingly rely on popular social networking sites such as Facebook or Twitter to stay connected with friends, families, and others [1]. Digital technologies have also had a significant influence on global activism, as they afford rapid mobilization of citizens around the world [2–4]. These changes in the nature of interactions between actors in society are consequences of decentralized information and a cooperative peer production characteristic of the networked information society [5–8]. Transnational Internet-based communication platforms empower individuals in disparate parts of the world to collaborate in producing and sharing information.

According to TeleGeography's annual surveys of Internet traffic and capacity, overall international Internet bandwidth has increased from less than 1 Tbps in 2002 to about 55 Tbps in 2011. International Internet bandwidth refers to the amount of data that can be transferred over the Internet, across national borders, in a given amount of time, and has been argued to be a good indicator of transnational Internet-traffic flows [9].

Springer

Previous studies analyzing global Internet bandwidth found that global Internet connectedness has grown significantly, and the global Internet network has become denser between 2002 and 2011 [10]. In the aggregate, countries have become more tightly connected. Countries that were central in the 2002 global Internet network with a larger number of direct connections with other countries and with relatively large amounts of international bandwidth mainly remained so in 2011.

While previous studies are helpful for understanding the global Internet network, there are several important questions that have not been empirically tested. For example, how, if at all, has the community structure of the global Internet network changed over the past decade? How has global Internet connectedness between regions and continents evolved? To what extent is the significance of geographical distance between countries receding as the Internet develops? Is there increasing concentration of bandwidth capacity at the high end of the distribution or is international Internet bandwidth becoming somewhat more evenly distributed? There are very few studies examining the over-time global distribution of Internet assets at country, regional, and continental levels, as most research has focused on identifying determinants of Internet diffusion or distribution [11–13]. We attempt to fill the gap in the literature by examining these and other issues in the area of global Internet connectedness.

Data and method

We analyze international bandwidth data curated by TeleGeography [14] and purchased from them. International Internet bandwidth is an upper bound on direct country-to-country Internet traffic flow [9]. Using a combination of bandwidth metrics and centrality indicators, we examine bandwidth data over the ten-year period from 2002 to 2011.

To construct the networks reported here, we first broke apart the country name pairs where a direct connection existed that year and for each country added the country, region, and continent ISO codes as used by the Statistics Division of the United Nations Secretariat. These data were read into *R* [15] and converted to edge lists for each year in the 2002–2011 period. This resulted, for each year, in *R* dataframes containing each dyad of countries with a direct bandwidth connection for that year together with the amount of bandwidth capacity shared by that connection. Countries with no reported direct connections were dropped from the analysis for that year. This resulted in 186 countries that shared bandwidth with at least one other country in 2002 and grew to 201 countries by 2011. The *R* package *igraph* [16] was used to build and analyze graphs from each of the edge lists. We report here on networks aggregated at UN region and continent levels as well as the country level networks.

Internet connections transmit data in either direction, and the resultant graphs are treated as undirected. In apportioning bandwidth to nodes (countries, regions, or continents) we adopted the convention of assigning one-half of dyad bandwidth to each node in the pair.

This procedure resulted in 30 network graphs—one for each year in the 2002–2011 period under study for countries, regions, and continents. We ask several key questions of each of the networks. These include which countries are connected to many other countries and which to only a few (degree); which countries have a lot of bandwidth and which have a relatively small amount (bandwidth distribution); and which countries are more *central* in the network. We also discuss global network properties such as network

size and density, and how an urn-based model can be applied to data at regional and continental levels. Finally, and most importantly, each of these questions are examined both for single time points and for possible patterned changes over time.

In the next section we describe properties of these graphs in detail especially as they relate to changes in the interconnectedness of the international Internet. All of this must be understood in the context of enormous international Internet bandwidth growth since 2002. As mentioned above, the total Internet bandwidth grew from less than 1 Tbps in 2002 to 55 Tbps in 2011. While all countries and regions have experienced significant bandwidth increases, the greatest absolute growth in raw bandwidth was generally enjoyed by the countries with the largest bandwidth shares in 2002, as will be shown in the Results section. Understanding patterns of over-time changes in bandwidth distribution requires careful examination of rates of increase of both bandwidth shares and absolute bandwidths at various points in the distributions.

Model

As an initial abstract model for the evolution of bandwidth distributions, we consider Polya's urn. A more complete discussion of our use of this approach is in [17]. The basic problem involves a finite number of urns (or bins) each containing one ball. At each new time point, with some probability q, a new urn is created and a new ball is placed in that urn. Or, with some probability $p = 1 - q$, the new ball is placed in an existing urn in such a way the probability it is placed in a particular urn is proportional to m^γ where m is the number of balls already in that urn. The formalization below follows closely that in [18].

> Assume fixed parameters, $\gamma \in \Re, 0 \leq q \leq 1$) and a positive integer $k > 1$. Imagine k urns each containing some number of balls. Let new balls be added one at a time. For each new ball with probability q add a new urn and place the new ball in it. With probability $p = 1 - q$, assign the new ball to an existing urn in such a way that the probability is proportional to m^γ where m refers to the number of balls already in that urn.

We use a special case of this model to look at the evolution of bandwidth by reinterpreting the urn problem in terms of countries and quanta of bandwidth. Think of each urn as being a country and each ball a quantum of international bandwidth and consider what happens when $q = 0$. Then new countries are never added. The m^γ component of the model can be thought of as capturing *preferential attachment,* and we focus on the simplest case where $\gamma = 1$. We then have what is known as a *finite Polya process*[1].

System evolution can be thought of as the probability density function (PDF) associated with the various paths given initial conditions. At every time point, each country will have some non-zero percentage of the total bandwidth quanta. After dividing by 100, these yield a vector of n probabilities. Each run of the process can be expected to yield different results. It is the probability distribution over these vectors that interests us.

In the case where $k > 2$ (there are three or more countries/urns), the distribution of PDFs is described by a Dirichlet mutinomial distribution with parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$ where α_i denotes the initial proportion of bandwidth/balls of the ith country/urn and where $\sum_{i=1}^{k} \alpha_i = 1.0$. The distribution of the individual PDFs (which is itself a PDF) depends upon the vector of α values.

The results are distributions that roughly mirror the initial probabilities. However, if we were to modify the assumption $q = 0$ and, instead, say that $0 < q < 1$ (that is, permit some new bandwidth quanta to be assigned to countries not previously in the network while continuing to assume $\gamma = 1$) we would get an infinite Polya process, and the bandwidth distribution would take the form of a power law. Seemingly slight changes in model assumptions can produce non-trivial modifications in behavior.

The basic urn model can be thought of as describing what happens with absent external shocks or significant internal policy changes. In such a case, under our assumption of no new countries, preferential attachment would drive the distribution of international bandwidth to a point where country shares would be constant. However, there are shocks such as wars and, on occasion, the addition or subtraction of countries. Moreover, as the importance of bandwidth has increased, more and more countries have introduced policies aimed in part at improving their relative position.

In [17] we term these policy changes as micro-processes which, in the aggregate, can affect country shares over time. In particular, we consider the case where the effects, at the country level, of these micro-processes are Gaussian distributed. Where the means of these distributions were ≈ 0 over the 2002–2011 period, we would expect bandwidth shares to remain fairly constant. We examined the 186 countries and, using a cross-validation approach, found statistical evidence for seven distinct groupings. Four of these had means very close to 0. These included countries below the 70th percentile. Countries above the 70th percentile but below the 99th had positive means and thus increasing bandwidth shares while the very top bandwidth countries had negative means. This is important as it suggests that policies can matter and that the result of these policies has been a modest reduction in concentration of international bandwidth.

Results

Changes in both compression technologies and uses of bandwidth suggest caution in interpreting raw bandwidth. As compression technology continues to improve, a fixed amount of data may require less bandwidth. At the same time, changing uses of bandwidth can result in demand shifts. For example, as video compression technology improves, demand for transmitting videos may increase thus requiring increased capacity. Furthermore, there are country-specific policies which may result in different demand patterns. For instance, at the time of this writing, China is blocking access to Google and related sites. Finally, it is important to keep in mind that our analysis focuses solely on international bandwidth as measured by direct country to country connections. This makes sense given our focus on global interconnectedness, but it also means that we are ignoring domestic bandwidth.
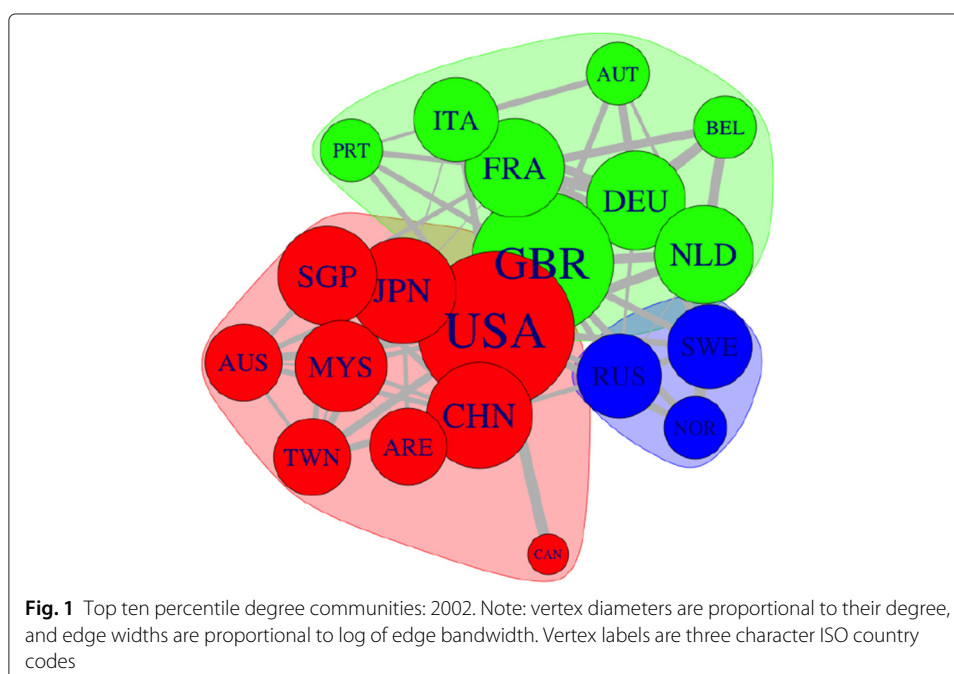
Country analysis

The country level network has countries as nodes with edges between countries where there is a direct international bandwidth connection between them. The bandwidth associated with each edge is treated as the edge's weight. This results in a weighted undirected country-level graph for each of the 10 years of our dataset. As reported above, the number of countries with at least one shared connection grew slightly from 186 in 2002 to 201 in 2011. The number of edges (direct connections) increased from 521 to 766. The median degree (number of direct connections) went from 2 (*mean* $= 5.57$) in 2002 to 4
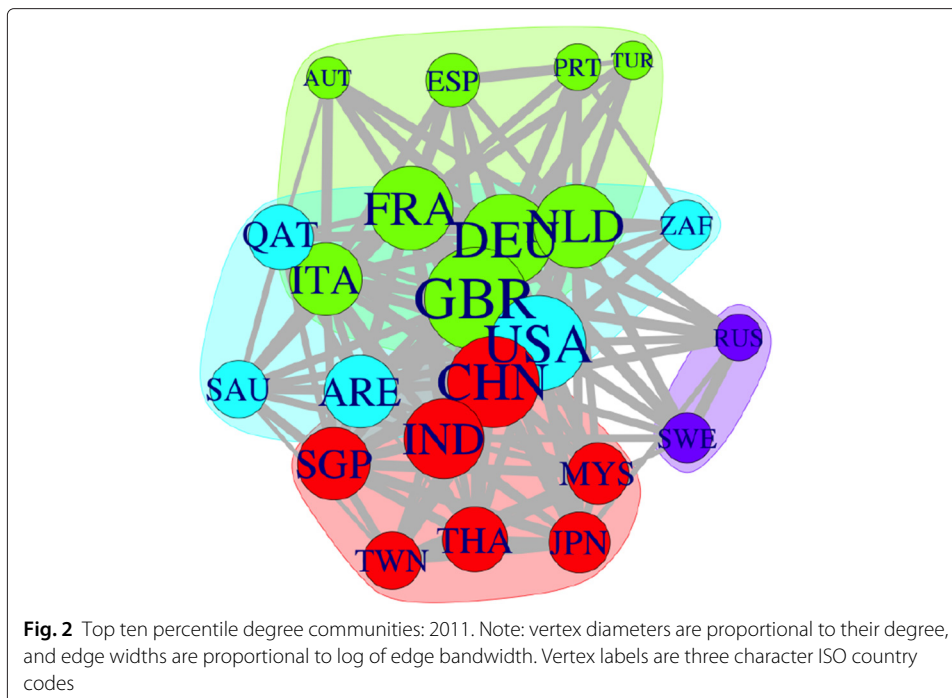
(*mean* = 7.62) in 2011. This is reflected in the increase in graph density from .03 to .08. Over the period, the correlation (Spearman ρ) between country degree rank and country bandwidth rank ranged between .75 and .81 ($p < .001$ in all cases) indicating that high ranking bandwidth countries also tended to be high ranking degree countries. The USA had the highest eigenvector centrality score [19], weighted by bandwidth, in 2002 and 2003. Great Britain had the highest eigenvector score from 2004 to 2011.

It is worth emphasizing the heavy right tail of the bandwidth distribution in each year. Countries at or above the 50th percentile in international bandwidth held over 97 percent of the total international bandwidth, and those in the top ten percentile enjoyed around 83 percent of all international bandwidth.

To visualize the structure of 2002 and 2011 tails, we extracted the subnetworks consisting only of the degree tail countries and then used the Walktrap algorithm with bandwidth as edge weights as implemented in igraph to identify communities within the subnetwork. The basic idea is that communities would be densely connected subnetworks of the degree tails. The Walktrap approach to identifying communities takes random walks (here of length 4) from nodes and identifies communities as those networks that are easily reached by those walks. In our case, we took into account the bandwidth of each connection in identifying the communities. We focused on countries ranking in the top ten percentile as measured by degree. Twenty countries were in the top ten percent (with regard to degree) in 2002 with the largest being the USA (degree = 124) and United Arab Emirates, Australia, Taiwan, Portugal, and Malaysia as the smallest (degree = 12). In 2011, there were 23 countries in the top ten percent. The largest was the USA (degree = 108). The smallest were Spain, Sweden, Turkey, Saudi Arabia, and Qatar (degree = 15).

Results are shown in Figs. 1 and 2. Node diameters are proportional to their degree in the subnetwork, and edge width is proportionate to the log of the bandwidth associated

Fig. 1 Top ten percentile degree communities: 2002. Note: vertex diameters are proportional to their degree, and edge widths are proportional to log of edge bandwidth. Vertex labels are three character ISO country codes

Fig. 2 Top ten percentile degree communities: 2011. Note: vertex diameters are proportional to their degree, and edge widths are proportional to log of edge bandwidth. Vertex labels are three character ISO country codes

with that edge. Node colors reflect community membership. Connections shown are only those within the subnetwork.

Several things are interesting. First, communities, not surprisingly, reflect spatial geography. Countries near one another tend to be in the same community. In 2002, three communities were identified—Asia Pacific, Western Europe, and a Nordic one consisting of Russia, Sweden, and Norway. It is noteworthy that in 2002, the USA and Canada were identified as being in a community with mostly Asia Pacific countries along with the United Arab Emirates. For 2011, we identify the following four communities: Asia Pacific, Western Europe, Nordic, and Middle East. The first three communities are similar to those found in 2002, but now there is also a largely Middle East community (United Arab Emirates, Qatar, and Saudi Arabia) together with the USA and South Africa.

This analysis of communities within the top ten percentile tail-degree countries provides important insight. The number of communities within the subnetwork increased from three to four between 2002 and 2011. While this finding is consistent with previous research showing growth of Internet connections in the Middle East [10, 20], it also provides additional empirical support for the emergence of these Middle Eastern countries as an important community (using node degree as the indicator) within the upper ten percentile tail of the global Internet. This may help us better understand political and social movements occurring in the region that may be facilitated by more tightly connected digital media-based collaborative networks with the larger international Internet.
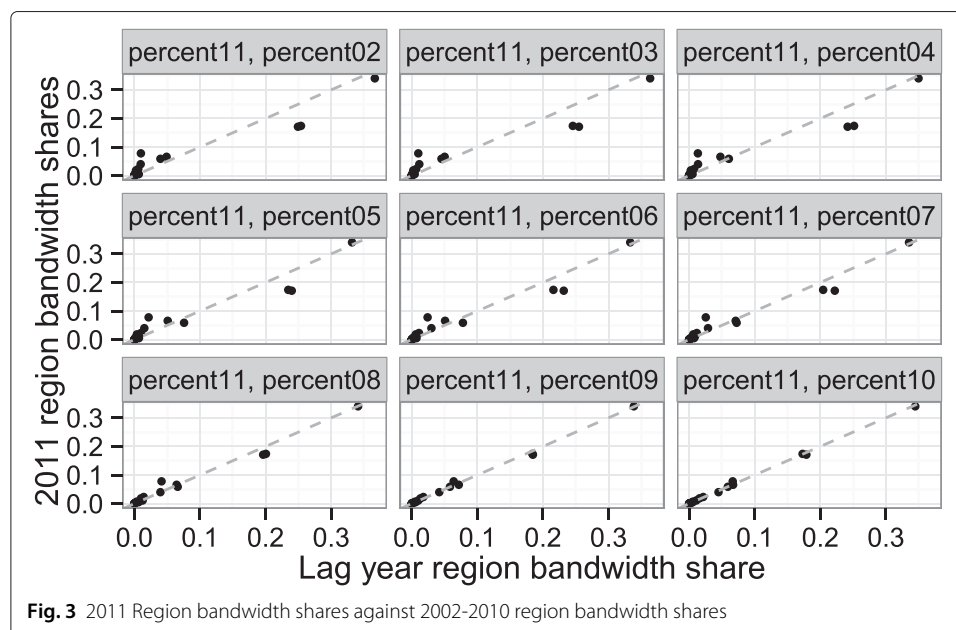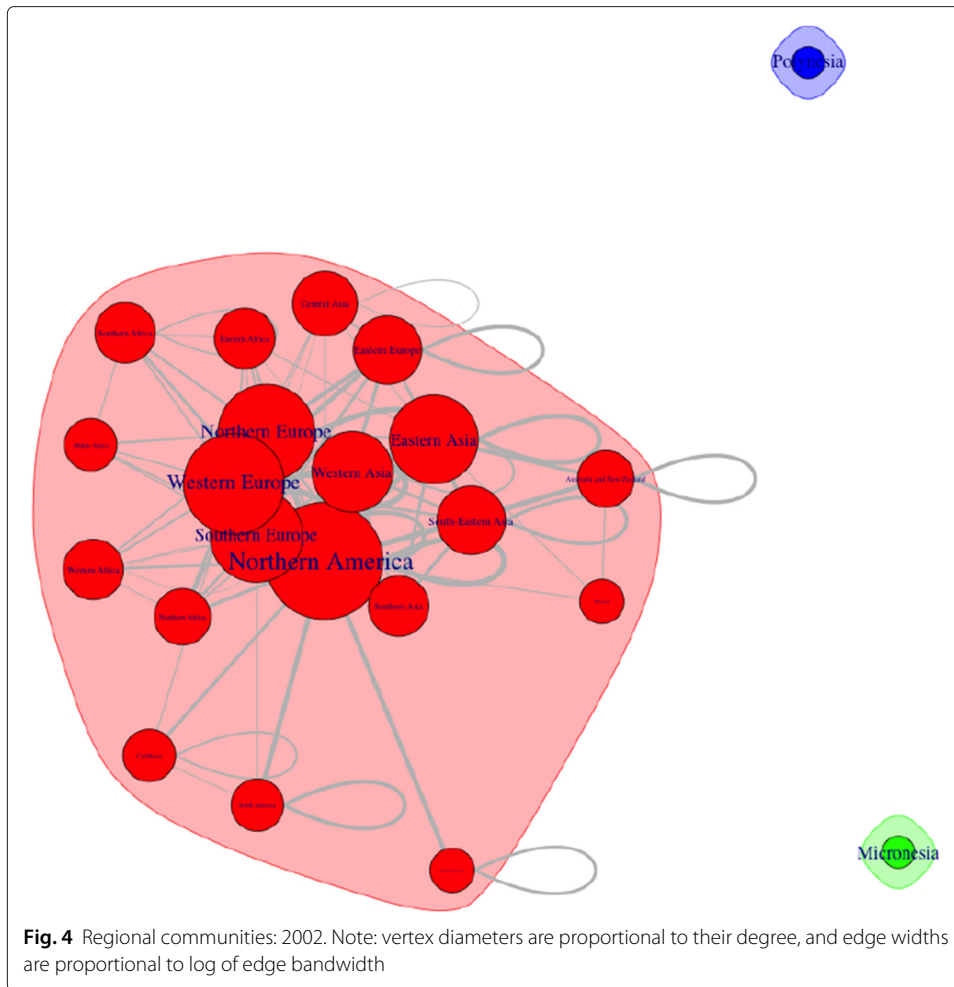
Region and continent analysis

To construct regional networks, we associated each country with its UN region. This resulted in 22 region nodes such as: Australia and New Zealand, Caribbean, Central America, Central Asia, Eastern Africa, Eastern Asia, Eastern Europe, Melanesia,

Micronesia, Middle Africa, Northern Africa, Northern America, Northern Europe, Polynesia, South America, South-Eastern Asia, Southern Africa, Southern Asia, Southern Europe, Western Africa, Western Asia, and Western Europe. For each region, we calculated both the bandwidth remaining within the region (for example, a Germany-Austria edge would result in bandwidth being assigned as internal to the Western Europe region) and the bandwidth going outside the region (the Germany-Sweden edge would connect the Western Europe and Northern Europe regions). In 2002, the regional network had 22 nodes and was interconnected by 91 edges. In 2011, there were the same 22 nodes and now 110 edges. Network density increased slightly from .36 (2002) to .43 (2011). The Western Europe region had the highest eigenvector centrality score, weighted by bandwidth, in both 2002 and 2011. Not surprisingly, regions are much more tightly connected than were individual countries.

Figure 3 shows the empirical relation between 2011 regional bandwidth shares and those of each year going back to 2002. The dashed line represents the prediction that would follow from the basic (that is, no micro-processes or external shocks) urn model. As can be seen, the 2009 and 2010 shares very closely resemble those in 2011. This is consistent with simple preferential attachment with no new countries. And, not surprisingly, the further back we go, the more deviations we see suggesting that other factors are in play. The substantive interpretation here is similar to what we reported at the country level [17]. The large regions, Northern Europe, Northern America, and Western Europe (despite the growth of Germany) showed reduced shares in 2011 as compared with 2002 while Eastern Europe and South America had the largest share gains over that same period.

Figures 4 and 5 provide a visualization of the community structures identified for 2002 and 2011. In 2002, there was essentially one large community consisting of all the regions except for Micronesia and Polynesia. By 2011 two large communities had emerged. One

Fig. 3 2011 Region bandwidth shares against 2002-2010 region bandwidth shares

Fig. 4 Regional communities: 2002. Note: vertex diameters are proportional to their degree, and edge widths are proportional to log of edge bandwidth
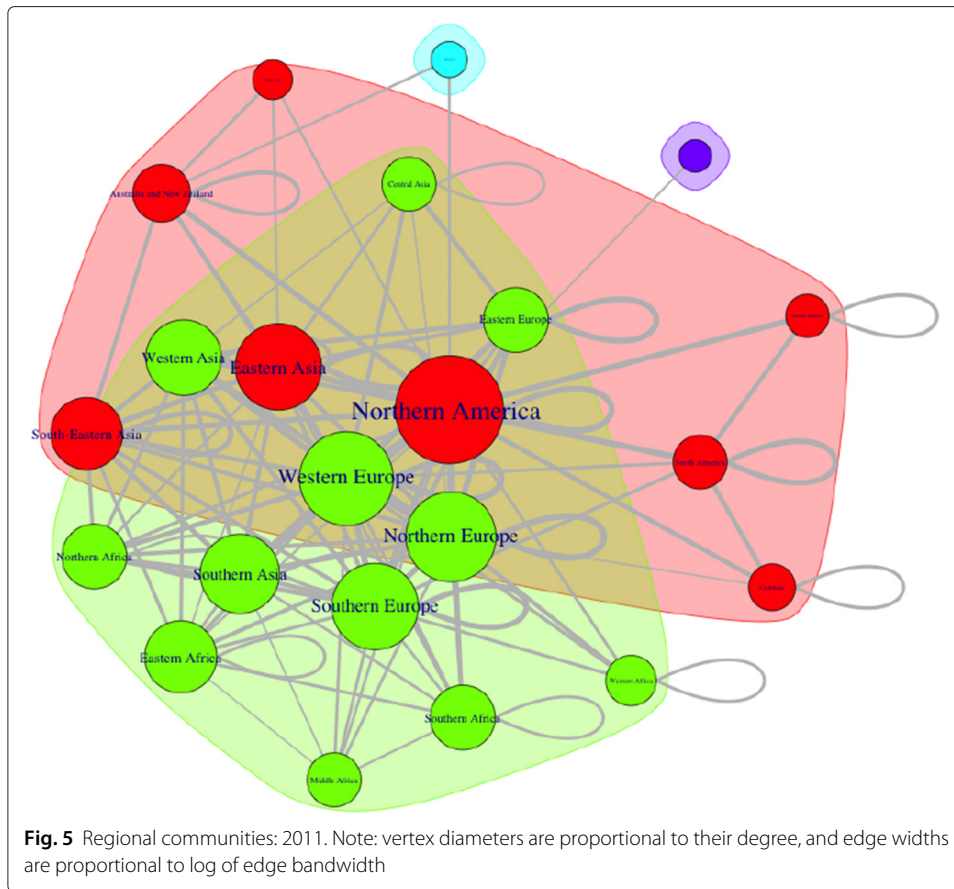
was comprised of the Americas, most of Asia, and Australia and New Zealand. The other included the European regions, Africa, and South Asia.

Continent networks were constructed in an analogous manner to region ones using each country's continent as assigned by the UN. This resulted in the following five nodes: Africa, Americas, Asia, Europe, and Oceania with 13 edges in 2002 and 14 by 2011 with the addition of an Oceania-Europe edge. Not surprisingly, given the level of aggregation, network density increased—going from .87 to .93. Also, as expected, no communities were identified at the continent level. Europe had the highest eigenvector centrality score, weighted by bandwidth, in both 2002 and 2011.

More interesting results can be seen in Table 1. The column labeled *Region* shows the proportion of total international bandwidth staying within the region. Note that this proportion *decreased* from .41 to .31 over the period. This suggests that as total bandwidth increased, an increasing proportion of it was being allocated to connections outside the region. This is consistent with increasing interconnectedness. In contrast, as reflected in the *Continent* column, at the continent level, the proportion of bandwidth staying within the continent actually increased slightly. On average, countries are increasingly directing bandwidth outside their local region. However, that bandwidth is generally remaining on the same continent. The Americas (.49 in 2002 to .58 in 2011) and Asia (.39 in 2002 to .47

Fig. 5 Regional communities: 2011. Note: vertex diameters are proportional to their degree, and edge widths are proportional to log of edge bandwidth

in 2011) had the largest increase in the proportion of their continent's international bandwidth being directed within the continent. Europe remained fairly constant going from .85 to .87 over the period.

Conclusion

Information and communication technologies (ICTs) have important implications for global political, economic, social, and cultural relationships. To help understand these

Table 1 International bandwidth proportion within region and continent

Year	Region	Continent	Capacity (Mbps)
2002	0.41	0.72	931424
2003	0.36	0.70	1759866
2004	0.34	0.71	2512052
2005	0.33	0.70	3548382
2006	0.33	0.71	5133354
2007	0.31	0.70	8714300
2008	0.31	0.73	14654850
2009	0.31	0.74	24082842
2010	0.31	0.75	37269263
2011	0.31	0.76	54855718

Note: The Region column shows the yearly proportion of international bandwidth within the U.N. regions and the Continent within the U.N. continents. The Capacity refers to the total international bandwidth each year

implications, we analyzed global Internet connectedness, operationalized by international bandwidth links, at country, regional, and continental levels. In particular, we analyzed longitudinal patterns of distribution of international Internet bandwidth that enable us to predict future growth patterns and understand consequences of these changes for future communication behavior. Taken together, our analyses offer several important results. First, our regional findings show that the proportion of international bandwidth allocated to connections outside the region has increased between 2002 and 2011 indicating increased Internet-based communications across regions compared with those within regions. In comparison, the proportion of international bandwidth allocated to connections outside the continent has slightly decreased. The greatest increase in global interconnectivity occurred within continents rather than between them for two of the bandwidth richest continents, the Americas and Asia. Europe, the most central (as measured by weighted eigenvector centrality), showed a constant proportion of bandwidth directed within the continent over the period.

Of particular interest is the emergence of two major regional communities by 2011—one Eurocentric and the other centered on Northern America and Eastern Asia. This is consistent with US policy pronouncements indicating a tilt toward Asia as well as policies in Eastern Asian countries including China, South Korea, and Japan directed at projecting a regional or global presence. Given the lead time involved in adding significant bandwidth, this may provide further support for the notion that shifting bandwidth patterns can serve as leading indicators of deeper socio-political changes.

While we identified interesting changes in both bandwidth distribution and community structure, these changes, with the exception of several Middle Eastern countries, took place largely within continental boundaries. More generally, the underlying network structure became a bit more complex and interconnected, as the international Internet saw large annual bandwidth capacity increases between 2002 and 2011. At the same time, high bandwidth regions—Northern Europe, Northern America, and Western Europe—saw their share of global bandwidth declining a bit while other regions saw modest share increases. This finding may have to do with both structural political change (Eastern Europe region) and sustained policies aimed at increasing ties to the global economy (Eastern Asia) [17].

Our analysis focused on bandwidth data aggregated for UN regions and continents. The supra-national aggregations are interesting given the packet-switched nature of the Internet. For example, a low bandwidth country can benefit from being geographically near high bandwidth ones while low bandwidth countries in low bandwidth regions and/or continents, for example Africa, are at an even greater relative disadvantage. While there has been modest change in the relative positions of some continents and regions, major change has not occurred among those at the very top of the bandwidth distribution. That said, our data do support the conclusion that bandwidth-share concentration has been decreasing over the period covered by this study.

This study contributes to understanding international communication networks and the global digital divide by examining changes in international Internet bandwidth distributions at country, regional, and continental levels. Results should inform studies examining determinants of global Internet growth and diffusion [11, 12]. Moreover, they can help policymakers better identify appropriate strategies for addressing global communication issues such as digital divide.

Endnote

[1] This is similar, though not identical to the situation in the global Internet where there were 186 countries in 2002 and 201 countries by 2002. In our empirical analysis we considered only countries with reported bandwidth for the entire 10 year period of study so we basically, if somewhat artificially, meet this condition.

Author details
[1] William Allen White School of Journalism and Mass Communications, University of Kansas, Lawrence, USA. [2] The Maxwell School, Syracuse University, Syracuse, New York, USA.

References
1. Duggan, M, Ellison, NB, Lampe, C, Lenhart, A, Madden, M: Social media update 2014. Pew Research Center. Washington D.C. (2014)
2. Chadwick, A: Internet politics: states, citizens, and new communication technologies, Oxford University Press, USA (2006)
3. Mefalopulos, P: Communication for sustainable development: applications and challenges. Media and glocal change. In: Rethinking communication for development, pp. 247–260, CLACSO, Buenos Aires, (2005)
4. Naude, A. M, Froneman, J. D, Atwood, R. A: The use of the internet by ten south african non-governmental organizations—a public relations perspective. Public Relations Rev. **30**(1), 87–94 (2004)
5. Barabási, AL: Linked: How everything is connected to everything else and what it means. Penguin Group, New York (2003)
6. Benkler, Y: The wealth of networks: how social production transforms markets and freedom, Yale University Press, New Haven (2006)
7. Castells, M: The information age: economy, society and culture. The Power of Identity, Vol. 2. John Wiley and Sons, West Sussex, UK (2010)
8. Sunstein, CR: Infotopia: How many minds produce knowledge, Oxford University Press, Oxford, UK (2006)
9. Barnett, GA, Park, HW: The structure of international internet hyperlinks and bilateral bandwidth. Ann. Telecommun. **60**(9), 1110–1127 (2005)
10. Seo, H, Thorson, SJ: Networks of networks: Changing patterns in country bandwidth and centrality in global information infrastructure, 2002–2010. J. Commun. **62**(2), 345–358 (2012)
11. Andrés, L, Cuberes, D, Diouf, M, Serebrisky, T: The diffusion of the internet: a cross-country analysis. Telecommun. Policy. **34**(5), 323–340 (2010)
12. Guillén, MF, Suárez, SL: Explaining the global digital divide Economic, political and sociological drivers of cross-national internet use. Social Forces. **84**(2), 681–708 (2005)
13. Beilock, R, Dimitrova, D. V: An exploratory model of inter-country internet diffusion. Telecommun. Policy **27**(3), 237–252 (2003)
14. TeleGeography: Global Internet geography. Tech. rep., PriMetrica, Inc. Washington D.C. (2011)
15. R Core Team: R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2013)
16. Csardi, G, Nepusz, T: The igraph software package for complex network research. Inter. J. Complex Syst, 1695 (2006)
17. Seo, H, Thorson, S: A mixture model of global Internet capacity distributions. J. Assoc. Inf. Sci. (in press)
18. Chung, F, Handjani, S, Jungreis, D: Generalizations of Polya's urn problem. Ann. Comb. **7**(2), 141–153 (2003)
19. Bonacich, P: Power and centrality: a family of measures. Am. J. Sociol **92**(5), 1170–1182 (1987)
20. Howard, PN: The digital origins of dictatorship and democracy: information technology and political Islam. Oxford University Press, Oxford, UK (2010)