Eurographics Conference on Visualization (EuroVis) (2015)
E. Bertini, J. Kennedy and E. Puppo (Editors)

Analyzing the Evolution of the Internet

Thienne Johnson[1,2], Carlos Acedo[2], Stephen Kobourov[1] and Sabrina Nusrat[1]

[1]Dept. of Computer Science, University of Arizona, USA
[2]Dept. of Elec. & Computer Engineering, University of Arizona, USA

Abstract

Existing representations of the Internet do not provide information on why countries have a bigger Internet presence (e.g., Internet Service Providers) than others. In this paper we evaluate four geo-economic parameters (area, population, GDP and GDP per capita), looking for clues of why some areas or countries have developed earlier/later, faster/slower than others. We use correlation studies to analyze which geo-economic variable leads to bigger development in the Internet infrastructure per continent, and cartograms to represent the growth of the Internet infrastructure around the world, in a sequence of 24 years. These representations make it possible to find interesting patterns and identify outliers.

1. Introduction

The Internet began as a research project in 1969 with four supercomputers, and evolved rapidly over the years to connect research and military institutions. In 1991, the WWW was presented to the public [Wik15], and in 1993 the Internet became available to the general public; in 2015 the number of Internet users around the world is calculated around 3 billion [Int14]. For Internet Service Providers (ISPs), anticipating and accommodating the rapidly shifting traffic demands has been a technological, economical, and political challenge [Ver13]. Thus far, this challenge has been met in an "organic" fashion, for the most part, based on unilateral actions of many different players such as ISPs, content providers, public policy makers, international organizations, and large enterprises. This symbiotic relationship among many, and often competing change factors, has led to a system of enormous complexity that was not a product of well-founded engineering principles. Considering this scenario, is it possible to characterize how the Internet evolved over time? Is there a strong correlation between economics and evolution of the Internet? What led some countries to develop their Internet infrastructure before others? Can we predict what happens next?

The Internet topology has been extensively analyzed, but as the Internet evolves over time, new studies are necessary to understand the Internet's infrastructure, the elements that compose and influence it, and the systematic new *phenomena* related to its expansion. The Internet infrastructure is composed, on a high level view, of Autonomous Systems (AS). ASes are networks under a single administrative, and often business, authority commonly referred to as ISPs; they provide Internet access to end users or data exchange between multiple ASes. At present, the Internet is composed of approximately 67,000 ASes [Eur15a]. The Internet topology is often visualized as a graph, with the ASes as nodes and the connections between ASes as edges; early attempts to model the Internet rely on graph models, studies of graph properties and metrics [MKF*06, OZZ07]. Some provide and high-level overview of the Internet topology at the AS level [FJS*14, CAI], while others aim for detailed views [BBP08, OLZ05], including user devices. Such representations do not provide information on possible reasons for why some countries have bigger Internet presence than others, or how a country or a continent has been developing Internet infrastructure over the years. In the Internet Maps (iMaps) of Fowler et al. [FJS*14], maps of the world are modified by moving countries around and changing their areas, in order to better represent the Internet infrastructure (nodes and links). That study raised questions such as: Why is a country bigger than others? What explains why a country with a small GDP has a big network infrastructure? Those questions lead us to investigate whether there are correlations between geo-economic parameters and infrastructure size. We chose contiguous cartograms as an effective and familiar (popular in news media and blogs) tool to represent our geo-referenced data.

In this paper we evaluate four geo-economic parameters (area, population, GDP and GDP per capita) of the

DOI: 10.2312/eurovisshort.20151123

world's countries and continents, looking for clues about why some areas have developed more than others, and earlier than others. Among the many geo-economic parameters that can be used (e.g., inequality of wealth, economic structure, demographics, access to education) we chose GDP, GDP per capita, area and population as they are commonly collected and available in census data for economic growth analysis. Correlation plots provide year by year information about the increasing or decreasing correlation between the geo-economic variable and the existing number of ASes. Cartograms embed information in the contours of a world map, using our existing assumptions and familiarity with the actual shape of the world to let us make inferences about the variability of the measured parameters [NAK15]. Cartograms showing the Internet and geo-economic growth, together with correlation plots, with multiple levels of information, help us analyze where, when, and how the Internet has evolved. Studies of the Internet topology growth usually rely on scatter plots, and bar graphs, and pie charts [MKF*06, HFC12, LHC*13]. While such visualizations are good for making comparisons, it is difficult to make geo-economic inferences since only the physical and logical Internet topologies are taken into consideration. Cartograms provide a great advantage when showing statistical information that has associated geographic location. Thus, while earlier studies report information on growth, patterns and trends, they cannot do a good job of showing why, where and when the related growth has happened.

The contributions of this paper are: (1) the use of correlation studies to analyze which geo-economic variable leads to bigger development in the Internet infrastructure per continent; (2) the use of cartograms to represent the growth of the Internet infrastructure around the world, in a sequence of 24 years, from 1990 to 2013 (datasets containing number of networks per country contain information since 1990); and (3) the dataset itself along with the tool to generate cartograms based on ASes and geo-economic parameters (available along with videos illustrating the evolution at our companion website `http://internetevolution.cs.arizona.edu/`).

2. Correlation

We compiled a dataset composed of yearly geo-economic variables for 195 countries (from the WorldBank website [Wor15]) and ASes statistics (number of ASes per country, per year) [Eur15a]. We parsed and merged both datasets, and excluded countries with missing information in our final dataset (available at the companion website). We then evaluated the correlation between number of ASes and the geo-economic variables by employing the Spearman's rank correlation (Spearman ρ) coefficient [spe08] used for non-parametric measurement correlation. It is used to determine the relation existing between two sets of data. There is a positive correlation when the large values of X have a tendency to be associated with large values of Y and small values of X with small values of Y. There is a negative correlation when large values of X have a tendency to be associated with small values of Y and vice versa. With this coefficient, we can analyze if a given geo-economic variable has a strong (e.g., the richest countries have the bigger number of ASes) or low correlation to the number of ASes, per year.

Fig. 1 shows the correlation plots for the world and 5 continents. Over the 24 years in our dataset the best correlation for Internet growth is the GDP, showing the (expected) tendency that the richer the country, the bigger its infrastructure (Fig. 1a). GDP has the highest correlation for North America (also includes Central America in this study), and Africa (Fig. 1b-c). We expected that GDP per capita would provide a good correlation, but this is not the case for any continent expect Oceania (graph not included), where it has a higher correlation (greater than 80%, and greater than 94% in the last years). The situation for the remaining continents is somewhat surprising. Europe (Fig. 1d) is better represented by population (the bigger the population, the bigger the number of ASes). The distribution of ASes in European countries did not have a high correlation until 1998, when the most populated countries were the ones having a bigger internet growth, and this tendency continues the same way today. For South America (Fig. 1e), it is interesting to note how the number of ASes and GDP, area and population have similar correlations over the evaluated years, demonstrating that the bigger countries have more networks, and are also the richest and populated (e.g., Brazil and Argentina) and the smaller ones are also the poorest and less populated. The correlation with all parameters increases after year 2000, as a result of the strong economic development in the region. The patterns in Asia (fig. 1f) are quite different from all the others. All parameters have low correlation coefficient with number of ASes, thus no parameter is a good predictor of bigger or smaller network infrastructure. The reasons behind such low correlations can range from economic (including costs to build long-distance physical connections to other Internet nodes), social/religious (low technology acceptance, thus no incentives to grow the networks) or political (governments restricting Internet access). There is an indication that in the last few years the correlation with population is getting stronger, thus countries are growing their infrastructure more proportionally to their population.

To understand better what happens in all regions, more information is needed to find explanations for the Internet evolution over the years. We next show how cartograms can complement the correlation plots.

3. Cartograms

A cartogram is a thematic representation of geographically distributed data on a planar map. Here geographic regions such as countries or provinces are scaled so that their areas are proportional to the data associated with them, while

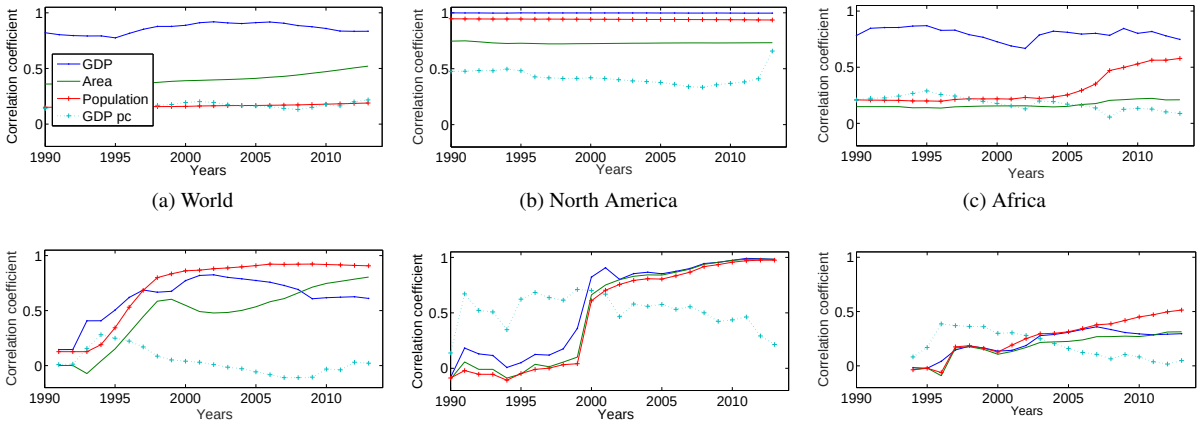


Figure 1: Plots showing correlation in different parts of the world over time

the overall map remains recognizable [Tob04]. This kind of visualization has been used for many years to represent census data (e.g., population or GDP) and to visualize election results and other geo-referenced statistical data. Contiguous cartograms stretch the boundaries of the original geographic map in order to realize the desired areas and were popularized by Gastner-Newman [GN04]. In this study, the cartograms are constructed using *d3.cartogram* [Eur15b] which is based on Dougenik et al. [DCN85]. The dataset from the previous section was used to create the cartograms, which have the following properties:

- The size of the countries reflect the number of ASes, and each country starts with its real physical area and distorts to reach its desired area, with respect to the percentage of number of ASes over the total number of Ases;
- A 4 color white-to-blue scale is used to represent the magnitude of the geo-economic parameter (Area, Population, GDP, GDP per capita).

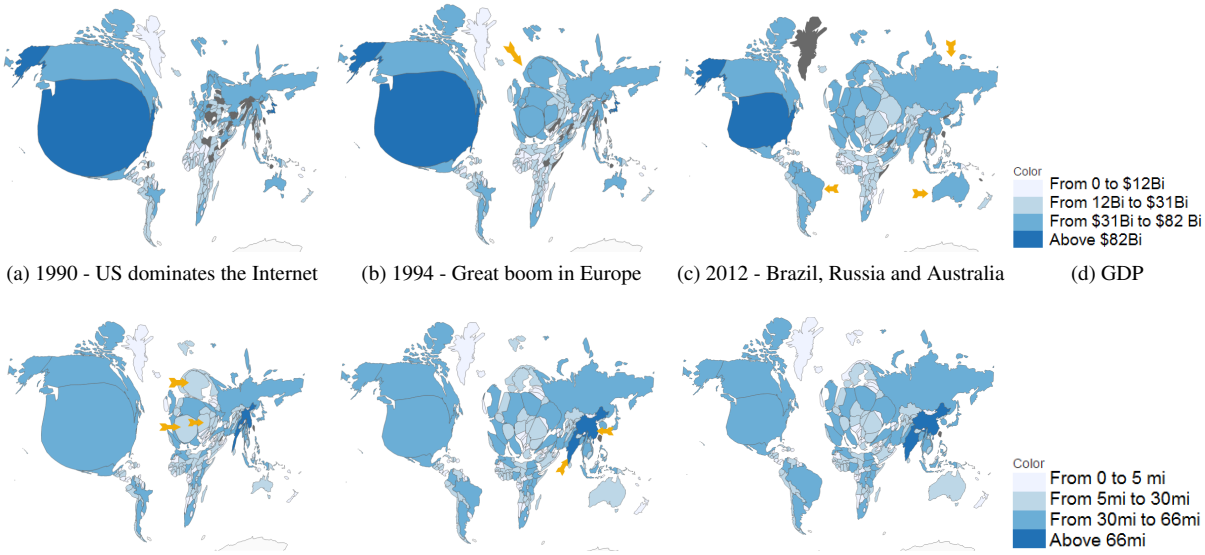
Using four fixed bins of equal size for coloring results in cartograms where most of the countries fall in just two of the bins. This is explained by the great number of countries with lower values for the geo-economic parameters (with very few countries represented at the other end of the scale). Thus, we use unequal size bins for colors, defined by the value ranges for all the geo-economic parameters.

Fig. 2 shows a subset of the generated cartograms. The companion website provides videos showing the complete sequences from 1990-2013 along with an online tool that generates the cartograms for a given year (or a sequence of all years) and the geo-economic parameter of choice. In every cartogram, the size of a country reflects the number of ASes, and each country starts with its real physical area and is distorted to get closer to its desired area, with respect to its percentage of the number of ASes over the total number of ASes. Colors represent the geo-economic parameter: the darker the country, the higher the value for the used parameter. Countries in gray indicate missing values. In the first row, country colors represent GDP. In 1990 (Fig. 2a), only a few countries have ASes such as US (the original Internet country with 389 ASes), Canada (33), Mexico (3), Panama(1) and South Africa(1), which explains the US big distortion. In 1994 (Fig. 2b) the Europe grows rapidly.Some countries with small GDP, such as Ukraine and Poland, have many new ASes, thus reducing the correlation to GDP, as seen in Sec. 2. In 2012 (Fig. 2c), some countries were experiencing better economic growth, which is reflected in the increase in number of ASes (Brazil, Russia, and Australia).

In the second row, country colors represent population. In 1994 (Fig. 2d), Europe experiences a big growth of some countries with large ASes and not very large populations (e.g., Sweden, Switzerland and Austria). It also shows India and China, big physical countries with big populations, with a proportionally small number of networks - thus leading to very small sizes and low correlation with the geo-economic parameters. Those two countries, for example, show an increase in number of ASes in the following years (Fig. 2e-f), but their country sizes in the cartogram continue to be smaller than their real physical country size. This shows that the number of networks in such countries is smaller than what would be expected for such large countries. The explanation for such low numbers needs to be investigated per country. Considering China as an example, possible causes may include the bigger control of the Internet by the government and a few approved ISPs.

4. Related Work

The study of Internet topology graph, where ASes are nodes and the logical connections between them are the links, in-



(a) 1990 - US dominates the Internet (b) 1994 - Great boom in Europe (c) 2012 - Brazil, Russia and Australia (d) GDP

(e) 1994 - Small correlation with Population (f) 2008 - Growth in China and India (g) 2013 - Correlation with Population in Africa, South America and Asia (h) Population

Figure 2: First Line: colors represent GDP. Second line: colors represent population.

volves the exploration of graph properties/metrics, such as average node degree, degree distribution, rich club connectivity, and betweenness centrality [MKF*06, HFC12, DF07, OZZ07, LHC*13, HFU*10, DCDc12]. Those metrics are then presented as *XY* plots for a given year, or over several years. With this type of approach it is difficult to make geo-economic inferences since only the physical and logical Internet topology is taken into account.

Existing Internet visualizations produce visual representations that often match the complexity of the original data, rather than make it easier to grasp and manage. Static node-link diagrams [BBP08, OLZ05, SMM13, BBGW05] have produced very complex visualizations. The AS Level Internet Graph [CAI] depicts the AS topology in polar coordinates by using out-degree of an AS to determine the distance from the center of a circle and its geographic location to determine its position around the circle. Cyclops [OLZ05] shows the internet as a graph, where each node size is drawn proportional to its connectivity degree to allow one visually differentiating big ISPs from small ones and edge thickness is proportional to the age of the link, thus separating edges that have existed for a long time from short-lived ones. VAST [OKB06] used quad-tree based visualization of AS Numbers depicting topological relationships in 3D. King et al [KHD*14] visualize the Internet with multiple coordinated view, including Hilbert's space-filling curves and animations to provide information on Internet traffic impact.

Geographical maps have also been used to overlay Internet activity of interest. Shavitt and Zilberman [SZ13] focus specifically in some ASes which are PoP (point of presence - locations owned by ISPs to place multiple networking equipment), and use a geographical map to visualize the patterns. CuttleFish [CAI15] provides an intuitive representation of geographically distributed Internet usage data with strong diurnal patterns. The Internet Map [Eni15] is a bi-dimensional presentation of links between websites on the Internet: every site is a circle on the map with its size determined by website traffic. WorldMapper [Wor06, DBN06] shows the distribution of Internet users in 1990 and 2002 with cartograms, making it easy to see countries with more users. iMap [FJS*14] represents the Internet topology using a map metaphor making it easy to identify countries with large AS presence. One major disadvantage of these approaches, for the tasks that we have in mind, is that these visualizations are static. In contract, we use the approach of coordinated geo-economic parameters, correlation plots and cartograms to visualize the Internet evolution.

5. Conclusions

With the joint use of correlations and cartograms it was possible to visually identify patterns of Internet growth along with some outliers. Countries in the Americas and Oceania have high correlation with GDP; in Europe there is a high correlation with population instead of GDP; the relatively poorer infrastructure in Asia results in low correlations with all of our geo-economic parameters. Our dataset and tool for generating customized cartograms based on ASes and geo-economic parameters are available online. A natural next step would be to use the observations made in order to model the underlying dynamics and provide a forecast for Internet growth in different countries and regions.

References

[BBGW05] BAUR M., BRANDES U., GAERTLER M., WAGNER D.: Drawing the AS graph in 2.5 dimensions. In *Proceedings of the 12th International Symposium on Graph Drawing* (2005), pp. 43–48.

[BBP08] BOITMANIS K., BRANDES U., PICH C.: Visualizing Internet evolution on the autonomous systems level. In *Proceedings of the 15th International Symposium on Graph Drawing* (2008), pp. 365–376.

[CAI] CAIDA: IPv4 and IPv6 AS core: Visualizing IPv4 and IPv6 Internet topology at a macroscopic scale in 2014. http://www.caida.org/research/topology/as_core_network/2014/.

[CAI15] CAIDA: Cuttlefish: Geographic visualization tool, January 2015. http://www.caida.org/tools/visualization/cuttlefish/index.xml.

[DBN06] DORLING D., BARFORD A., NEWMAN M.: Worldmapper: the world as you've never seen it before. *Visualization and Computer Graphics, IEEE Transactions on 12*, 5 (2006), 757–764.

[DCDc12] DHAMDHERE A., CHERUKURU H., DOVROLIS C., CLAFFY K.: Measuring the evolution of internet peering agreements. In *IFIP Networking* (May 2012), vol. 7290, pp. 136–148.

[DCN85] DOUGENIK J. A., CHRISMAN N. R., NIEMEYER D. R.: An algorithm to construct continous area cartograms. *The Professional Geographer 37*, 1 (1985).

[DF07] DONNET B., FRIEDMAN T.: Internet topology discovery: a survey. *Communications Surveys & Tutorials, IEEE 9*, 4 (2007), 56–69.

[Eni15] ENIKEEV R.: The internet map, January 2015. http://www.caida.org/tools/visualization/cuttlefish/index.xml.

[Eur15a] EUROPEANA LABS: D3-cartogram, January 2015. http://www-public.it-sudparis.eu/~maigron/RIR_Stats/RIR_Delegations/World/ASN-ByNb.html.

[Eur15b] EUROPEANA LABS: D3-cartogram, January 2015. http://labs.europeana.eu/apps/D3cartogram/.

[FJS*14] FOWLER J. J., JOHNSON T., SIMONETTO P., LAZOS L., KOBOUROV S., SCHNEIDER M. L., ACEDO C.: IMap: Visualizing network activity over Internet maps. In *In Proc. VizSec* (2014).

[GN04] GASTNER M. T., NEWMAN M. E. J.: Diffusion-based method for producing density-equalizing maps. In *Proc. of the National Academy of Sciences* (2004), vol. 101, pp. 7499–7504.

[HFC12] HUFFAKER B., FOMENKOV M., CLAFFY K.: *Internet Topology Data Comparison*. Tech. rep., CAIDA University of California, San Diego, 2012.

[HFU*10] HADDADI H., FAY D., UHLIG S., MOORE A., MORTIER R., JAMAKOVIC A.: Mixing biases: Structural changes in the AS topology evolution. In *Traffic Monitoring and Analysis*. Springer, 2010, pp. 32–45.

[Int14] INTERNET LIVE STATS: Internet users, January 2014. http://www.internetlivestats.com/internet-users/.

[KHD*14] KING A., HUFFAKER B., DAINOTTI A., ET AL.: A coordinated view of the temporal evolution of large-scale Internet events. *Computing 96*, 1 (2014), 53–65.

[LHC*13] LUCKIE M., HUFFAKER B., CLAFFY K., DHAMDHERE A., GIOTSAS V.: AS relationships, customer cones, and validation. In *IMC'13* (2013), pp. 243–256.

[MKF*06] MAHADEVAN P., KRIOUKOV D., FOMENKOV M., DIMITROPOULOS X., VAHDAT A., ET AL.: The Internet AS-level topology: three data sources and one definitive metric. *ACM SIGCOMM Computer Communication Review 36*, 1 (2006), 17–26.

[NAK15] NUSRAT S., ALAM M. J., KOBOUROV S. G.: Evaluating cartogram effectiveness. *CoRR abs/1504.02218* (2015).

[OKB06] OBERHEIDE J., KARIR M., BLAZAKIS D.: VAST: visualizing autonomous system topology. In *Proc. VIZSEC* (2006).

[OLZ05] OLIVEIRA R., LAD M., ZHANG L.: Visualizing Internet topology dynamics with cyclops. In *Proc. VIZSEC* (2005).

[OZZ07] OLIVEIRA R. V., ZHANG B., ZHANG L.: Observing the evolution of internet AS topology. *ACM SIGCOMM Computer Communication Review 37*, 4 (2007), 313–324.

[SMM13] SALLABERRY A., MUELDER C., MA K.-L.: Clustering, visualizing, and navigating for large dynamic graphs. In *Proceedings of the 20th International Symposium on Graph Drawing* (2013), pp. 487–498.

[spe08] Spearman rank correlation coefficient. In *The Concise Encyclopedia of Statistics*. Springer New York, 2008, pp. 502–505.

[SZ13] SHAVITT Y., ZILBERMAN N.: The internet geographical pop level maps. In *Proc. of the European Conference on Complex Systems* (2013), Springer, pp. 189–194.

[Tob04] TOBLER W.: Thirty five years of computer cartograms. *Annals of Association of American Geographers 94* (2004), 58–73.

[Ver13] VERIZON: Unbalanced peering, and the real story behind the Verizon/Cogent dispute, June 2013. http://goo.gl/FYVqEj.

[Wik15] WIKIPEDIA: Arpanet, January 2015. http://en.wikipedia.org/wiki/ARPANET.

[Wor06] WORLDMAPPER: Internet users 2002, January 2006. http://www.worldmapper.org/posters/worldmapper_map336_ver5.pdf.

[Wor15] WORLDBANK: The worldbank:data, January 2015. http://worldbank.com/.