Hidden Internet topologies info: Truth or Myth?

Sofía Silva Berenguer*, Esteban Carisimo†, J. Ignacio Alvarez-Hamelin†
and Francisco Valera Pintor*
* UC3M, Madrid, Spain. {ssilva,fvalera}@it.uc3m.es
† INTECIN (UBA-CONICET), Buenos Aires, Argentina. {carisimo,ihameli}@cnet.fi.uba.ar

ABSTRACT

Internet mapping projects usually get information from several routing data collectors or vantage points. The accuracy of maps relies on the amount and location of these collectors, which are usually near the backbone or at large developed regions, such as ARIN's or RIPE NCC's. The lack of vantage points in Latin America makes these maps not really show the current actual status of the network in this region. For this reason, in this work we have added data from some local sources and measured how much information was missing without them.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Network topology*

Keywords

Latin America, Network topology, traceroute

1 Introduction

The AS-level Internet topology has been widely studied for years using many different methodologies. The earliest studies about the Internet AS-level topology [6, 10] are from between 1997 and 1999. For this kind of studies, the amount and location of the vantage points is crucial. A low number of vantage points in some regions may lead to the representation of a topology for that region being incomplete. This is probably one of the reasons behind the fact that most of the known studies have a world-wide scope and not a regional scope. An example of a regional study is [7], which evaluates the topology interconnecting ISPs based on Africa. But unfortunately, there are not many other examples of regional studies.

LANCOMM, August 22-26, 2016, Florianopolis , Brazil

DOI: http://dx.doi.org/10.1145/2940116.2940118

On the other hand, it has been proven that a considerable amount (at least 35 %) of the links between ASes are not being discovered and that most of them are peer-to-peer links [8]. The reason for this is that peering relationships are usually not announced to transit providers or other peers, and therefore, a vantage point needs to be located at one of the peering ASes or at one of the peers' customer cones [9] in order to be able to discover the peering link. The inevitable consequence of this is that the Internet topology maps that can be created are incomplete, mostly for those regions where the number of vantage points is low (e.g., Africa and Latin America).

According to [3], AS maps can be enhanced by adding local routing information from Looking Glasses (LGs) and routing policy information from Internet Routing Registry (IRR) databases. The main purpose of this work is to add local routing information and measure the improvement accomplished.
For this article, we generate the typical Internet graphs representing the Autonomous Systems (ASes) as vertices and their corresponding relationships as edges. With the aim of analysing the impact on these graphs of adding local routing information to the publicly available routing information, we create graphs at three different levels of scope: country, region and world. Finally, we measure the enhancement on the graphs when adding local routing information.

2 Methodology

The graphs used as a reference point when making the comparisons are built from the BGP information offered by the RIPE RIS and RouteViews projects. The local routing information added in order to enhance the topology diagrams is BGP data obtained from different Looking Glasses in the Latin America and the Caribbean (LAC) region (CABASE (Argentina), PTT Metro (Brazil), etc.) and from *show ip bgp* outputs provided by some operators in the region. The lack of vantage points in Latin America has led to some projects, such as PladMeD [2], to develop their own platform to gather data from this region. PladMeD, which is a traceroute-based platform to study the

	$\|V\|$	$\|E\|$	$< d >$	$\max(d)$	$< cc >$	$< k >$	$\max(k)$
BO	17	17	2	6	0	1.2	2
BO+	19	32	3.37	8	0.44	2.2	3
LAC	4835	23056	9.54	811	0.42	4.8	27
LAC+	4926	30824	12.52	865	0.65	4.9	40
World	52688	204462	7.8	4777	0.27	3.9	75
World+	54270	222561	8.2	5133	0.30	4.2	76

Table 1: Analysed parameters for different scales, and their enhanced versions (plus sign).

improvements of the first Bolivian IXP, provides us detailed information about Bolivian ASes ecosystem. For the graphs at the country level, Bolivia is used as a case of study and apart from adding to the national graph the local routing information mentioned above, we also added relationships inferred from an AS paths' set from PladMed.
Relationships between ASes are inferred using CAIDA's AS Relationship inference algorithm [5]. Although the main focus of this algorithm is to identify the type of relationship between the ASes (Transit or Peering), the type of relationship is not used for this work, just the fact that a relationship between two ASes exists.
It is important to note that mainly two different criteria can be used to restrict an Internet topology graph to a certain area: 1) use the information of the country where the organisation to which the AS was assigned is based; 2) use geolocation information in order to determine the countries in which an AS is active, i.e., the countries to which the prefixes being announced by an AS are geolocated. Taking into account that an AS is not necessarily used in the country to which it was assigned, the first option is not used. Instead, the prefixes announced by all the active ASes were geolocated using RIPEstat Data API in order to determine the countries where each AS is active. To create the national graphs for Bolivia and the regional graphs for the LAC region, we filtered the World graphs in order to include just the ASes that are active in the area of interest and the relationships in which these ASes are involved.

Geolocation tools determinate which foreign ASes operate in Latin America and which Latin American ASes operate abroad. However, Latin American ASes rarely operate overseas or even in ARIN's region. In our study, inaccuracies could be at country-level graph with ASes that operate in several Latin American countries.

Although criterion 2) is more accurate than 1), there are some misleading entries on every geolocation database. In this case, RIPEstat shows that AS701, AS1239 and AS10434 are active in Bolivia but no dataset has shown a link between these ASes and other ASes truly participant in the Bolivian network. Due to this mistake, Bolivian graphs are not completely connected.

Looking for comparing the graphs we used the following parameters: $|V|$ (number of vertices), $|E|$ (number of edges), $< d >$ (average degree), $\max(d)$ (maximum degree), $< cc >$ (average clustering coefficient), $< k >$ (average shell index) and $\max(k)$ (maximum k-core). The average degree is computed as $< d >= \frac{2|E|}{|V|}$ and it measures the number of relationships in which an AS is involved on average. The average clustering coefficient as $\frac{1}{|V|}\sum_{1\leq i\leq |V|} cc(i)$, where $cc(i) = \frac{m(i)}{d(i)(d(i)-1)}$ is the clustering coefficient of vertex i, $d(i)$ is the degree of vertex i and $m(i)$ is the number of edges between the neighbours of i. The cc measures the level of interconnection within a node's neighbourhood. The k-core is a subgraph where all vertices have at least degree k. The k-shell-index is the maximum core a vertex belongs to. We use LaNet-vi [1] to compute the k-core decomposition (i.e, the computation of vertices belonging to each shell), and also to verify which vertices (or ASes in this case) verify the core-connectivity property.

3 Analysis

In order to study the improvement on the graphs introduced by the added Latin American vantage points, we analyse the changes on the three levels: *BO* is the Bolivian network, *LAC* is the Latin American network, *World* is the whole Internet. For each of them, we have the initial version (RIPE RIS and RouteViews) and the enhanced one (adding Looking Glasses and vantage points) distinguished with a plus sign. Table 1 displays the measured parameters for each graph. In this particular case, the six graphs show the *core-connectivity* [1] property, that is, a vertex in a k-core has at least k different paths to another vertex in the same k-core. This property was expected because the world ASes graphs have always verified this along the time. Therefore, a network with a large maximum k-core (or large $< k >$, the average shell-index) implies robustness because there are more paths to interconnect its ASes. We created a website [1] where the graphs generated by LaNet-vi are shown.

Regarding Table 1, we can affirm that when adding more local information about interconnection between ASes, the observed parameters increase. For instance, the average degree increases by 68% in *BO*, by 32% in LAC and by just 6% in World. Similar results have been found for the maximum degree: 33%, 6% and 7% respectively; clustering coefficient: *exists*, 23% and 9%

[1] http://cnet.fi.uba.ar/ASes_topology_LatAm/

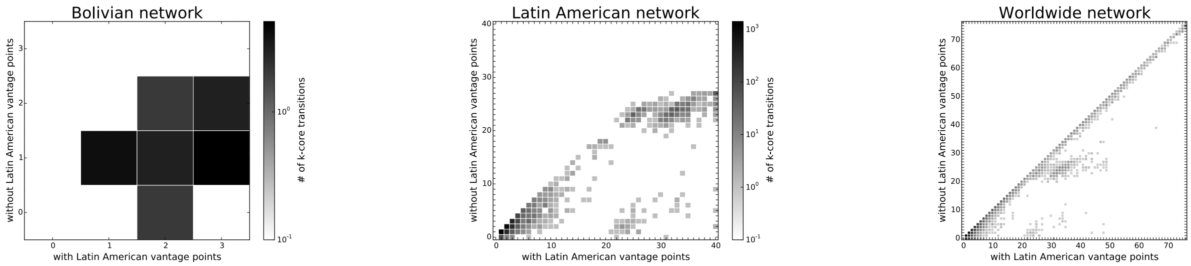

Figure 1: Improvement of k-shell-index sets (connectivity).

respectively; average shell-index (connectivity): 83%, 2% and 8% respectively; and finally maximum k-core: 50%, 48% and 1% respectively. This shows that the increment in the interconnection that was detected has a great impact on the national networks as BO, some impact on the regional ones as LAC, and less impact on the whole Internet. Moreover, the robustness detected, measured by the average shell-index $< k >$ (all graphs are core-connected, i.e., there are at least k different paths between ASes in a k-core subgraph) and the maximum k-core (the backbone of such network), has a great impact in national and regional networks, and much less in the World case.

Figure 1 shows the probability an AS has of changing from shell k to shell k' (the darker, the more probable) for the BO, LAC and $World$ graphs. We observed that shells maintain or increase their value (i.e., $k' \geq k$), which is expected since new edges were added. The greatest impact is at the national and regional levels (BO and LAC networks), which is an expected result as we added national and regional routing information. It can also be noticed that the improvement at these levels has an impact on the $World$ network (the reasons for which the LAC graph is improved also apply to the $World$ graph as the first one is a subset of the latter). Moreover, this figure highlights the improvements are in the low and high shells for the LAC case, and the whole BO network. The impact on the $World$ network shows the relative importance of the LAC region in the World network.

4 Conclusions

As suggested by [4], it seems that adding local routing information is highly relevant. For this reason, we wanted to verify if there is actually hidden information in regular Internet graphs. We have effectlively found that there is, as it is shown by the variation of the different computed metrics. We have also verified that the regional hidden information has a global impact, although a deeper analysis of the implications of the different metrics should be carried out as further study.

Acknowledgements. This work was partially founded by UBACyT 2014 (20020130200122BA). E.C. acknowledges CONICET Argentina for a PhD fellowship.

5 References

[1] M. G. Beiró, J. I. Alvarez-Hamelin, and J. R. Busch. A low complexity visualization tool that helps to perform complex systems analysis. *New J. Phys*, 10(12):125003, 2008.

[2] E. Carisimo, H. Galperin, and J. I. Alvarez-Hamelin. A new intrinsic way to measure ixp performance: an experience in bolivia. *arXiv e-print*, abs/1505.00837, May 2015.

[3] H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger. Towards Capturing Representative AS-level Internet Topologies. *Comput. Netw.*, 44(6):737–755, Apr. 2004.

[4] L. Dall'Asta, J. I. Alvarez-Hamelin, A. Barrat, A. Vázquez, and A. Vespignani. Exploring networks with traceroute-like probes: Theory and simulations. *Theor. Comput. Sci.*, 355(1):6–24, 2006.

[5] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, G. Riley, et al. As relationships: Inference and validation. *SIGCOMM Comput. Commun. Rev.*, 37(1):29–40, 2007.

[6] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *SIGCOMM Comput. Commun. Rev.*, 29(4):251–262, Aug. 1999.

[7] R. Fanou, P. Francois, and E. Aben. *PAM 2015, New York, NY, USA*, chapter On the Diversity of Interdomain Routing in Africa, pages 41–54. Springer International Publishing, Cham, 2015.

[8] Y. He, G. Siganos, M. Faloutsos, and S. Krishnamurthy. Lord of the links: A framework for discovering missing links in the internet topology. *IEEE/ACM Trans. Netw.*, 17(2):391–404, Apr. 2009.

[9] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, and k. claffy. As relationships, customer cones, and validation. In *IMC 2013, Proceedings*, IMC '13, pages 243–256, New York, NY, USA, 2013. ACM.

[10] D. Magoni and J. J. Pansiot. Analysis of the autonomous system network topology. *SIGCOMM Comput. Commun. Rev.*, 31(3):26–37, July 2001.