Silva Berenguer, Sofía; Valera Pintor, Francisco (2018). Radiography of Internet Autonomous Systems Interconnection in Latin America and the Caribbean, *Computer Communications*, v. 119, pp.: 15-28.

DOI: https://doi.org/10.1016/j.comcom.2018.01.008

# Radiography of Internet Autonomous Systems Interconnection in Latin America and the Caribbean

Sofía Silva Berenguer*, Francisco Valera Pintor

*University Carlos III of Madrid*

*Av. Universidad, 30, 28911*

*Leganés (Madrid), Spain*

**Abstract**

Lots of studies about the Internet Autonomous System (AS) level topology have been carried out during the last twenty years, most of them analyzing this topology on a world-wide scale, a lot of them based on routing information from the Border Gateway Protocol (BGP). However, studies focusing on a specific region and making comparisons between regions are not that popular and in fact, most world-wide studies are not valid in some particular regions. This work is targeting this particular problem of the regional or country topology analysis by enhancing regular AS-level graphs where to apply different connectivity metrics. The focus is set on Latin America and the Caribbean (the LAC region) which exhibits appropriate conditions for this type of analysis and where we show that a basic metric comparison may not be good enough so as to realize that there is a connectivity problem in the region. After concluding that the situation in the LAC region in terms of interconnection is even worse than expected, we perform some country-level studies finding correlations between graph characteristics and some socioeconomic indicators. We then use these correlations to identify countries in which it would be worth pushing for the deployment of an Internet Exchange Point (IXP), as simulating the creation of an IXP there has a great impact on the interconnection level and on the robustness of the regional Internet.

*Keywords:* Interconnection, Autonomous System, Internet topology, Graph, Latin America

## 1. Introduction

The Internet has continuously been studied from every angle and perspective since it was created. Its own evolution in terms of size, traffic patterns, applications, hardware improvements, etc. makes it necessary to keep on performing regular measurements to understand this hugemongous network. The authors of [1] make a good prospective of the different studies performed so far during the last 20 years. As it is explained in this survey, one of the perspectives that has been commonly adopted to understand the Internet connectivity is the Autonomous System (AS) level topology because the performance of the Internet highly depends on the quality of the existing paths between these pieces that together form the Internet jigsaw puzzle. There are other different alternatives but, as it is described in [1], they are more focused on the physical topology than in the logical topology (connectivity) which is the main purpose if this article.

The earliest studies on AS topology are from the late nineties [2, 3]. All these studies are normally based on the use of graphs, representing ASes as nodes and the relationships between them as edges and they are usually analyzing the AS-level topology of the Internet on a world-wide scale based on BGP routing information (e.g. [3]). Works involving regional-level analysis (country level or even continental) or comparisons amongst regions are not that popular and in fact, most global Internet topology studies are not valid in some particular regions.

The main reason for this is that in general, this type of global Internet studies tends to use routing information from projects like RIPE NCC's Routing Information Service (RIS) [4] or University of Oregon's RouteViews (RV) [5], which collect and make publicly available BGP routing data from several locations around the world. However, the goal of these projects is not to provide information to infer AS-level topologies, because the BGP protocol was not designed with this purpose in mind and the topologies that can be obtained from this information are usually not complete [6], [7]. It is a fact for instance that a number of ASes relationships, in particular peer to peer ones, are not revealed in this topology map because in many occasions they are not announced as the rest of the relationships (at least 35 % according to [8]). One of the main drawbacks of the topological studies is that they require a considerable amount of vantage points (collectors) properly located so as to be able to infer a reasonable AS-level topology and this is normally not the case when the focus is set on par-

*Corresponding author

*Email addresses:* `ssilva@it.uc3m.es` (Sofía Silva Berenguer), `fvalera@it.uc3m.es` (Francisco Valera Pintor)

ticular regions where the connectivity between ASes is not completely developed instead of keeping it at a world-wide scope. Latin America, Africa or even Asia are the type of regions whose AS topological maps would probably be incomplete because of the commented reasons (see [9] for instance for a particular regional study about the African region).

In this article we provide evidences to show this effect in the Latin American and Caribbean region (LAC region), although it will be seen that the methodology, the considered metrics and the conclusions can be easily applied to any region.

The LAC region is quite heterogeneous with regards to Internet infrastructure and to the connectivity between networks. It is usually mentioned that the LAC region is a poorly interconnected region with low local traffic exchange [10]. One of the reasons for this is the presence of very few cables in the region due to its geographic characteristics [11], which leads to the dependence of the LAC region on the infrastructure of the operators from North America. And it is also widely known that there are not many collectors in the LAC region, making it even harder to have a realistic view of this region [12].

Because of these characteristics, the LAC region is a good choice to perform our analysis that will initially get a basic graph of the region using common routing information. This graph will be later on enriched using looking glasses in the region or directly requesting data from local Internet Exchange Points (IXPs) or Internet Service Providers (ISPs), as it is typically required for an AS level topological study in order to better approach the real topology and as expected, it will be shown that many peering relationships are now exposed. These hidden peering relationships have been particularly analyzed in [13] for Bolivia and some general conclusions were also obtained for the continent and whole world.

Later on the article focuses on the application of different metrics to the LAC graphs and diverse comparisons with the graphs of the other regions. This type of regional AS topology analysis will lead to a better understanding of the Internet performance in a certain area and will provide criteria to determine which countries of the region have greater needs to enhance their interconnection degree, the deployment status of the Internet in the region and the situation of the Internet infrastructure in the region. Network operators may get the incentives to enhance the interconnection level, thus improving the robustness, the security and the performance of the Internet regionally and even creating new business opportunities by promoting content innovation [14].

In addition, the analysis may be used to derive consistent criteria to make decisions about the creation of IXPs by the governments and organizations involved in these decisions and thus be able to better plan the locations for new IXPs. The creation of new IXPs helps reducing transit costs, improves Internet service quality and promotes infrastructure investment in smaller markets [15].

The rest of the article is organized around two main sections. Section 2 introduces the methodology that has been followed, including the data sources that have been used, the different metrics considered and related work. Section 3 presents all the results and explains their implications. This section is also including an analysis on the implications of adding an IXP in particular countries that are identified applying socioeconomic trends. Section 4 shows the main conclusions of the work.

## 2. Methodology

The development of this work was divided into five stages: (1) Data collection; (2) Topologies construction; (3) Topology Enhancement; (4) Metrics Computation and (5) Topologies Analysis. This process is shown in Figure 1. This section describes the different tasks performed during each of these stages, mentioning the methodology usually applied by similar projects and commenting on the differences with the methodology that we used.

### *2.1. Data Collection*

In order to complete the stages of this project, various information was collected: routing data, ASes assignment and geolocation information, information about existence of IXPs in the countries in the LAC region, information about presence of ASes in those IXPs and economic, transport and tourism indicators for the countries in the LAC region. All this information was collected between April 2015 and February 2016. In general, as AS-level topologies represent business relationships between ASes and these relationships are rather static, we can assume that the inferred topology has not changed much since the moment we performed the data collection.

There are some studies focused on the temporal evolution of the AS-level topologies of interest [16], and therefore they collect routing data periodically. In this case, we wanted to take a snapshot of the Internet AS-level topology and taking into account that the local routing information for the LAC region required manual work to be done in order to collect it, we did not perform a periodic collection of routing data. We used routing data from RIPE NCC's RIS [4], University of Oregon's RouteViews (RV) [5] and Packet Clearing House (PCH) [17] collectors. Additionally, routing information from Looking Glasses (LGs) in the LAC region (from CABASE[1], PTT Metro[2], NAP Chile[3] and Orange Chile[4]) and some *show ip bgp* outputs provided by operators in the region (Access Haiti, GTD Internet (Chile) and LACNIC) was also used.

Assignment information about the ASes was obtained from Team Cymru's WHOIS service and from IANA's web page[5], while geolocation information for ASes (information

[1] http://looking.cabase.org.ar
[2] http://ix.br
[3] http://lg.nap.cl
[4] http://pit.orange-business.cl/lg
[5] http://iana.org

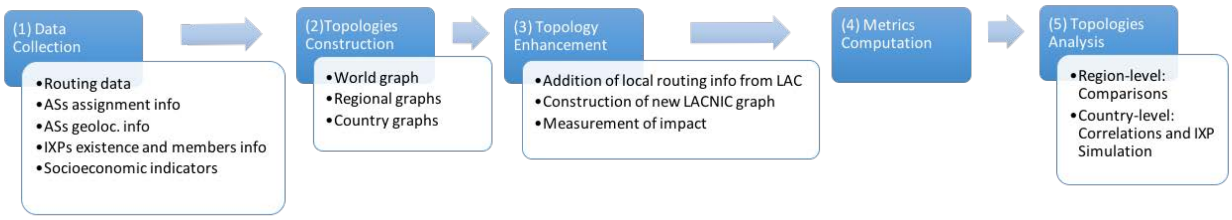



Figure 1: Stages of the Methodology

about the countries where an AS is active) was obtained from the API offered by the RIPEstat project [18].

Information about the existence of IXPs in the different countries of the LAC region and about the presence of ASes in these IXPs was obtained by doing manual research (performing Google searches and visiting IXP's web pages).

Finally, economic, transport and tourism indicators for the countries in the LAC region were downloaded from The World Bank's web page [19] and from the CIA's World Factbook [20]. The most recent information provided by The World Bank at the time of the data collection process corresponded to the year 2014. The information provided by the CIA's World Factbook at the time of the data collection process corresponded to the year 2013.

### *2.2. Graphs Construction*

In order to generate a world graph, CAIDA's AS Relationship Inference algorithm [21, 22] was executed, using routing data from the different collectors of the RIPE NCC's RIS [4] and University of Oregon's RouteViews (RV) [5] projects as an input to infer relationships between the different ASes. This graph represents the way in which the ASes are interconnected. In this sense, the nodes of the graph represent ASes and the edges of the graph represent the relationships between two ASes. These relationships can be Peer to Peer (P2P) or Provider to Customer (P2C) relationships. For most of the analysis performed as part of this project, no distinction is made between these two types of relationships and therefore, undirected graphs are used.

AS topology studies usually consider a world graph as a whole and do not represent region-level or country-level topologies, therefore no reference of how the world graph should be filtered in order to obtain region-level or country-level graphs was found. That is why we had to design some criteria specifically for this work. Two different approaches were considered:

- **Criterion 1 - Active Edges:** only the relationships that are active in the area of interest are included in the graph, i.e. those relationships involving at least one AS active in the region, and all the ASes involved in these relationships
- **Criterion 2 - Active Nodes:** all the ASes active in the region and all the relationships between them are included in the graph.

Figure 2 shows schematically how these two criteria are applied. Nodes filled with dots represent ASes active in the area of interest; nodes without dots represent ASes not active in this area.

Graphs defined by Criterion 1 include all the ASes active in the area of interest and additionally, they also include all the ASes involved in relationships with ASes active in the area of interest, hence graphs defined by Criterion 2 are a subset of the graphs defined by Criterion 1. Criterion 2 graphs provide information about how the ASes are interconnected within the area, while Criterion 1 graphs also include some information about how that area is connected to other areas. The focus of this work is set in the internal interconnection, that is why, for the sake of brevity, we will not include the results for both criteria but we will focus on the graphs defined by Criterion 2. We applied the same methodology to analyze the graphs obtained when applying criterion 1 and we got similar results. All the plots and datasets corresponding to both criteria can be found at [23].

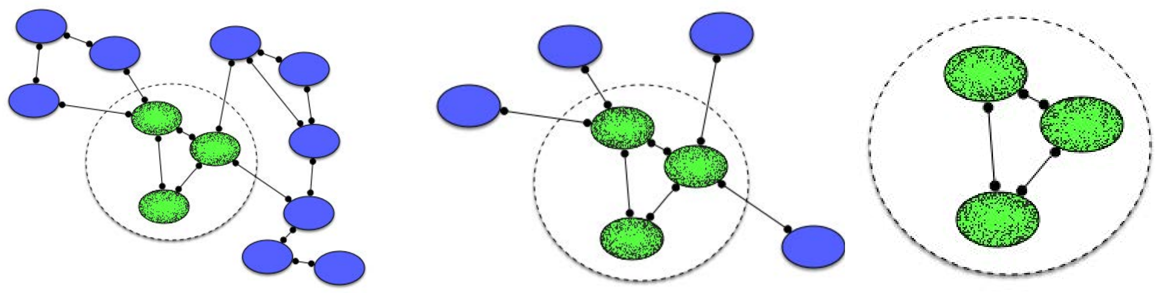
For Internet resources' management matters, the world is administratively divided into five regions covered by the five Regional Internet Registries (RIRs). AfriNIC[6] for the African region, APNIC[7] for the Asia-Pacific region, ARIN[8] for the North American region and part of the Caribbean, LACNIC[9] for the Latin American region and the rest of the Caribbean and RIPE NCC[10] for Europe and the rest of Asia. These are the regions considered when generating region-level graphs.

ASes are geolocated using the API offered by RIPEstat ([18]). An AS is considered to be active in a country if it is originating at least one IP prefix that is geolocated to that country. Then, the AS is considered to be active in a region if any of the countries it is active in are part of the coverage area of the corresponding RIR.

### *2.3. Graph Enhancement*

In general, routing collection projects like RIS and RV have been proven to be incomplete, offering a limited view

[6] https://www.afrinic.net
[7] https://www.apnic.net
[8] https://www.arin.net
[9] http://www.lacnic.net
[10] https://www.ripe.net



(a) World Internet graph simplified (b) Criterion 1 (c) Criterion 2

Figure 2: Criteria to define area-specific graphs

of the Internet AS-level topology [24, 7, 8, 12, 25]. There have been several approaches to complete this topology through the usage of active [26, 27]; passive [28, 29, 30] and a combination of active and passive [31] measurements.

Approaches based on active measurements perform *traceroutes* from probes distributed world-wide to different destination nodes and infer AS relationships from the paths obtained. Passive measurements approaches consist in the utilization of routing information from Routing Collectors (RCs), Looking Glasses (LGs), Internet Routing Registries (IRRs), WHOIS services offered by the RIRs and other databases. RCs are devices connected to different ASes that collect feeds of BGP table dumps. LGs are publicly available servers/routers offering a web interface or telnet access from which different *show* commands can be executed. IRRs are databases that include routing information registered manually by AS administrators. WHOIS servers are publicly available servers that can be queried to get information about domain names, IP addresses, ASes, etc.

Apart from the general incompleteness of the AS-level topologies that can be generated using different approaches, it is widely known that there are not many RCs in the LAC region, making it even harder to have a realistic view of this region [12]. In order to compensate this and thus be able to improve the inferred topology for the LAC region, we wanted to add local routing information to the already used inputs (RIS and RV), in order to augment the set of AS-relationships inferred for the LAC graph. Adding local routing information would obviously reveal hidden links for any region. However, taking into account the high number of RCs in the ARIN and RIPE NCC regions, it would probably not have an impact as high as the impact of adding local routing information for the other regions, nevertheless, it is important to bare in mind that the situation in the AfriNIC and in the APNIC regions is similar to the situation in the LAC region with regards to RCs and that we are not adding local routing information there because our focus is on the LAC region.

Active measurements approaches to complete AS-level topologies are usually more costly and are not always available due to the lack of probes. That is why for this article, an approach based on passive measurements was preferred. Amongst the options based on passive measurements, we decided to use routing information from those Packet Clearing House (PCH) RCs located in the LAC region, routing information from LGs located in the LAC region and adding some *show ip bgp* outputs provided by operators in the region which, despite of not being publicly available, could be considered as LGs. It is believed that using PCH RCs on top of RIS and RV does not make much difference ([25]), however, the situation for the LAC region differs from the global situation as RIS and RV only have RCs in Brazil, while PCH has RCs in other countries. This is why the authors expect that adding routing information from PCH RCs would enhance the LAC graphs obtained from RIS and RV.

The decision of using RCs and LGs was based on the fact that the information about AS-relationships offered by IRRs, WHOIS or other databases mostly depends on the input by network operators and therefore it could be inaccurate, incomplete, out-of-date and/or false, while the information offered by RCs and LGs comes from BGP tables and is much more reliable as a consequence.

We will refer to the routing data from RIS and RV as **RIS+RV** dataset and to the data enriched with local routing information as **RIS+RV+LAC** dataset. Throughout this article we will refer to the LACNIC graph created using the RIS+RV dataset as "LACNIC" and to the graph created using the RIS+RV+LAC dataset as "LACNIC+".

We considered [28] as an example of an approach based on passive measurements. In this work, Khan et al. designed a tool to automatically query a list of LG servers, using the *show ip bgp summary* command to obtain the neighboring routers of the LG and constructed an AS topology of the whole Internet based on this information. They then compare this topology with the AS topologies generated from other datasets and analyze the overlapping and unique AS links included in each topology.

In our case, instead of comparing topologies from different datasets, we first generate an AS topology from the RIS+RV dataset, then we augment this dataset adding routing information from the LAC region, obtaining the RIS+RV+LAC dataset, and generate a new AS topology from this dataset. We finally compare the two AS topologies to measure the impact of adding the local routing information from the LAC region. As the second topology is generated from a dataset that is an augmented version of the first dataset, all the relationships discovered from the RIS+RV dataset will also appear in the second topology and therefore, instead of analyzing overlapping and unique links, we just compute the newly discovered relationships.

Apart from counting the new relationships inferred when adding routing information, both [28] and [29] compare the Complementary Cumulative Distribution Function for the Node Degree in order to detect the ranges of values of the Node Degree for which new edges are discovered. For this article, we also perform this comparison.

### *2.4. Metrics Computation*

With the aim of analyzing the interconnection level in the LAC region and comparing it to the level in other regions, we selected those graph metrics that would allow us to characterize and properly compare the different graphs of interest. Based on [32] and [33] we made comparisons using the characteristics that are explained in the following subsections. Metrics are grouped by their interpretation, being the most important those related to the level of interconnection and the robustness of the graphs, and the rest of them providing interesting additional information. All the considered characteristics were computed using Python's library Networkx. An analysis and interpretation of the values of these metrics are provided in the Results section (section 3).

#### *2.4.1. Dimension of a graph*

**Order:** The order measures the number of nodes. In this case it represents the number of ASes included in the analyzed graph.

**Giant component:** The giant component is the order of the biggest connected component of the graph. The graphs defined by Criterion 2 include all the nodes active in the area of interest but we have to take into account that there are cases of nodes active in the area of interest that are not connected to any other node active in the same area and therefore, these nodes will appear to be disconnected in these graphs. This is why the graphs will consist of one giant component and a set of disconnected nodes and/or small components.

**Disconnected nodes:** Disconnected nodes are nodes that are not involved in any relationship in the graph of interest, i.e. they have a degree value of zero. As explained above, the graphs defined by Criterion 2 include all the nodes active in the area of interest but we have to take into account that there are cases of nodes active in the area of interest that are not connected to any other node active in the same area and therefore, these nodes will appear to be disconnected in these graphs. For example, ASes involved in relationships that were not inferred from the routing information we are using could be in this situation. For these disconnected nodes we performed an analysis checking whether they have any common neighbors with other nodes active in the region and therefore are indirectly connected to the graph.

#### *2.4.2. Interconnection between ASes*

**Size:** The size of the graph measures the number of edges. In this case, it represents the number of relationships included in the analyzed graph.

**Node Degree:** The degree of a node is the number of edges the node has to other nodes in the graph. In our graphs, the degree represents the number of AS relationships an AS is involved in.

**Degree Distribution:** The degree distribution provides the probability that a randomly selected node in the network has degree k. Numerous studies [2, 34, 35] have shown that Internet graphs are scale-free networks, which means that their degree distribution follows a power-law, i.e. they have a lot of nodes with very small degree and a few nodes with very high values of degree, therefore the degree distribution will be long-tailed.

**Average Node Degree:** The average node degree of a graph is the average of the node degrees of all the nodes included in the graph. It is not recommended to use this metric for Internet graphs [32], as they usually have a power-law degree distribution and therefore the deviation around the average degree could be arbitrary large. Nevertheless, the average node degree equals the number of edges divided by the number of nodes, therefore it can be thought of as the size of the graph normalized by its order, offering an idea of the number of relationships per AS on average. This is why we decided to use this metric to make comparisons anyway.

#### *2.4.3. Correlation of Degrees of Connected Nodes*

**Average Degree Connectivity:** The average degree connectivity, also called neighbor connectivity, is the average of the degrees of the neighbors of the nodes with degree k. This metric provides information about the correlation between the degree of connected nodes. A decreasing average degree connectivity plot means nodes with high degree usually get connected to nodes with low degree, while an increasing average degree connectivity plot means that relationships are usually between nodes with similar degree. The average degree connectivity is related to the assortativity: assortative graphs have increasing average degree connectivity plots, while disassortative graphs have decreasing average degree connectivity plots.

**Assortativity:** The assortativity of a graph is the Pearson correlation coefficient between the degrees of pairs of nodes that are linked. Graphs with positive assortativity are called *assortative*. In these graphs, relationships are usually between nodes with similar degree. Graphs with negative assortativity are called *disassortative*. In these graphs, nodes with high degree usually get connected to nodes with low degree. The assortativity is related to the average degree connectivity: assortative graphs have increasing average degree connectivity plots, while disassortative graphs have decreasing average degree connectivity plots. Two versions of the assortativity were computed: one denoted 'Local' which considers the degree of the node in the graph of interest and another one denoted 'Global' which considers the degree of the node in the world graph. The argument behind this is that we wanted to know whether the motivation for two ASes to get interconnected depends on how well interconnected each AS is globally or on how well interconnected each AS is at the area of interest.

**Joint Degree Distribution:** The Joint Degree Distribution (JDD) or Node Degree Correlation matrix shows the probability that a randomly selected edge connects two nodes with a certain combination of degree values [33]. Authors of [33] found that this metric appears to fundamentally characterize Internet AS topologies.

#### *2.4.4. Local Interconnection between ASes*

**Average Clustering Coefficient vs Degree:** According to [32], the clustering coefficient (CC) captures the degree to which the neighbors of a given node link to each other, measuring the network's local link density. The plots of Average CC versus Degree show the dependence of this local link density on the node's degree k.

#### *2.4.5. Interconnection and Robustness*

**K-Core Decomposition:** A k-core or a core of order k is a subgraph of the graph of interest for which, each node's degree at the subgraph is higher or equal to k, and the subgraph is a maximum subgraph with this property. The k-core decomposition of a graph provides information about the hierarchical properties of large scale networks, focusing on the regions of increasing centrality and connectedness properties. More central cores are indeed more strongly connected, with larger number of possible distinct paths between nodes. [36]

Applying the algorithm described in [37], a k-core decomposition can be performed. After this, the core number of a node is the highest order of a core that contains this node in the decomposition. We call this core number the Shell Index of the node. We use as a graph metric the Maximum Shell Index in the decomposition, which gives us the number of cores in which the graph is decomposed and therefore, measures how hierarchical the graph is. AS-topology graphs have verified along the time the core-connectivity property, that is, a node in a k-core has at least k different paths to another node in the same k-core. [38] Therefore, a network with a large maximum k-core is very robust because there are more paths to interconnect its ASes.

This metric has already been successfully used by the authors in another article ([13]) to measure the impact of adding local routing information on graphs at different scales.

In order to better compare the graphs in terms of robustness, we also analyze the distributions of the Shell Index variable for the nodes of the different graphs to be able to study the sizes of the different cores.

### *2.5. Region-level Analysis: Comparison of LACNIC graph with other region's graphs*

In order to make the comparisons amongst the different regions the graph characteristics that can be summarized as a single value (Order, Size, Giant Component, Average Node Degree, Disconnected Nodes, Maximum Shell Index, Local Assortativity and Global Assortativity) are presented in Table 3 sorted by the Average Node Degree column. The maximum absolute value for each column is written in **bold** letters, while the minimum absolute value for each column is written in *italics*. The rest of the characteristics (Degree Distribution, Average Clustering Coefficient vs Degree, Average Degree Connectivity and Joint Degree Distribution) are compared using the corresponding plots. In the table and in the plots both LACNIC and LACNIC+ graphs are included so that the impact of adding local routing information is further understood.

When making comparisons amongst the five regions, it is important to take into account that not only the LAC region has a low number of RIS and RV route collectors (two in Brazil), but also the African (one in Kenya) and the Asia-Pacific (one in Australia and two in Japan) regions are in this situation. The regions covered by ARIN and RIPE NCC have 17 and 12 collectors respectively. The scarcity of collectors in LAC was compensated by the addition of local routing information, as we explained in the *Graph Enhancement* section (section 2.3), but no local routing information was added for the regions covered by AfriNIC and APNIC, therefore we will focus on the comparisons between the LACNIC graphs and the ARIN and RIPE NCC graphs, which we consider are in similar conditions in terms of the accuracy of the topology.

The same methodology applied for this work could be applied to enhance the AfriNIC and the APNIC graphs to compare them to the other regions.

### *2.6. Country-level Analysis*

After the detailed analysis on regional connectivity based on the different graph metrics, it is important to go further and try to determine in which areas of the region the connectivity could be improved by deploying IXPs. This critical infrastructure is usually deployed in countries where the opportunities are present (e.g. there is political will and/or there are enough resources). Deploying an IXP always has a big impact for the country, however, the impact on the regional interconnection level that can be caused by the creation of a new IXP may be negligible for some countries.

The development of the Internet infrastructure in a country in general evolves in parallel with barely evident criteria like the economy (GDP) but it probably also evolves with other criteria that may be not that evident (e.g. number of tourists). For this reason we have addressed this topic in a relatively holistic way, looking for correlations between parameters associated with graph connectivity and socioeconomic parameters, and once we have found these correlations, we have relied on them to simulate the creation of an IXP in specific countries and validated the results comparing the new connectivity figures with the old ones and also with other countries with other socioeconomic values.

### 2.6.1. *Correlations with Socioeconomic Indicators and Outliers Detection*

In order to analyze whether there exists any correlation between Internet metrics and some socioeconomic indicators, two groups of variables were created:

1. **Internet Metrics** group: includes the graph metrics computed for the national graphs and the variable IXPsNum, which contains the number of IXPs present in each country.
2. **Socioeconomic Indicators** group: includes population, economic, transport and tourism indicators with information about number of inhabitants, GDP, number of passengers, number of flights, number of airports, number of air carriers, number of arrivals and tourism receipts.

We then calculated the Pearson's correlation coefficient (r) between all the possible pairs of variables formed by one variable of the Internet Metrics group and one variable of the Socioeconomic Indicators group and the corresponding F-test p-value.

For the sake of brevity, a detailed description of all the above listed indicators is not provided. Those variables involved in interesting correlations will be explained in the Results section 3.

After computing all the correlations, we selected those that are significant. We consider significant those correlations that, apart from having a p-value lower than a threshold of 0.05, have an absolute value of the correlation index higher than 0.7.

Then, we made plots of one of the variables versus the other one, coloring the points based on whether there is one or more IXPs in the country or not, in order to detect outliers. We consider outliers those countries that do not have an IXP but in the plots appear located near groups of countries that do have at least one IXP. We understand this as a country that should probably have an IXP since there are countries with quite similar indicators that already have one.

### 2.6.2. *Study of IXP creation impact*

For those countries detected as outliers as explained in the previous section, the creation of an IXP was simulated by adding to the corresponding graph all the relationships that could be established as a consequence of the presence of this new IXP in the country. In order to determine which relationships could be created, some assumptions were considered:

1. Mandatory Multilateral Peering. We assume that for each member of the IXP, peering with the other members of the IXP is mandatory. We know that this is a quite optimistic assumption as only 19 out of 58 IXPs in the LAC region (about one third) have this policy.
2. ASes that are already connected to other IXPs in the region and that are active in the country, will get connected to the simulated IXP. We base this in the assumption that ASes that already peer at some IXP have an open peering policy and are aware of the benefits of peering and if they are already active in the country, they will probably be interested in peering at the new IXP as well.
3. The first 10% of the biggest ASes (with highest Degree) active in the country get connected to the IXP, even if they are not already present in any other IXP. Analogously to what was explained in the previous item, we assume that ASes that already have a high number of peering relationships have an open peering policy.
4. Google and Akamai get connected to the IXP. Google and Akamai are present in five and four countries respectively, out of a total of fifteen countries in the LAC region that have at least one IXP. Although these proportions are not very high, these content providers have an open peering policy and are usually actively looking for locations to install caches, that is why we assume that they would get connected to a new IXP in a near future.

It is very important to bear in mind that these assumptions are rather optimistic, with the results of the simulations being as a consequence, best case scenarios. It is also important to note that, in practice, the main difficulty to make the interconnection at an IXP a reality is the absence of political and/or high management will to interconnect with a competing organization, which makes it almost impossible to guess which ASes would get connected to a new hypothetical IXP. For this reason, we decided to make optimistic but reasonable assumptions on the aspects that can be modeled, considering that the fact of the same assumptions being applied to all the countries where the creation of an IXP wanted to be simulated, lets us estimate the potential impact of this new IXP.

After creating the new relationships, the impact of the simulated IXP was measured by recomputing regional and national graph metrics and comparing them to their former values to determine the absolute and the relative variations both at the regional level and at the country level.

## 3. Results

This section presents the results obtained throughout the different stages of this work. We first comment the impact of adding local routing information to the LACNIC graph. After that we present the results of the comparisons between LACNIC and the other regions and the analysis made at the country level. Some plots are shown in order to better visualize the results obtained, however, not all the plots we generated are included in this article. All the plots and datasets can be found at [23]. The reader can check Table 1 in order to recall the considered graph metrics' definitions.

### *3.1. Impact of addition of local routing information to topologies*

Several studies ( [24, 7, 8, 12, 25]) have shown that inferred Internet AS level topologies are usually incomplete and that it is necessary to increment the number of monitors used to get input data for the inference algorithm. According to [7], using a snapshot from the RouteViews project as input, approximately 70 % of the total links are discovered, so based on this reference we assume that if we use RouteViews and in addition RIS, we should be able to discover as well at least 70 % of the total links. These two data sources provide 3,581 monitor ASes, from which 595 are active in the LAC region, which represents a 12 % from the total of ASes active in the LAC region. Furthermore, when adding local routing information from the LAC region, we add 251 monitor ASes, having now a total of 3,832 monitors, from which 826 are active in the LAC region, which represents 16.6 % of the total ASes active in the region. Incrementing the proportion of monitor ASes in the LAC region from 12 % to 16.6 %, we expect to be even closer to the ground truth in this region than without considering the this local routing information.

After adding local routing information from the LAC region to the inputs considered for the AS relationships inference, we observed some variation in the metrics of the graph for this region. In first place, 7,768 new relationships were discovered, incrementing the number of relationships (Edges) from 23,056 to 30,824, which represents an increment of 33.7 %. Most of these new relationships are peer-to-peer (P2P) relationships as expected taking into account the results of [39, 29, 31], which show that provider-to-customer (P2C) relationships are easy to discover, while P2P relationships usually remain hidden. As a consequence of the increment in the number of edges, the Average Degree and the Maximum Degree were also incremented, from 9.27 to 12.4 and from 811 to 865 respectively. It can also be noted that the maximum shell index also increased considerably. The shell index depends largely on the amount of discovered edges and lateral connectivity, therefore, as new edges were discovered after adding local routing information, it is reasonable that this metric has increased. These variations are clearly shown in Table 2.

Apart from observing the graph metrics, we also analyzed how the Degree Distribution was modified in order to determine whether the newly discovered relationships involved nodes with degree values within a certain range. It can be deduced from Figure 3 that, although increments are present at most values of the degree (k), the highest increment is for nodes with k between 20 and 500 approximately. This is explained by the fact that the additional routing information was obtained from medium-sized ASes. There are some very relevant peaks present in these degree distributions that will be explained in section 3.2.2.

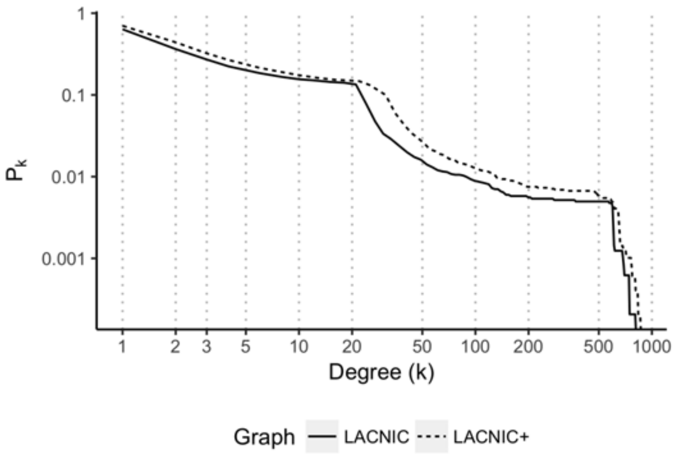



Figure 3: Complementary Cumulative Degree Distributions for LACNIC and LACNIC+ graphs

### *3.2. Comparisons between LAC and other regions*

This section describes the interpretation of the comparisons amongst the regions that were performed. Table 3 presents all the considered graph characteristics that can be summarized to a single value, sorted by the Average Degree. Figures 5(a), 6, 8, 9, 10 and 12 show the rest of the characteristics. They include the LACNIC graph both before and after adding local routing information, tagged LACNIC and LACNIC+ respectively. The following subsections present an analysis for the different characteristics.

#### *3.2.1. Dimension of the graphs*

**Number of nodes:** It can be observed that the LACNIC graphs are the second smallest in terms of number of nodes after the AfriNIC graph. This could be thought to be due to more restrictive Autonomous System Number (ASN) assignment policies, however, pre-RIR ASN assignment policies applied to all the regions and RIR policies are almost the same for the five RIRs. Therefore, the reason behind this is probably the fact that that the Internet started in North America and Europe in the 1950s and 60s and Asia joined in the 1970s [40], but it did not get to Africa and Latin America until around the 1990s [41, 42, 43]. From historical ASN assignment data ([44] and [45]) we could see that before 1990, 292 ASN assignments were made for the ARIN region, while only 5 ASN assignments were made for the LACNIC region. It is important to note that the number of nodes included in the regional graphs does not depend on the availability of route collectors in the regions but just depends on the number

| **Characteristic** | **Short Definition** |
|---|---|
| *Order* | Number of nodes. |
| *Size* | Number of edges. |
| *Giant Component* | Biggest connected component of the graph. |
| *Node Degree* | Number of edges of a node. |
| *Degree Distribution* | Probability that a randomly selected node in the graph has degree k. Internet graphs have power-law distributions. |
| *Average Node Degree* | Number of edges per node on average. |
| *Disconnected Nodes* | Nodes not involved in any relationship. Nodes with k = 0 |
| *Maximum Shell Index* | Number of cores in which the graph can be decomposed. Measures how hierarchical the graph is and provides measure of robustness. |
| *Average Clustering Coefficient vs Degree* | Average clustering coefficient for all the nodes with degree k plotted versus the degree k. |
| *Average Degree Connectivity* | aka Neighbor Connectivity. Average of the degrees of the neighbors of the nodes with degree k plotted versus the degree k. |
| *Joint Degree Distribution (JDD)* | aka Node Degree Correlation matrix. Probability that a randomly selected edge connects two nodes with a certain combination of degree values. |
| *Assortativity* | Pearson's correlation coefficient between the degrees of pairs of nodes that are connected. Local assortativity considers degree in regional graph. Global assortativity considers degree in world graph. |

Table 1: Summary of Considered Graph Metrics

| | **LACNIC** | **LACNIC+** | **Absolute increment** | **Relative Increment (%)** |
|---|---|---|---|---|
| Edges | 23,056 | 30,824 | 7,768 | 33.7 |
| P2P Edges | 14,178 | 20,655 | 6,477 | 45.7 |
| P2C Edges | 8,878 | 10,169 | 1,291 | 14.5 |
| Avg Degree | 9.27 | 12.4 | 3.13 | 33.8 |
| Max Degree | 811 | 865 | 54 | 6.7 |
| Max Shell Index | 27 | 40 | 13 | 48.1 |

Table 2: Impact of adding local routing info to the LACNIC graph

| Graph | **Average Degree** | Size (# of Edges) | Order (# of Nodes) | Giant Component | Disconnected Nodes | Maximum Shell Index | Local Assortativity | Global Assortativity |
|---|---|---|---|---|---|---|---|---|
| LACNIC+ | **12.40** | 30,824 | 4,973 | 4,926 | 47 | 40 | -0.47 | -0.2 |
| RIPE NCC | 9.46 | **112,849** | **23,866** | **23,778** | 88 | **71** | -0.32 | -0.24 |
| LACNIC | 9.27 | 23,056 | 4,973 | 4,835 | 138 | 27 | **-0.48** | -0.2 |
| APNIC | 6.5 | 25,196 | 7,752 | 7,628 | 118 | 30 | -0.3 | -0.11 |
| ARIN | 6.25 | 60,100 | 19,232 | 19,077 | **150** | 45 | -0.27 | **-0.28** |
| AfriNIC | *4.71* | *2,693* | *1,143* | *1,087* | *46* | *12* | *-0.23* | *-0.05* |

Table 3: Graph metric comparison amongst regions sorted by **Average Degree** column. Maximum absolute values are in **bold** letters; minimum absolute values are in *italics*

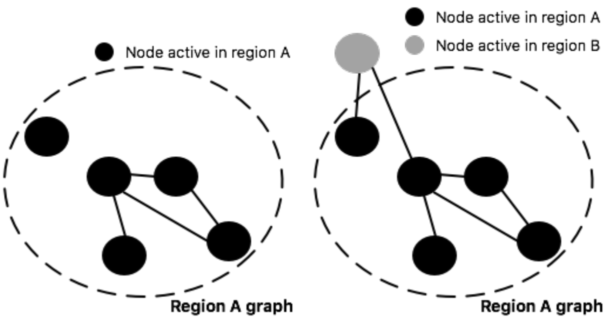
of ASes that are geolocated to each region. AS geolocation is based on the countries from which the AS originates prefixes as explained in 2.2.

**Giant Component and Disconnected Nodes:** Some nodes in the regional graphs may appear to be disconnected because not all the relationships they are involved in could be inferred from the routing information we are using as an input. Another reason could be wrong geolocation information, causing the AS to be geolocated to a region different from the region where it is actually being used, or it could also happen that the AS is being geolocated only based on the country to which it was assigned, but it is being used in another country.

We can see that the ARIN graph is the graph with the highest number of disconnected nodes. Analyzing these disconnected nodes we found that 35 (23.3 %) of them are legacy, reserved or unassigned ASes. Probably there is not enough information to geolocate them and they are being wrongly geolocated as a consequence.

Vervier et al in [46] confirmed the existence of campaigns of malicious BGP hijacks using registered but unannounced IP address space. Therefore, it is not strange that the ARIN graph is the one with the highest number of disconnected nodes as the Internet started in North America and, at the beginnings of the Internet, lots of Universities and other organizations got assigned big blocks of IP addresses that nowadays are probably not being used by the holder organization, making this address space attractive to hijackers who could use it in some other region.

It can also be observed that there are a lot of disconnected nodes for the LACNIC graph, but this number dramatically drops (from 138 to 47) when we add local routing information, as a lot of new relationships between ASes active in the LAC region are discovered.




(a) Graph of Region A with a disconnected node

(b) Graph of Region A with node connected through common neighbor

Figure 4: Some nodes that appear to be disconnected could be connected to the rest of the graph through a node that is active in another region.

We were curious about the real situation of these disconnected nodes and we thought some of them could be not really disconnected from the graph, but connected through some AS that is not active in the region, as shown schematically by Figure 4. Hence, we decided to remove from the sets of disconnected nodes those that have a common neighbor with any of the other nodes of the graph. After doing this we could see that the number of disconnected nodes for the graphs was reduced from 150 to 64, from 138 to 32. from 118 to 51, from 88 to 20, from 47 to 32 and from 46 to 31 for the ARIN, LACNIC, APNIC, RIPE NCC, LACNIC+ and AfriNIC graphs respectively. A possible cause for some nodes not being connected to the rest of the graph (directly or even indirectly) could be that the corresponding ASes are being misgeolocated and actually operate in other region/s, although an exhaustive analysis of the cause goes beyond the scope of this article.

#### *3.2.2. Interconnection between ASes: Average Degree, Number of Edges and Degree Distribution*

Although the number of edges of a graph is one of the important characteristics to analyze, it is an absolute value not really useful as a measure of interconnection when making comparisons between graphs of different sizes. This is why we preferred to sort Table 3 by the average degree variable, which provides a normalized measure of size.

Regarding the degree distributions for our graphs, Barabasi, in Chapter 4 of his book [32], recommends to use log-log plots with logarithmic binning to show degree distributions of Internet graphs, as the visualization of this kind of distributions with regular scales and regular binning is quite useless. Following this recommendation, the Complementary Cumulative Distribution Function (CCDF) of the node degree of Internet graphs should be a straight decreasing line as shown in [35]. In order to compare the different regions' distributions, the degree (k) in the x axis was normalized dividing it by the average degree of the corresponding graph (¡k¿), annotating it in the plots as $k/<k>$.

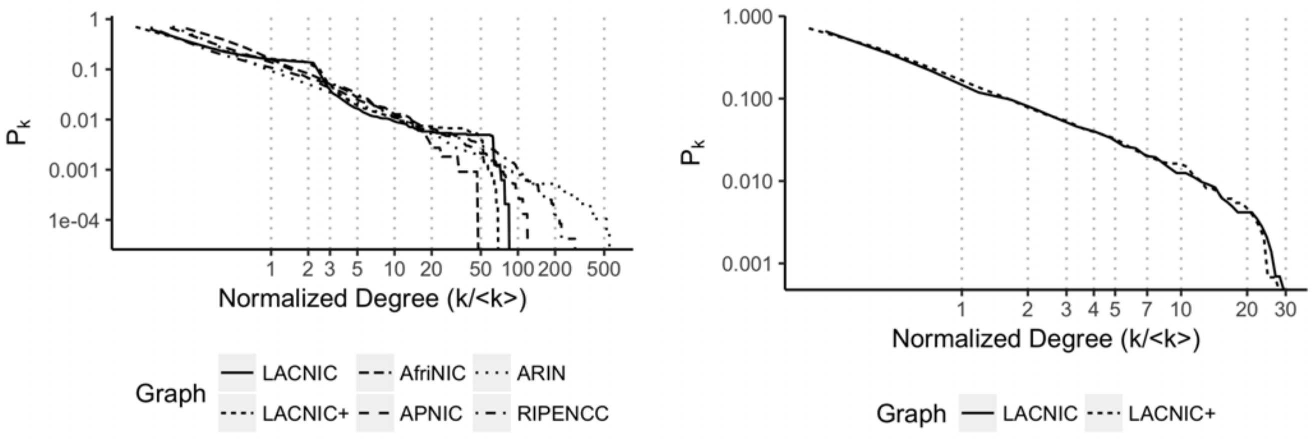
From Table 3 it may be surprising to see that the LACNIC+ graph has the highest average degree (12.4). However, the LACNIC+ graph is the enriched graph, so it could be claimed this is why it has the highest average degree. What in fact is surprising is that even before adding local routing information (LACNIC graph), the average node degree was quite high (9.27), similar to the value for the RIPE NCC graph and quite dissimilar to the values of the rest of the graphs (the next one is the APNIC graph with an average degree of 6.5).

One could conclude from this that the interconnection in the LAC region is not as bad as expected.

However, observing the Degree Distributions in Figure 5(a), we can see that most graphs seem to have a power-law distribution, but LACNIC graphs present two interesting peaks approximately around $k/<k>=2$ and $k/<k>=50$. After analyzing in detail the nodes with degree values in the proximities of the corresponding values of k, we found that more than 90 % of them are from Brazil.

It is widely known that Brazil is an outlier in Latin America in terms of population, economy, etc., even in terms of "Latin Americanness" self-perceived by citizens [47, 48, 49]. Internet interconnection is not an exception for this as Figure 14 shows. For this reason, we decided to remove the nodes active in Brazil and plot the degree distribution for the LAC region again.

Figure 5(b) shows the degree distributions for the LACNIC graphs excluding the nodes active in Brazil. It can be observed that now there are no big peaks and the curves are more similar to a power-law distribution. We conclude from this that Brazil is an outlier in the LAC region, having a behavior in terms of interconnection that does not follow the behavior of the rest of the countries in the region. After removing the ASes active in Brazil, the average degrees decrease from 12.4 to 4.99 and from 9.27 to 4.19 for the LACNIC+ and the LACNIC graphs respectively, which shows that the interconnection level in the LAC region is one of the lowest amongst all the regions.




(a) CCDFs for regional graphs

(b) CCDFs for LACNIC without nodes active in Brazil

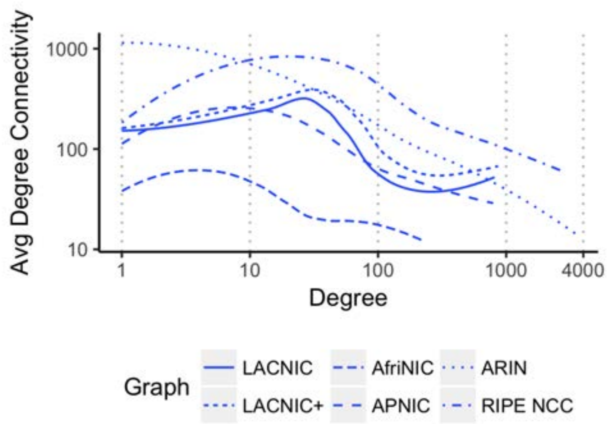
Figure 5: Complementary Cumulative Degree Distribution Functions

#### *3.2.3. Correlation of Degrees of Connected Nodes: Assortativity, Average Degree Connectivity and Joint Degree Distribution*

In first place we analyze the Assortativity of the graphs and we see that all the graphs are disassortative (negative assortativity). This means that most links connect nodes of dissimilar degrees. We found an interesting result for the Global Assortativity (the Assortativity computed using the degree of the nodes in the world graph (global degree)), which is very low for all of the graphs (the maximum absolute value is 0.28 and corresponds to the ARIN graph), meaning that the correlation between the global degree of interconnected nodes is very weak. This shows that the total number of neighbors (peers, providers or customers) of an AS in the world graph does not have an influence on the size of the ASes it connects to.

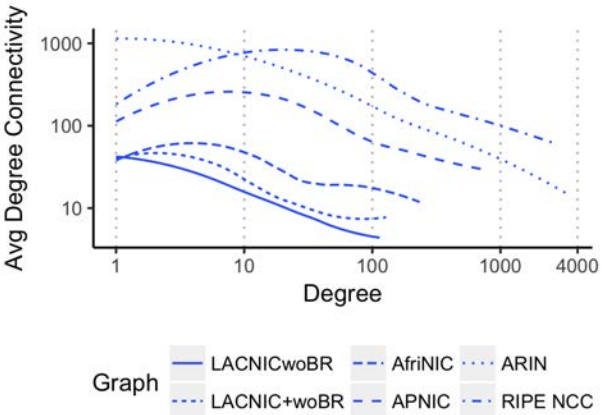
On the other hand, in general the absolute values for the Local Assortativity (the assortativity computed using the degrees of the nodes in the regional graph) are higher. In particular, the LACNIC graphs (both before and after adding the local routing information) are the most disassortative, suggesting that in the LAC region it is more likely that an AS with a high number of neighbors in the region gets connected to an AS with a low number of neighbors in the region. However, it is worth mentioning that again Brazil has a strong influence in this result, being active in this country almost 86 % of the nodes involved in relationships with nodes of very different degree. When the nodes active in Brazil are removed from the LACNIC graphs and the Local Assortativity is recomputed, both for LACNIC and for LACNIC+ the new value is -0.33, which is more similar to the values in the other regions.

Regarding the Average Degree Connectivity, all the graphs being disassortative, one would expect the curves corresponding to the Average Degree Connectivity versus the Degree for all the graphs to be mostly decreasing. Observing Figure 6 we can see that this holds true for degree values higher than approximately 30 for all the graphs except for the LACNIC graphs, which show an increase after approximately k=200. Once again we are witnessing the Brazil-effect: We analyzed in detail the ASes in the LACNIC graphs that have a degree value higher than 200 and all of them are active in Brazil. After removing the nodes active in Brazil, recomputing this metric and plotting again, we now got decreasing plots for the LACNIC graphs as shown by Figure 7 and the lowest Average Degree Connectivity amongst all regions.




***Note:*** *The average degree connectivity shows a high variance between consecutive values of the degree, obtaining as a consequence plots with too much noise. As we are interested in the general tendency of the average degree connectivity, we apply an approximating function in order to plot a smoothed version of the curves.*

Figure 6: Regions Average Local Degree Connectivity




***Note:*** *The average degree connectivity shows a high variance between consecutive values of the degree, obtaining as a consequence plots with too much noise. As we are interested in the general tendency of the average degree connectivity, we apply an approximating function in order to plot a smoothed version of the curves.*

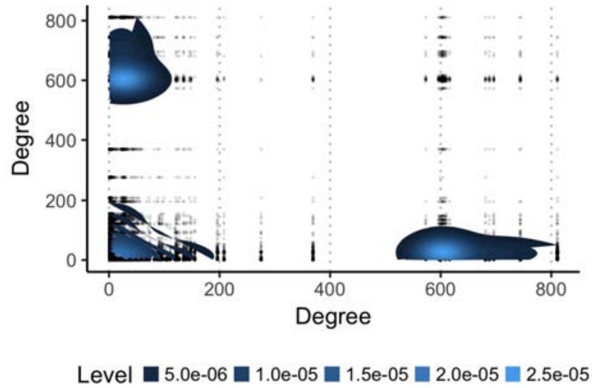
Figure 7: Regions Average Local Degree Connectivity removing nodes active in Brazil from LACNIC graphs

It is also worth mentioning that the JDD plots for the LACNIC and LACNIC+ graphs (shown in Figure 8 are quite peculiar in comparison to the plots corresponding to the other regions. The other JDD plots show the highest levels of concentrations for edges between nodes of low degree (as shown by Figure 9), while both LACNIC and LACNIC+ plots also show a high level of concentration of edges connecting nodes with very high degree (around $k = 600$) with nodes with very low degree. Again Brazil is the responsible for this as all the nodes with degree higher than 200 are nodes active in Brazil and the areas of high concentration of edges mentioned represent edges between these nodes.

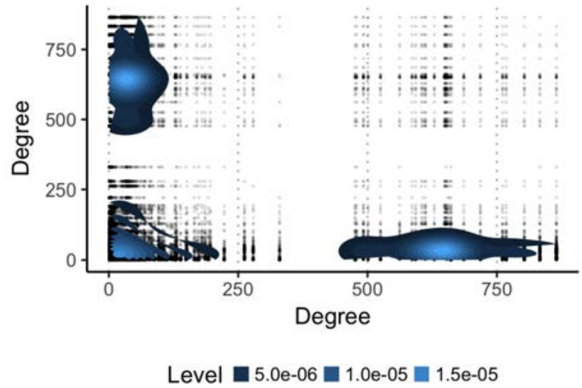
### *3.2.4. Local Interconnection between ASes: Average CC vs Degree*

The average clustering coefficient (CC) for the nodes with a certain value of degree measures how close the neighbors of those nodes are to forming a complete graph (i.e. a fully interconnected graph), giving us an idea of how dense the neighborhoods of those nodes are. Clustering expresses local robustness in the graph and thus has practical implications: the more dense the neighborhood of a node is, the higher the local path diversity around the node [33].

The plots of Average CC versus Degree show the dependence of this local link density on the node's degree k. To create these plots the CC of all the nodes with degree k is averaged for the different values of k. By definition, the CC is zero for nodes with degree values of zero or one, therefore disconnected nodes and nodes with only one neighbor are not considered.




(a) Deg-Deg Correlation LACNIC




(b) Deg-Deg Correlation LACNIC+

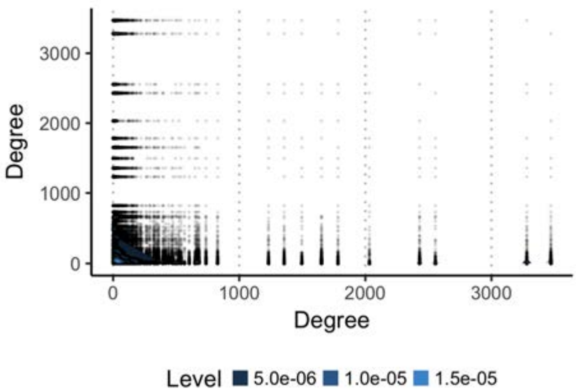
***Note:*** *These are scatter plots combined with heatmaps that show the level of concentration of edges that connect nodes with the different combinations of the degree values (each dot of the plot represents an edge of the graph and the density curves show the concentration of dots in the corresponding area of the plot).*

Figure 8: Correlation between degree of connected nodes for LACNIC and LACNIC+ graphs

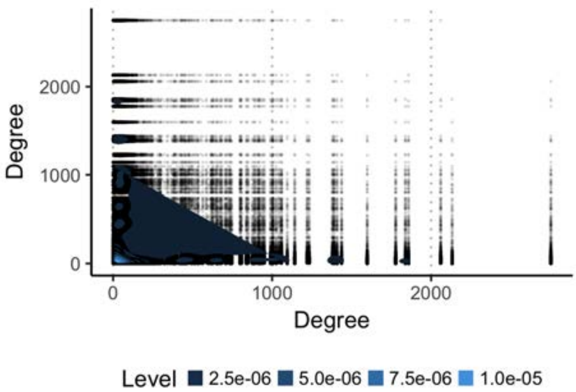
As explained in the previous section, the graphs we are working with are disassortative and for this reason, we would expect the plots of average CC vs Degree to be decreasing as nodes with high degree usually get connected to nodes with lower degree and therefore, the neighbors of nodes with high degree are less likely to be interconnected. In other words, we would expect that the highest the degree value of a node is, the more sparse its neighborhood will be. This holds true for the APNIC, ARIN and AfriNIC graphs as shown by Figure 10, however, we can see that the LAC region has a behavior similar to the behavior of the region covered by RIPE NCC, showing a peak for medium-size nodes ($10 < k < 50$ approximately), which appear to be in much more dense neighborhoods.

Analyzing those medium-size nodes for the LAC region, we found out that about 97 % of them are active in Brazil, showing again the consequences of the Brazil effect. After removing the nodes active in Brazil, recomputing the CC for all the nodes, recomputing the Average CC for the different values of Degree and plotting again, we can see in Figure 11 that now the Average CC vs Degree plots for the LAC region are decreasing as expected.

The situation for the RIPE NCC graph is harder to explain as there is not an obvious outlier as in the case of the LAC region. We analyzed the nodes with a degree value between 10 and 50 and a high CC ($CC > 0.4$), that are the ones that generate the peak, and we could not find any outlier country. However, AMS-IX, one of the largest IXPs in the world, is based in this region and is most likely one of the main influencers on the regional metrics. After analyzing the list of nodes that meet the above mentioned conditions and the relationships these nodes are involved in, we found out that all of them are connected to AMS-IX or to an AS that is connected to AMS-IX, explaining why the local interconnection is so high for medium-size ASes in this region and confirming the high impact of an IXP in the AS topology as studied by [50].



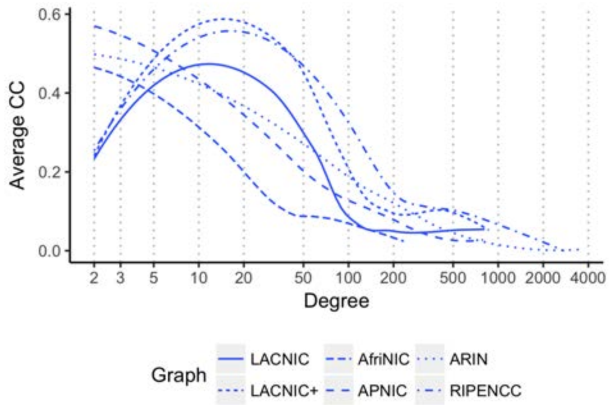

(a) Deg-Deg Correlation ARIN




(b) Deg-Deg Correlation RIPE NCC

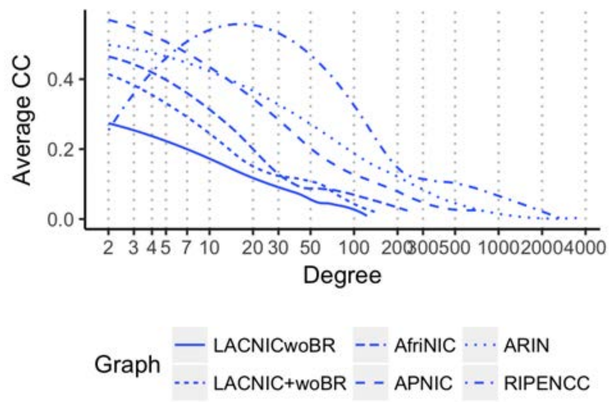
***Note:*** *These are scatter plots combined with heatmaps that show the level of concentration of edges that connect nodes with the different combinations of the degree values (each dot of the plot represents an edge of the graph and the density curves show the concentration of dots in the corresponding area of the plot).*

Figure 9: Correlation between degree of connected nodes for ARIN and RIPE NCC graphs




***Note:*** *Similarly to the average degree connectivity, there is a high variance in the average CC for consecutive values of the degree, obtaining as a consequence plots with too much noise. As we are interested in the general tendency of the average CC, we apply an approximating function in order to plot a smoothed version of the curves.*

Figure 10: Regions Average Clustering Coefficient vs Degree




***Note:*** *Due to the high variance in the average CC for consecutive values of the degree, the original plots have too much noise. As we are interested in the general tendency of the average CC, we apply an approximating function in order to plot a smoothed version of the curves.*

Figure 11: Regions Average Clustering Coefficient vs Degree removing from LAC nodes active in Brazil

#### 3.2.5. Interconnection and Robustness: K-Core Decomposition

We finally analyze the k-core decomposition of the different graphs in order to compare them in terms of robustness.

We can see from Table 3 that the RIPE NCC graph has the highest maximum shell index (71), with a great distance to the second one (the ARIN graph with a maximum shell index of 45) and in the third place, the LACNIC+ graph (40), not that far from the second. As mentioned in section 2.4.5, the higher the maximum shell index is, the more robust the network is, because there are more paths to interconnect its ASes. Therefore, we can conclude that the RIPE NCC graph is the most robust and that LACNIC's robustness is similar to that of ARIN.

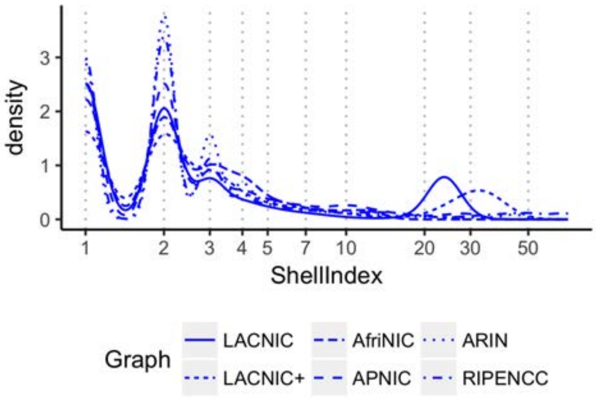



Figure 12: Regions Shell Index Density (Size of cores)

Besides, analyzing the Shell Index distributions shown in Figure 12 we can observe that, in general, shells with low index are usually bigger than higher shells. However, once again LACNIC graphs are an exception, showing peaks for high shells. This means there are more nodes in higher shells, which could be interpreted as a better robustness in the LAC region than in the rest of the regions.

However, we want to highlight that again Brazil has a fundamental influence in this result, disrupting not only this but all the different considered metrics. Analyzing more in detail the Shell Index for the nodes corresponding to the LAC graphs, we observed that, from the nodes with Shell Index higher than 20, more than 90 % are active in Brazil.

Although each region could have their own outliers, this work is focused on the LAC region, making very important the fact of detecting a country in this region that so strongly affects the regional metrics. In order to have a better idea of the real situation in the LAC region, we removed the nodes active in Brazil from the LACNIC+ graph and recomputed the k-core decomposition of the graph and found out that now the Shell Index distributions for all the regions are similar, as shown by Figure 13, confirming that Brazil is an outlier in the LAC region.

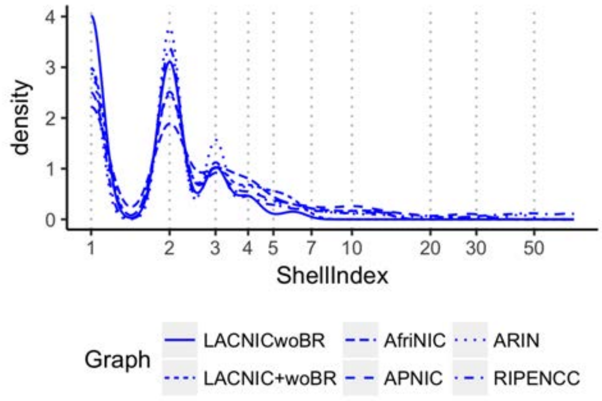



Figure 13: Regions Shell Index Density (Size of cores) - Removing nodes active in Brazil from the LACNIC graphs

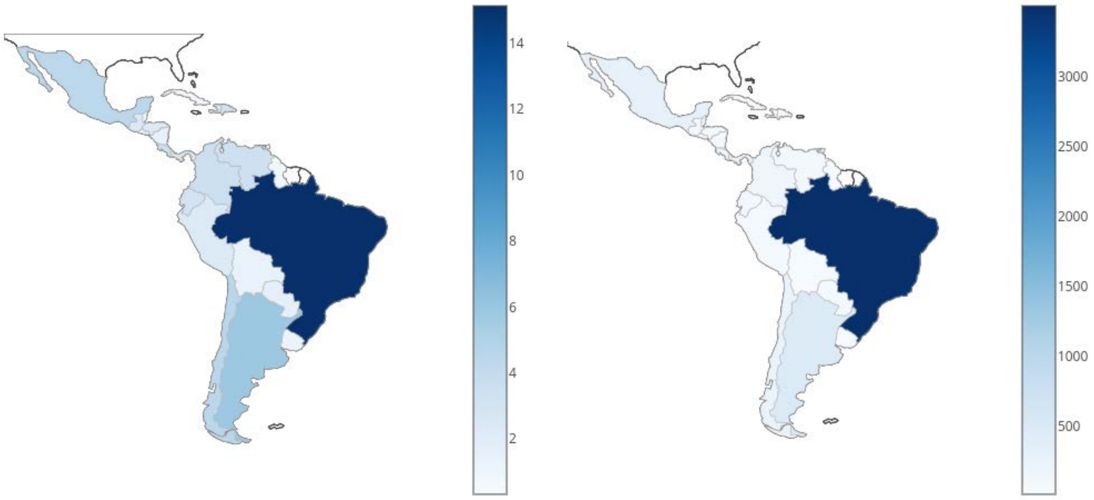
When further studying the effect on the LAC graph of removing the nodes active in Brazil, we could see that the maximum shell index dropped from 40 to just 6. This leads us to a remarkable conclusion: the robustness of the LAC region is even worse than the robustness of Africa when Brazil is not considered.

### 3.3. Country-level Analysis

Comparisons similar to those we performed for regions could be made as part of the country-level analysis using an analogous methodology. For this work the goal is to focus on the LAC region, which includes around 30 countries. Comparing so many countries through a table or through a plot including all the corresponding distributions would be quite confusing and is beyond the scope of this work. As a way of exemplifying how the different characteristics for the countries in LAC could be displayed and compared, in Figure 14 we present geographical plots for the Average Degree and for the Number of Nodes of the countries in LAC respectively. Geographical plots for all the computed metrics for country graph can be found in [23]. We can clearly see once again that Brazil is an outlier in the LAC region.

#### 3.3.1. Correlations with Socioeconomic Indicators and Outliers Detection

So far we have discovered that, if Brazil is not taken into consideration, the LAC region is very poorly interconnected. We now want to determine in which countries it would be useful to deploy an IXP. As we have already mentioned, we computed the correlation between metrics strictly associated to graphs and socioeconomic variables in order to verify whether it is possible to somehow predict the most adequate country to locate an IXP. With this purpose we computed the correlation indexes between the pairs of variables formed by one "Internet Metrics" variable and one "Socioeconomic Indicators" variable and filtered the significant correlations ($|r| > 0.7$ and




(a) Average Node Degree (b) Number of Nodes

Figure 14: Country Metrics

$p-value < 0.05$). We found six pairs of variables strongly correlated, which are listed below:

- Nodes and Number of Airports
- P2C Edges and Number of Airports
- Maximum Shell Index and Air Passengers Carried
- Maximum Shell Index and International Tourism Receipts
- Maximum Shell Index and Total Population
- Maximum Degree and Number of Airports

The "Number of Airports" variable gives the total number of airports or airfields recognizable from the air [20]. The "Air Passengers Carried" variable includes both domestic and international aircraft passengers of air carriers registered in the country. The "International tourism receipts" variable measures expenditures by international inbound visitors, including payments to national carriers for international transport. The "Total Population" variable is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship– except for refugees not permanently settled in the country of asylum, who are generally considered part of the population of their country of origin. [19].

It is quite reasonable that the number of airports correlates with the considered graph parameters as in some way, the airports are built in locations that are neuralgic, either for commercial or for tourist reasons, or that are strategic geographic points for transit.

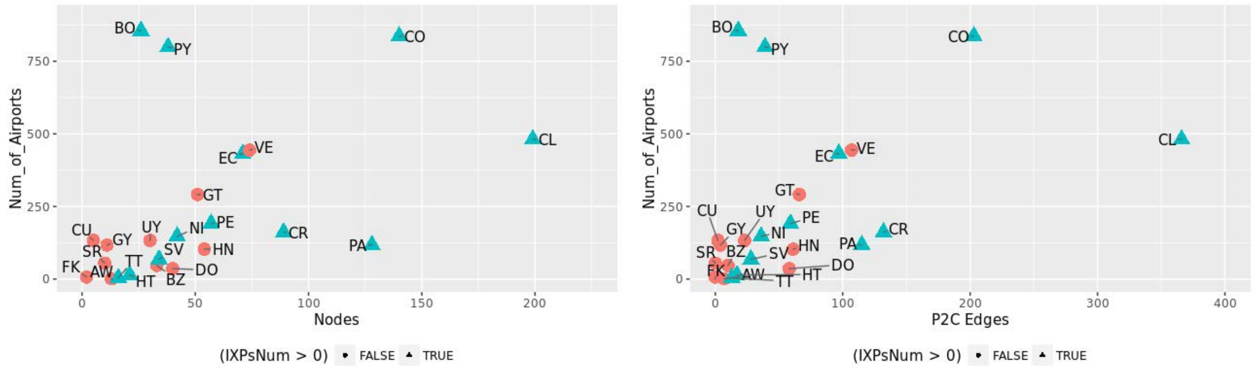
Figure 15 shows plots for the two most correlated pairs, both with correlation indexes (r) of 0.73. Plots are zoomed in the area where most countries are concentrated in order to show clearly the countries of interest. Observing these and the rest of the plots corresponding to significant correlations, we saw that two countries that appear as outliers in most of the plots are Venezuela (VE) and Guatemala (GT). We consider them outliers because although they appear in the plots surrounded by countries that have at least one IXP, they do not have one. Plots for all the possible combinations of the variables of the two groups can be found in [23].

*3.3.2. Study of IXP creation impact depending on location*

Applying the assumptions for the simulation of the creation of an IXP that were mentioned in the Methodology section (section 2), we added to the national graphs and to the LACNIC graph the resulting sets of relationships. It is interesting to note that, although the mandatory multilateral peering assumption is quite optimistic, most of the countries around the detected outlier countries in which we run the simulation (GT and VE) have this policy.

Table 4 shows the relative and the absolute variations of some graph metrics both at the regional level and at the country level. We can observe that the regional impact is not very high but the impact at the country level is considerable. For example, we can see that the average degree for the countries shows a 51.4 % and a 65.9 % of increment for GT and VE respectively.

In particular, it is interesting to analyze the impact of the new IXP in the shell index of the nodes. At the regional level we can see that there is no change in the maximum shell index in GT or in VE, but the corresponding average shell indexes show an increment, meaning that, although the total number of cores in the graphs is the same, the new relationships made some nodes move to higher cores. At the country level we can observe that the increment in the maximum shell indexes is huge for both countries,




(a) Number of Airports vs Nodes (r=0.85)

(b) Number of Airports vs Number of P2C Edges (r=0.73)

Figure 15: Correlations for Country Metrics

| | Abs Δ Edges | Rel Δ Edges (%) | Abs Δ Avg Deg | Rel Δ Avg Deg (%) | Abs Δ Avg Shell Index | Rel Δ Avg Shell Index (%) | Abs Δ Max Shell Index | Rel Δ Max Shell Index (%) |
|---|---|---|---|---|---|---|---|---|
| **GT Regional Impact** | 33 | 0.11 | 0.013 | 0.1 | 0.007 | 0.11 | 0 | 0 |
| **GT Country Impact** | 39 | 57.35 | 1.37 | 51.4 | 1.37 | 84.34 | 6 | 200 |
| **VE Regional Impact** | 86 | 0.28 | 0.035 | 0.28 | 0.016 | 0.26 | 0 | 0 |
| **VE Country Impact** | 95 | 70.37 | 2.4 | 65.9 | 2.59 | 126.1 | 11 | 275 |

Table 4: Regional and Country-level impact of IXP creation in Guatemala and in Venezuela

i.e. the country graphs can be decomposed in much more cores when the IXP is present, which translates into much more robustness.

We also run the simulation for two countries that do not have an IXP yet and are not outliers: Cuba (CU) and Guyana (GY). When measuring the impact of the creation of an IXP in these countries, we found out that, although the impact at the country level is quite high (e.g. an increment of 50 % and 60 % respectively in the number of edges), the regional impact is negligible (0.003 % or less for all the considered metrics and for both countries).

It is worth comparing the regional impact of the creation of an IXP at an outlier country against the creation of an IXP at a non-outlier country. For example, creating an IXP in Guatemala or in Venezuela causes an increment of 0.1 % and 0.28 % respectively on the average degree of the regional graph, while creating an IXP in Cuba or Guyana causes an increment of only 0.003 % in both cases on this metric. These results suggest that our holistic criteria may be useful to choose countries in which it would be worth investing to deploy an IXP, although they are not enough to draw general conclusions.

Something important to mention is that we are just measuring the individual impact on the region of creating a single IXP in only one out of more than 30 countries in the LAC region. It is expected that the creation of more IXPs in more countries would have a bigger impact than the sum of the individual impacts, but a more complex simulation should be run in order to confirm this.

## 4. Conclusions and Future Work

Throughout this article we explained in detail how we created graphs representing the Internet AS topology, worldwide and at the region and country levels (for countries in Latin America). Our main focus for this work was to better understand the interconnection level in the LAC region as compared to other regions, so, in order for the LAC graph to be comparable to the other graphs, we added local routing information from looking glasses and other local sources in the region, compensating the lack of vantage points. We showed that a lot of relationships (specially peering relationships) were hidden as many studies suggested, and we exposed and used them to improve our graphs.

We then compared the different graphs, based on various graph metrics, and found that the LAC region does not seem to be as badly interconnected as we suspected. Nevertheless, we detected that the high level of interconnection in this region was strongly influenced by autonomous systems from Brazil, therefore we decided to exclude them from the analysis. Comparing the different graphs, it seems that the LAC region connectivity is comparable to the one in other regions. However, after detecting that the metrics that are normally used are strongly influenced by ASes in Brazil, we concluded then that the LAC region is really very poorly interconnected and has a very weak robustness (even worse than AfriNIC, for instance).

In order to evaluate the room for improvement we performed a country level analysis correlating interconnection and socioeconomic metrics to try to find the proper places to locate IXPs and validate the approach by means of simulations that suggest that these correlations are in fact good indicators for our purposes. We observed that the impact on the regional interconnection level and robustness for the candidate countries is quite high, while the impact for the non candidate countries is negligible. In the case of the national impact, it is quite big both for the candidate countries and for the non candidate countries (from 50 % to more than 100 % of increment in different metrics). This shows that the deployment of an IXP in any country considerably improves the interconnection level and the robustness of the national Internet topology, with all the benefits this carries along.

We hope that, with these results in mind, governments and other organizations will understand the importance of improving the national interconnection level, e.g. by deploying Internet Exchange Points, not only to enhance the national Internet but also the regional Internet as a whole.

Apart from extracting these interesting conclusions, we have identified some areas in which further work could be done. First of all, it would be interesting to perform some tasks oriented to disambiguate in which country the interconnection between two ASes happens. There are some big ASes that are active in lots of countries and so far we just know that they have relationships with other ASes but we do not know at which of those countries these ASes get interconnected. PeeringDB[11] information could be used, although this information could be out of date or incomplete. Another option would be to use information about prefixes being announced by the ASes.

It could also be interesting to use information from CAIDA's AS-Org mapping project [12] to merge all the nodes that correspond to the same organization and work with Org-level graphs in stead of AS-level graphs.

In relation with the metrics analysis and graph comparison, it would be interesting to add measures of centrality (betweenness, closeness, etc.) to the graph metrics considered and, at the country-level studies, to include the delay between countries measured by SIMON project [13] (LACNIC) as part of the "Socioeconomic Indicators" variables when looking for correlations at the country level. We would also like to apply the methodology described in [51] to compare the regional graphs using the normalized Laplacian spectrum (NLS), which has been shown to be a powerful tool for comparing graphs with different sizes.

Finally, it would be interesting to design a more complex, realistic and complete simulation of the IXP creation, taking into consideration, for example, the simultaneous creation of IXPs in more than one target country, and perform more simulations in order to be able to draw general conclusions.

## 5. Acknowledgements

The work done by Francisco Valera has been partially granted by the European Commission project LEONE (From local measurements to global management, grant number FP7-317647). The work done by Sofía Silva has been partially funded by IMDEA Networks. The authors would like to thank Marcelo Bagnulo and Andra Lutu for their help, and for providing ideas, comments and support during the initial phases of this work.

[11] https://www.peeringdb.com/

[12] https://www.caida.org/data/as-organizations/

[13] https://simon.lacnic.net

*Vitae

**Sofía Silva Berenguer (⋆1986)**

Sofía Silva Berenguer holds a degree in Telematics Engineering from Universidad de Montevideo. During her professional career Sofía has worked with different technologies and has performed research related to topics such as the Internet, Routing, Security, the Domain Name System (DNS), the Resource Public Key Infrastructure (RPKI), Domain Name System Security Extensions (DNSSEC), Anycast, and Internet Exchange Points.
Sofía started studying her Master's degree working as a Research Assistant for IMDEA Networks (Madrid) and then as an Assistant Professor and Researcher for University Carlos III of Madrid. Before joining IMDEA Networks and up to December 2014, Sofía worked as Senior Security and Stabiity Specialist at the Internet Address Registry for Latin America and the Caribbean (LACNIC). She originally joined LACNIC's Registration Services Department as Hostmaster/Policy Officer, and later went on to work as Infrastructure, Systems and Security Engineer within the organization's Technical Department. Before joining LACNIC, Sofia worked in IBM Uruguay's Networking department and was Head of Computer Security for an outsourcing project by IBM.

**Francisco Valera Pintor (⋆1974)**

Prof. Dr. Francisco Valera received the Telecommunication Engineering degree in 1998 from the Technical University of Madrid (UPM), and the Ph.D. in Telecommunications in 2002 from the Univ. Carlos III de Madrid (UC3M), where he is currently a tenured associate professor. He has been involved in several national and international research projects and contracts related with experimental facilities, protocol design, inter-domain routing, protocol engineering, next generation networks and multimedia systems, serving there as PI, work package leader and also as coordinator. Some of the recent research projects in which he has participated, are: IST Trilogy, IST LEONE, DRONE, IST MUSE or IST E-NEXT. Dr. Valera has published over 50 papers in the field of advanced communications in magazines and conferences. He has also has participated in the scientific committee, organization and technical review in different national and international conferences.

## 6. References


[1] R. Motamedi, R. Rejaie, W. Willinger, A survey of techniques for internet topology discovery, IEEE Communications Surveys Tutorials 17 (2) (2015) 1044–1065. doi:10.1109/COMST.2014.2376520.

[2] M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the internet topology, SIGCOMM Comput. Commun. Rev. 29 (4) (1999) 251–262. doi:10.1145/316194.316229.
URL http://doi.acm.org/10.1145/316194.316229

[3] D. Magoni, J. J. Pansiot, Analysis of the autonomous system network topology, SIGCOMM Comput. Commun. Rev. 31 (3) (2001) 26–37. doi:10.1145/505659.505663.
URL http://doi.acm.org/10.1145/505659.505663

[4] R. NCC, Routing information service, https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/routing-information-service-ris, [Online; accessed 16-March-2016] (2001).

[5] RouteViews, University of oregon route views project, http://www.routeviews.org/, [Online: accessed 16-March-2016] (2000).

[6] M. Roughan, W. Willinger, O. Maennel, D. Perouli, R. Bush, 10 lessons from 10 years of measuring and modeling the internet's autonomous systems, IEEE Journal on Selected Areas in Communications 29 (9) (2011) 1810–1821. doi:10.1109/JSAC.2011.111006.

[7] R. Oliveira, D. Pei, W. Willinger, B. Zhang, L. Zhang, The (in)completeness of the observed internet as-level structure, IEEE/ACM Transactions on Networking 18 (1) (2010) 109–122. doi:10.1109/TNET.2009.2020798.

[8] R. Cohen, The internet dark matter: on the missing links in the as connectivity map, in: in Proc. IEEE INFOCOM, 2006.

[9] R. Fanou, P. Francois, E. Aben, PAM 2015, New York, NY, USA, Springer International Publishing, Cham, 2015, Ch. On the Diversity of Interdomain Routing in Africa, pp. 41–54. doi:10.1007/978-3-319-15509-8_4.
URL http://dx.doi.org/10.1007/978-3-319-15509-8_4

[10] O. Messano, International internet connectivity in latin america and the caribbean, https://www.itu.int/en/ITU-D/Regulatory-Market/Documents/International%20Internet%20Connectivity%20in%20Latin%20America%20and%20the%20Caribbean.pdf, [Published by ITU. Online: accessed 11-May-2016] (March 2013).

[11] G. F. Treverton, L. Mizell, The future of the information revolution in latin america: Proceedings of an international conference, http://www.rand.org/pubs/conf_proceedings/CF166-1.html, [Santa Monica, CA: RAND Corporation. Online: accessed 11-May-2016] (2001).

[12] L. Sani, [Advisor: Lenzini, Prof. Luciano. Coadvisor: Gregori, Ing. Enrico . pp. 111. [IMT PhD Thesis]] (2014). [link].
URL http://e-theses.imtlucca.it/142/1/Sani_phdthesis.pdf

[13] S. Silva Berenguer, E. Carisimo, J. I. Alvarez-Hamelin, F. Valera Pintor, Hidden internet topologies info: Truth or myth? (may 2016). doi:10.1145/2940116.2940118.
URL http://dx.doi.org/10.1145/2940116.2940118

[14] ISOC, An introduction to internet interconnection concepts and actors, http://www.internetsociety.org/sites/default/files/bp-interconnection.pdf, [Internet Society Briefing Paper. Online: accessed 01-Jun-2016] (2012).

[15] H. Galperin, Connectivity in latin america and the caribbean: The role of internet exchange points, .http://live.v1.udesa.edu.ar/files/AdmTecySociedad/14_galperin%20rev.pdf, [Universidad de San Andres/CONICET (Argentina). Online: accesses 01-Jun-2016] (2013).

[16] R. V. Oliveira, B. Zhang, L. Zhang, Observing the evolution of internet as topology, SIGCOMM Comput. Commun. Rev. 37 (4) (2007) 313–324. doi:10.1145/1282427.1282416.
URL http://doi.acm.org/10.1145/1282427.1282416

[17] P. C. House, Packet clearing house web page, https://www.pch.net/, [Online; accessed 07-June-2016].

[18] R. NCC, Ripestat, https://stat.ripe.net, [Online; accessed 16-March-2016] (2016).

[19] W. Bank, World bank open data, http://data.worldbank.org/, [Online; accessed 07-June-2016] (2014).

[20] C. I. Agency, The world factbook, urlhttps://www.cia.gov/library/publications/the-world-factbook/index.html, [Online: accessed 13-Jun-2016] (2015).

[21] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, k. claffy, G. Riley, AS Relationships: Inference and Validation, ACM SIGCOMM Computer Communication Review (CCR) 37 (1) (2007) 29–40.

[22] CAIDA, As relationships, http://www.caida.org/data/as-relationships/, [Online: accessed 01-Jun-2016].

[23] S. Silva Berenguer, Radiography of Internet ASes interconnection in the LAC region (jun 2016).
URL http://shiny.netcom.it.uc3m.es/ASinterconnectionLAC/

[24] M. Roughan, J. Tuke, O. Maennel, Bigfoot, sasquatch, the yeti and other missing links: What we don't know about the as graph.

[25] E. Gregori, A. Improta, L. Lenzini, L. Rossi, L. Sani, On the incompleteness of the as-level graph: A novel methodology for bgp route collector placement, in: Proceedings of the 2012 ACM Conference on Internet Measurement Conference, IMC '12, ACM, New York, NY, USA, 2012, pp. 253–264. doi:10.1145/2398776.2398803.
URL http://doi.acm.org/10.1145/2398776.2398803

[26] K. Chen, D. R. Choffnes, R. Potharaju, Y. Chen, F. E. Bustamante, D. Pei, Y. Zhao, Where the sidewalk ends: Extending the internet as graph using traceroutes from p2p users, in: Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies, CoNEXT '09, ACM, New York, NY, USA, 2009, pp. 217–228. doi:10.1145/1658939.1658964.
URL http://doi.acm.org/10.1145/1658939.1658964

[27] Y. Shavitt, E. Shir, Dimes: Let the internet measure itself, SIGCOMM Comput. Commun. Rev. 35 (5) (2005) 71–74. doi:10.1145/1096536.1096546.
URL http://doi.acm.org/10.1145/1096536.1096546

[28] A. Khan, T. Kwon, H.-c. Kim, Y. Choi, As-level topology collection through looking glass servers, in: Proceedings of the 2013 Conference on Internet Measurement Conference, IMC '13, ACM, New York, NY, USA, 2013, pp. 235–242. doi:10.1145/2504730.2504758.
URL http://doi.acm.org/10.1145/2504730.2504758

[29] H. Chang, R. Govindan, S. Jamin, S. J. Shenker, W. Willinger, Towards Capturing Representative AS-level Internet Topologies, Comput. Netw. 44 (6) (2004) 737–755. doi:10.1016/j.comnet.2003.03.001.
URL http://dx.doi.org/10.1016/j.comnet.2003.03.001

[30] V. Giotsas, S. Zhou, Improving the discovery of ixp peering links through passive bgp measurements, in: INFOCOM, 2013 Proceedings IEEE, 2013, pp. 3249–3254. doi:10.1109/INFCOM.2013.6567146.

[31] Y. He, G. Siganos, M. Faloutsos, S. Krishnamurthy, Lord of the links: A framework for discovering missing links in the internet topology, IEEE/ACM Trans. Netw. 17 (2) (2009) 391–404. doi:10.1109/TNET.2008.926512.
URL http://dx.doi.org/10.1109/TNET.2008.926512

[32] A.-L. Barabási, Network Science, 2012.
URL http://barabasi.com/networksciencebook/

[33] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, k. claffy, A. Vahdat, The Internet AS-Level Topology: Three Data Sources and One Definitive Metric, ACM SIGCOMM Computer Communication Review (CCR) 36 No 1 (2006) 17–26.

[34] A. Medina, I. Matta, J. Byers, On the origin of power laws in internet topologies, SIGCOMM Comput. Commun. Rev. 30 (2) (2000) 18–28. doi:10.1145/505680.505683.
URL http://doi.acm.org/10.1145/505680.505683

[35] G. Siganos, M. Faloutsos, P. Faloutsos, C. Faloutsos, Power laws

and the as-level internet topology, IEEE/ACM Trans. Netw. 11 (4) (2003) 514–524. doi:10.1109/TNET.2003.815300.
URL http://dx.doi.org/10.1109/TNET.2003.815300

[36] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, A. Vespignani, k-core decomposition: a tool for the analysis of large scale internet graphs, CoRR abs/cs/0511007.
URL http://arxiv.org/abs/cs/0511007

[37] V. Batagelj, M. Zaversnik, An o(m) algorithm for cores decomposition of networks, CoRR cs.DS/0310049.
URL http://arxiv.org/abs/cs.DS/0310049

[38] M. G. Beiró, J. I. Alvarez-Hamelin, J. R. Busch, A low complexity visualization tool that helps to perform complex systems analysis, New Journal of Physics 10 (12) (2008) 125003.
URL http://stacks.iop.org/1367-2630/10/i=12/a=125003

[39] R. Bush, O. Maennel, M. Roughan, S. Uhlig, Internet optometry: Assessing the broken glasses in internet reachability, in: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, IMC '09, ACM, New York, NY, USA, 2009, pp. 242–253. doi:10.1145/1644893.1644923.
URL http://doi.acm.org/10.1145/1644893.1644923

[40] ISOC, Brief history of the internet, http://www.internetsociety.org/brief-history-internet, [Internet Society Brief History of the Internet. Online: accessed 10-Jun-2016] (2012).

[41] ISOC, History of the internet in africa, http://www.internetsociety.org/what-we-do/where-we-work/africa/history-internet-africa, [Internet Society History of the Internet in Africa. Online: accessed 10-Jun-2016] (2003).

[42] A. Cassinelli, J. Fernandez, Better late than never, http://cs.stanford.edu/people/eroberts/cs201/projects/2006-07/latin-america/, [Stanford University. Online: accessed 10-Jun-2016] (2006).

[43] W. Thomasson, James; Foster, L. Press, The diffusion of the internet in mexico, http://lanic.utexas.edu/project/etext/mexico/thomasson/thomasson.pdf, [Online: accessed 10-Jun-2016].

[44] Arin's historical assignment data, http://ftp.arin.net/pub/stats/arin/delegated-arin-extended-latest, [Online: accessed 07-Dec-2016].

[45] Lacnic's historical assignment data, ftp://ftp.lacnic.net/pub/stats/lacnic/delegated-lacnic-latest, [Online: accessed 07-Dec-2016].

[46] P.-A. Vervier, O. Thonnard, M. Dacier, Mind your blocks: On the stealthiness of malicious bgp hijacks (feb 2015).
URL http://dx.doi.org/10.14722/ndss.2015.23035

[47] V. . Derek Hill (Arlington, Latin america: R&d spending jumps in brazil, mexico, and costa rica, National Science Foundation, Division of Science Resources Studies 00 (316).

[48] J. Onuki, F. Mouron, F. Urdinez, Latin American Perceptions of Regional Identity and Leadership in Comparative Perspective, Contexto Internacional.
URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-85292016005002106&nrm=iso

[49] B. Research, Little respite for sovereign risk in 2016, Key Analysis.

[50] K. Xu, Z. Duan, Z.-L. Zhang, J. Chandrashekar, On Properties of Internet Exchange Points and Their Impact on AS Topology and Relationship, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 284–295. doi:10.1007/978-3-540-24693-0_24.
URL http://dx.doi.org/10.1007/978-3-540-24693-0_24

[51] B. Jiao, J. mai Shi, Graph perturbations and corresponding spectral changes in internet topologies, Computer Communications 76 (2016) 77 – 86. doi:http://dx.doi.org/10.1016/j.comcom.2015.11.011.
URL http://www.sciencedirect.com/science/article/pii/S0140366415004570