A Multilayer Graph Model of the Internet Topology

Georg Tilch [1], Tatiana Ermakova [2] and Benjamin Fabian [3,*]

[1] Humboldt University Berlin, georg.tilch@gmail.com
[2] University of Potsdam; ermakova@uni-potsdam.de
[3] Humboldt University Berlin, bfabian@wiwi.hu-berlin.de;
Hochschule für Telekommunikation Leipzig (HfTL); fabian@hft-leipzig.de
* Correspondence: bfabian@wiwi.hu-berlin.de; Tel.: +49 341 3062 225

Abstract

Purpose: Despite intensive research during the last two decades, the detailed structural composition of the Internet is still opaque to researchers. Nevertheless, due to the importance of Internet maps for the development of more effective routing algorithms, security mechanisms, and resilience management, more detailed insights are required. This article advances the understanding of the Internet structure by integrating data from different large-scale measurement campaigns into a set of comprehensive Internet graphs at different abstraction levels, and analyzes them in terms of important statistics and graph measures.

Design/methodology/approach: This study follows the topology measurement framework suggested by Gunes and Sarac (2009), involving three phases: topology collection, topology construction, and topology analysis.

Findings: An integrated data set of Internet graphs at different abstraction layers is provided that can serve as a baseline for future research on Internet analytics. Furthermore, results of important graph metrics are presented and power-law relationships for the degree distributions on every level of the current Internet are substantiated.

Research limitations/implications: By necessity, the integrated graphs provide a snapshot of the Internet topology. In future work, repeated measurements and automated data integration could lead to a better understanding of Internet dynamics.

Practical implications: Due to increasing dependency on the Internet as a critical global infrastructure, studying Internet connectivity is more important than ever for both companies and Internet service providers. The data set will be made publically available for network research.

Social implications: Understanding the structure of Internet serves as a fundamental step in improving the robustness, security, and privacy of any online service.

Originality/value: By carefully integrating six different traceroute datasets such as iPlane, CAIDA, Carna, DIMES, RIPE Atlas, and RIPE IPv6L, this paper presents the Internet graphs of a substantially larger and thus solid scale than previously known, at well-established abstraction levels such as the IP interface, router, Point of Presence (PoP), Autonomous System (AS), and Internet Service Provider (ISP). Furthermore, by employing a broad diversity of graph measures, this study creates a more exhaustive snapshot of the global Internet topology than earlier works.

Keywords:
Internet · Internet topology · Computer networks · Graph analysis · Complex networks

1 Introduction

The Internet has radically been changing many aspects of modern society, from personal relations to novel businesses. However, this has also created a great dependency on the Internet infrastructure, which has been experiencing massive growth (Ho et al., 2007). Researchers have been striving to understand its properties and evolution through studying its topology in both their static and dynamic aspects. *Topology* refers to the edge-based and structural attributes of a network or a graph (Zaki and Meira, 2014, p. 93): the various entities, such as routers, and their interconnections. The study of Internet topology enables the evaluation of performance and vulnerabilities of the Internet infrastructure and individual services in the case of failures or intentional attacks, with increasing importance in light of present cybercrime and cyber warfare (Albert and Barabási, 2000; Cohen et al. 2000; Cohen et al. 2001; Doyle et al., 2005; Xiao et al., 2008; Sterbenz et al. 2010; Sterbenz et al. 2013; Doerr and Kuipers, 2014; Baumann and Fabian, 2014; Fabian et al. 2015). Internet maps are also important for the development of more effective routing algorithms and security and privacy mechanisms.

Due to its decentralized architecture, massive scale, and constantly changing nature, obtaining a comprehensive model of the Internet's structure is very challenging. The study on Internet topology is hampered by the fact that there is no single authority overlooking its development. Furthermore, organizations that have information about parts of the Internet, such as Internet Service Providers (ISP), are reluctant to unveil it due to privacy and security concerns. The fear of losing competitive advantage is another driver for the confidentiality of local topologies. For these reasons, researchers usually collect massive amounts of paths through the Internet with a traceroute application and *infer* the actual topology from these measurements. Nevertheless, there are countless possible paths through the network and time and money constraints force researchers to focus their efforts, affecting the accuracy of the inferred topology maps.

The purpose of this research is to address some of those drawbacks by combining six major traceroute data sets into a unified set of integrated graph models of the Internet for a recommended period of two weeks (Huffaker et al., 2012; Shavitt and Zilberman, 2010). These include those ones provided by iPlane (Madhyastha et al., 2006), Center for Applied Internet

Data Analysis (CAIDA, 2013), the globe-spanning "Carna" botnet (Botnet, 2013a, 2013b), DIMES (Shavitt and Shir, 2005; Donnet and Friedman, 2007), Regional Internet Registry (RIR) for Europe, the Middle East, and Central Asia (RIPE Atlas, and RIPE IPv6L). The investigated levels involve the IP interface, router, Point of Presence (PoP), Autonomous System (AS), and Internet Service Provider (ISP). This "snapshot" serves as a foundation for the calculation of graph statistics and measures describing the Internet topology in this article. We will make this data set publically available for future network research.

This study follows three common phases of topology measurement (Gunes and Sarac, 2009): 1) topology collection, 2) topology construction, and 3) topology analysis. The remainder of this article is organized as follows: First, an overview on the Internet topology is provided in Section 2. Then in Section 3, the different data sources are presented and evaluated. Section 4 describes the methodology for topology construction. Section 5 discusses important graph measures and prepares the subsequent analysis of different topological levels presented in Section 6. Section 7 concludes with a discussion of limitations and future work.

2 Internet Topology

The Internet topology can be analyzed at five granularity or abstraction levels illustrated in Figure 1. Those are the IP interface, router, Point of Presence (PoP), Autonomous System (AS), and Internet Service Provider (ISP) levels.

The most fine-grained resolution available in public data is the *IP interface level* (black dots in Figure 1). Each router has by definition at least two interfaces while backbone routers may have many more. Each interface is assigned with one or multiple IP addresses. At this granularity, each IP address appears as a node in a graph, whereas an edge refers to a network hop (at layer 3). This implies that each router appears multiple times in an IP interface graph.

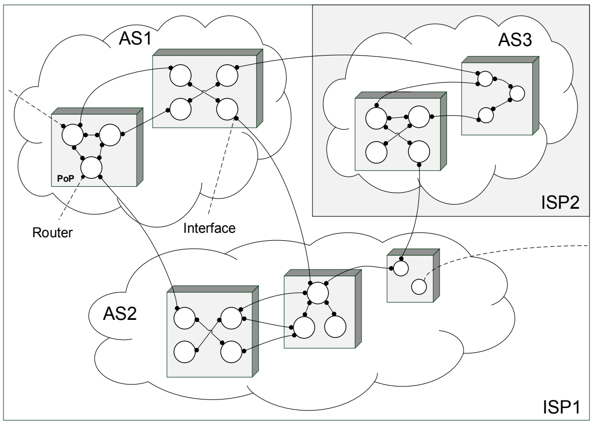

Figure 1: The different levels of Internet topology

Note that this topology granularity disregards devices and connections working at lower protocol layers such as hubs or switches. It is unfortunately not trivial to infer lower layer topologies, and therefore network research has mainly concentrated on the IP interface level and above, which can be remotely measured. The *traceroute* tool can be used to infer the connections between two IP addresses and therefore has become a standard in topology research. It uses increasing time-to-live (TTL) values in a series of outgoing packets to a particular destination and analyzes the returning error messages for information on intermediate hops.

The *router-level* topology (white circles in Figure 1) is the result of aggregating all interfaces belonging to a single router in a process called *alias resolution* (AR) (Spring et al., 2004). Its quality has significant influence on the resulting topology. Next in granularity is the level of *Point of Presence* (PoP) (rectangles in Figure 1). A PoP is typically defined as a physical location where an AS accommodates a group of routers providing connectivity to the users of the AS in the area and connecting to higher hierarchy levels (Willinger and Roughan, 2013). Usually the clustering of routers into PoPs is achieved by investigating the affiliation of IP addresses to an AS and their spatial proximity.

An even higher level in the taxonomy is the *Autonomous System* (AS) (clouds in Figure 1), which can be one network or a group of networks under the authority of a single institution (e.g., universities, Internet Service Providers, banks). This level represents a logical view on the Internet, where nodes are ASes and edges represent *data forwarding relationships* among entities. This representation is also useful in determining the importance of sub-networks since some ASes are at the core while others primarily provide connectivity to end-users. Furthermore, this granularity is used in many research articles since it is less prone to distortions and significantly simplifies the analysis (Chang et al., 2004).

There is no one-to-one relationship between an AS and the organization that administers it. It is possible that one ISP can control several ASes, for example, because of mergers or acquisitions. Conceptually, the *ISP level* considers only connections between different organizations (shaded areas in Figure 1). This level is useful for analyzing actual business relationships between organizations.

3 Data Sources

Our work integrates six traceroute datasets from five different measurement projects. Observed nodes and links are merged for a specific measurement period (Govindan and Tangmunarunkit, 2000; Huffaker et al., 2002). A larger time window results in more detected links; however, some of these connections are already disintegrated at the end of the observation window (He et al., 2009; Huffaker et al., 2002). Recent works agree to use a period of two weeks (Huffaker et al., 2012; Shavitt and Zilberman, 2010), which we also adopt in our study. Like most measurement projects we focus on the IPv4 topology. During the "IPv6 Launch" on June 6, 2012, major ISPs permanently enabled IPv6 for their services (Internet Society, 2012). This is the main motivation why our work considers data collected during the timeframe of June, 7–20, 2012 as observation period. This sampling period allows obtaining a comprehensive view of the IPv4 Internet without large distortions by the new routing protocol. Furthermore, it might be the last opportunity to examine the "old" address space before too many devices use the new protocol exclusively. Table 1 gives an overview on the data sources used for this time period.

Table 1: Summary of the traceroute data sources

	iPlane	CAIDA	Carna	DIMES	RIPE Atlas	RIPE IPv6L
Size of raw data	45.9 GiB	86.2 GiB	17.8 GiB	30.7 GiB	20.3 GiB	30.5 GiB
Number of files	2,106	1,154	1	7	35	1
Number of records	264.6 mn.	203.3 mn.	67.0 mn.	21.0 mn.	20.9 mn.	10.3 mn.
Vantage points	299	56	266,604	783	4,780	56
Destination IPs	127,566	195.7 mn.	63.0 mn.	2.3 mn.	39	4,323
Number of traces	112.9 mn.	105.6 mn.	41.8 mn.	15.1 mn.	4.1 mn.	1.8 mn.

Madhyastha et al. (2006) introduced iPlane in 2006. The distributed iPlane architecture runs on various nodes of PlanetLab. Although the data account for 264.1 million records over our time period, the number of *unique* traces is much smaller: 57.31% of the records are duplicates, which results in 112.9 million unique traces. This is still the largest number among all datasets, which is evidence of an elaborate probing strategy.

One of the most renowned institutions for Internet topology research is the Center for Applied Internet Data Analysis (CAIDA). They focus on a collaborative approach for data collection and encourage sharing of data (CAIDA, 2013). After removing duplicate paths (48.05%), 105.6 million unique traces remain.

On March 17, 2013, another extraordinary dataset was published online by an anonymous hacker (Botnet, 2013a, 2013b). Both the collection method and the scale of the data are unprecedented: The dataset comprises a collection of results of various probing techniques and has a decompressed size of 9 TB. The author asserts that these data are the result of measurements conducted with the globe-spanning "Carna" botnet, which was created solely for an "Internet Census", i.e., a complete scan of the entire IPv4 Internet in 2012. While the different scanning routines are technically appropriate, scanning *without* consent or knowledge of the owners of the monitors can be considered unethical. However, the data is “out there” and can be considered as *pre-existing public data* (Department of Homeland Security, 2012; Dittrich et al., 2014; Krenc et al., 2014). It is possible to substantiate the validity of the dataset via reverse engineering (Krenc et al., 2014) or by a comparison with similar projects conducted at the same time (Le Malécot and Inoue, 2014). This consideration and the fact that the Carna botnet was presumably not used in malicious activity led to the inclusion of those measurements. From the raw traceroute records, only 37.63% were duplicates and 41.8 million unique traces remain, which cover many different parts of the Internet.

DIMES is a globally distributed topology measurement project (Shavitt and Shir, 2005). The goal is to study the large-scale topology at different levels with the help of voluntary contributors who install software agents (Donnet and Friedman, 2007). Though the project stalled in early 2012, we are grateful for having obtained raw data for our observation period directly from the DIMES team, contributing 15.1 million unique traces.

The RIPE NCC is the Regional Internet Registry (RIR) for Europe, the Middle East, and Central Asia. RIPE hosted two topology measurement projects at the time of the observation period: It collected traceroute data in the *Test Traffic Measurement* (TTM) and during the *IPv6 Launch Day* (IPv6L). Both datasets contain only marginal amounts of information since even together they only constitute less than 3% of the final IP graph. Both RIPE datasets show a vast amount of duplications due to the measurement strategy that repeatedly probed the same destinations to infer changes in performance. After the removal of traces that traversed the exact same path, 1.84 million records are left.

The data sources are combined to enable an Internet graph analysis at unprecedented scale. Merging data from diverse topology discovery projects could convert individual drawbacks to advantages because the information gathered from various points and with diverging methods provide a complemented view on the Internet. A first indication of the quality of the combined dataset can be observed through exploring the traceroute characteristics: The individual raw data sum to vast 231.4 GiB and include 587.3 million records for the duration of the observation period. The number of unique monitors adds up to 272,505. In Figure 2 all monitors are plotted on a world map.

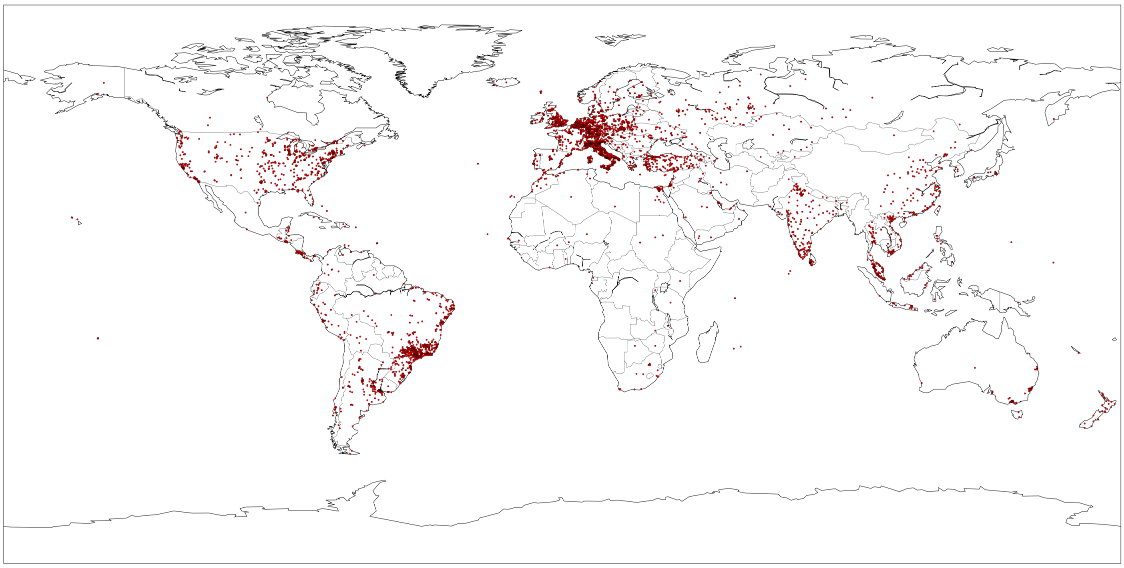

Figure 2: Geographic location of all vantage points

The 257.7 million unique destination IPs in the combined dataset targeted *all* 14.4 million routable /24 prefixes. Overall, the average length of traceroutes is a small: 13.29 with a standard deviation of 6.04 hops.

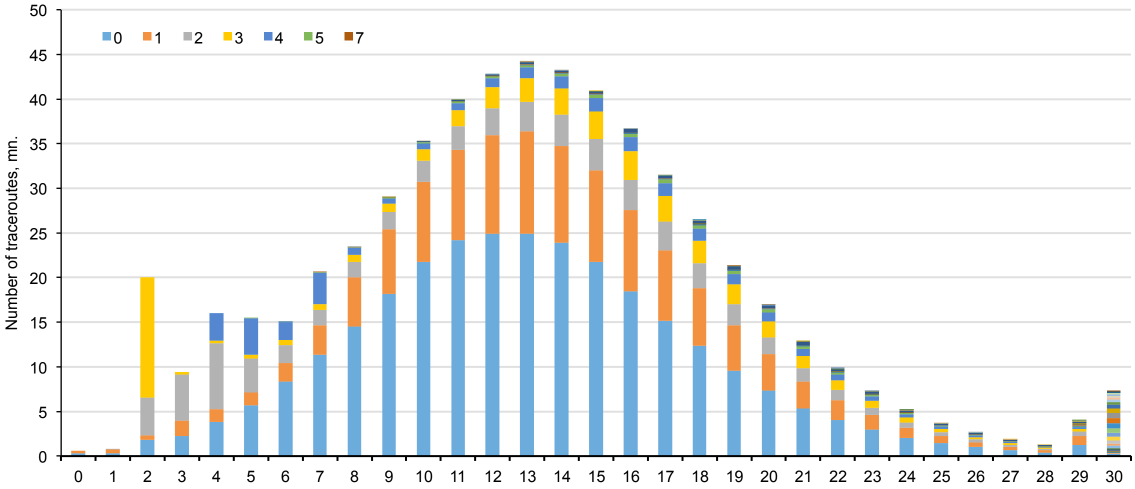

Figure 3: Distribution of trace lengths in hops, combined dataset (occurrences of anonymous interfaces colored)

There is a smooth hop distribution (Figure 3) demonstrating the underlying principle in an illustrative way: The individual datasets may have distortions or peculiarities due to shortcomings in the traceroute implementation, the choice of packets, or the location of sources and targets. However, abnormalities "even out" after superimposing the individual components. Consequently, the features of the combined dataset produce results that are probably closer to reality than those of any of the individual approaches alone.

Finally, the acquisition, preprocessing, and integration of the data result in a combined dataset with 281.5 million unique traces for the two-week observation period. To our knowledge, this is the largest and the most encompassing IPv4 dataset in a traceroute-based topology discovery project so far; it establishes a thorough foundation for the following graph extraction and analyses.

4 Topology Construction

The next phase involves *topology construction.* Adjacencies in traces are interpreted as direct edges of the IP graph. Moreover, the graphs of higher topology levels can be constructed through aggregation. The assumption is that layer 3 information contains knowledge about the higher levels of the Internet; akin to the statistical physics approach that links "microscopic dynamics and interactions […] to the statistical regularities of macroscopic physical systems" (Pastor-Satorras and Vespignani, 2007). Data from all sources are processed in *exactly* the same fashion to alleviate the risk that differences in data aggregation influence the properties of the graphs.

A general question is whether more traceroute measurements actually result in better information about the Internet structure. One point of view is that additional probes suffer from diminishing returns since additional probes capture the same paths over and over again. While the decreasing marginal utility is undisputed, the mass under the long tail of traceroute probing may provide significant additional information (Shavitt and Shir, 2005). Moreover, the approach followed in this article moves beyond merely adding new measurements but actually imitates a completely different probing structure when data from different projects are integrated, which yields a more comprehensive picture of the actual Internet topology and can alleviate individual measurement biases.

Table 2: Overview on the size of the extracted graphs

	IP	Router	PoP	AS	ISP
Unique nodes	3,358,491	2,893,862	56,385	33,756	31,034
Unique edges	8,636,410	5,109,228	104,558	122,563	113,491

After the correction of traceroute peculiarities (such as loops), the actual IP edges are extracted directly from the processed traces. In total, 3.93 billion IP edges are extracted. The resulting *unique edgelist* has considerably fewer records, which was expected because of multiple traversings of sub-paths: 99.71% of edges were duplicates, which results in 11.2 million unique IP edges.

Revealing for the quality of the input data is the number of unique edges for the individual datasets. Initially iPlane had the most traces, followed by CAIDA, Carna and DIMES. This has shifted after edge extraction: DIMES became the second largest dataset in terms of unique edges. This is a strong indicator of the quality of their destination selection algorithms and the fact that intelligent measurement design determines the usefulness of collected edges.

This article focuses on undirected graphs in order to make a cross-level comparison possible. It is straightforward to construct the undirected graph by sorting edges and subsequently removing duplicates, which results in 8,636,410 undirected IP edges**.** Some examinations, such as the position of monitors in Figure 2, require knowledge about the geographic location of Internet resources. After some consideration, MaxMind GeoLite (MaxMind, 2015) was selected for geolocation because it is available for the historic period and is regularly used by other researchers as well.

For converting an IP level graph into a router topology, it is necessary to cluster IP addresses of routers. Because of the time passed since the data was gathered, a passive AR technique was applied to the datasets for which we adopted the *kapar* tool of CAIDA (Keys, 2010). The procedure resulted in 2,893,862 nodes and 5,109,228 undirected router edges.

We then proceeded to the AS-level because PoP graph construction is based on AS inference. The mapping of IP addresses to their origin AS should be conducted at the time this AS administered the respective prefix since the connections among ASes change. The BGP table dumps collected by RouteViews are preprocessed by CAIDA into historical prefix-to-AS mappings (CAIDA, 2016b). The final AS graph contains 33,756 unique nodes and 122,563 edges**.**

CAIDA collects quarterly dumps from WHOIS servers and processes them into an "Inferred AS to Organization Mapping Dataset" (CAIDA, 2016a). Using this, it is straightforward to further aggregate the data into the ISP-level. After sorting and removing of duplicates, 31,034 ISP nodes and 113,491 ISP edges remain.

PoP-level maps allow for identifying important locations of a network, supporting important purposes such as vulnerability assessments (Huffaker et al., 2012; Shavitt and Zilberman, 2010). Researchers developed several methods of how to infer PoP-level maps. The rationale of our approach is to investigate the affiliation of IP addresses to the owning ASes and their spatial proximity. With this, 25.6% of IP edges could be converted into PoP edges. After the removal of internal edges and duplicates, the undirected graph has 56,385 PoP nodes and 104,558 undirected PoP edges**.**

5 Graph Measures

Our process of data preprocessing and integration results in one integrated undirected graph per topological level. In the following, the *largest connected component* (LCC) is analyzed for each graph. The extracted graphs are further evaluated and compared at different levels in terms of important characteristics that quantify local and global properties of network graphs.

The graph analysis is executed by applying major graph analysis frameworks for Python, one of which is NetworKit, a high-performance graph analysis software introduced by researchers of the Karlsruhe Institute of Technology (KIT) (Staudt et al., 2014; Karlsruhe Institute of Technology, 2015). The unique advantage of this software suite is that the algorithms are highly parallelized and provide the ability for massive scaling.

5.1 Graph Size and Node Degree

A graph is formally defined as a pair of a finite nonempty set of vertices or nodes *N* and a set of edges *L*: *G = (N, L).* The *number of nodes* in graph *G*, denoted as |N| = *n*, is called the *order* of *G*. The *number of edges* in graph *G*, denoted as |L| = *m*, is called the *size* of *G* (Zaki and Meira, 2014). Real-world graphs typically have much less edges than maximally possible and are therefore referred to as *sparse*. This property can be quantified with the network's *density*, or in other words the fraction of present edges in all possible edges in the given graph (Barabási, 2016):

$$\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)} \tag{1}$$

Density ranges between $0 \leq \rho \leq 1$ and is almost 0 for sparse graphs.

The main local property of a vertex is its *degree k*, which is the number of edges incident with it, indicating the connectivity of that vertex towards the rest of the network. The entire graph can be characterized through the *average degree* $\langle k \rangle$ (Pastor-Satorras and Vespignani, 2007) and the extreme values *minimum degree* k_{min} and *maximum degree* k_{max} can give further insights. Graphs with a higher average degree are generally better connected and are therefore more robust towards failures (Huffaker et al., 2012). However, the explanatory value of the average degree is limited because the actual network structure of two graphs with the same $\langle k \rangle$ can be extremely different (Mahadevan et al., 2005).

Furthermore, the entire *degree distribution P(k)* specifies the probability that any randomly chosen vertex has degree k. It can also be seen as the fraction of nodes that share the same number of neighbors in a network. The degree distribution of *G* is given by the number of nodes of degree k, n_k, divided by the total number of nodes n (Pastor-Satorras and Vespignani, 2007; Zaki and Meira, 2014):

$$f(k) = P(X = k) = \frac{n_k}{n} \quad (2)$$

The degree distribution is a pivotal network metric because it contains all information with respect of degrees. The general structure can be excellently described with the degree distribution and many network characteristics can be inferred from the shape of *f(k)*. A common assumption, based on the findings of Barabási and Albert (1999), is that the degree distribution of the Internet follows a *power-law distribution* of the form:

$$P(k) \sim k^{-\gamma} \quad (3)$$

where γ is a positive exponent (Barabasi, 2016).

A typical property of these distributions is that they are heavy-tailed, i.e., there are many low-degree nodes and few that have high-degree nodes. Power-law degree distributions however do not necessarily follow relation (3) strictly over their whole range (Barabási, 2016). They could deviate from it for small degrees, but usually follow it in the tail after a *minimum power-law degree* k_{min}^{PL}.

The value of the power-law exponent is at the center of Internet topology research because the average degree as summary statistic of the degree distribution has no explanatory value due to the scale-freeness. Most works have found the exponent to be between 2 and 3 (e.g., Faloutsos et al., 1999). A distinct property of scale-free networks is the prevalence of "hubs", i.e., nodes with extremely high *k* that are pivotal for the structural integrity of these networks (Willinger et al., 2009). Since the power-law properties are important for the evaluation of the network characteristics, much effort was made in our work to infer the exponent γ by maximum-likelihood estimation (Clauset et al., 2009), using the Python package "powerlaw" (Alstott et al., 2014).

5.2 Clustering Coefficient

A local metric that goes beyond the characteristics of a single node is *clustering*, which indicates the tendency of neighbors of a vertex to connect to each other (Pastor-Satorras and Vespignani, 2007). The *local clustering coefficient* c_i measures the probability that the neighbors of a vertex are connected (Newman, 2013). Given the subgraph induced by the neighbors of the vertex, the metric represents the fraction of present edges in all possible edges (Zaki and Meira, 2014).

$$c_i = \frac{m_i}{\binom{n_i}{2}} = \frac{2m_i}{n_i(n_i - 1)}$$

The local clustering coefficient ranges between 0 and 1; a value of, for instance, 0.4 would indicate that there is a 40% chance that two neighbors of a vertex are connected (Barabási, 2016). This local metric is aggregated for the whole graph with the *average clustering coefficient* $\langle c \rangle$, which is just the average c_i of all nodes in the network (Watts and Strogatz, 1998). The average clustering coefficient is an indicator for the local robustness of the whole network (Mahadevan et al., 2005).

There is also a second metric for assessing the clustering of the complete graph, which is the *global clustering coefficient* C (Newman, 2013), or *transitivity* (Zaki and Meira, 2014):

$$C = \frac{3 \times Number\ of\ Triangles}{Number\ of\ Connected\ Triples} \tag{4}$$

C investigates the whole network without a need to calculate each local metric. Even though the average local clustering coefficient $\langle c \rangle$ and the global clustering coefficient *C* are conceptually related, they can yield significantly different values. Calculating both measures may complement each other and can thus improve inferences about robustness.

5.3 Assortativity

The interconnections of vertices can be investigated based on assortative mixing, which refers to the tendency of nodes to be connected to nodes with similar properties (Newman, 2003). Informative for the assessment of the topology of a network is whether it is *assortatively mixed by degree*. This refers to the extent to which nodes are connected to other nodes with similar degree (Newman, 2013). In an assortative network, high degree nodes tend to connect to each other. With respect to the Internet this would signify that major hubs cluster together, resulting in a core of high degree vertices and a periphery of low-degree nodes (Newman, 2013). Conversely, in a disassortative network, hubs tend to connect to vertices with low degrees without a core.

5.4 Distance Measures

The *path length* is the number of hops a path traverses from one vertex to another. The *shortest path* (or distance, geodesic path) between two vertices *i* and *j* is represented by d_{ij} (Barabási, 2015; Newman, 2013). Distances from one individual node give an idea about its position in the network: The *eccentricity* ε_i of a node i is the maximum distance starting from i to any other node. It is thus the longest shortest path originating from i. The *radius* is the minimum eccentricity for any node within the whole network (Mahadevan et al., 2005) but more commonly used it the opposite metric: the maximum eccentricity for any node, referred to as *diameter* d_{max}. The diameter thus refers to the longest shortest path within the whole network (Barabási, 2016). Since shortest path and eccentricity are local metrics for individual nodes, usually averaged distance measures are calculated in order to make a statement about the entire network.

5.5 Centrality

Given all the shortest paths between all pairs of nodes through a network, some vertices are traversed more often than others, what is quantified by *betweenness centrality* b_i (Pastor-Satorras and Vespignani, 2007). This measures the number of shortest paths that pass through a certain vertex i. The values of betweenness centrality depend on the size of the network, which is why the measure is often normalized by $n(n-1)$ (Mahadevan et al., 2005). The measure estimates the traffic through nodes and can thus assess their potential "monitoring" congestion (Zaki and Meira, 2014, p. 103).

A second centrality measure that uses shortest paths is *closeness centrality*, which measures the average distance from a node to all other nodes (Pastor-Satorras and Vespignani, 2007). It is the inverse sum of the length of the distances from vertex i to all other vertices, corrected by $n-1$ and normalized to the interval [0, 1]. The higher closeness centrality values, the less separated (on average) a node is from others. Often, the importance of one node increases with the importance of its neighbors. *Eigenvector centrality* is a measure of node importance "proportional to the sum of the scores of its neighbors" (Pastor-Satorras and Vespignani, 2007). It is calculated as the leading eigenvectors of the graph's adjacency matrix, which indicates the existence of edges between nodes in the graph. The eigenvector centrality has the characteristic that the importance of a node is higher if it has more or more important neighbors.

6 Results

Table 3 gives an overview over the results of all graph measures, which will be discussed in detail in the following sections.

Table 3: Results of graph measures

	IP	Router	PoP	AS	ISP
Number of nodes, n	3,255,088	2,806,857	53,348	33,752	31,030
Number of edges, m	8,544,788	5,039,348	102,591	122,561	113,489
Density, ρ	1.61e-06	1.27e-06	7.21e-05	2.15e-04	2.35e-04
Avg. degree, $\langle k \rangle$	5.25011	3.590741	3.846105	7.262444	7.314792
Minimum degree	1	1	1	1	1
Maximum degree	14,023	13,874	4,329	5,376	6,593
Exponent, γ	3.22154	2.13352	2.41165	2.21073	2.23117
Standard Error for γ	0.02885	0.00524	0.05981	0.01758	0.01698
k_{min}^{PL}	219	23	39	7	6
Avg. local clustering coefficient, $\langle c \rangle$	0.04357	0.11366	0.23312	0.50603	0.520577
Global clustering coefficient, C	0.02594	0.03237	0.01158	0.01864	0.017814
Avg. shortest path length, $\langle d \rangle$	-	-	4.28843	3.20602	3.16390
Diameter, d_{max}	46	45	15	7	7
Avg. Eccentricity, $\langle \varepsilon \rangle$	-	-	9.57100	5.55036	5.547599
Radius	-	-	8	4	4
Avg. betweenness centrality	-	-	1.2328e-04	1.3072e-04	1.3948e-04
Avg. betweenness centrality for IXP ASNs	-	-	-	1.7210e-04	-
Avg. closeness centrality	-	-	0.23805	0.31572	0.32007
Avg. closeness centrality for IXP ASNs:	-	-	-	0.32213	-
Avg. eigenvector centrality	1.7539e-05	3.9305e-05	1.2791e-03	1.9353e-03	2.0943e-03
Avg. eigenvector centrality for IXP ASNs	-	-	-	2.6601e-03	-

6.1 Graph Size

The size, number of edges, density, and the average degree for the largest connected components (LCCs) are shown in Table 3. Both the size and the number of edges decrease with rising aggregation level of the data. The graphs' densities of near zero indicate that the Internet graphs at all levels are sparse.

There are theoretically some 4.3 billion unique values for IPv4 addresses. However, since not all prefixes are allocated and some are reserved, the maximum possible number of unique IPv4 addresses is smaller, with around 3.7 billion possible addresses (for 2012). The question of how many of the allocated IP addresses are actually in use is the objective of *Internet Cen-*

suses (Heidemann et al., 2008). It is important to refer to the same period when comparing the number of used IP addresses since the IPv4 address space is exhausting continually. The largest census, by the Carna botnet, determined for 2012 that 456 million IP addresses were "definitely in use" (Botnet, 2013b). If the Carna figures can be trusted, then our IP graph covers 0.71% of all usable IP addresses. While this appears to be a small share, one needs to consider that the IP graph refers to a transit topology, i.e., end hosts are by construction not included in the dataset.

To further investigate the composition of the IP graph, a graphical visualization of the IP address space utilization is created by mapping IP addresses into a 2-dimensional picture using a *Hilbert curve*. Figure 4 (left) depicts addresses of the combined IP nodelist. Each pixel represents a subnet with up to 16,384 different hosts, and the color refers to its utilization.

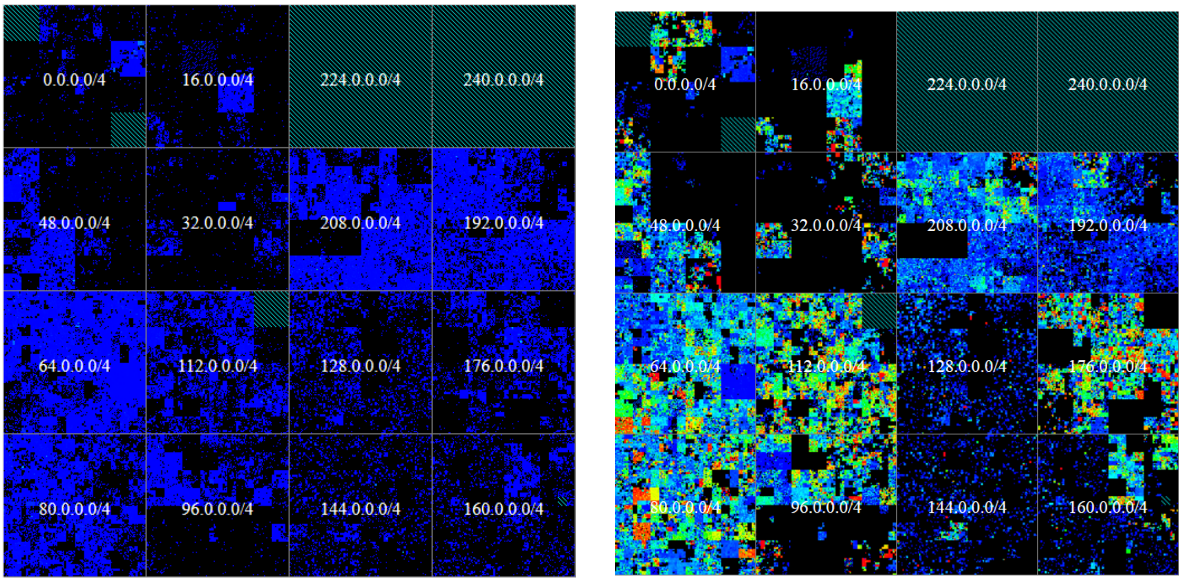

Figure 4: Address utilization of our IP dataset (left) and of the Internet Census 2012 (Botnet, 2013b) (right)

This is very similar to other images investigating the address space (University of Southern California, 2015; The Measurement Factory, 2009). It particularly reflects almost the same populated areas as in the comprehensive Internet Census 2012. Divergences emerge from different probing strategies; the very similar pattern, however, indicates that the IP graph is actually a representative sample of the entire Internet backbone.

The nodes of the IP graph are geolocated and plotted on a world map in Figure 5 (top). As one can directly see, some areas are more occupied than others. An important driver behind the evolution of Internet infrastructure is economic demand: First, the more people, the higher the necessity for Internet resources (Figure 5, bottom). The second aspect is the advancement of a region. That is why in economically developed areas the correlation is even stronger up to the point where Internet demand is proportional to population density (Pastor-Satorras and Vespignani, 2007; Yook et al., 2002).

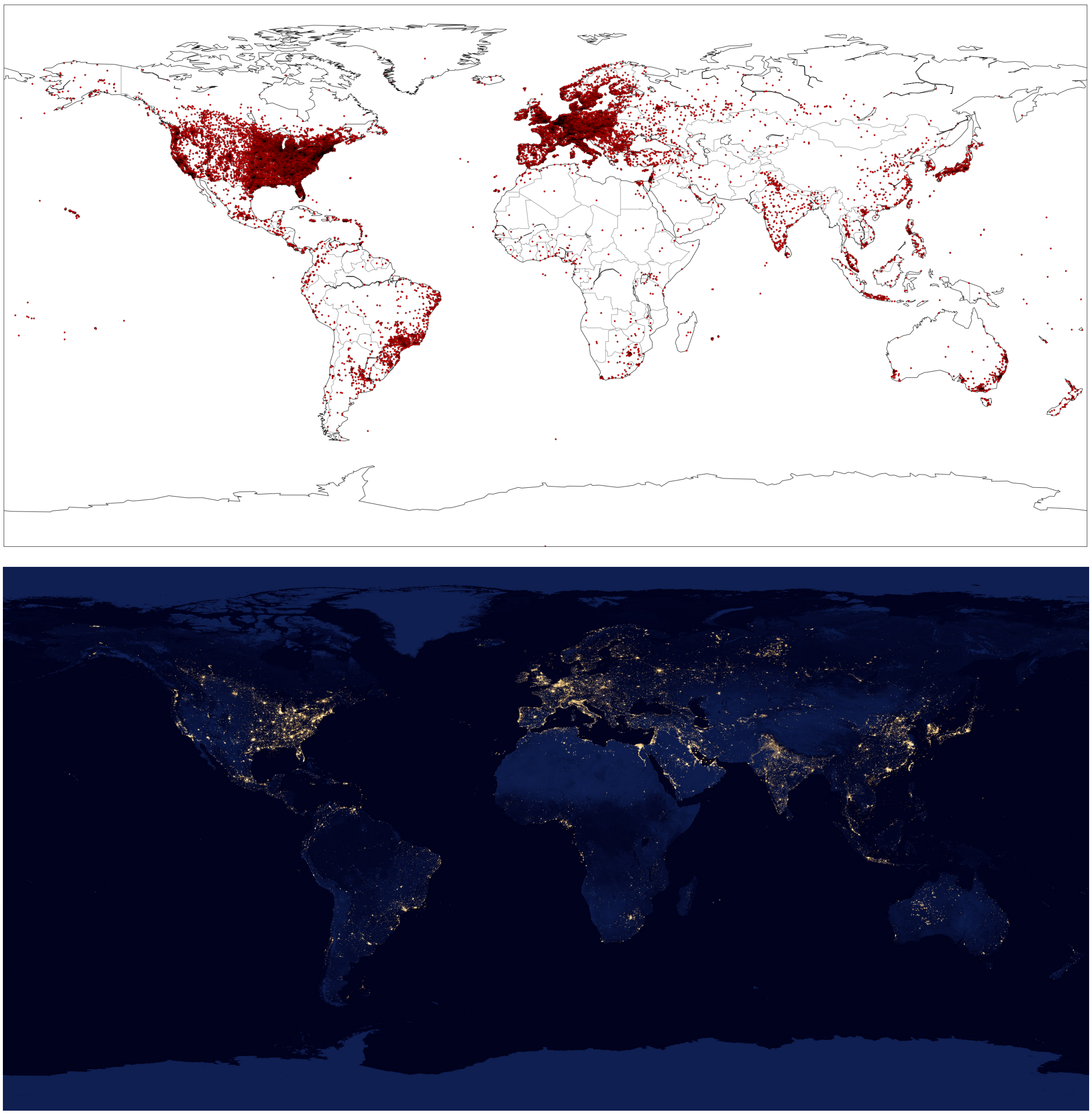

**Figure 5: Geographic location of all IP addresses (top)
NASA Earth Observatory 2012 (bottom)[1]**

The size of the router level graph is 13.8% smaller than the IP graph and has 41% fewer edges. The two contemporary router topologies (Huffaker et al., 2002) are smaller in terms of nodes and edges. In the PoP graph, the numbers of nodes and edges decrease dramatically in comparison to the router graph.

An advantage of the AS level is that there has been much research conducted. The BGP table dumps from June 29, 2012, determine 57,353 advertised ASes (CAIDA, 2016a). The number of nodes in the AS graph (33,752) are thus 41.1% short from detecting the whole set

[1] http://earthobservatory.nasa.gov/Features/NightLights/page3.php

of advertised ASNs. Nevertheless, for a traceroute-based approach this is successful: None of the AS graphs in Huffaker et al. (2012) detected more than 28,000 ASes only with traceroute probing. The number of edges differs even more, which is the reason for a higher average degree in the AS graph. It is assumed that there are more AS links than are known because backup links are hidden (Augustin et al., 2009).

One comprehensive dataset, which also refers to 2012, is the AS graph created by (Baumann and Fabian, 2014). They merged numerous types of sources (IRR, BGP, TR). In total, 33,399 nodes were present in both datasets. That accounts for 98.95% of the nodes in our combined AS graph and for 75.23% of the nodes in their dataset.

The ISP graph is closely related to the AS level. Size (31,030) and number of edges (113,489) are yet again smaller due to further aggregation of the data. Since the maximum possible number of ISPs is 50,788 (CAIDA, 2016a), the ISP graph captures lower 61.1% of all possible organizations.

6.2 Node Degree

The AS and ISP graphs display the highest connectivity, followed by the IP graph. Interestingly, the router and PoP levels have a lower degree than the remaining graphs. In general, the calculated values for all levels are in the range of findings of other works. The average degree of the AS level, however, is the second highest in the literature, only exceeded by (Baumann and Fabian, 2014). This is due to the relatively high number of discovered AS edges.

Based on their DNS entries, the IP nodes with the largest and the third largest degrees are presumably part of the Tor (The Onion Router) network. This network aims at providing privacy to the participants by routing their data packets over several dedicated machines until the traffic leaves Tor over an exit node towards the “normal” Internet. Therefore, an end node of this network works effectively as a gate towards many users, which is the reason for their extremely high degree. The other extremely high-degree IP nodes can presumably be explained by MPLS activated networks, which produce a similar pattern. The tenth “largest” IP node with a degree of 3,020 has as advertising ISP the DE-CIX at Frankfurt, the largest *Internet Exchange Point* (IXP) worldwide (DE-CIX Management GmbH, 2015), which gives a validation of the importance of this node.

The largest degree in the AS graph is 5,376; this is higher than any other measurement in the literature, which could be related to the newly detected edges. The ranking shows that the high degree AS and ISP nodes actually represent the top of the Internet hierarchy, showing their importance to the routing portion of the Internet. In order of decreasing ranking, the best-connected ASes belong to Level3, Cogent, Primus, and AT&T.

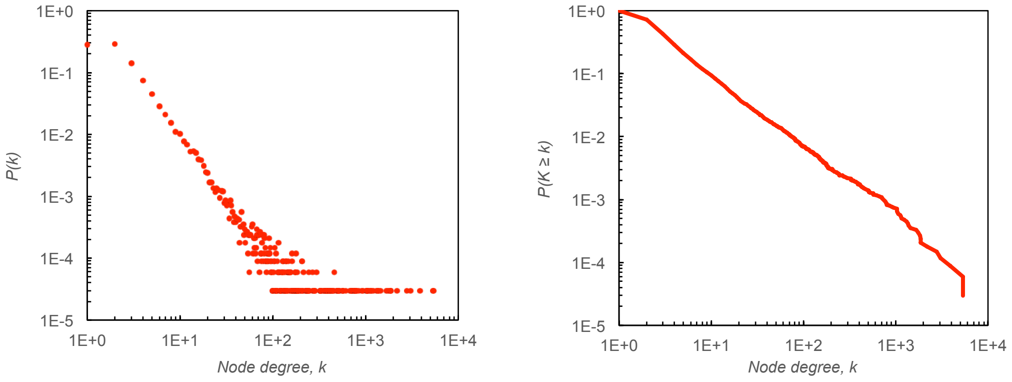

Figure 6: Degree Distribution (left) and CCDF (right) for the combined AS graph

One of the most commonly studied properties of graphs is the degree distribution. The heavy-tailed nature makes it impractical to directly visualize power-law distributions. That is why the degree distribution is plotted in a double-logarithmic fashion, such as the one in Figure 6 (left). Usually, the complimentary cumulative distribution function (CCDF) (Newman, 2013) is used, which returns the probability that any randomly chosen vertex has degree k *or greater*. Distributions that follow a power-law show as a straight line when the CCDF is plotted on a log-log scale, a pattern clearly visible in Figure 6 (right).

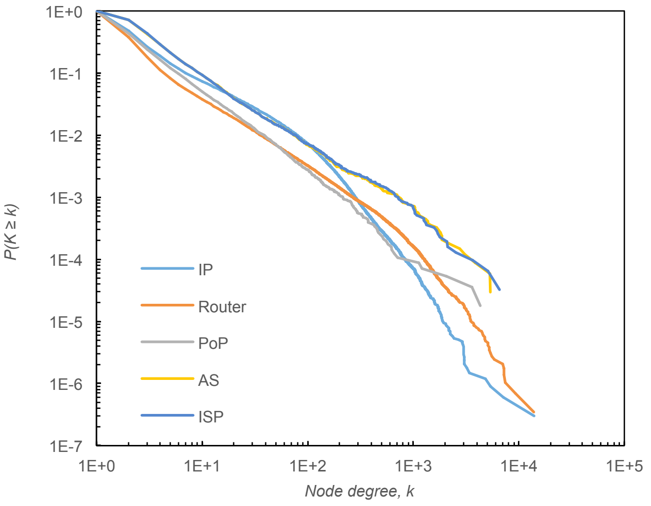

Figure 7: CCDF for the graphs at different levels

Figure 7 plots the degree CCDF for all topology levels in one diagram. There is a strong visual evidence for the power-law relationships for the degree distributions on every level. Most of the nodes have low degrees, while there are miniscule shares of nodes with extreme degrees at all levels. This property is quantified by calculating the power-law exponent γ. Statistically, all graphs follow the power-law relationship but with a diverging closeness, in accordance with the visual trend of the CCDFs (see Table 3).

An additional metric to evaluate the heavy-tailed behavior is the minimum degree k_{min}^{PL} above which the power-law regime holds. Almost all degree distributions start following the power-law relationship already at small degrees.

The power-law exponent γ of the IP graph (γ = 3.22) is somewhat different from what might be expected; but the low prevalence of hubs in the IP graph can also be seen in the plotted CCDF. The PoP graph resembles the IP graph with respect to the course of the degree distribution. The second largest exponent (γ = 2.41) indicates that there is a relatively large number of hubs. The exponent for the router level graph (γ = 2.13) is the smallest among all levels. Many researchers interpreted the scale-free property at the router level as evidence for highly connected hubs that make up the Internet's core. However, the interpretation of high degrees and power-laws for router topologies can be prone to some misperceptions about the implications of the results of graph analyses (Motamedi et al., 2015; Willinger et al., 2009).

The AS graph shows a larger value (γ = 2.21) but still at a comparatively low level. The power-law exponent of around 2.2 is a persistent result for AS level graphs, even over time. The exponent of the ISP level graph is just minimally higher (γ =2.23). Overall, it can be confirmed that the "heavy-tailed behavior observed in real mapping experiments is a genuine feature of the Internet" (Dall'Asta et al., 2006).

6.3 Clustering Coefficient

Clustering coefficients evaluate the connectedness of a graph. Consider Table 3, where both the average local clustering coefficient $\langle c \rangle$ and the global clustering coefficient C for the Internet graphs are shown. Obvious is that $\langle c \rangle$ increases with a higher topological aggregation. The lowest value of average clustering is at the IP level with only a 4.3% chance that two neighbors of a node are connected. In the AS and ISP graphs, over 50% of all possible connections within nodes' neighborhoods exist. As one can see in Table 3, the values of the global clustering coefficient differ significantly from the average of the local clustering coefficients. It has been shown that substantial differences between average local clustering and the global clustering coefficient are inherent to disassortative graphs (Mahadevan et al., 2005).

6.4 Assortativity

Accordingly, the differences between average local and global clustering coefficients can be explained by the assortative property of the graph. And in fact, the assortativity coefficient is negative at all topology levels (see Table 3). The interpretation of a negative assortativity coefficient is that low-degree nodes tend to link to high-degree nodes: The Internet is a disassortative network. This matches with the findings of other works (Newman 2003; Zhou, 2004). Assortativity is smallest for the AS graph and is then increasing towards lower topology levels, which corresponds to the maximum degrees at the respective tiers and the differences in the power-law exponent.

6.5 Distance Measures

Distance measures investigate graphs from a global "flow of information" perspective and are particularly interesting for traceroute-collected topologies since the measurement approach itself resembles path taking through the network. The main drawback of distance metrics is that their calculation is very resource demanding. That is why some distance metrics could not be calculated so far for either the IP or the router graph, even on powerful hardware and more than a month of calculation time.

The most interesting metric is the average shortest path length $\langle d \rangle$. The same pattern as for the other metrics is also visible here: The metric decreases with increasing aggregation. The deviations in the average path lengths can be explained with the values of the other structural metrics (Mahadevan et al., 2005): First, the relative number of edges in the network is larger, making it in general more densely connected. Second, the AS graph shows a smaller power-law exponent and has thus relatively more hubs present. Third, the assortativity coefficient is smaller for the AS graph as well, implying that here low-degree nodes rather tend to connect to high degree nodes (i.e., the hubs). These hubs act thus as "shortcuts" that enable traversing the graph in just few hops. This quality of scale-free networks is called "small-world phenomenon" or for the social sciences "six degrees of separation" (Milgram, 1967). The average path length is for the present technological networks considerably smaller than six. Furthermore, no pair of nodes in the AS and ISP graphs is further apart than 7 hops.

6.6 Centrality

The final measures concern the centrality of nodes in the networks, i.e., their relative importance based on three different metrics. The average centrality values increase with topological aggregation. It was possible to calculate the average eigenvector centrality for all five Internet graphs and this pattern holds – an increasing structural importance of neighbors.

One subset of the AS level graph consists of the 389 ASNs, which could be definitely determined to represent an IXP. For this subset, the centrality metrics were calculated separately and their average is also reported in Table 3. Evidently, average centrality measures are higher than the averages of all AS nodes. These figures indicate that IXPs display in fact a higher structural importance to the AS graph. Especially the larger average betweenness centrality supports the assumption that IXPs are crucial for the exchange of traffic.

Finally, the most important nodes of the ISP graph are ranked in a similar order for all centrality metrics, and their values decrease quickly. To further assess the importance of actual economic entities in future work, one could therefore refer to the centrality values of the novel ISP graph.

7 Limitations, Outlook, and Conclusion

As other research on the Internet topology, our work has some potential limitations. The first limitation concerns the quality of the input data. Artifacts in the data can result in a low-quality topology inference and analysis. Examples for these distortions are anonymous routers, multiple response hops, and loops in traces. A different handling of these anomalies will result in different topologies. This project, however, directly processes the *raw* traceroute data in an identical way and is thus less prone to limitations emerging from distortions in the different datasets. Another limitation is whether the used aggregation procedures are accurate. In particular, the router and PoP graph inferences are subject to this concern because on those levels some compromises had to be made.

A conceptual limitation is the question whether it is valid to merge datasets from different projects. While it can be assumed that no individual dataset is decisive and thus misrepresents the actual picture of the network, we argue that for a most possible comprehensive investigation all collective evidence *has* to be exploited. Individual issues are likely to be removed by combining different and diverse datasets into one. The rationale behind that is that individual inconsistencies “cancel out” or put differently, that they are aggregated to such an extent that only the actual structural properties remain. In that case, the described drawbacks can be overcome and the topological properties observed in the combined data sources are representative of the ground truth.

Another limitation is that the traceroute tool is not suitable to detect the actual physical infrastructures that make up the Internet in detail. This relates to the issue that lower (layer-1/2) structures, such as switches, MPLS tunnels, or ATM circuits are unobservable to layer-3 (IP) measurements. The actual topology might hence look different from what is inferred by traceroute. Especially Willinger at al. (2009) argue that TTL-probing necessarily results in wrongfully inferred maps and that, in particular, the detected power-law degree distributions are the result of inflated node degrees due to opaque layer-2 clouds. They argue from a network engineering perspective that high degree routers in the network core are nonsensical because they would be beyond technological possibilities or uneconomical – a point of view they exaggerated in the tendentious term of a "scale-free Internet myth" (Willinger et al., 2009). If their view would be adopted, traceroute-based probing would be per se doomed to infer wrong topologies.

However, this criticism can be tackled by several arguments. While their technological constraint argument for high degrees of routers cannot be denied, Willinger et al. (2009) extend their argument to *all* levels of the Internet topology. This conclusion can be rejected by both theoretical presumptions and empirical data. For the higher levels (e.g., ASes), there are no technological constraints on node degree whatsoever. The power-law relationships actually become more prevalent when aggregating the data beyond the router level. An MPLS tunnel over several ASes is very unlikely (despite not impossible for ASes of one ISP) and transitions between networks are thus correctly detected. Another argument, based on the results of this study, is that even the Carna dataset displays the idiosyncratic scale-free behavior – despite the monitors presumably being *within* the networks and thus being less prone to MPLS use. Lastly, their theoretical presumption was not supported with any empirical proof. MPLS tunneling was determined to be not employed to an extent that would alter the topological measurements (Feldman et al., 2012), and based on the pervasiveness of the scale-free behavior for studies at all levels it is highly unlikely that this is just a result of the measurement approach.

In general, an exhaustive validation of inferred maps remains in the absence of ground truth data an open question. That is why researchers need to be aware of these pitfalls when applying traceroute-like measurements. Despite its limitations, traceroute is still the tool of choice for topology discovery and researchers have chosen trust its outcomes.

The geographic properties of graphs could facilitate validation with an inference of actual physical infrastructure. There are some projects that attempt to collect information about the physical infrastructure available from ISPs, e.g., the Internet Topology Zoo (Knight et al., 2011) or the Internet Atlas (Durairajan et al., 2013). Interesting questions would for example be: Are high degree nodes close to network intersections or to the bottlenecks of submarine cable landing stations? Are PoP edges spatially close to physical cables laid along roads (Durairajan et al., 2014)? Are there any long-haul edges that might thus represent MPLS tunnels?

The information of the several levels could be combined into one integrated graph of the Internet at *all* levels, that is to annotate IP nodes with their router, PoP, AS, and ISP affiliation. It would then be possible to investigate whether IP cliques refer to routers, router cliques refer to PoPs etc., or in other words: Are PoPs *internally* denser meshed than between each other?

By necessity, the integrated graphs provide a snapshot of the Internet topology at a given time. In future work, repeated measurements and automated data integration could lead to a better understanding of Internet dynamics. To complete the possibilities for future work, using our data set and the computed centrality values can improve network control simulations and vulnerability assessments such as percolation analyses.

References

Alstott, J, Bullmore, E, Plenz, D (2014). Powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions. PLoS ONE 9(1): e85777.

Augustin, B, Krishnamurthy, B, Willinger, W (2009). IXPs: Mapped? In: Feldmann A (ed) Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference (IMC '09). ACM, New York, NY, USA, pp 336–349.

Barabási, A, Albert, R (1999). Emergence of Scaling in Random Networks. Science 286(5439): 509–512.

Albert, R, Jeong, H, Barabási AL (2000). Error and Attack Tolerance of Complex Networks. Nature 406(6794): 378-382.

Barabási, A (2016). Network Science. Cambridge University Press. http://barabasi.com/networksciencebook/ (accessed 31 March 2017)

Baumann, A, Fabian, B (2014). How Robust is the Internet? – Insights from Graph Analysis. In: Proceedings of the 9th International Conference on Risks and Security of Internet and Systems (CRiSIS 2014), Trento, Italy, Springer LNCS 8924.

Carna Botnet (2013a). Full Disclosure: Port scanning /0 using insecure embedded devices. http://seclists.org/fulldisclosure/2013/Mar/166 (accessed 31 March 2017)

Carna Botnet (2013b). Internet Census 2012. Port Scanning /0 Using Insecure Embedded Devices. http://internetcensus2012.bitbucket.org/paper.html (accessed 31 March 2017)

Chang H, Govindan R, Jamin S, Shenker, SJ, Willinger, W (2004). Towards Capturing Representative AS-level Internet Topologies. Computer Networks 44(6): 737–755.

Clauset, A, Shalizi, CR, Newman, MEJ (2009) Power-Law Distributions in Empirical Data. SIAM Rev. 51(4): 661–703.

Cohen, R, Erez, K, Ben-Avraham, D, Havlin, S (2000). Resilience of the Internet to Random Breakdowns. Physical Review Letters 85(21): 4626.

Cohen, R, Erez, K, Ben-Avraham, D, Havlin, S (2001). Breakdown of the Internet under Intentional Attack. Physical Review Letters 86(16): 3682.

Cooperative Association for Internet Data Analysis (CAIDA) (2013). About CAIDA. http://www.caida.org/home/about/index.xml (accessed 31 March 2017)

Cooperative Association for Internet Data Analysis (CAIDA) (2016a). Inferred AS to Organization Mapping Dataset. http://www.caida.org/data/as-organizations/ (accessed 31 March 2017)

Cooperative Association for Internet Data Analysis (CAIDA) (2016b). Routeviews Prefix to AS mappings Dataset for IPv4 and IPv6. http://www.caida.org/data/routing/routeviews-prefix2as.xml (accessed 31 March 2017)

Dall'Asta, L, Alvarez-Hamelin, I, Barrat, A, Vazquez, A, Vespignani, A (2006). Exploring Networks with Traceroute-like Probes: Theory and Simulations. Theoretical Computer Science 355(1): 6–24.

DE-CIX Management GmbH (2015). About. https://www.de-cix.net/about/ (accessed 31 March 2017)

Department of Homeland Security (2012). The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. http://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803.pdf (accessed 31 March 2017)

Dittrich, D, Carpenter, K, Karir, M (2014). An Ethical Examination of the Internet Census 2012 Dataset: A Menlo Report Case Study. In: Proceedings of the 2014 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS 2014).

Doerr, C, Kuipers, FA (2014). All Quiet on the Internet Front? IEEE Communications Magazine 52(10): 46-51.

Donnet, B, Friedman, T (2007). Internet Topology Discovery: A Survey. IEEE Communication Surveys and Tutorials 9(4): 56–69.

Doyle, JC, Alderson, DL, Li, L, Low, S, Roughan, M, Shalunov, S, Tanaka, R, Willinger, W (2005). The "Robust Yet Fragile" Nature of the Internet. Proc. Natl. Acad. Sci. U.S.A. 102(41): 14497–14502.

Durairajan, R, Ghosh, S, Tang, X, Barfold, P, Eriksson, B (2013). Internet Atlas: a Geographic Database of the Internet. In: Hui P (ed) Proceedings of the 5th ACM Workshop on HotPlanet (HotPlanet '13). ACM, New York, NY, USA, pp 15–20.

Durairajan, R, Sommers, J, Barford, P (2014). Layer 1-Informed Internet Topology Measurement. In: Williamson, C (ed) Proceedings of the 2014 ACM Conference on Internet Measurement (IMC '14). ACM, New York, NY, USA, pp. 381–394.

Fabian, B, Baumann, A, Lackner, J (2015). Topological Analysis of Cloud Service Connectivity, Computers & Industrial Engineering, 88 (October):151-165.

Feldman, D, Shavitt, Y, Zilberman, N (2012) A Structural Approach for PoP Geo-Location. Computer Networks 56(3): 1029–1040.

Faloutsos, M, Faloutsos, P, Faloutsos, C (1999). On Power-Law Relationships of the Internet Topology. ACM SIGCOMM Computer Communication Review 29: 251–262.

Govindan, R, Tangmunarunkit, H (2000). Heuristics for Internet Map Discovery. In: Sidi M (ed) Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2000): 1371–1380.

Gunes, MH, Sarac, K (2009). Resolving IP Aliases in Building Traceroute-Based Internet Maps. IEEE/ACM Trans. Networking 17(6): 1738–1751.

He ,Y, Siganos, G, Faloutsos, M, Krishnamurthy, S (2009). Lord of the Links. A Framework for Discovering Missing Links in the Internet Topology. IEEE/ACM Transactions on Networking (TON) 17(2): 391–404.

Heidemann, J, Pradkin, Y, Govindan, R, Papadopoulos, C, Bartlett, G, Bannister, J (2008). Census and Survey of the Visible Internet. In: Papagiannaki K (ed) Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement (IMC '08). ACM, New York, NY, USA, p 169.

Ho, S-C, Kauffman, RJ, Liang, T-P (2007). A Growth Theory Perspective on B2C e-Commerce Growth in Europe: An Exploratory Study. Electronic Commerce Research and Applications 6: 237–259.

Huffaker, B, Fomenkov, M, Claffy, K (2012). Internet Topology Data Comparison. CAIDA Tech. Rep.

Huffaker, B, Plummer, D, Moore, D, Claffy, K (2002). Topology Discovery by Active Probing. In: Proceeding of the 2002 Symposium on Applications and the Internet Workshops (SAINT 2002 Workshops). IEEE, Los Alamitos, CA, USA, pp 90–96.

Huffaker, B, Fomenkov, M, Claffy, K (2012). Internet Topology Data Comparison. http://www.caida.org/research/topology/topo_comparison/ (accessed 31 March 2017)

Internet Society (2012). World IPv6 Launch. http://www.worldipv6launch.org/ (accessed 31 March 2017)

Karlsruhe Institute of Technology (2015). NetworKit. https://networkit.iti.kit.edu/ (accessed 31 March 2017)

Keys, K (2010). Internet-Scale IP Alias Resolution Techniques. ACM SIGCOMM Computer Communication Review 40(1): 50–55.

Knight, S, Nguyen, HX, Falkner, N, Bowden, R, Roughan, M (2011). The Internet Topology Zoo. IEEE J. Select. Areas Commun. 29(9): 1765–1775.

Krenc, T, Hohlfeld, O, Feldmann, A (2014). An Internet Census Taken by an Illegal Botnet – A Qualitative Assessment of Published Measurements. ACM SIGCOMM Computer Communication Review 44(3): 103–111.

Le Malécot, E, Inoue, D (2014). The Carna Botnet Through the Lens of a Network Telescope. In: Danger, J, Debbabi, M, Marion, JY, Garcia-Alfaro, J, Zincir Heywood, N (eds) Foundations and Practice of Security. Lecture Notes in Computer Science 8352, Springer, Cham: 426-441.

Madhyastha, HV, Isdal, T, Piatek, M. Dixon, C, Anderson, T, Krishnamurthy, A, Venkataramani, A (2006). iPlane: An Information Plane for Distributed Services. In: Bershad B (ed) Proceedings of the 7th Symposium on Operating Systems Design and Implementation (OSDI '06). USENIX Association, Berkeley, CA, USA, pp 367–380.

Mahadevan, P, Krioukov, D, Fomenkov, M, Huffaker, B, Dimitropoulos, X, claffy, kc, Vahdat, A (2005). Lessons from Three Views of the Internet Topology: Technical Report. http://www.caida.org/publications/papers/2005/tr-2005-02/tr-2005-02.pdf (accessed 31 March 2017)

Mahadevan, P, Krioukov, D, Fomenkov, M, Huffaker, B, Dimitropoulos, X, claffy, k, Vahdat, A (2006). The Internet AS-level Topology: Three Data Sources and One Definitive Metric. ACM SIGCOMM Computer Communication Review 36(1): 17-26.

MaxMind (2015). GeoIP2 Databases. https://www.maxmind.com/en/geoip2-databases (accessed 31 March 2017)

The Measurement Factory (2009). IPv4 Heatmaps: Gallery. http://maps.measurement-factory.com/gallery/index.html (accessed 31 March 2017)

Milgram, S. (1967) The Small World Problem. Psychology Today 2(1): 60–67.

Motamedi, R, Rejaie, R, Willinger, W (2015). A Survey of Techniques for Internet Topology Discovery. IEEE Communication Surveys & Tutorials 17(2): 1044-1065.

Newman, MEJ (2003). Mixing Patterns in Networks. Phys. Rev. E 67(2): 026126.

Newman, MEJ (2013). Networks: An Introduction. Oxford Univ. Press, Oxford.

Pastor-Satorras, R, Vespignani, A (2007). Evolution and Structure of the Internet: A Statistical Physics Approach. Cambridge University Press.

Shavitt ,Y, Shir, E (2005). DIMES: Let the Internet Measure Itself. ACM SIGCOMM Computer Communication Review 35(5): 71–74.

Shavitt, Y, Zilberman, N (2010). A Structural Approach for PoP Geo-Location. In: Allman, M (ed) Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement (IMC '10). ACM, New York, NY, USA.

Spring, N, Mahajan, R, Wetherall, D (2004) Measuring ISP Topologies with Rocketfuel. IEEE/ACM Trans. Networking 12(1): 2–16.

Staudt, C, Sazonovs, A, Meyerhenke, H (2014). NetworKit: An Interactive Tool Suite for High-Performance Network Analysis. arXiv:1403.3005. http://arxiv.org/abs/1403.3005 (accessed 22 April 2017)

Sterbenz, JP, Hutchison, D, Çetinkaya, EK, Jabbar, A, Rohrer, JP, Schöller, M, & Smith, P (2010). Resilience and Survivability in Communication Networks. Computer Networks 54(8): 1245-1265.

Sterbenz, JP, Cetinkaya, EK, Hameed, MA, Jabbar, A, Qian, S, & Rohrer, JP (2013). Evaluation of Network Resilience, Survivability, and Disruption Tolerance. Telecommunication Systems 52(2): 705-736.

University of Southern California (USC) (2015). ANT Censuses of the Internet Address Space. https://ant.isi.edu/address/ (accessed 22 April 2017)

Watts, DJ, Strogatz, SH (1998). Collective Dynamics of 'Small-World' Networks. Nature 393(6684): 440–442.

Willinger, W, Alderson, D, Doyle, JC (2009). Mathematics and the Internet: A Source of Enormous Confusion and Great Potential. Notices of the AMS 56(5): 586–599.

Willinger, W, Roughan, M (2013). Internet Topology Research Redux. In: Haddadi, H, Bonaventure, O (Eds.) Recent Advances in Networking. ACM SIGCOMM, pp. 1–59. http://sigcomm.org/education/ebook/SIGCOMMeBook2013v1.pdf (accessed 22 April 2017)

Xiao, S, Xiao, G, & Cheng, TH (2008). Tolerance of Intentional Attacks in Complex Communication Networks. IEEE Communications Magazine 46(1): 146-152.

Yook, S, Jeong, H, Barabási, A (2002). Modeling the Internet's Large-scale Topology. Proc. Natl. Acad. Sci. U.S.A. 99(21): 13382–13386.

Zaki, MJ, Meira, W (2014). Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press, New York.

Zhou, S (2004) Accurately Modeling the Internet Topology. Phys. Rev. E 70(6): 066108.