UNIVERSITÀ DI PISA

Scuola di Dottorato in Ingegneria “Leonardo da Vinci”

Corso di Dottorato di Ricerca in
INGEGNERIA DELL’INFORMAZIONE

Tesi di Dottorato di Ricerca

On the Analysis of the Internet from a Geographic and Economic Perspective via BGP Raw Data

Autore:

Alessandro Improta ________________________

Relatori:

Prof. Luciano Lenzini ________________________

Prof. Gigliola Vaglini ________________________

Ing. Enrico Gregori ________________________

Anno 2013
SSD ING-INF/05

Sommario

Internet è ormai parte integrante della vita di ciascuno di noi, e ricoprirà un ruolo sempre più importante in futuro. Già oggi stiamo assistendo ad una crescita esponenziale degli utenti che fruiscono della Rete grazie all'evoluzione e alla crescente pervasività di sistemi di trasmissione dati e di dispositivi mobili (e.g. tablet e smartphones) che sono costantemente connessi ad Internet, di cui fanno parte. L' estrema pervasività ed affidabilità di Internet hanno portato un numero sempre maggiore di persone ad appoggiarsi alla rete stessa per la gestione di molti aspetti della propria vita quotidiana, siano essi di carattere personale o lavorativo, come ad esempio la possibilità di accedere al proprio conto in banca, o di discutere di lavoro in videoconferenza con un collega dall'altra parte del mondo. Tuttavia, poche persone hanno consapevolezza di ciò che accade ai propri dati una volta inviati dai propri dispositivi verso Internet e soltanto una ristretta cerchia di ricercatori ha una approssimativa visione d'insieme della reale infrastruttura di Internet. Tali ricercatori hanno tentato negli ultimi anni di scoprire più in dettaglio le caratteristiche di Internet, in modo da poter creare un modello su cui poter identificare e colmare le debolezze della Rete. Nonostante i continui sforzi in questa direzione, al momento non è noto in letteratura alcun modello capace di rappresentare efficacemente la reale infrastruttura di Internet, soprattutto a causa della mancanza di dati e del non appropriato livello di dettaglio applicato dagli studi prodotti fino ad oggi. Questa tesi affronta entrambe le problematiche considerando Internet come un grafo i cui nodi sono rappresentati da Sistemi Autonomi (AS) e le cui connessioni sono rappresentate da connessioni logiche fra AS. In prima istanza questa tesi ha l'obiettivo di fornire nuovi algoritmi ed euristiche allo scopo di studiare Internet ad un livello di granularità adeguato alla realtà, introducendo nell'analisi elementi di carattere economico e geografico che limitano il numero di possibili percorsi fra i vari AS che i dati possono intraprendere. Sulla base di tali euristiche viene inoltre fornita una metodologia innovativa idonea a quantificare la completezza dei dati a disposizione al fine di poter identificare gli AS da coinvolgere nella raccolta di dati di routing per ottenere una visione completa e reale del cuore di Internet. Nonostante i risultati di tale metodologia evidenziano che attualmente i sistemi di raccolta dati non riescono a ottenere informazioni riguardanti la maggioranza degli AS facenti parte del cuore di Internet, è tuttavia possibile maggior numero degli AS individuati dalla metodologia.

Abstract

The Internet is nowadays an integral part of the everyone's life, and will become even more important for future generations. Proof of that is the exponential growth of the number of people who are introduced to the network through mobile phones and smartphones and are connected 24/7. Most of them rely on the Internet even for common services, such as online personal bank accounts, or even having a video-conference with a colleague living across the ocean. However, there are only a few people who are aware of what happens to their data once sent from their own devices towards the Internet, and an even smaller number – represented by an elite of researchers – have an overview of the infrastructure of the real Internet. Researchers have attempted during the last years to discover details about the characteristics of the Internet in order to create a model on which it would be possible to identify and address possible weaknesses of the real network. Despite several efforts in this direction, currently no model is known to represent the Internet effectively, especially due to the lack of data and the excessive coarse granularity applied by the studies done to date. This thesis addresses both issues considering Internet as a graph whose nodes are represented by Autonomous Systems (AS) and connections are represented by logical connections between ASes. In the first instance, this thesis has the objective to provide new algorithms and heuristics for studying the Internet at a level of granularity considerably more relevant to reality, by introducing economic and geographical elements that actually limit the number of possible paths between the various ASes that data can undertake. Based on these heuristics, this thesis also provides an innovative methodology suitable to quantify the completeness of the available data to identify which ASes should be involved in the BGP data collection process as feeders in order to get a complete and real view of the core of the Internet. Although the results of this methodology highlights that current BGP route collectors are not able to obtain data regarding the vast majority of the ASes part of the core of the Internet, the situation can still be improved by creating new services and incentives to attract the ASes identified by the previous methodology and introduce them as feeders of a BGP route collector.

Contents

List of Figures

List of Tables

This thesis is dedicated to my family
and to my best friends,
for their endless support

1

Introduction

The Internet is a complex system that evolved over the last few decades from a small network confined to the U.S. – i.e. ARPANET, 1969 [75] – to the current worldwide network of networks that, during the last few years, became more and more part of the everyday life of billions of people. Thanks to the Internet, it is now possible to do things that just a couple of decades ago would have been impossible, such as receiving mail from the other side of the world just a couple of seconds after the mail was sent, checking one's bank account from home and retrieving any kind of information at any time with just a couple of clicks. Despite its increasing pervasiveness, very little is known about the real structure of the Internet and what happens to data once they have left the home router. This is a major issue, since it means that is impossible to detect any structural problem or Achilles' heels until an outage occurs. To address this problem, several researchers started to analyse the Internet in a topological perspective in the hope of revealing any potential weaknesses at any level, such as the physical level [23] and the overlay level [99, 108] and the network level. The latter is the most relevant level in order to perform analyses about the inter-domain relationships existent on the Internet, and can be subdivided furthermore in the following four level of abstractions:

- *IP interface-level*: each node is represented by an IP interface, while edges are IP connections between pair of interfaces. Data is typically gathered via Traceroute probes [2, 3, 11, 12].
- *Router-level*: each node is represented by a router, while edges are IP connections between pair of routers. Data is typically gathered by applying an heuristic to aggregate IP interfaces on the IP interface-level graph [62, 70], by setting peculiar IP options in Traceroute probes [105], by exploiting ad-hoc probes [57] and analysing the IP ID values of the probes sent [20, 71].
- *PoP-level*: each node is represented by a collection of routers located in the same points of presence (PoPs), while edges are connections between pair of PoPs. Data is typically gathered by applying reverse DNS lookups [107] or by looking for peculiar characteristics in the IP interface-level graph [47].

- *AS-level*: each node is represented by an AS, i.e. an organization (or part of it) that manages a certain amount of IP subnets, while edges are connections between pairs of ASes, established via Border Gateway Protocol (BGP-4) [97]. Data is typically gathered via Route Collectors (see Chapter 2) or by applying IP-to-AS techniques to infer an AS-level topology from the IP interface-level graph [29, 79, 80].

Each of these levels of abstraction have been widely studied and analysed in the recent past, and led researchers to identify Internet graph properties (e.g. [14, 26, 46, 58] and [106]) aiming to create a proper model of the Internet. Unfortunately, the objective is still far from being achieved, mainly due to the extremely poor knowledge of the Internet real characteristics. This poor knowledge is not fully imputable to researchers, but relies mostly on the poor coverage of current Internet measurement infrastructure [60, 87] and on the reluctance of ISPs to publicly announce their routing information. As a consequence, current efforts in modeling the Internet either rely on the random graph model of Erdös and Rényi [22], on the small world graph model of Watts and Strogatz [112] or on scale-free network models based on the preferential attachment [116]. However, each of them fails to correctly model the Internet inter-domain connectivity [113]. In particular, there is an interesting and enlightening story about the latter approach that may let understand the current status of modeling efforts about the Internet. In 1999 it was firstly discovered that the node degree distribution of the router-level of the Internet – basing only on the available measurements – is scale-free and follows a power-law distribution with α-parameter between 1 and 2 [46]. This result was mostly caused by the type of dataset used [89] and by its biases [14] and is far from the reality [113]. Nevertheless, mathematicians used these results to create a network growth model able to fit them [16, 19] and claimed that the Internet display a high degree of tolerance against random failures, but are extremely vulnerable to targeted attacks due to the presence of important hub nodes in the network [17]. Reality is that Internet is extremely robust by design thanks to the IP protocol [34, 93], that reacts immediately to failures by re-routing traffic around them [113]. A neater summary about the current status of research is provided in [27] and [113].

Nevertheless, the real lesson that this anecdote teach us is that it is of fundamental importance the deep knowledge of data used. This thesis focuses in particular on data collected from several public route collectors deployed across the world, that are collecting routing information from volunteer ASes via the Border Gateway Protocol (BGP). In the following, it will be described the major topics and the most relevant contributions that the author provided in order to shed some light about BGP routing data and to draw inferences about the Internet ecosystem.

1.1 Towards enhancing AS-level Internet measurements

One of the largest problems encountered in developing a proper model of the Internet is that its structure is not driven merely by scientific metrics (e.g. minimization of the number of hops/organizations to reach a destination [53]), but mostly by economic [40, 42, 52, 59, 67, 87, 109] and geographic [61, 85] factors. This means that the largest number of paths extracted from the undirected graph of the Internet – at any level of abstraction – do not exist and, thus, a mere undirected graph cannot be used to represent the Internet topology, but it must be introduced an *enhanced* graph, where each connection is tagged with a proper label that may reflect the routing choices among each node. Another problem concerning the scarce knowledge about the Internet structure is due to the low amount of routing information that is currently collected from Route Collectors [61, 87], that are often used by large ISPs to advertise their connectivity [61].

This thesis aims to contribute in filling the gaps present in the knowledge of the Internet by focusing on the AS-level of abstraction of the Internet and, more specifically, by analysing and inferring useful information from raw BGP data gathered via BGP Route Collectors. More details and related works about the remaining levels of astraction can be found in [43]. The main contributions provided by this thesis can be summarized as follows:

1.1.1 Inter-AS economic relationships

The Internet AS-level topology as it is cannot provide much useful information about the real interaction among ASes, since the IP address space exchanged between each pair of ASes strictly depends on the type of economic agreement established among them, which is technically implemented by applying outbound route filters described via BGP export policies. Despite the large number of possible economic agreements, inter-AS relationships can still be categorized into three main classes on the basis of the set of routes that each AS announces to the other: provider-to-customer (p2c) – or customer-to-provider (c2p) – peer-to-peer (p2p) and sibling-to-sibling (s2s) [52], that will be analysed in detail in Chapter 3. This granularity is typically considered to be consistent with the reality [102], despite there exist some exceptions [84] mainly caused by policies established on a geographic basis [60, 85]. The existence of BGP export policies implies that the largest amount of graph paths that can be extracted from the undirected Internet AS-level graph do not exist in reality. This means that the bare undirected topology cannot be used as it is to properly study or model the real Internet routing behavior.

In this context, the knowledge of inter-AS economic relationships plays a fundamental role, and any realistic Internet AS-level analysis has to take them into account. The common approach is to enhance the undirected AS-level topology into an economic AS-level topology where each edge is tagged with a proper economic label that reflects the type of relationship existing among the involved ASes and the related BGP export policy applied. However, despite (or due to) their key role in business environment, details about inter-AS economic relationships are usually not publicly available, and researchers needed to develop heuristics [40, 42, 52, 59, 67, 87, 109] to infer them, typically exploiting data gathered via BGP Route Collectors. A well-known drawback of this type of data is that it contains several spurious entries caused by router misconfigurations [78] and showing up during path exploration phenomena [88] that can potentially affect the correctness of the inferences drawn by researchers. Despite some of the heuristics tried to minimize the impact of these routes by limiting the impact of short-lived routes [87], no definitive method has been developed to get rid of them.

In this thesis are proposed two type of algorithms to fill up this gap. The first is a time-aware tagging algorithm, which exploits the AS paths gathered from BGP monitors and takes into account their lifespan during the inference of the tags, thus preserving backup links. The second is a spuriousness-free tagging algorithm which exploits a preliminary data hygiene filter where spurious routes are identified and purged from BGP data available and, still exploiting the valley-free principle [52], tags each connection with an economic label.

1.1.2 Geography

The pervasive evolution of the Internet did not occur homogeneously all around the world for obvious historical, economic and political reasons. The result is that the Internet today is the composition of loosely connected groups of networks identifiable by some geographic boundaries [111, 120], each with its particular pricing models, business contexts and regulatory environments [85]. Typically each organization has a particular role and specific economic behavior in each region of the world where it is located, which strictly depend on the connectivity and performance that it can provide for its customers in that region. For example, an intercontinental AS may be widespread in its home region – in terms of the number of connections and services offered – while it may be not competitive outside that region. This different level of pervasiveness may lead the same AS to establishing economic relationships in those regions with different criteria.

Most research in the Internet topology analysis have considered ASes as homogeneous entities, each with a global set of metrics and characteristics, regardless of their heterogeneity. Those works (e.g. [48, 73, 82]) that tried to get a better insight in the geographic distribution of ASes relied on `traceroute` and packet delays to infer geographic information but none of them explicitly focused on inferring regional AS topologies. Moreover from `traceroute` data is not trivial to obtain an AS-level topology, due to dealiasing and router-to-AS mapping issues [118], and it is almost impossible to have a precise view of the dynamic evolution of each AS path, that is fundamental to correctly infer the economic nature of each connection [59]. These problems can be easily solved exploiting the BGP data collected by Route Collectors (RCs).

In this thesis will be provided a methodology to infer geography from raw BGP data, and it will be shown that the Internet actually consists of regional and independent ecosystems, which differ greatly in terms of both topological and economic properties and introduce particular characteristics that are hidden in a global-level analysis. These differences should be taken into serious consideration by every research based on the AS-level topology of the Internet in order to rely on a more realistic structure of the Internet, instead of on a coarse and potentially-misleading representation.

1.1.3 Data incompleteness

Although several works have highlighted problems in using traceroute data to infer AS-level information – such as aliasing [70], biasing [14] and router-to-AS mapping [66, 119] – only a few have investigated the limitations of BGP raw data and, in doing so, have mainly focused on the incompleteness of data gathered via BGP route collectors. To fill the gap, this thesis provides a deep analysis of BGP data currently gathered by the most important route collector projects deployed around the world, highlighting that their view is extremely *narrow* – due to the extremely low number of ASes that are actively feeding the route collectors – and *biased* – due to the nature of the feeding ASes, mostly managed by worldwide ISPs. This top-down view does not allow the route collector infrastructure to discover a large set of p2p connections that may be established among ASes that are part of the lower part of the Internet hierarchy, as already highlighted in [33, 35, 64, 87]. Moreover, this thesis introduces an innovative metric – named *p2c-distance* – which allow to provide an analysis able to take into account the presence of BGP decision processes and BGP export policies crossed by BGP UPDATE messages before reaching a route collector and, as a consequence, provides a better understanding of the level of completeness of the data gathered. Thanks to this new metric it is also possible to formulate a tailored set cover problem in order to understand which AS should actively participate to a Route Collector problem in order to retrieve a complete view of the Internet core, i.e. the *transit* ASes. Even though the MSC is NP-complete [54], this methodology exploits the low density of the p2c-distance matrix to reduce the size of the problem via mathematical techniques [37, 55, 83, 94, 95], finally retrieving an optimal solution. The application of this problem to the global and regional AS-level topologies highlighted that parts of the world, such as Africa, completely hide their p2p connectivity from the current route collector infrastructure.

This thesis is organized as follows. Chapter 2 describes BGP raw data and the basic techniques used to extract them from Route Collectors repository and extract from them AS-level connectivity information. Chapter 3 describes the role of economy in the Internet environment and how to retrieve economic information from raw BGP data. Chapter 4 describes the impact that geography has on the Internet topology, providing a methodology to infer geographic information from raw BGP data. Chapter 5 analyses the completeness of BGP data available and provides a methodology to identify the ASes that should announce their full routing tables to the Route Collectors to have a complete view of the Internet core. Finally, Chapter 6 concludes the thesis.

2

BGP data

One of the most fundamental aspects in any kind of research is to know in detail the data on which the analyses are performed, in order to not draw any wrong conclusion lead by poor knowledge of data itself (see for example [17]). Without any doubt, BGP messages are the best source of data to infer the Internet AS-level topology, since AS-level information is directly contained in the `AS_PATH` BGP attribute and no further heuristics have to be applied. BGP route collectors receive BGP routing information from cooperating ASes, hereafter *BGP feeders*, to which they establish a BGP session. Thanks to the collected routing data it is possible to re-create the dynamics of the inter-domain routing as perceived from customers of the BGP feeders. In this chapter are described the BGP route collector projects currently active and publicly available, and the processes required to handle BGP data correctly. The procedure described will be used to retrieve parsable BGP data on which will be applied the heuristics described in the following chapters.

2.1 A brief overview on the BGP protocol

The Border Gateway Protocol (BGP) [97] is nowadays the de-facto inter-domain routing protocol. It was firstly conceived in the late 80s [77] to replace the Exterior Gateway Protocol (EGP) [100] by overcoming its limits – e.g. it forced a treelike topology onto the Internet [91] – and reflecting the new characteristics of the growing Internet, due to the newly introduced Autonomous System (AS) architecture [76]. However, BGP success came only in 1994 with its fourth version (BGP-4) [96] with the introduction of the support for the Classless Inter-Domain Routing (CIDR) and of the route aggregation to reduce the size of the routing table, but mostly thanks to the proliferation of Internet companies subsequent to the end of the sponsorship of the Internet backbone by the National Science Foundation, that made all the Internet traffic to transit on commercial networks [75]. In this work, it will be used the acronym BGP referring to the fourth version of the BGP protocol.

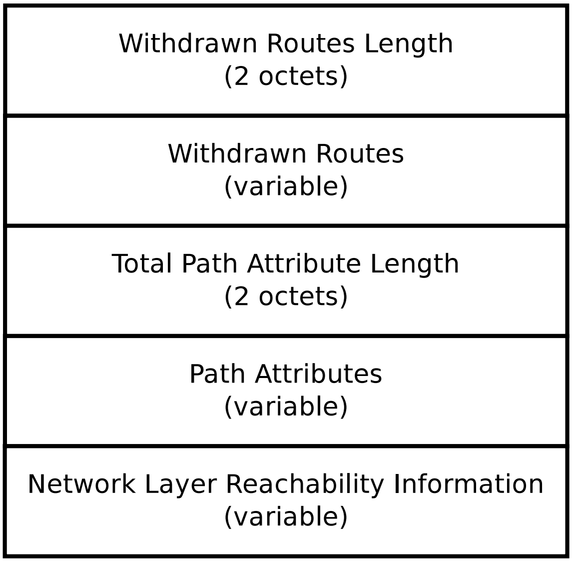


Figure 2.1: BGP UPDATE message format

The primary function of the BGP protocol is to *“[...] exchange network reachability information with other BGP systems, [...] and that includes information on the list of Autonomous Systems (ASes) that reachability information traverses.”* [97]. To achieve that, BGP relies on the exchange of UPDATE messages (Fig. 2.1) between pair of BGP speakers once a BGP session is set up. The purpose of UPDATE messages is to communicate to the other BGP party which part of the network can be reached through the considered BGP connection, by announcing or withdrawing a list of routes. In this context, a route is considered to be “a unit of information that pairs a set of destinations with the attributes of a path to those destinations” [97].

In the Internet AS-level analysis perspective, the most interesting path attribute that is exchanged between BGP speakers is the `AS_PATH` attribute. This attribute is well-known and mandatory – i.e. it must appear in every UPDATE message and it must be supported by every BGP software implementation – and contains the sequence of ASes that will be traversed to reach the announced destinations if the receiving BGP speaker decides to route traffic to those destinations via that inter-domain connection. This attribute is originally created by the AS border router (ASBR) that owns the announced prefixes contained in the Network Layer Reachability Information, and it is modified every time another ASBR propagates the route on the Internet. For example, consider the scenario depicted in Fig. 2.2. The first UPDATE messages are created by `AS 1` by announcing to the `AS 2` and `AS 3` the reachability of prefix

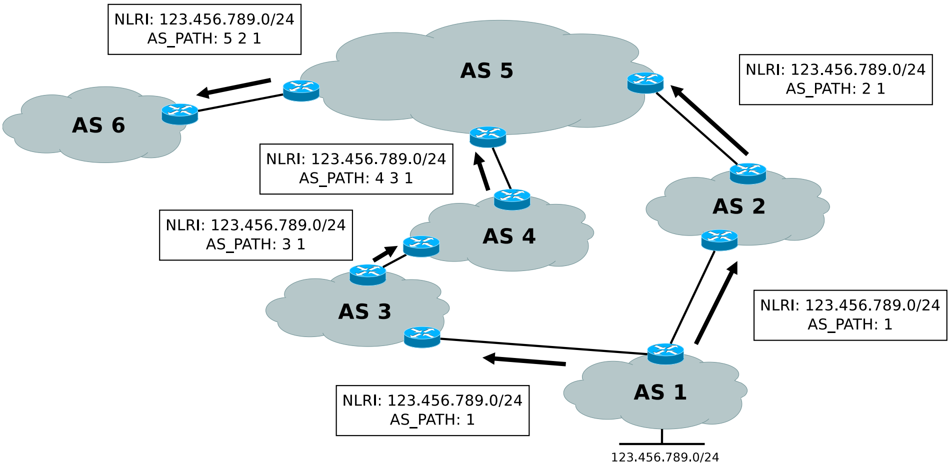


Figure 2.2: AS path creation example scenario

`123.456.789.0/24`. The two ASes will then propagate the UPDATE message to `AS 4` and `AS 5` by prepending their AS number in the `AS_PATH` attribute, as well as `AS 4`, that will propagate the message to `AS 5`. `AS 5` will then select, on the basis of its own criteria, the best way to reach prefix `123.456.789.0/24` and will announce it to `AS 6`.

The routes obtained from each neighbor are stored in its RIB (Routing Information Base) tables. More in detail, each route is firstly maintained in the Adj-RIBs-In tables – one per BGP neighbor – representing unprocessed routes that are available as input to the local BGP decision process. The decision process runs every time an UPDATE message is received by assigning a degree of preference (e.g. preferring shortest AS paths) to each route in each Adj-RIB-In and by applying the policies in the local Policy Information Base (PIB) to the routes stored in the Adj-RIBs-In (Adjacent RIB, incoming), updating the local RIB (LOC-RIB) and placing the chosen routes in the Adj-RIBs-Out (Adjacent RIB, outgoing) to advertise them to the other neighbors. Back to the example in Fig. 2.2, the decision process run by the router in `AS 5` chose to prefer the route announced by `AS 2` over the route announced by `AS 4` over a well defined criterium, e.g. the AS path length announced by `AS 2` is shorter than the AS path length announced by `AS 4`. The `AS_PATH` attribute has also another fundamental role in the decision process: the prevention of routing loops. It is indeed possible to use the information contained in the `AS_PATH` attribute to prune routing loops by excluding from the final phases of the decision process all those routes that carry an `AS_PATH` attribute where is present the AS number of the local system.

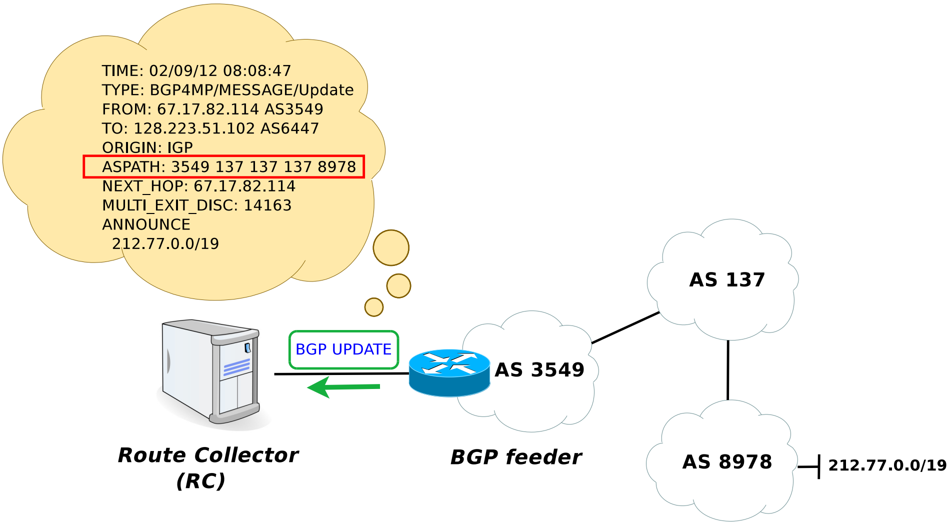


Figure 2.3: Example of route collector

2.2 Route collectors

The most common data sources exploited to analyse the Internet AS-level topology are provided by research projects which deployed a set of *route collectors* (Fig 2.3) around the world aiming at collecting as much information as possible about the Internet routing. A route collector is a device that mimics a BGP router with the sole purpose to gather BGP UPDATE messages from cooperating ASes – i.e. *feeders*. Collected messages contain announcements of new routes and/or withdrawals of previously announced routes and are used by the route collectors to update their RIBs. However, unlike a real BGP router the route collectors do not announce any route back to their feeders. These projects make publicly available, typically in MRT format [21], periodic snapshots of the RIB of each route collector and periodic dumps of the related BGP UPDATE messages, which can be exploited to study the evolution of the RC RIB offline [117].

2.3 BGP Route Collector Projects

There are four main projects that collect BGP data and make it publicly available on the web:

- **BGPmon (BGP Monitoring System)** [1] is a project conceived at the Colorado State University to monitor BGP UPDATE messages and routing tables in real-time. BGPmon makes publicly available parsed MRT data in XML format since August 2012.

- **PCH (Packet Clearing House)** [7] is a non-profit research institute that supports operations and analysis in the areas of Internet traffic exchange, routing economics and global network development. Since July 2010, PCH has been making BGP data available on its website, collected by several route collectors deployed on distinct Internet Exchange Points (IXPs).

- **RIS (Routing Information System)** [9] is a project developed by the Réseaux IP Européens Network Coordination Center (RIPE NCC) – the European Internet Registry – which collects and stores Internet routing data from several locations around the globe deployed on the largest IXPs. It offers also several tools that allow the Internet community to easily read and use BGP data. Data is available in MRT format since July 2000.

- **RouteViews** [10] is a project conceived at the University of Oregon as a tool for Internet operators to obtain real-time information on the global routing system from the perspectives of several different backbones and locations. Since its birth in 1997, this project has provided an invaluable amount of BGP data through its route collectors. Data is available in MRT format since October 2001.

Table 2.1 details the number of BGP feeders per route collector and their geographical location during the month of January 2013. As can be seen, route collectors are mostly located in Europe and North America, potentially introducing geographic biases in the collecting data. A detailed analysis on the impact of geographic factors on BGP data will be provided in Chapter 5.

2.4 Dealing with BGP data

The availability of data in MRT format and the presence of several tools crafted to parse MRT data (e.g. libbgpdump[1] provided by RIPE RIS) ease the BGP data analy-

---

[1] http://www.ris.ripe.net/source/bgpdump/

Route collector (Country)	*# of feeders*
BGPmon	
netsec.colostate (US)	33
PCH	
ams.pch.net (NL)	284
atl.pch.net (US)	48
ber.pch.net (DE)	16
bur.pch.net (US)	29
cai.pch.net (EG)	2
cdg.pch.net (FR)	24
cpt.pch.net (ZA)	12
dac.pch.net (BD)	13
dub.pch.net (IE)	13
equinix-paris.pch.net (FR)	41
eze.pch.net (AR)	2
fra.pch.net (DE)	187
gnd.pch.net (GD)	1
hkg.pch.net (HK)	35
iad.pch.net (US)	80
icn.pch.net (KR)	4
jax.pch.net (US)	3
jpix.pch.net (JP)	16
ktm.pch.net (NP)	17
lax.pch.net (US)	24
lga.pch.net (US)	48
lhr.pch.net (UK)	194
lonap.pch.net (UK)	36
mia.pch.net (US)	36
mnl.pch.net (PH)	4
mpm.pch.net (MZ)	12
muc.pch.net (DE)	12
nl-ix.pch.net (NL)	56
nrt.pch.net (JP)	10
nyiix.pch.net (US)	75
ord.pch.net (US)	36
paix-sea.pch.net (US)	14
pao.pch.net (US)	67
per.pch.net (AU)	7
sea.pch.net (US)	77
sfinx.pch.net (FR)	36
sgw.pch.net (SG)	42
sna.pch.net (US)	79
syd.pch.net (AU)	16
tie-ny.pch.net (US)	24
tll.pch.net (EE)	4
tmp.pch.net (US)	8
trn.pch.net (IT)	21
vie.pch.net (AT)	61
waw.pch.net (PL)	14
wlg.pch.net (NZ)	8
yow.pch.net (CA)	3
yyz.pch.net (CA)	31
zrh.pch.net (CH)	40
RIS	
rrc00 (NL)	25
rrc01 (UK)	64
rrc03 (NL)	75
rrc04 (CH)	12
rrc05 (AT)	42
rrc06 (JP)	5
rrc07 (SE)	14
rrc10 (IT)	21
rrc11 (US)	25
rrc12 (DE)	59
rrc13 (RU)	19
rrc14 (US)	16
rrc15 (BR)	13
RouteViews	
route-views2 (US)	32
route-views4 (US)	19
route-views6 (US)	14
route-views.eqix (US)	16
route-views.isc (US)	14
route-views.jinx (ZA)	8
route-views.kixp (KE)	1
route-views.linx (UK)	26
route-views.saopaulo (BR)	15
route-views.sydney (AU)	8
route-views.telxatl (US)	5
route-views.wide (JP)	5

Table 2.1: List of active route collectors (January 2013)

ses, but it still remains an extremely time consuming procedure. Thus, the best choice is to create ad-hoc software tools extremely focused on the type of analysis to perform. In this perspective, even though a *route* is formally defined as "*a unit of information that pairs a set of destinations with the attributes of a path to those destinations*" [97], in this thesis a route is considered to be only composed by the attributes that allow to identify a BGP session and the AS path toward the destination, i.e. the attribute triplet `<destination, BGP session ID, AS_PATH>`. The BGP session ID is not directly available in MRT data, but it can be inferred by taking into account the pair composed by the MRT fields `Peer IP Address` and `Peer AS`. The evolution of each route is then obtained by extracting the related entries from the first RIB available in the time period considered and from each subsequent UPDATE message. At the end of the time period it is also introduced an artificial withdrawn to each route still present in the RIB in order to perform a proper time-limited analysis.

Despite the routing information contained in the data repositories are considered to be reliable since directly collected from devices that are effectively participating in the inter-domain routing process, a proper evolution analysis tool cannot be considered to be reliable if it does not take into account several flaws that are still contained in raw data. Here in the following are listed the most relevant flaws in BGP raw data, and the approach used to limit their presence.

Unusable AS numbers

The Internet Assigned Numbers Authority (IANA) [5] reserved a well defined set of AS numbers that should not be announced in the Internet. Among the others, private use AS numbers (`AS64512–AS65535`) are part of this set. These AS numbers are typically used if an AS is only required to communicate via BGP with its unique provider and should not be announced in the Internet and should not be part of any AS path. However, some of them are still erroneously announced in the Internet, and in data repositories can be found a non negligible number of routes whose AS path terminates with a private AS number. The analysis should get rid of them by removing them from AS paths.

BGP support for 4-octet AS numbers

In order to handle the two-octet AS numbers exhaustion, IANA has introduced new AS numbers coded on four-octets allowing a gradual transition from 2-octet to 4-octet AS numbers [110]. In the very same RFC, it is introduced also the `AS_TRANS` (`AS 23456`) that should be used in the interaction between four-octet-capable (new) and two-octet-capable (old) BGP speakers. In detail, whenever a new BGP speaker announces to an old BGP speaker an UPDATE message, the new BGP speaker must send the AS path information encoded with two-octet AS numbers to the old BGP speaker, using `AS_TRANS` to represent the presence of a 32-bit AS number in the `AS_PATH` and by sending the four-octet AS path information in the `AS4_PATH` attribute. It is thus

Error type	**(Real) Example**
Lack/excess of a trailer digit	3561 26821 **27474 2747 27474**
Lack/excess of a header digit	286 3549 9731 **38077 8077 38077**
Lack/excess of a middle digit	13030 1273 **9329 929 9329**
Missing space between ASNs	. . . 2152 3356 **35819 3581935819 35819 35819**. . .
Error on a digit	13030 22212 19024 **25782 25785 25782**
Error on two digits	11686 4436 3491 **23930 23390 23930** 7306
Missing digit cause split ASN	6939 5603 **21441 21 41 21441**

Table 2.2: Loops caused by human errors

possible to receive BGP messages including `AS_TRANS` in their AS paths, and the analyses should be aware of that. In this thesis, since `AS_TRANS` do not represent a single AS, every connection involving it is dropped, while the AS path is considered to be still valid.

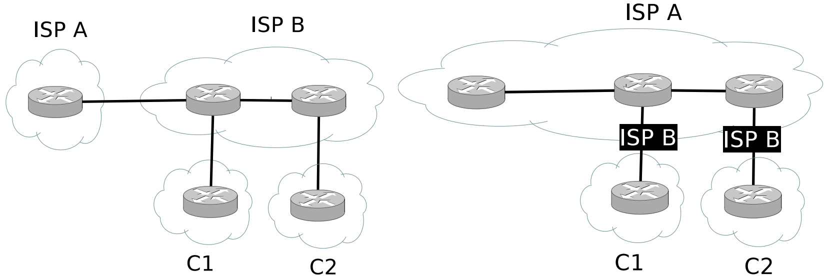
BGP state information missing

Only a few route collectors record the state information about their BGP sessions, and thus it is not straightforward to understand when the routes contained in their local RIBs have to be invalidated due to a session failure. The lack of these entries leads routes that are withdrawn during the BGP session downtime to be considered still active after the session is set back. More details can be found in [50]. In absence of state information recording, it is however still possible to detect BGP session failures by identifying full routing table transfers in the stream of UPDATE messages – which occur right after a BGP session is (re)established [97] – exploiting one of the algorithms available in literature (e.g. [31, 32]) and then considering each route in the RIB as withdrawn whenever a session failure is found. In the analyses performed in this thesis it has been considered to be withdrawn each route in the RIB whenever a session failure is found via the Minimum Collection Time (MCT) algorithm described in [31].

Loop correction

Some loops can be found in AS paths even if the default behavior of BGP is supposed to prevent their formation. Most of them are caused by typos in configuring ASBRs export policies, as already highlighted in [52], and these may introduce false connections in the analysis. There are three major causes of loops in BGP data:

- *a) Human error during AS path prepending.* When a BGP router sends an announcement it must prepend its local AS number to the AS path field. BGP allows the manipulation of the AS path length by *prepending* the owned AS number multiple times to influence the routing decision of neighboring ASes. This feature is



(a) Start scenario (b) Migrated Scenario

Figure 2.4: Network migration scenario

implemented through rules to identify which announcement must be affected by the manipulation of the AS path and how many times the prepending must be done. Typically these rules are set manually by administrators, thus during this setup errors may be introduced generating loops. Table 2.2 uses real AS paths in order to highlight seven different kinds of human errors.

- *b) Network Migration.* Consider ISP `A` and `B`, as represented in Fig.2.4a. Suppose that `A` purchases `B`, then customers of `B` become customers of `A` (Fig.2.4b). The external BGP (eBGP) peering sessions with customers of `B` have to be re-configured, requiring significant coordination and planning efforts. The Cisco Local-AS feature allows the *migrated* routers (routers formerly owned by `B`) to participate in AS `A` while impersonating AS `B` towards (previous) customers of AS `B`. Routers using the Local-AS feature retain the information that the BGP routes have passed the local AS in the AS path. They prepend *B* in inbound eBGP updates and prepend both current ASN *A* and *B* in outbound eBGP updates. In this environment, some loops can be introduced if (previous) customers of `B` exchange UPDATE messages with each other passing through AS `A`.

- *c) Split ASes.* A split AS is an AS that is divided into two (or more) islands. An AS can be split steadily or could split due to an internal temporary network failure. An example of a steady split AS is `AS 48285` (Robtex.com). Consider the following AS path: `[44581 48285 16150 5580 48285]`. `AS 16150` and `AS 5580` are respectively located in Sweden and the Netherlands. After private conversation with the Robtex.com administrator and it was found that `AS 48285` is indeed *steadily* split. Thus, to obtain the connectivity between the two islands, the path needs to pass through other ASes (in this case `AS 16150` and `AS 5580`).

Note that AS paths matching case *a*) can introduce *false* connections, thus it is critical to fix them so as to obtain a reliable topology. Thus, each time an AS path is showing a loop caused by an human error, the loop is removed from the AS path. On the other hand AS paths falling in class *b*) and *c*) should not be fixed since they reflect a real-world situation.

3

BGP and economics

The undirected graph of the Internet inferred through the analysis of the `AS_PATH` attribute of BGP data collected via route collectors is not sufficient to model the real inter-domain routing of the Internet. Contractual agreements override scientific metrics [53] and, as a result, the largest amount of paths that can be extracted from the undirected AS-level graph of the Internet are not actually feasible [52]. The common approach to solve this issue and map the real routing choices of ASes is to enhance the undirected AS-level topology into an economic AS-level topology where each edge is tagged with a proper economic label that reflects the type of relationship existing among the involved ASes and the related BGP export policy. However, despite (or due to) their key role in business environment, details about inter-AS economic relationships are usually not publicly available, and researchers needed to develop heuristics [40, 42, 52, 59, 67, 87, 109] to infer them. The availability of a topology of the Internet that would be aware of the economic relationships among each AS has several practical implications in the real world. For example, a Content Data Network (CDN) could use this knowledge to select the best places in which to deploy replicas of its server, and a new regional Internet Service Provider (ISP) could select the best upstream ASes through which it could connect in a shorter number of hops to the rest of the Internet [109].

AS relationships are thus fundamental to yield a better insight into the business choices that lead to the creation of the current Internet structure. Despite the large number of possible economic agreements, inter-AS relationships can still be categorized into three main classes (Fig. 3.1) on the basis of the BGP export policy applied by each AS, i.e. the set of routes that each AS announces to the other: provider-to-customer (p2c) – or customer-to-provider (c2p) – peer-to-peer (p2p) and sibling-to-sibling (s2s) [52]. In p2c and c2p relationships, the provider announces to the customer routes required to reach every Internet destination, selecting them among the routes that the provider obtained from its customers, providers and peers (if any) plus the routes owned by the provider itself. The customer, on the other hand, announces back only routes related to its own IP prefixes and routes obtained from its customers (if any) in order to avoid the possibility to become a transit AS for its provider. In p2p relationships, each of the peers announces routes related to its own IP prefixes and routes obtained from its customers to the other peer. This relationship is typically free-of-charge and is established between ASes for mutual benefit. For example, two ASes may decide to establish a p2p connection between them to lower the amount of traffic directed to their providers and, as a consequence, lowering their expenses and their dependence on their providers. Finally, in s2s relationships each of the siblings acts like a provider for the other by announcing its full routing table to the other, and is typically applied between ASes owned by the same organization. This level of granularity is typically considered to be consistent with the reality [102], despite there exist some exceptions [84] mainly caused by policies established on a geographic basis [60, 85].

The first works about inter-domain economic characteristics of the Internet appeared during the late 90s [56, 68], but only a couple of years later the BGP export policies related to the most common economic inter-AS relationships [52] were described for the first time. These relationships, together with the related BGP export policies, were defined by identifying the routes announced from an AS to its neighbors, leading to the definition of the valley-free property that every AS path should respect. This property has been widely used in a large number of algorithms aimed at inferring the economic relationships that lie behind each AS connection [40, 42, 67, 87, 109]. The major drawback of these algorithms is that the valley-free property does not always hold, mainly due to the presence of misconfigured export policies on BGP border routers [78], that show up in the form of no-valley-free AS paths announced during BGP path exploration phenomena [88]. This is a well-known issue that may affect the results of inferences about economic relationships based on on the analysis of raw BGP data.

3.1 On the effects of BGP misconfigurations in an economic inference perspective

Entries in BGP data are actually containing AS paths announced in the Internet, but they still cannot be used as they are in every type of inference. In particular, inter-AS economic relationship inferences suffer the presence of *spurious* entries, that may lead to wrong results. These entries are typically caused by BGP export policy misconfigurations [78] and show up during BGP path exploration phenomena [88], that occurs when an AS Border Router (ASBR) experiences a network failure. In detail, before declaring the destination network unreachable, an ASBR tries to use several different routes available in its Adj-RIBs-in tables for which a withdrawn message has not been received yet, very likely due to time propagation delays. As a consequence, during the lifetime of this process several UPDATE messages that contain routes not

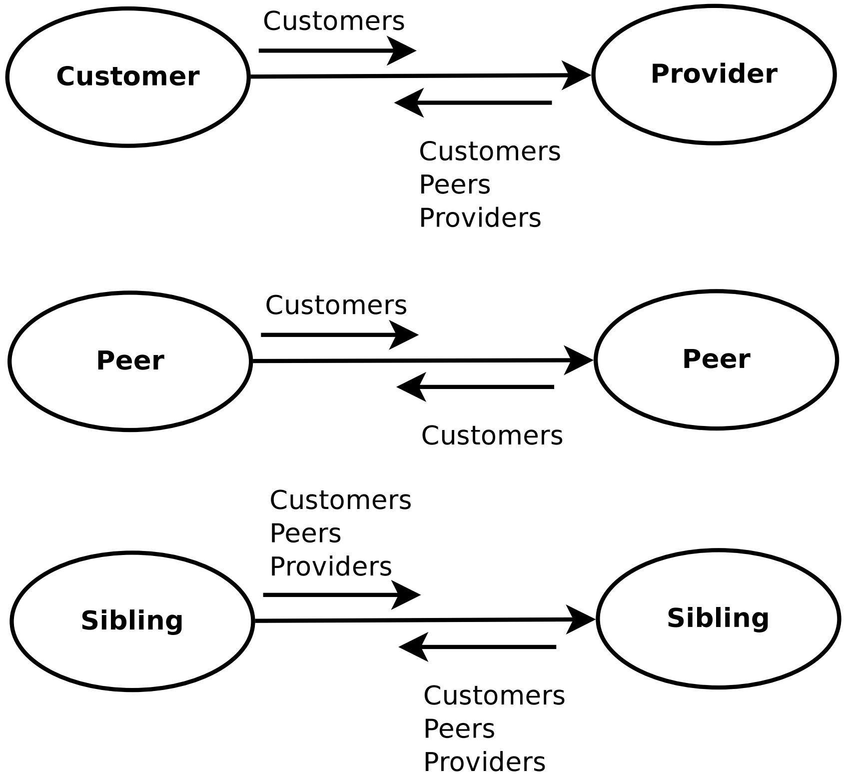


Figure 3.1: Inter-AS economic relationships

commonly used are generated and propagated. Among them, there may also exist messages containing routes which violate the business relationship agreement among ASes – i.e. violate the valley-free principle described in [52] – typically due to human errors in defining BGP export policies on ASBRs. In [78] are described two different classes of BGP misconfigurations that can potentially induce these AS paths: *origin* and *export* misconfiguration. In the former, an AS accidentally announces a prefix that should not be announced, while in the latter an AS sends an announcement to a neighboring AS which violates the commercial agreement between them. More details about the nature of these misconfigurations can be found in [59, 78] and [88]. These spurious routes are then propagated through the Internet up to the route collectors and, thus, appear in public datasets. As a result, several economic inferences drawn using BGP public data are potentially wrong, since AS paths contained in spurious entries interfere with those obtained from stable AS paths. Note that these routes, due to their ephemeral nature, do not affect the network performances of the

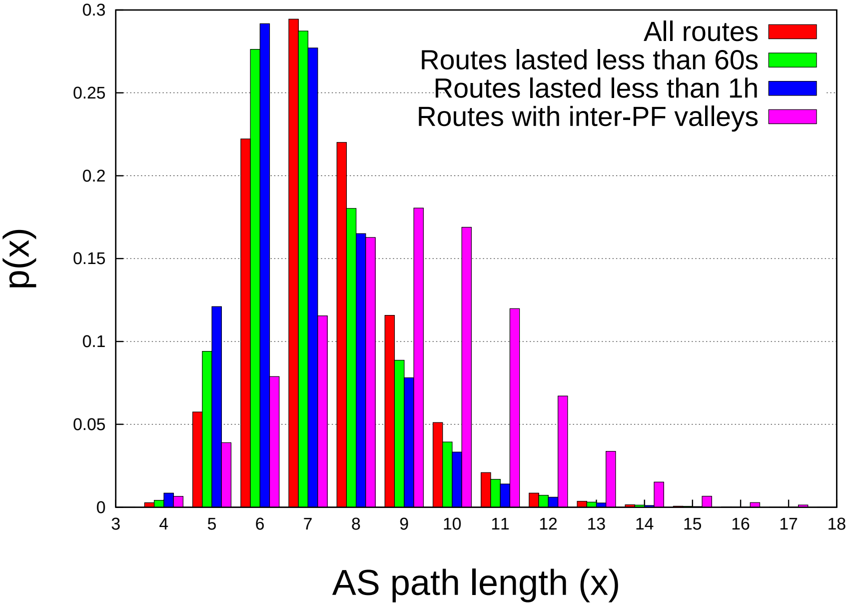


Figure 3.2: AS path length distribution (January 2013)

AS owning the misconfigurated ASBR, thus it is very likely that several misconfigurations are actually gone unnoticed by network operators. For the very same reason, it is unlikely (but still possible) that an AS unconsciously maintains a BGP misconfiguration for a long period of time, since an AS that transits non-planned traffic for at least one of its providers (or peers) perceives a degrade in its network performance.

The presence of spurious entries is a very well-known issue. The common approach to solve it is to consider every transient route as a potential source of problem for the inferences, applying heuristics that limit their effect on the final tagging results [52, 59, 87]. This is not completely correct, since not every transient route carries no-valley-free AS paths. For example, backup routes may have a relatively short lifespan, but they carry perfectly legitimate AS paths from which they may be inferred previously unnoticed p2c releationships between pair of ASes. One of the most evident characteristics that differentiates a spurious entry from a normal transient route is related to its AS path length. As a proof of that, in Fig. 3.2 is depicted the Probability Distribution

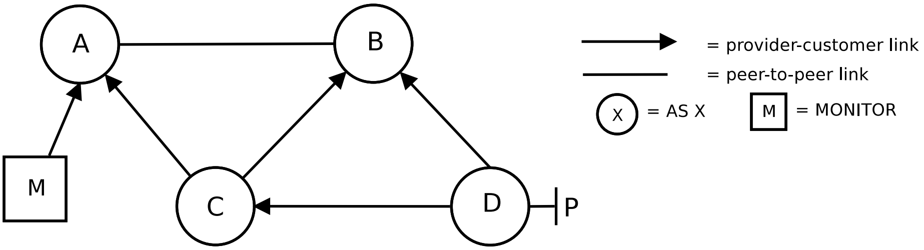


Figure 3.3: Scenario

Function (PDF) of the AS path length[1] of four set of routes, composed by: *a*) every route, *b*) every route which lasted less than 60 seconds, *c*) every route which lasted less than one hour and *d*) routes containing easily spottable no-valley-free AS paths. Note that the four sets are not disjoint, e.g. routes in set *b* are also included in set *a* and *c*. To identify the routes belonging to set *d* it is exploited the a priori knowledge about the provider-free property owned by a limited set of ASes, i.e. those ASes that do not need to buy transit from any other AS to reach every Internet destination. A list of these ASes can be found on Wikipedia [13]. Then, it is selected every AS path that includes two (or more) provider-free ASes not directly connected to each other, i.e. every AS path in which a third AS transited traffic for one of the provider-free ASes. In the rest of this work these routes are referred as *inter-PF* routes. As can be seen, inter-PF routes have a greater probability to have longer AS paths than routes in the other sets, due to BGP misconfigurations on ASBRs showing up during BGP path exploration phases. This is not the general behavior of short-lived routes (sets *b* and *c*), that can be generated by a plethora of other reasons – e.g. backup connections, route flapping, traffic engineering issues – and that typically carry perfectly legitimate AS paths. Moreover routes belonging to set *d* are usually short-living with respect to the total set of routes, as shown in the Complementary Cumulative Distribution Function (CCDF) depicted in Fig. 3.4, confirming their extremely transient nature. Thus, it seems reasonable to *assume* that spurious routes can be identified by their abnormal AS path length and by their short-living appearances in BGP datasets. In an economic relationship inference perspective, it is also worth to note that the AS paths of the routes in set *d* involve a total of 11,265 connections over the 170,204 found in BGP data, i.e. the inferences related to at least the 6.61% of AS connections found on the Internet are potentially affected by spurious routes, stressing the need for an appropriate filtering mechanism of input data.

---

[1] In computing the AS path length of each route each AS path was compressed by removing the effect of AS path prepending. During this process are also removed from AS paths the effects of human errors in AS prepending as described in the previous chapter.

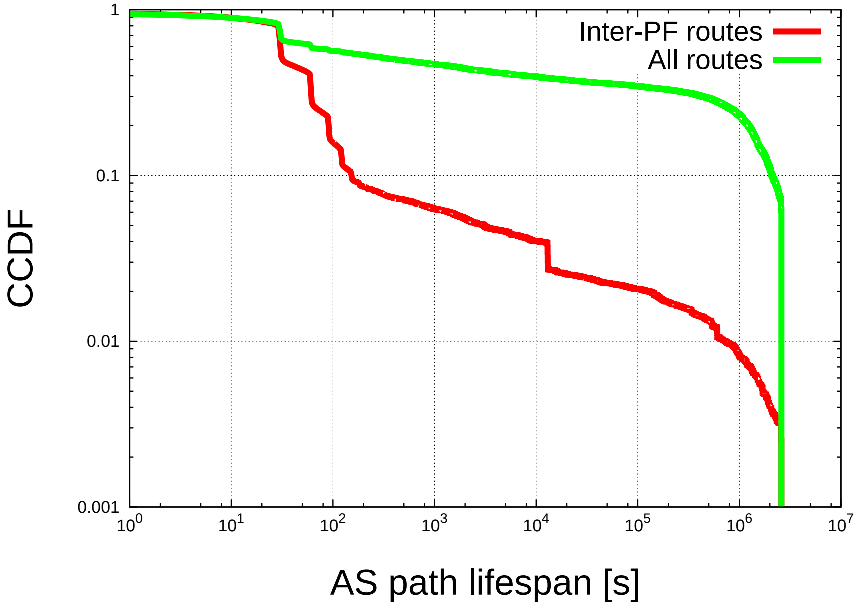


Figure 3.4: AS path lifespan CCDF

Note also that the largest amount of long-lived routes belonging to this set can still be considered as legitimate, since it is largely composed by routes in which the provider-free ASes are separated by a third AS that can be reconduced to organizations owning one of the involved provider-free ASes. Indeed, it is a common practice of worldwide ISPs to own more than a single AS number, both because of traffic engineering reasons (e.g. Verizon Communications manages AS 701 for North-American routing, AS 702 for European, Middle Eastern and African routing and AS 703 for Asian and Pacific region routing) and because of mergers and acquisitions [24]. The peak of lifetimes around 30 seconds in Fig. 3.4 can be explained by the fact that 30 seconds is the suggested value for the *MinRouteAdvertisementIntervalTimer* (MRAI timer) of the BGP protocol [97]. This timer indicates the minimum amount of time that should elapse between two consecutive announcements regarding the same route. This default value is used by the vast majority of routers, and several short-lasting no-valley-free paths are replaced at least every 30 seconds.

One further plausible explanation for several of paths in set *d* is that they are a consequence of the co-effect of the convergence of BGP protocol upon a network failure and the usage of a particular type of outbound policy operated by one of the ASes involved. BGP allows the filter of outbound announcements to be set up on a prefix basis. The filter thus prevents an announcement carrying a network prefix that belongs to one of its providers from being propagated to its peers and providers. However this filter has a drawback that can be seen as soon as a BGP connection is closed, either due to a temporary network failure or to the end of an agreement. Consider the scenario in Fig.3.3. In this scenario, `C` uses the AS path `[D]` to reach prefix `p`, which belongs to its customer `D`. However, `C` has stored in its Adj-RIB-In tables also the AS paths `[A, B, D]` and `[B, D]`, received from its providers `A` and `B` respectively. Now suppose that due to a network failure in `p`, `D` sends a withdrawal to `B` and `C`. `C`'s BGP decision process will remove from its RIB the AS path `[D]` to reach network `p` and will search for another way to reach it before declaring to its neighbors that network `p` has been withdrawn. Since it has not yet received any withdrawal message from `B` concerning `p`, the direct consequence is that `C` will select `[B, D]`[2]. If `C` performs an outbound filtering implemented as described above, it will then announce to `A` the route `[B, D]` to reach `p`, even if it is clearly in contrast with the p2c agreement signed with `B`. This is because network `p` appears in the list of networks that can be advertised to all the providers. Furthermore, since an AS typically prefers a route toward a customer over a route toward a peer or a provider, `A` will select the AS path `[C, B, D]` to reach `p` and it will announce it to the monitor, thus causing the creation of a no-valley-free path. In practice, the monitor sees that `C` is transiting traffic between its providers `A` and `B` for a short time since the network `p` will be withdrawn from the Internet at the end of the convergence of BGP protocol. Note that this is an issue for the tagging algorithm, but not for the AS-level topology discovery tools. The connections among ASes that appear during these spurious routes actually exist, even though the traffic typically does not pass via the given AS path.

3.2 A time-aware tagging algorithm

The first proposed approach to infer inter-AS economic relationships exploits the list of provider-free ASes available in Wikipedia [13], denoted in the rest of this work by T_{list}, and relies upon the following basic principle: every AS should be able to reach every Internet network, thus there must exist at least one AS path including the considered AS and an AS part of T_{list}. A fundamental characteristic of this algorithm is that it computes the economic relationships by also considering the lifespan of the AS path: tagging decisions made via long lasting paths must not be affected by the presence of spurious and transient AS paths that are not used to transit IP traffic, since they can lead to misleading results in the tagging algorithm. The main piece of information that can be exploited to infer AS relationships is the set of AS paths that can be gathered

---

[2] For simplicity's sake, the length of AS paths is considered to be the only relevant decision factor in BGP decision process

```
1  foreach AS path
2      foreach direct connection (A,B)
3          if (T_1 $\in T_{list}$ follows (A,B) in the AS Path)
4              Tag(A,B) = c2p
5          if (T_1 $\in T_{list}$ precedes (A,B) in the AS Path)
6              Tag(A,B) = p2c
7          if (does not exist any T_1 $\in T_{list}$ )
8              Tag(A,B) = p2p
```

Figure 3.5: Time-aware economic algorithm: step a) Inferences per direct connection

from the BGP data stored by monitors and their lifespan. The proposed algorithm assigns an economic tag to each connection of the topology, which also indicates the level of reliability for each of them. The algorithm is organized in the following three steps:

a) Inference of every possible economic relationship for each AS connection.

In this step, the algorithm analyses every single AS path separately from the others and assigns a raw economic tag to each direct connection[3] found in each AS path. To do that it exploits the presence of ASes in T_{list} in order to highlight which AS is transiting traffic for another AS. For each direct `(A,B)` connection inside the path considered the algorithm proceeds as reported in Fig.3.5. Consider an AS path such as $[\ldots\ A\ B\ \ldots\ T_1\ \ldots]$. The algorithm infers that B is a provider of A (line 3), because B is announcing to A routes retrieved from T_1. In other words, B is *providing* A with connectivity to a portion of the Internet.

Consider now an AS path such as $[\ldots\ T_1\ \ldots\ A\ B\ \ldots]$. In this case the algorithm infers that A is provider of B (line 5), because their relationship cannot be neither p2p nor c2p. This because if A and B established a p2p relationship, then A would transit traffic between one of its providers or peers (T_1.) and another peer (B), thus violating the export rules imposed by the p2p agreement. Note that by definition T_1 is provider-free, thus it cannot be a customer of the neighboring ASes. The same argument can be applied to show that A and B cannot have a c2p relationship. If the AS path considered does not contain any of the T_{list} ASes following or preceding `(A,B)`, the algorithm infers that A and B have a p2p relationship because there is no proof that either A or B uses the other as a provider (line 7). Note that the aim of the algorithm is not to infer the relationships among ASes included in T_{list}, since they are assumed

---

[3] It is denoted as *direct connection* a connection in which the direction *is* relevant, i.e. the direct connection `(A,B)` is different from the direct connection `(B,A)`.

```
1 foreach direct connection (A,B)
2     foreach Tag(A,B) from the longest-lasting to the shortest-lasting
3         if (exists dir_Tag(A,B))
4             if ( lifespan [dir_Tag(A,B)] $\sim$ lifespan [Tag(A,B)])
5                 dir_Tag(A,B) = dir_merge(dir_Tag(A,B), Tag(A,B))
6         else
7             dir_Tag(A,B) = Tag(A,B)
```

Figure 3.6: Time-aware economic algorithm: step b) Inference per connection

to be p2p. This step of the algorithm also maintains for each tagged connection the maximum lifetime of the AS paths used to infer it, that will be useful in the following steps.

b) Inference of a single economic relationship for each direct AS connection.

In this step, the algorithm uses the results of the previous step to infer for each *direct* connection `(A,B)` a unique economic tag among all the tags collected for the same direct connection. The algorithm first orders the tags inferred for the direct connection `(A,B)` in descending order of lifespan. Then, it analyzes each tag from the longest-lasting to the shortest-lasting, as illustrated in Fig.3.6. The lifespan of each tag plays a major role, since the algorithm allows the current examined tag (Tag`(A,B)`) to affect the *current* resulting tag for the same direct connection (direct_Tag`(A,B)`) iff its lifespan does not differ by more than N_{MAG} order of magnitude from the longest-lasting tag found for the same direct connection, i.e. iff the two lifespans are *comparable* (line 4). Note that the algorithm examines each tag in descending order of lifespan, thus `dir_Tag(A,B)` contains the longest-lasting tag for the direct connection `(A,B)`. With this solution, the algorithm simply ignores the transient paths to infer the economic relationship for those connections that are found both in transient and stable paths, while it still analyzes connections that are found only as short-living since it assumes that they are backup connections. The merging rules used in this step are reported in Table 3.1a and can be justified by the export policies described at the beginning of this Chapter (Fig. 3.1). If `A` and `B` have a p2c relationship – or `B` and `A` have a c2p relationship – this means that `A` can reach only the customers of `B` while `B` can reach the customers, peers and providers of `A`. On the other hand, if `A` and `B` have a s2s relationship, `A` can reach the customers, peers and providers of `B` and vice versa, while if `A` and `B` have a p2p relationship, they only reach their respective customers. Thus, the merge of a s2s tag with either p2c or p2p tags results in a s2s tag and the merge of a p2c (c2p) tag with a p2p tag results in p2c (c2p) tag.

```
1 foreach direct (A,B) connection
2     get dir_Tag(A,B)
3     if (exists dir_Tag(B,A))
4         if ( lifespan [dir_Tag(B,A)]∼lifespan [dir_Tag(A,B)])
5             confirmed_Tag(A,B) = inv_merge(dir_Tag(A,B), dir_Tag(B,A))
6         else
7             unconfirmed_Tag(A,B) = dir_Tag(A,B)
8     else
9         unconfirmed_Tag(A,B) = dir_Tag(A,B)
```

Figure 3.7: Time-aware economic algorithm: step c) Final tagging and two-way validation

c) Final tagging and two-way validation.

In the final step, the algorithm uses the results of previous step to infer an economic relationship for each indirect connection[4], as illustrated in Fig.3.7. For each *direct* `(A,B)` connection the algorithm first gets its resulting tag (`dir_Tag(A,B)`) and checks if there is a resulting tag for the direct connection `(B,A)` (`dir_Tag(B,A)`). If both exist then it checks whether their lifespans are comparable (line 4) and if necessary merges the two tags (`inv_merge(dir_Tag(A,B),dir_Tag(B,A))`). The merging rules, listed in Table 3.1b, are very similar to those introduced in the previous step, but they also take into account that a p2c for connection `(A,B)` is equal to a c2p for the `(B,A)` connection. In this step it is also found which tag has a *two-way validation*, i.e. which tag is inferred from both the tags of the direct connections `(A,B)` and `(B,A)`. Then, a label is assigned for each tag which indicates if it is two-way validated or not, distinguishing between *confirmed* and *unconfirmed* tags.

	[A, B]			
[A, B]	**p2c**	**p2p**	**c2p**	**s2s**
p2c	p2c	p2c	s2s	s2s
c2p	s2s	c2p	c2p	s2s
p2p	p2c	p2p	c2p	s2s
s2s	s2s	s2s	s2s	s2s

(a) Direct Merge

	[B, A]			
[A, B]	**p2c**	**p2p**	**c2p**	**s2s**
p2c	s2s	p2c	p2c	s2s
c2p	c2p	c2p	s2s	s2s
p2p	c2p	p2p	p2c	s2s
s2s	s2s	s2s	s2s	s2s

(b) Inverse Merge

Table 3.1: Time-aware economic algorithm: rules to merge tags

---

[4] It is denoted as *indirect connection* a connection in which the direction *is not* relevant, i.e. the indirect connection `(A,B)` is equal to the indirect connection `(B,A)`.

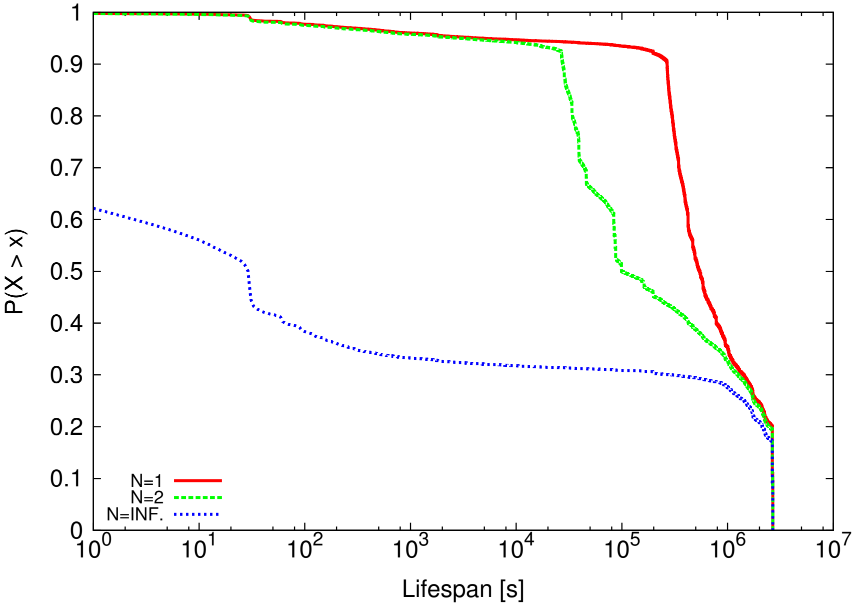


Figure 3.8: Time-aware economic algorithm: CCDF of the minimum value of lifespan of AS paths exploited

3.2.1 Results

For sake of semplicity in this subsection are presented only the results of the tagging algorithm obtained using N_{MAG} = 1, 2, ∞ [5], while the results related to the other values of N_{MAG} can be found at [4]. The results are shown in Table 3.2. Note that the number of *confirmed* tags is ranging from 4.77% ($N_{MAG} = 1$) to 6.70% ($N_{MAG} = \infty$) for raising N_{MAG}. However, if the value of N_{MAG} is increased there is more probability that transient paths will affect the final tag decision, thus lowering the reliability of the algorithm itself. This is confirmed by Fig.3.8 that shows the CCDF of the minimum lifespan of AS paths that participate to the tag of each connection. As can be seen from this figure, if the lifespan is not considered then a large portion of short-living AS paths are contributing to the tag of each connection, possibly leading to a wrong result. The results also highlight the presence of a large percentage of unconfirmed p2p relationships, which decrease for increasing values of N_{MAG}. This is caused mainly by lack of information [61, 87], but also by the fact that the reverse connection that could confirm the tag is short lasting, and thus not considered by the algorithm.

---

[5] $N_{MAG} = \infty$ represents the case where the lifespan is not considered, i.e. every AS path contribute to the tagging algorithm outcome.

	Tag type	Total	Unconfirmed	
			Total (%)	Involving Stubs (%)
$N_{MAG} = 1$	*p2c*	89,138	86,103 (96.59 %)	64,340 (72.18 %)
	p2p	79,378	75,542 (95.16 %)	27,427 (34.55 %)
	s2s	1,375	−− (−− %)	−− (−− %)
$N_{MAG} = 2$	*p2c*	89,898	86,641 (96.37 %)	64,651 (71.91 %)
	p2p	78,389	74,526 (95.07 %)	27,115 (34.59 %)
	s2s	1,604	−− (−− %)	−− (−− %)
$N_{MAG} = \infty$	*p2c*	95,958	90,678 (94.49 %)	66,030 (68.81 %)
	p2p	70,741	67,260 (95.07 %)	25,735 (36.37 %)
	s2s	3,192	−− (−− %)	−− (−− %)

Table 3.2: Time-aware economic algorithm: tag results

A particular scenario created due to lack of information is depicted in Fig. 3.9. The figure shows that, due to the effect of BGP decision processes, the monitors cannot gather every possible AS path. Thus the results of our algorithm can be misleading. Consider `S` to be the source AS, `T` an AS included in T_{list} and `M` the monitor. Consider also that `A`, `B`, `C` and `T` are the transit providers of `S`, that in this case acts as a stub AS, i.e. it does not transit any IP traffic for any other AS, but only for end users. Supposing that `T` decides that the best path to `S` is via `A`, the AS paths that are gathered by `M` in the steady state of this scenario will be `(T, A)`, `(T, B)`, `(T, C)`, `(T, A, S)`, `(B, S)` and `(C, S)`. Applying our algorithm to this scenario leads to a single unconfirmed p2c relationship for `(A, S)`, while the couples `(B, S)` and `(C, S)` will be interpreted as unconfirmed p2p. This could be solved by connecting the BGP collector to the stub AS, since there would thus be a higher probability to gather AS paths involving `T`, i.e. `(S, A, T)`, `(S, B, T)` and `(S, C, T)`, which would then transform the relationships `(S, B)` and `(S, C)` into an unconfirmed p2c. The same scenario could be applied to small and medium ISPs, since it is the effect of BGP decision processes, and this phenomenon also introduces a large number of unconfirmed p2p which, in the real world, are p2c relationships. More details about the effects and the amount of incompleteness in BGP data is provided in Chapter 5. Note however that this affect mostly the reliability of p2p connections, since whenever a tag is labelled with p2c (c2p) there has been a tangible proof in BGP data of the fact that one of the two ASes transited traffic for the other.

Some of these connections can be spotted by analyzing stub ASes. A typical stub AS has several p2c relationships to transit ASes due to multihoming and it does not develop any p2p relationship. Particular real world examples of this class of ASes are ASes owned by banks, car manufacturers, universities and local ISPs. It is possible to find these kinds of ASes by analysing every AS path available and considering

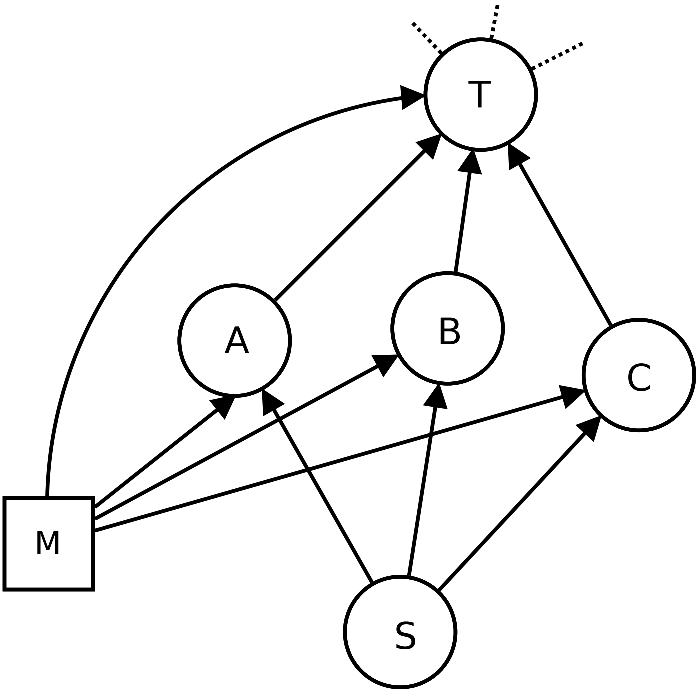


Figure 3.9: BGP monitor placement pitfall example scenario

as stub AS each AS that only compare as last hop in the AS paths, i.e. those ASes that do not transit any IP traffic for other ASes. Due to their nature these ASes are likely to be customers in the relationships that involve them, thus there is a higher probability that unconfirmed p2p relationships involving them could be turned into p2c relationships. As reported in Table 3.2, the large number of unconfirmed p2p relationships involving stub ASes supports this rationale. The probability of being a real p2c relationship is very high also for unconfirmed p2c relationships. Note that if the unconfirmed p2p connections involving stubs and unconfirmed p2c are considered as confirmed p2c, the percentage of confirmed tags would raise to near 60% for all the N_{MAG} considered.

3.3 Towards spuriousness-free economic inferences

So far, economic tagging algorithm in literature have tried to mitigate the effects of spurious routes minimizing the number of no-valley-free routes by formulating an optimization problem [40, 42, 67, 109]) or limiting their presence in the final inferences using thresholds, either on the number of entries used to infer transit relationships [52] or on the lifespan of each route, e.g. cutting off every route lasted less than two days [87] or every route whose lifespan is not comparable with those long-lasting [59], as it was done in the algorithm proposed in Section 3.2. In both cases, the choice is arguable. On one side, the minimization of the number of no-valley-free paths during the optimization algorithm forces the property to hold, regardless of the data available and of the real routing behaviors. On the other side, it is hard to identify the optimal time threshold able to discard transient routes from legitimate routes, and its choice can potentially affect the inferences of the algorithms. Moreover, as will be shown in the following sections, not every short-lived route carries spurious routes, and discarding each of them indiscriminately can thus lead to a loss of potentially valuable routing information. To the best of the author's knowledge, no methodology aimed to identify spurious routes has been developed yet, although several methodologies have been proposed to identify BGP instabilities (e.g. [49, 38]). In this section it is proposed a new algorithm that is able to infer inter-AS economic relationships by removing spurious entries within an initial set of data thanks to a dedicated preliminary data hygiene phase, rather than relying on arbitrary and debatable thresholds.

3.3.1 Preliminary data hygiene phase

In Section 3.1 it was highlighted that spurious routes are characterized by a short lifespan and an abnormally long AS path. To identify candidate spurious routes, it has been devised a three-step filter (Fig. 3.10) which exploits the following consideration: during a significant observation time – in this case a whole month – an AS tends to select predominant routes to proficiently reach a destination [51, 98], i.e. the AS paths used by an AS tend to have predominant AS path length values to reach a given destination. In particular, it is possible to apply data binning to the time-based AS path length distribution in order to retrieve a binned normal distribution centered around the mode of the original distribution. As a consequence, AS paths which do not fall in the the normal binned behavior can be marked as outliers by applying the three-sigma rule, commonly used to identify outliers in normal distributions [74].

More in detail, step *a*) of the algorithm (Fig. 3.11) aims to create the AS path length distribution experienced during the whole month by each pair `<feeder IP f, destination p>` and to compute the proper bin value to consider this distribution as a

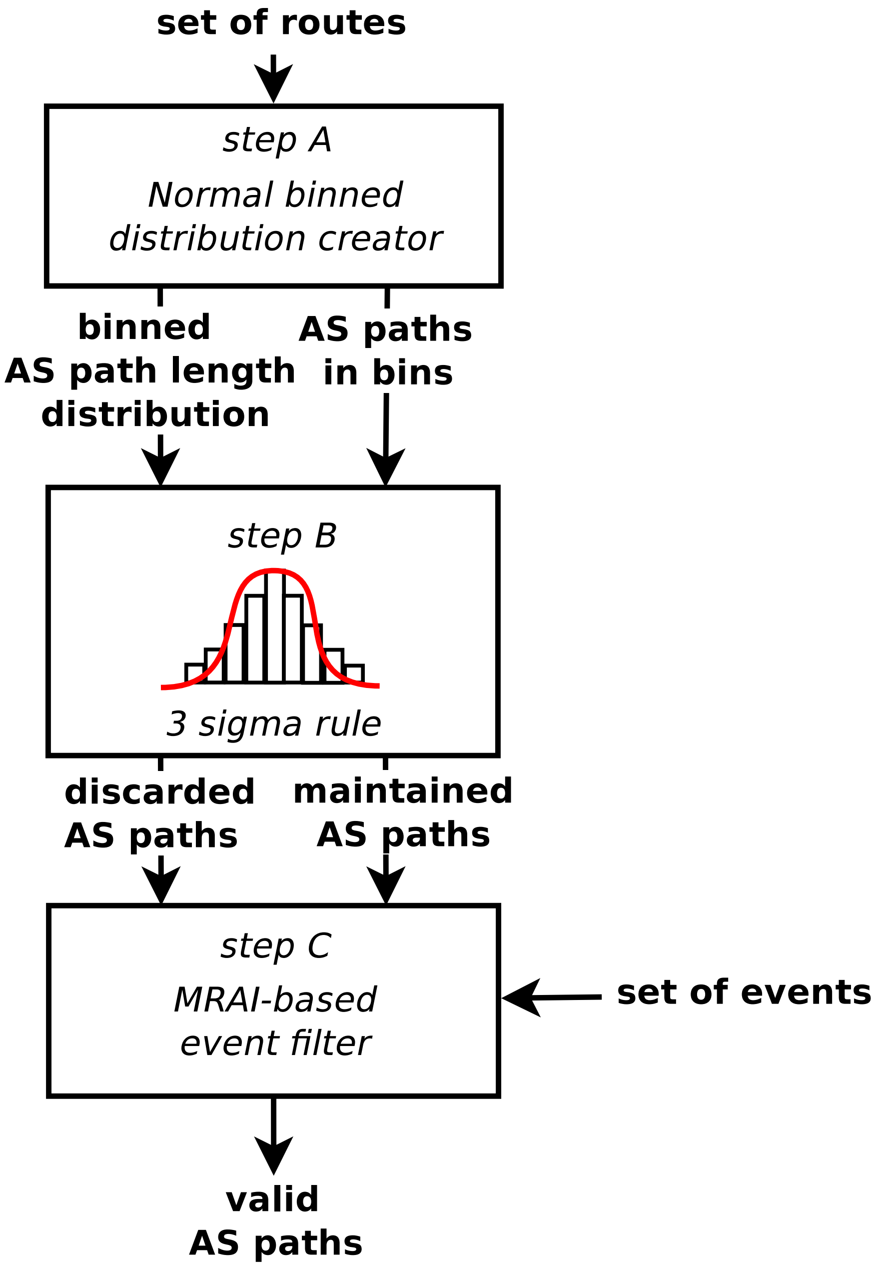


Figure 3.10: Spuriousness-free economic algorithm: Data hygiene phase filters schema

```
Input: feeder f, destination d, routes$_{f,d}$
------------------------------------------------
foreach r in routes$_{f,d}$
    as_path = r.as_path
    as_path = remove_prepending(as_path)
    as_path_length = length(as_path)
    length_as_paths[as_path_length].insert(as_path)
    length_distr[as_path_length] += lifespan(as_path)

left_edge = right_edge = compute_mode(distr)
ranked_length_distr = sort_by_lifespan_desc(length_distr)
for(i = 1; i < size(ranked_length_distr); i++)
    bin_size = right_edge − left_edge +1
    binned_distr$_{f,d}$ = length_in_bins = $\emptyset$
    foreach (length, lifespan) in length_distr
        if (length < left_edge)
            bin = (length−left_edge)/bin_size)
            if ((length−left_edge)%bin_size)
                bin−−
        if (length > right_edge)
            bin = (length−right_edge)/bin_size)
            if ((length−right_edge)%bin_size)
                bin++
        binned_distr$_{f,d}$[bin] += lifespan
        as_paths_in_bins$_{f,d}$[bin].insert(length_as_paths[length])

    1st_quartile = compute_1st_quartile(binned_distr$_{f,d}$)
    3rd_quartile = compute_3rd_quartile(binned_distr$_{f,d}$)
    if(1st_quartile == 3rd_quartile)
        break
    if(ranked_length_distr[i] < left_edge)
        left_edge = ranked_length_distr[i]
    if(ranked_length_distr[i] > right_edge)
        right_edge = ranked_length_distr[i]
    bin_size = right_edge − left_edge


------------------------------------------------
Output: binned_distr$_{f,d}$, as_paths_in_bins$_{f,d}$
```

Figure 3.11: Spuriousness-free economic algorithm: step a) Binned AS path length distribution creation

```
Input: f, d, binned_distr$_{f,d}$, as_paths_in_bins$_{f,d}$
------------------------------------------------------------
$\mu_{f,d}$ = compute_mean(binned_distr$_{f,d}$)
$\sigma_{f,d}$ = compute_std_dev(binned_distr$_{f,d}$)
foreach bin in binned_distr$_{f,d}$
    if bin $> \mu_{f,d} + \sigma_{f,d}$
        $\mathcal{D}_{f,d}$.insert(as_paths_in_bins$_{f,d}$[bin])
    else
        $\mathcal{M}_{f,d}$.insert(as_paths_in_bins$_{f,d}$[bin])
------------------------------------------------------------
Output: discarded AS paths $\mathcal{D}_{f,d}$, mantained AS paths $\mathcal{M}_{f,d}$
```

Figure 3.12: Spuriousness-free economic algorithm: step b) Three-sigma rule filtering

well-approximated normal distribution. To do that, the RIB dynamics related to the pair `<f,p>` are recreated and the *AS path lengths* experienced between the first and the last announcement of `p` are sampled every second, i.e. it is assigned to each length found the total amount of time (in seconds) in which a route with an AS path with that length was found to be active for the pair `<f,p>` during the sampling period (lines 3-8). Then, it is computed the proper size of the bin and the distribution just computed is transformed in a binned distribution centered on the mode of the original distribution, i.e. bin zero contains *at least* the predominant length of the distribution. The size of the bin is determined by `left_edge` and `right_edge`, which are initially set equal to the mode of the distribution (line 10). Then, each length value is assigned to the appropriate bin (lines 15–25) and it is checked if the values of the first quartile and of the third quartile of the binned distribution are equal (line 29), i.e. the mode of the *binned* distribution is predominant on the other bin values and the binned distribution can be considered as normal with a good approximation. Otherwise, the procedure is repeated once again by enlarging the bin size, which will include the next most lasting length by updating accordingly the value of the left or right edge (lines 32–35).

In step *b)* (Fig. 3.12), it is then applied the three-sigma rule [74] on the just created binned distribution, by considering as outlier every AS path included in a bin whose value is larger than $\mu_{f,d} + 3\sigma_{f,d}$, where $\mu_{f,d}$ and $\sigma_{f,d}$ are respectively the temporal mean and the standard deviation of the binned distribution created during step *a*), and are computed as follows:

$$\mu_{f,d} = \frac{\sum_{i=1}^{N_{f,d}} b_i \cdot w_i}{\sum_{i=1}^{N_{f,d}} w_i} \tag{3.1}$$

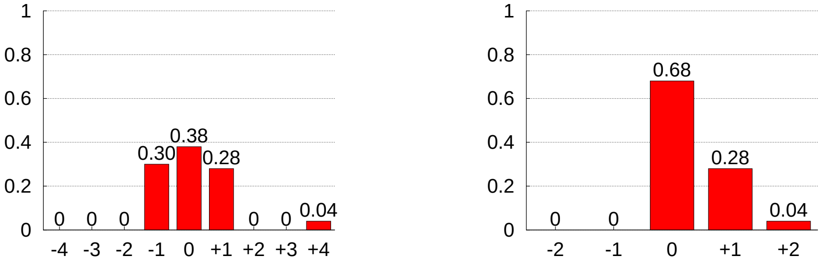
$$\sigma_{f,d} = \sqrt{\frac{\sum_{i=1}^{N_{f,d}} w_i \cdot (b_i - \mu_{f,d})^2}{\sum_{i=1}^{N_{f,d}} w_i}} \tag{3.2}$$

```
**Input:** f, d, routes$_{f,d}$, $\mathcal{D}_{f,d}$, $\mathcal{M}_{f,d}$
------------------------------------------------------------
**foreach** r in routes$_{f,d}$
    **if** (r.as_path $\in \mathcal{D}_{f,d}$)
        curr_r = r
        **while**(**exists** previous route)
            prev_r = *previous_route*(curr_r)
            **if** (*birth*(curr_r) $-$ *birth*birth(prev_r) $\leq$ MRAI)
                to_check.remove(prev_r.as_path)
            **else**
                **break**
            curr_r = prev_r

        $\mathcal{V}.insert$(to_check)
        to_check = $\emptyset$
    **else**
        to_check.**insert**(r.as_path)

$\mathcal{V}$.**insert**(to_check)

------------------------------------------------------------
**Output:** valid AS paths $\mathcal{V}$
```

Figure 3.13: Spuriousness-free economic algorithm: step c) MRAI-based event filtering

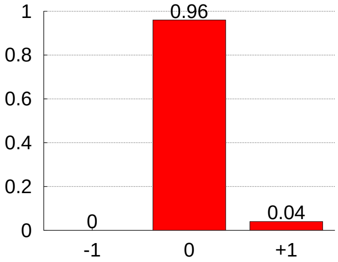
where $N_{f,d}$ is the number of bins and b_i and w_i are respectively the value of the i-th bin and its temporal weight. Note that $\mu_{f,d} + 3\sigma_{f,d}$ is still a threshold, but its value is not arbitrary. Rather, it has both a real and statistical significance, since it depends on the AS paths length that `f` uses to reach `d`. As a result of this filtering step each AS path is classified as *dropped* or *mantained*.

It is possible though that some AS paths generated during a path exploration phenomenon are still not correctly discarded, due to their limited AS path length. The aim of step *c)* (Fig. 3.13) is exactly to identify these remaining routes by checking if any route was announced less than thirty seconds earlier than any spurious route identified in the previous step. Thirty seconds is not an arbitrary value, but instead it reflects a protocol operational standard value, since it is the the default value of the MRAI timer [97] described in Section 3.1, that is typically used to limit the amount of announced routes during BGP transients to improve BGP routing convergence [72, 90, 103]. In this step the algorithm takes thus into consideration each route containing an AS path declared as dropped in the previous step, and checks if any routes has been an-



(a) Bin size = 1

(b) Bin size = 2



(c) Bin size = 3

Figure 3.14: Spuriousness-free economic algorithm: binned distribution creation example related to routes collected for <$\overline{\mathtt{d}}$, $\overline{\mathtt{f}}$>

nounced no more than thirty seconds earlier. In this case, the closest precedent route is flagged as part of the path exploration phenomenon too (line 9) and the control is applied once again starting from it, until no more routes are found in the analysed time interval. Every AS path passing through this last filter is then declared as *valid* and can be used to draw economic inference. Note that network operators typically leave the MRAI parameter at the default value, but some of them applied an even lower value to reduce even more the BGP convergence times [44]. Since the chosen values are typically equal or lower than the default value, this not affect the results of the algorithm.

To better understand the full methodology, consider as example the evolution of the routes related to <$\overline{\mathtt{d}}$, $\overline{\mathtt{f}}$>, where $\overline{\mathtt{d}}$`=2a03:2040::/32` and $\overline{\mathtt{f}}$`=(2001:43f8:1f0::29, AS 6968)` during November 2012. During that month, this pair involved at least a route carrying one inter-PF AS path as described in Section 3.1. Table 3.3 summarizes the related distribution of AS path length and the lifespan of each AS path length found.

<table>
<tr><th>AS path length</th><th>lifespan [s]</th><th>final bin</th><th>status</th></tr>
<tr><td>5</td><td>220,858</td><td>0</td><td>valid</td></tr>
<tr><td>6</td><td>282,158</td><td>0</td><td>valid</td></tr>
<tr><td>7</td><td>211,402</td><td>0</td><td>valid</td></tr>
<tr><td>10</td><td>26,774</td><td>+1</td><td>dropped</td></tr>
<tr><td colspan="4">$\mu_{binned} = 0.036$
$\sigma_{binned} = 0.03$
$\mu_{binned} + 3\sigma_{binned} = 0.09$</td></tr>
</table>

Table 3.3: Spuriousness-free economic algorithm: AS path length distribution of routes related to <$\overline{\texttt{d}}, \overline{\texttt{f}}$>

Firstly, the methodology computes the correct bin value to consider the distribution as normal. The steps of this procedure are depicted in Fig. 3.14. The distribution is initially centered around its mode value (*length* = 6) and analysed with a bin size equal to one. Since the binned distribution is not showing a predominant bin, the bin size is enlarged to two, i.e. the difference between the first (*length* = 6) and the second (*length* = 5) modes plus one. The resulting distribution, even if more peaked than the previous, still does not satisfy the condition of normality required. Thus, the bin size has to be furthermore enlarged (*bin size* = 3) to include also the third mode (*length* = 7). The resulting binned distribution is now extremely peaked, and step *b*) cuts off every AS path included in the final $+1$ bin, including the inter-PF AS path. In this particular example the path exploration events were all temporally distant from announcements of stable paths, thus step *c*) does not affect the results.

3.3.2 Economic inference phase

Now that a methodology able to get rid of a large part of spurious routes is available, it is only required a proper algorithm able to exploit the filtered routes to infer inter-AS economic relationships. Our choice fell on enhancing the common technique described in Section 3.2 [59] and in [87] due to their strong bonds with collected raw data and the low amount of hypotheses and assumptions supporting the heuristic. As previously shown, the original algorithms rely on the valley free property of AS paths [52] and on the a priori knowledge of a set of provider-free ASes. The valley-free property implies that in a given AS path: *a*) there exists at maximum a single p2p connection, *b*) no c2p connections can follow a p2p connection, and *c*) no c2p connections can follow a p2c connection. By considering also the provider-free property of a well-identifiable set of ASes, this also means that *d*) every connection that appears before a provider-free AS can be considered as c2p, and *e*) every connection that appears after a provider-free AS can be considered as p2c. Note that connections among provider-free ASes are considered p2p by definition.

```
Input: set of preliminary tags $\mathcal{P}$
--------------------------------------------------
foreach p in AS path tagged with $\mathcal{P}$
    foreach direct connection [A,B] in p
        if ($\mathcal{P}_{[A,B]}$ == p2p)
            if ($[A,B]$ precedes any c2p $\in \mathcal{P}$ in p)
                if (exists($\mathcal{T}_{[B,A]}$))
                    $\mathcal{T}_{[B,A]}$ = s2s
                else
                    $\mathcal{T}_{[A,B]}$ = ($\mathcal{T}_{[A,B]}$ == s2s) ? s2s : c2p
            if ($[A,B]$ follows any p2c $\in \mathcal{P}$ in p)
                if (exists($\mathcal{T}_{[B,A]}$))
                    $\mathcal{T}_{[B,A]}$ = s2s
                else
                    $\mathcal{T}_{[A,B]}$ = ($\mathcal{T}_{[A,B]}$ == s2s) ? s2s : p2c

foreach $[A,B]$ in $\mathcal{P}$
    if (!exists $\mathcal{T}_{[A,B]}$)
        $\mathcal{T}_{[A,B]}$ = $\mathcal{P}_{[A,B]}$


--------------------------------------------------
Output: set of tags $\mathcal{T}$
```

Figure 3.15: Spuriousness-free economic algorithm: tagging algorithm enhanced step

The main differences between the two algorithms are related to their choices in handling spurious routes. Since the handling of problematic routes are now demanded to the preliminary step, the algorithm proposed is exactly the same of the algorithm described in [87] deprived of the two-day time threshold filtering (or, equivalently, to the algorithm proposed in [59] with $N_{MAG} = \infty$, i.e. with no time threshold) *plus* an enhancement step, that allow to refine the quality of the inferences drawn furthermore.

The enhancement step is described in Fig. 3.15 and consists in exploiting the inferences drawn by the original algorithm to infer additional p2c (or c2p) relationships *also* from those AS paths that do not contain any provider-free AS. In the original algorithm, these AS paths led only to simple p2p connections, that were potentially overwritten by p2c (c2p) relationships whenever it was found a path containing a provider-free AS and the proof that one of the two ASes was providing transit to the other. However, despite the lack of provider-free ASes, these AS paths *may* still contain useful information to infer further transit relationships. In particular, there may appear connections inferred to be p2c (or c2p) because they were found in at least another AS

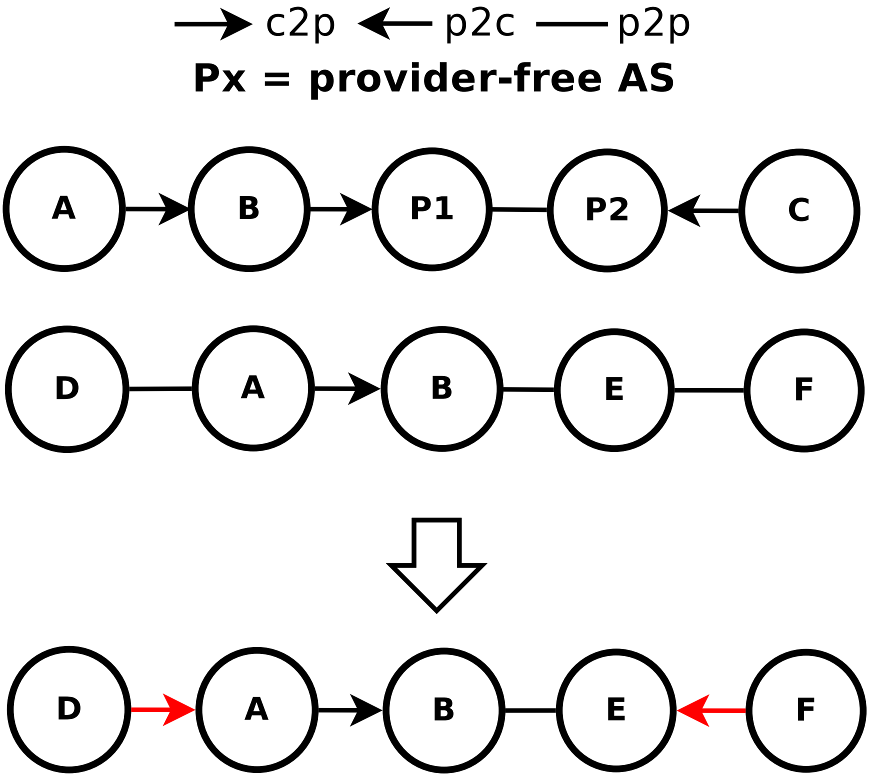


Figure 3.16: Spuriousness-free economic algorithm: enhanced tagging step application example

path containing a provider-free AS. This information is a further proof to infer p2c relationships, thus it can be exploited to force these AS paths to be compliant to the valley-free property, by turning some p2p relationships present in AS paths without provider-free ASes into c2p (or p2c) relationships exploiting the same rationale of the original algorithm. More in detail, if a set of p2p connections precedes a c2p connection, each of them is converted in c2p. Likewise, if any p2p connection follows a p2c connection, each of them is converted in p2c. In both cases, it must be taken into account the presence of other inferences made on different AS paths, in order to spot the presence of a s2s connection.

The main rationale behind this algorithm is that a p2c label is a tangible proof that an AS announced to the other some network received from one of its peers or providers, while a p2p label only proves that the ASes exchanged routing information related to each other customers. Thus, if multiple different tags are found for the same

	p2c	p2p	s2s	Total
with spuriousness	107,900 (63.46%)	58,925 (34.65%)	3,192 (1.89 %)	170,017
without spuriousness	94,639 (57.72%)	67,727 (41.30%)	1,591 (0.87%)	163,957

Table 3.4: Spuriousness-free economic algorithm: impact of spuriousness on tagging results

connection, a p2p label is overruled by a p2c label, while contrasting p2c and c2p labels lead to a s2s label, i.e. each AS is provider of the other. For example, consider the situation depicted in Fig. 3.16. Despite the AS path `[D, A, B, E, F]` does not contain any provider-free AS, it is still possible to infer that `D` is customer of `A` and that `F` is customer of `E` since, from the AS path `[A, B, P1, P2, C]`, there is proof that `A` and `E` announced them routing information related to their providers.

3.3.3 Results

It can be found that 40.44% of the total number of different AS paths collected during January 2013 appears only in spurious routes. In particular, the data hygiene phase was able to get rid of the 90.78% of inter-PF routes that lead existing tagging algorithm to unrealistic s2s relationships. In Table 3.4 it is shown a comparison of the results of the proposed tagging algorithm by using every AS path available and by using only the AS paths purged by spurious routes. Despite the large number of AS paths discarded, their effects on economic inferences is rather limited. This can be explained by the fact that only a limited set of ASes has a misconfiguration on ASBRs and contribute in creating no-valley-free AS paths. Moreover, during the last years several techniques were proposed to reduce the propagation of spurious routes [69] in order to limit the BGP traffic volume due to excessive UPDATE messages, lowering also the number of spurious routes generated. Nevertheless, the percentage of tag changed is not negligible (9.04%), and cannot be ignored by any application which relies on this data. It must also be noted that 6,063 connections out of 170,017 – i.e. the 3.56% out of the total – have been found to appear only in BGP spurious routes, and due to their ephemeral nature are not tagged by our methodology.

Particularly interesting is also the quantification of the quality of the inferences made thanks to this new approach. Given the absence of a proper ground truth, the quantification of the quality of the inferences can be performed by computing the amount of different AS paths that concur to infer every single inter-AS relationships. The distribution of this value is depicted in Fig. 3.17. Not surprisingly, the tags that were inferred from a larger number of AS paths – and thus that are more reliable – are related to p2c and s2s relationships. On the other hand, only a small number of

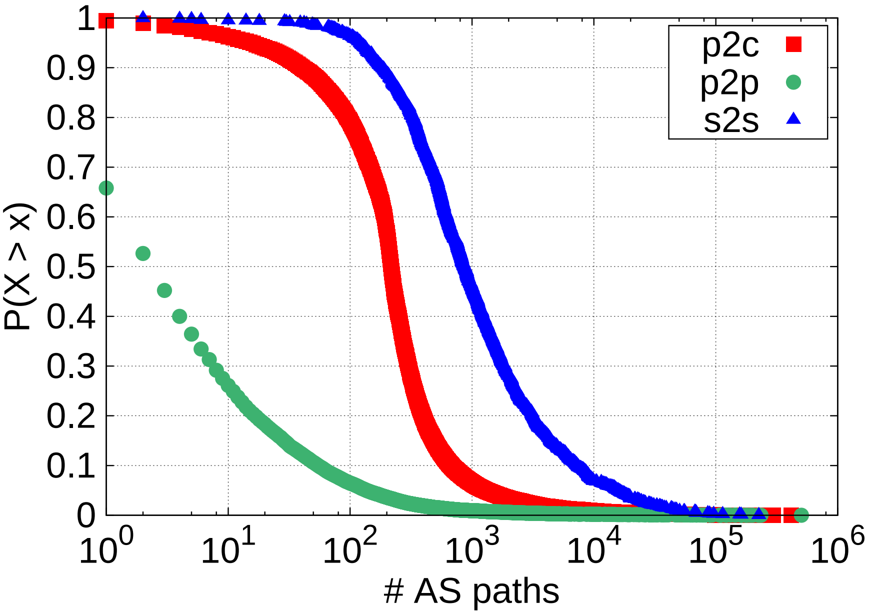


Figure 3.17: Spuriousness-free economic algorithm: CCDF of the number of valid paths involving each tag

p2p connections were inferred by a reasonable number of AS paths. In particular, it can be seen that about the 75% of p2p connections have been found in less than 10 different AS paths, as opposite to the 95% of p2c connections and the 99% of s2s connections. This is likely caused by the well-known incompleteness of BGP data repositories. In [61] has been shown that a large part of the BGP route collectors currently deployed are connected to ASes part of the top of the Internet hierarchy. This means that the current BGP infrastructure is unable to discover a large amount of p2p connections, and that a large number of p2p connections inferred are likely to be p2c connections failed to be correctly inferred due to the lack of AS paths proving that one of the ASes transit traffic for the other. On the other hand, it must be noted as well that both the p2c and s2s connections can be considered to be reliable, since it is found at least a tangible proof of their existence in spuriousness-free BGP data. The quality of the inferences will improve however in presence of a more complete dataset.

		p2c	**p2p**	**s2s**
Algorithm Sect. 3.2	$N_{MAG} = 1$	89,138	89,258	1,375
	$N_{MAG} = 2$	89,898	78,509	1,604
	$N_{MAG} = 3$	90,865	77,299	1,847
	$N_{MAG} = 4$	92,441	75,356	2,214
	$N_{MAG} = 5$	95,395	71,568	3,048
	$N_{MAG} = 6$	95,685	71,182	3,144
	$N_{MAG} = 7$	95,791	71,053	3,167
	$N_{MAG} = \infty$	95,958	70,861	3,192
Oliveira et al. [87]		87,406	67,304	1,361
Algorithm Sect. 3.3		94,639	67,847	1,361

Table 3.5: Comparison among the results of economic tagging algorithms

3.4 Economic algorithm comparisons

In Table 3.5 are shown the comparisons between the proposed algorithms and the algorithm proposed in [87], that relies on thresholds to get rid of spurious entries. As can be seen, the enhancement step introduced in the algorithm described in Section 3.3 allows to infer an amount of p2c connections that is typically larger than other cases, still maintaining the spuriousness-free characteristic required. It also must be noted that an arbitrary time threshold is not a good solution. As can be seen from Table 3.5, the algorithm originally proposed by in [87] lacks of a large amount of p2c connections inferred by our algorithm exclusively due to their choice to apply a two-day threshold on input routes. This happens also in the results obtained with the most conservative N_{MAG} values[6] of the algorithm proposed in Section 3.2. On the other hand, by using less conservative N_{MAG} values, some of the wrongly inferred p2c connections rise up, but the number of wrongly inferred s2s relationships increase as well, due to the presence of a large number of short-living no-valley-free AS paths in raw data. In particular, the larger the value of N_{MAG}, the larger is the number of inter-PF AS paths that participate in the s2s generation as well.

---

[6] The smallest is the value of N_{MAG}, the narrowest is the transient filter applied by the algorithm, and the smallest are the effects of transient routes on the final tagging. Small values of N_{MAG} are thus considered as conservative cases.

4

BGP and geography

The Internet is a complex system that evolved over the last few decades from a small network confined to the U.S. (i.e. ARPANET, 1969) to the current worldwide network of networks. It now consists of thousands and thousands of networks, under the administrative control of more than 40,000 ASes. This pervasive evolution did not occur homogeneously around the world for obvious historical, economic and political reasons. This different level of pervasiveness may lead the same organization to establish economic relationships in different regions with different criteria, but most research in the Internet topology analysis have considered ASes as homogeneous entities, each with a global set of metrics and characteristics, regardless of their heterogeneity.

BGP data do not contain any straightforward geographic information. The well-known mandatory `AS_PATH` attribute contains only the sequence of ASes that the traffic crosses to reach the announced subnet, but no information on the geographic regions where the traffic actually flows. The aim of this chapter is to identify a methodology able to infer geographic information from AS paths and to analyze local properties of the Internet, with a special focus on AS topology properties and economic relationships. The methodology proposed consists in inferring the geolocation of each AS by exploiting the possibility to geolocate its prefixes thanks to the availability of a geographic IP database [6]. These data are then exploited to produce a set of geographically tagged AS paths in which each connection is geolocated. This set is used to derive regional AS-level topologies that are analyzed both from a statistical and economic perspective. In particular, it is provided a methodology to adapt the algorithms proposed in Chapter 3 to deal with geographic AS paths.

4.1 AS Geolocation

Knowledge regarding the geographic range of an AS is one of the fundamental parameters for decisions concerning the establishment of a settlement-free peering or a transit type of relationship between ASes. Several Tier-1 ASes – identified as provider-free in the previous chapter – include in their peering requirements at least one geographic constraint for candidate peers that need to be fulfilled. Just to name a few, AT&T (`AS7018`)[1] requires a list of the countries served by the candidate peer in the peering request submission; Verizon (`AS701`)[2] requires a minimum number of served countries in the region where the peering is requested and that candidates own a "geographically-dispersed network"; and TeliaSonera (`AS1299`)[3] requires that the candidate peer is present and able to exchange traffic and to be interconnected in a minimum number of cities in two out of three regions (Europe, North America and/or Asia Pacific/Oceania). To geolocate each AS it can be exploited its formal definition:

> "*an AS is a connected group of one or more IP prefixes run by one or more network operators which has a single and clearly defined routing policy. [63]*"

Given this definition, it is straightforward that an AS is geolocated if its own prefixes are geolocated. The list of all the active prefixes of a given AS can be collected by parsing the BGP raw data provided by route collector projects. Each prefix can be geolocated in turn by geolocating each IP prefix inside it, using one of the IP geolocation databases available [104]. Consider a generic route `x.y.z.0/24-[A B C D]`. It is possible to claim that the last element of the AS path[4] owns at least a network – and thus it may or may not be present – in the region(s) where the prefix is geolocated. This approach is correct for any given geographic region (e.g. countries, continents) iff the granularity of the geolocation tool is fine enough and iff the route does not carry the `AGGREGATOR` and the `ATOMIC_AGGREGATE` attribute. The `AGGREGATOR` is an optional transitive attribute and the `ATOMIC_AGGREGATE` is a well-known discretionary attribute of the BGP protocol and may be included in UPDATE messages by a BGP speaker which performs route aggregation. If one of these attributes is present, it is possible that part of the real AS path is missing, hidden by the aggregating router. In this case, it is not possible to state that the considered prefix belongs to the last element of the AS path, but additional confirmation is needed from the `WHOIS` service provided by the Internet Routing Registries (IRRs)[5]: the prefix is considered to belong to the last AS of the AS path if that AS is the owner of the announced prefix also according to the `WHOIS` response. For example consider the route shown in Figure 4.1 – which is the textual representation in MRT data of a route collected by the route collector `rrc12` of RIPE RIS – whose prefix is entirely geolocated in Europe. Given

---

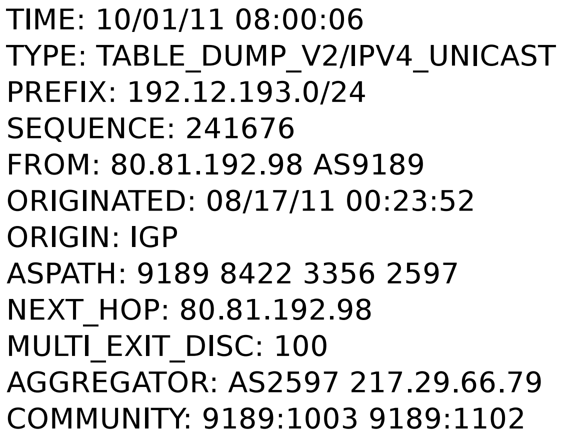
[1] http://www.corp.att.com/peering/

[2] http://www.verizonbusiness.com/terms/peering/

[3] http://www.teliasoneraic.com/Ourservices/IP/IPTransit/index.htm

[4] The last AS of an AS path is the AS that has originated the BGP UPDATE.

[5] A list of available `WHOIS` locations can be found at `http://www.irr.net/`



```
TIME: 10/01/11 08:00:06
TYPE: TABLE_DUMP_V2/IPV4_UNICAST
PREFIX: 192.12.193.0/24
SEQUENCE: 241676
FROM: 80.81.192.98 AS9189
ORIGINATED: 08/17/11 00:23:52
ORIGIN: IGP
ASPATH: 9189 8422 3356 2597
NEXT_HOP: 80.81.192.98
MULTI_EXIT_DISC: 100
AGGREGATOR: AS2597 217.29.66.79
COMMUNITY: 9189:1003 9189:1102
```

Figure 4.1: Textual representation of a route in MRT format

the presence of the `AGGREGATOR` attribute, it is required an additional query to the `WHOIS` service. Since the response state that the prefix belongs to AS 2597, it can be concluded that `AS2597` is present in Europe.

Results presented in this thesis are obtained using the Maxmind GeoIPLite database [6] and the following regional division: 1) Africa, 2) Asia-Pacific (Asia and Oceania), 3) Europe, 4) Latin America (the Caribbean, Central America, Mexico and South America) and 5) North America (Bermuda, Canada, Greenland, Saint Pierre and Miquelon, USA). In the national level analysis, for sake of readiness, are considered only those countries that are currently participating in the G8 forum, i.e. U.S.A. (US), Japan (JP), Germany (DE), France (FR), United Kingdom (UK), Italy (IT), Canada (CA) and Russia (RU). Note that information related to the missing countries can be found in [4].

4.2 Introduction of geography in BGP data

Geolocation of ASes by itself is not enough to extract geographic information from AS paths. An AS can have a geographic range that spread across multiple regions, thus it is not possible to infer where each AS connection forming an AS path is located. To overcome this problem, in this work is proposed a three-step algorithm which, based on the geolocation of each AS, is able to geolocate each AS connection of the AS paths.

a) Enhanced routes from BGP raw data

In this step is obtained an *enhanced route* – defined as the triplet `{SOURCE, ASPATH, DESTINATION}` – for each route available in the BGP data. `SOURCE` is the region where the ASBR that announced that route to the route collector is located and can be obtained by geolocating its IP address (`FROM` field in Fig. 4.1). `ASPATH` is the content of the homonym BGP attribute cleaned of private/reserved/unallocated ASNs and of `AS23456`, as described in Chapter 2. `DESTINATION` is the region where the prefix announced is located. Since a prefix could be geolocated in more than one geographic region, more than one enhanced route could be created from a single route, one for each region where the destination is found to be located. Consider again the route reported in Fig. 4.1. Both the IP address of the BGP speaker (`80.81.192.98`) and the prefix announced (`192.12.193.0/24`) are located in Europe, thus we obtain the enhanced route `{EU, [9189 8422 3356 2597], EU}`.

b) Detection of Single Region Located Transit Points (SRLTPs) in each enhanced route

In this step the set of SRLTPs are extracted from each enhanced route. This set contains regional intermediate points where the traffic needs to flow. The `SOURCE` and the `DESTINATION` of each enhanced route are by definition part of this set, since they are both geolocated in a single region. This set also includes two classes of ASes that can be found in the `AS_PATH`. The first class of ASes that fits in this definition is represented by ASes that own prefixes only located in a single region, i.e. ASes that do not own an inter-regional infrastructure. Another class of ASes that fits this definition are those ASes that have a single region in common with a neighboring AS. The basic idea is that typically ASes follow a *regional principle* to route their traffic. Inter-regional ASes tend to subdivide their ASes into different areas by exploiting the features of IGP protocols such as OSPF and IS-IS in order to maintain traffic as regional as possible to maximize the performances. Thus, an inter-regional AS does not exploit its inter-regional infrastructure when it is not needed, by representing a SRLTP under certain circumstances. For example consider the enhanced route extracted from the route shown in Fig. 4.1. Geolocating each AS using the methodology described in the previous section, we obtain that `AS9189`, `AS8422` and `AS2597` are located only in Europe, while `AS3356` is located in every continent. Since also `DESTINATION` is located in Europe, each AS represent a SRLTP.

c) Geographic AS paths

In this step the information just extracted are exploited to geolocate the connections of the AS path of each enhanced route. Following the same regional principle introduced above, inter-regional ASes are typically interconnected on every location where they are co-located, trying to maintain the inter-AS traffic as regional as possible. This means that, for example, if the traffic flows from a source to a destination both located

```
 1 for each enhanced route R
 2   region = SOURCE;
 3   for(i=0; i < length(ASPATH); i++)
 4     if (AS$_i$ $\in$ SRLTP && region $\notin$ regions(AS$_i$))
 5       region = regions(AS$_i$);
 6       for(j = i; j > 0; j−−)
 7         if (region $\in$ regions(AS$_j$))
 8           add(AS$_{j-1}$, AS$_j$) to GEO_PATH(Region);
 9         else
10           break;
11     elsif (region $\in$ regions(AS$_i$))
12         add(AS$_i$, AS$_{i-1}$) to GEO_PATH(Region);
13     else
14       i = index of next SRLTP;
15       region = regions(AS$_i$);
16       for(j = i; j > 0; j−−)
17         if (region $\in$ regions(AS$_j$))
18           add(AS$_{j-1}$, AS$_j$) to GEO_PATH(Region);
19         else
20           break;
21   if (region $\neq$DESTINATION)
22     region =DESTINATION;
23     for(j = length(ASPATH); j > 0; j−−)
24       if (region $\in$ regions(AS$_j$))
25         add(AS$_{j-1}$, AS$_j$) to GEO_PATH(Region);
26       else
27         break;
```

Figure 4.2: Geographic tagging algorithm

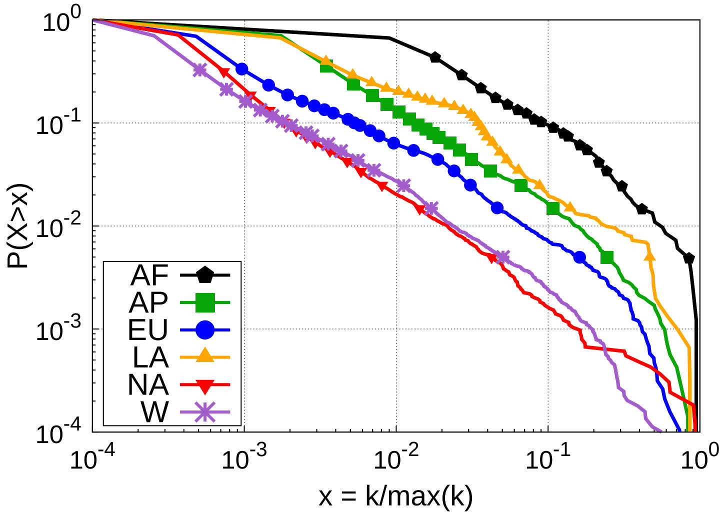
in region R through ASes located in R, the traffic is very likely to remain confined in R even if these ASes are co-located in other regions. By exploiting these considerations and the SRLTPs identified in the previous step it is possible to complete the geographic tagging algorithm, that is presented in Fig. 4.2. The algorithm aims to create a set of geolocated AS connections from each enhanced route. Each enhanced route together with its set of geolocated connections is referred in this work as a *Geographic AS path*. To achieve this, it is required the analysis of each AS in `AS_PATH` and check whether the connection with its neighboring ASes can be established in the considered region. The procedure starts by considering `SOURCE` as the initial region (line 2), and every AS connection that may belong to the considered region added to the set of AS connections located in that region (line 8). The considered region will changed if an SRLTP (line 4) or if a multi-regional AS (line 13) not located in that region is found. In this latter case the change is preceded by a jump to the next SRLTP (line 14). The output of the geographic tagging algorithm is composed by the set of Geographic AS paths, each inferred from the related enhanced routes. Considering the route in Fig. 4.1 and its characteristics shown so far, it is possible to infer that the full AS path is located in Europe.

4.3 Undirected Graph Analysis

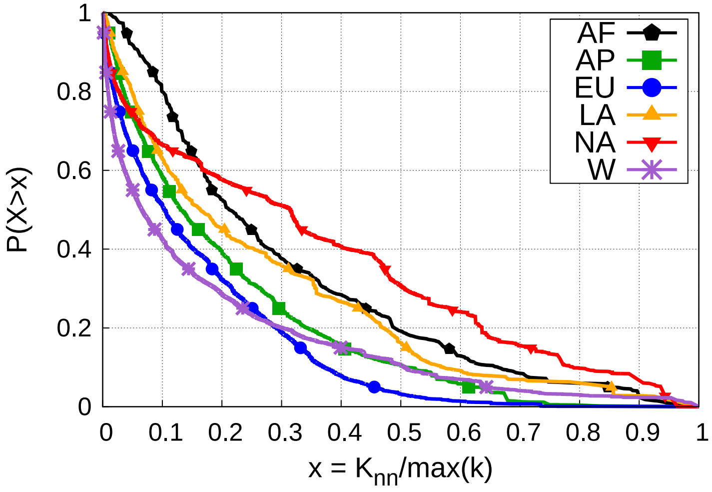
Thanks to the algorithm described it has been possible to extract from each geographic AS path the geolocated connections to create continental and national AS-level topologies, as well as a global AS-level topology composed by every connection found in AS paths, independently from any geographic information. The result is that 43,686 out of 44,389 ASes appear in at least one regional topology. The missing ASes are in 558 out of 703 cases due to missing IP prefixes in the Maxmind database. In the remaining cases the ASes, although being geolocated, do not appear in any regional topology because they do not share any region with the neighboring ASes in any AS path in which they appear. This may be due to the use of BGP multihop sessions – where ASes are actually located in different regions – or due to the mistaken/partial geolocation of prefix(es) by the Maxmind database. In both cases the geographic tagging algorithm is not able to infer where the connection is geolocated. For the same reason, the geographic tagging algorithm is not able to assign a region to 7,860 connection out of 170,204.

The first important result is that the regional topologies extracted overlap only slightly, as highlighted by the Jaccard similarity indices computed between pairs of continents for nodes (J_{nodes}) and connections (J_{edges}) and reported in Table 4.1. In detail, the Jaccard similarity index J_i is an index that allows to quantify the similarity between pair of sets and is defined as $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$. This poor overlap is confirmed by the fact that only the 4.63 % of ASes are located in more than one continent and only the 1.30% in more than two, and indicates that the Internet can be subdivided in five independent macro-regions with their own peculiar properties (Fig. 4.3a and Fig. 4.3b). This poor overlap also means that the regional principle has been applied by the algorithm only on a small set of connections and thus the largest part of connections is correctly geolocated at the continental level.

The situation changes slightly if the Internet is analyzed at a national level. Despite only 7.99% of ASes are located in more than one country and only 3.2% in more than two, the national topologies extracted by this methodology do not show the same degree of independence found in the continental scenario. In particular, the smaller is a national topology – in terms of nodes – the lower is the probability to find in that topology ASes whose networks do not cross national borders (Fig. 4.4). This is quite



(a) CCDF of the Node degree k



(b) CCDF of the Average neighbor degree k_{nn}

Figure 4.3: Geographic node properties per continent

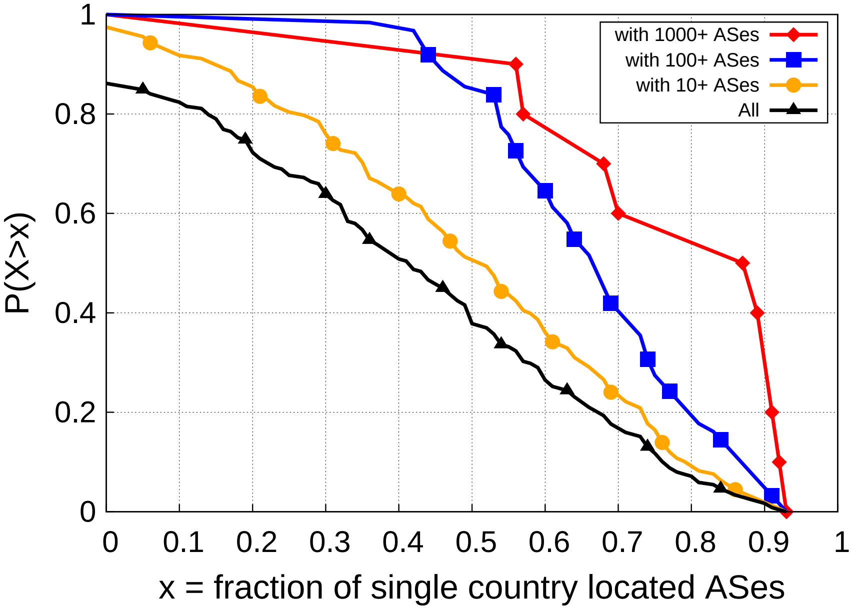


Figure 4.4: CCDF of the number of single country located ASes

intuitive, since the more a national topology is populated, the higher is the probability that the related country is economically and technologically advanced and home of a large number of local-scoped organizations with a network infrastructure sufficiently developed to meet the AS number assignment criteria [63]. On the opposite, less developed countries are likely to demand their Internet reachability requirements to the few inter-regional ISPs, increasing thus the degree of overlapping with other country topologies.

	AF	**AP**	**EU**	**LA**	**NA**
AF	–,–	0.016,0.023	0.009,0.008	0.026,0.025	0.010,0.014
AP	0.016,0.023	–,–	0.022,0.032	0.019,0.029	0.036,0.078
EU	0.009,0.008	0.022,0.032	–,–	0.012,0.018	0.022,0.030
LA	0.026,0.025	0.019,0.029	0.012,0.018	–,–	0.022,0.030
NA	0.010,0.014	0.036,0.078	0.031,0.055	0.022,0.030	–,–

Table 4.1: Geographic Jaccard similarities indices $J = (J_{nodes}, J_{edges})$ between continents

A particular scenario is depicted by countries that are representing critical communication hubs for the Internet and strategic points for IP transit market, such as United Kingdom and Germany, that are currently hosting LINX[6] and DE-CIX[7], two of the most populated IXPs of the entire world. The topologies related to these countries are indeed showing a large number of ASes located in more than a single country as well as a large average node degree value (Table 4.3), reflecting the dense interconnectivity tendency of those ecosystems. Further evidence regarding the differences between regional topologies is provided by the properties of the regional graphs summarized in Table 4.2 and Table 4.3. The sizes of the topologies differ greatly in terms of nodes, reflecting the different degrees of economic and technological development of the regions.

It is particularly interesting to compare North American and European topologies, which have a quite similar number of nodes but differ significantly in terms of edges. The CCDF of the node degree shows that the European region is more densely connected than the North American region where, on the other hand, there are ASes with a larger degree and where the number of ASes with a small degree is higher. This suggests quite a hierarchical structure in North America versus a flatter structure in Europe. This is confirmed by the CCDF of the normalized Average Neighbor Degree, which shows that in Europe, ASes tend to connect to ASes with a similar degree, while in North America they tend to connect to ASes with a very large degree. The differences between these ecosystems reflect the Internet's historical evolution in the respective regions [85]. In North America, especially in the U.S., a small set of large ISPs (e.g. AT&T, Centurylink and Verizon Communications) provide connectivity to all the states. In Europe on the other hand, each country is typically characterized by the presence of a national telco (e.g. Deutsche Telekom and Telecom Italia) which usually own a large part of the national Internet infrastructure, and by the presence of at least one IXP that encouraged the establishment of settlement-free peering connections among local ISPs. More details on the role of IXPs in the development of

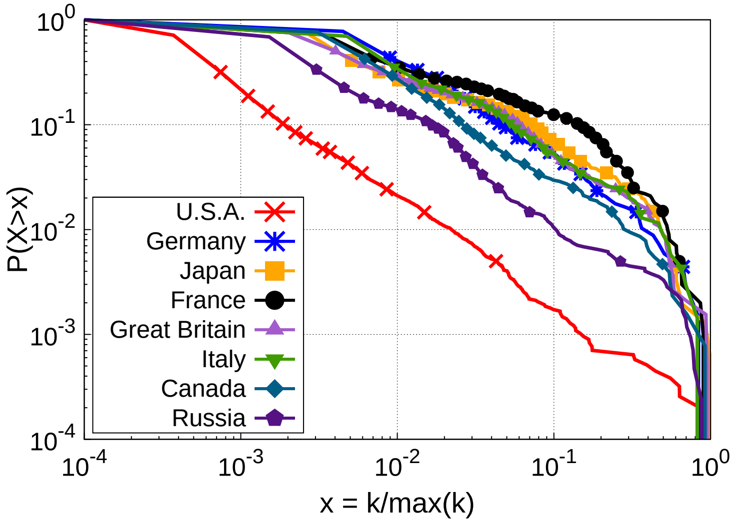
Continent	Nodes	Edges	Average degree	Max degree	Single continent located ASes
AF	824	2,189	5.31	111	599 (72.69 %)
AP	7,064	20,931	5.93	574	6,084 (86.12 %)
EU	19,109	94,824	9.92	2,076	17,828 (93.29 %)
LA	3,025	13,686	9.05	580	2,547 (84.19 %)
NA	16,408	46,491	5.67	2,731	14,738 (89.82 %)
World	44,389	170,204	7.67	3,925	– (– –)

Table 4.2: Geographic topology statistics per continent

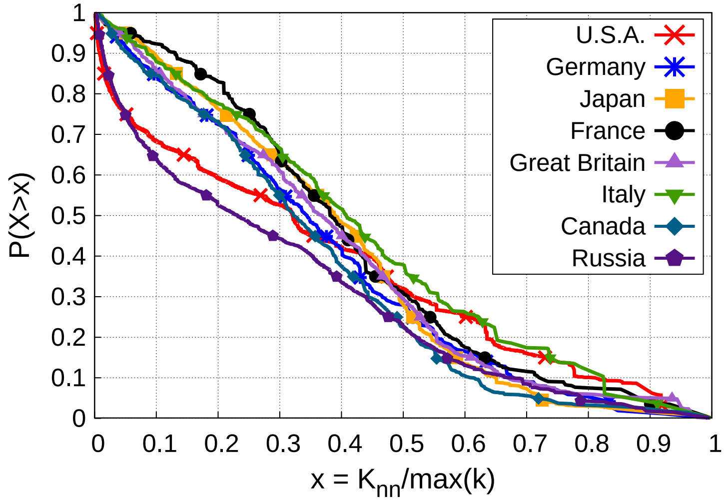
---

[6] `http://www.linx.net`

[7] `http://www.de-cix.net`



(a) CCDF of the Node degree k



(b) CCDF of the Average neighbor degree k_{nn}

Figure 4.5: Geographic node properties per G8 country

Country	Nodes	Edges	Average degree	Max degree	Single country located ASes
US	15,623	44,554	5.70	2,693	13,736 (87.92 %)
RU	4,236	13,856	6.54	658	4,045 (95.49 %)
UK	1,935	13,155	13.60	496	1,127 (58.24 %)
DE	1,546	9,548	12.35	393	1,125 (72.76 %)
CA	1,282	4,112	6.41	324	760 (59.28 %)
FR	1,003	6,635	13.23	291	679 (67.69 %)
IT	695	2,263	6.51	210	553 (79.56 %)
JP	680	2,237	6.58	223	515 (75.53 %)

Table 4.3: Geographic topology statistics per G8 country

the Internet can be found in [15, 18, 25, 58] and [115]. In particular, the impact of national telcos can be ascertained by comparing the European normalized k_{nn} curve with the normalized k_{nn} curves related to the European countries in the G8 forum (Figure 4.5b). In the former, ASes seem to prefer to connect with ASes with a small-medium degree value rather than with the ASes with the largest degree. In the latter, it is possible to see that ASes are likely to connect to ASes with a degree value close to the maximum value in each country. Since the number of large degree ASes in each country is rather small (Fig. 4.5a), it is possible to conclude that there are ASes that are extremely important in the national connectivity, while they have not the same importance in the continental scenario.

4.4 Geography and inter-AS business relationships

The analysis of the undirected graph of the Internet highlighted just part of the extreme complexity of the Internet ecosystem, since it completely lacks the real business model of each element. The Internet consists of a network managed by thousand of different organizations. Some of these organizations, e.g. small or large ISPs, live on the sale of Internet transit to other organizations, others, such as CDNs and search engines, just aim to offer content to end users. Most organizations, though, just care about Internet connectivity. In order to highlight that heterogeneity and to get a better insight into the Internet, economic algorithms are required (see Chapter 3). Although the proposed approaches were devised for the global Internet, they are still valid in a regional analysis, since the Internet is nothing more than the sum of the regional economic ecosystems that it is made up of. It is now presented a methodology to infer the regional economic relationships between ASes, by showing how to adapt the algorithm proposed in Section 3.2.

```
foreach (Region R)
  foreach (Geographic AS path G)
    foreach connection [A,B]
      if (T$_1$ $\in T_{list}$ follows [A,B] in the ASPATH)
        Tag[A,B] = c2p
      if (T$_1$ $\in T_{list}$ precedes [A,B] in the ASPATH)
        Tag[A,B] = p2c
      if (does not exist any T$_1$ $\in T_{list}$ )
        Tag[A,B] = p2p
      if (R $\notin$ Regions[A,B])
        Tag.remove([A,B])
```

Figure 4.6: Geographic time-aware economic algorithm: step a) enhanced

The original algorithm requires a list of AS paths as input together with their maximum lifespan, defined as the time interval during which there is at least one active route in a RC that includes that AS path in its attributes. It consists of three steps. In the first step all the possible tags for each connection `(A,B)` are computed by applying the knowledge of the list of Tier-1 ASes to infer economic relationships in each of the AS connection of the path. In the second step, all the tags inferred for each connection `(A,B)` are merged to obtain a single economic relationship for each connection. Finally, in the last step, the economic relationship obtained in the second step for the specular connections `(A,B)` and `(B,A)` are merged to infer a single economic relationship between AS `A` and AS `B`. It should be noted that both the merging phases are based on the parameter N_{MAG}: economic tags with lifespans that differ by more than N_{MAG} orders of magnitude are not merged together in order to avoid transient – and potentially wrong [78] – AS paths from distorting the results. The lower the N_{MAG} value, the lower the probability that transient information will affect the results. On the other hand, a low N_{MAG} value also reduces the number of two-way validated[8] economic relationships.

The input of the algorithm can no longer be an AS path together with its lifespan. An AS path may be gathered from multiple BGP routers located in different parts of the world and may be used to reach different locations. In this case, some of the AS connections that make up the AS path are likely to refer to different physical links and the lifespan of each AS path may differ depending on the destination region. For example, the AS Path `A B C` used to reach subnets in Asia may be affected by the

---

[8] An economic relationship between AS `A` and `B` is considered to be *two-way validated* if an economic tag for the connection `(A,B)` and for the reverse connection `(B,A)` has been found.

failure in the link that interconnects `B` and `C` in Tokyo, while the same AS Path used to reach subnets in Europe may not, since `B` and `C` are also interconnected in Paris. To overcome this problem, it is exploited the concept of geographic AS path just introduced which, together with its lifespan, represents the enhanced algorithm input. The lifespan of a geographic AS path is computed similarly to the lifespan of an AS path, i.e. it is the maximum period of time in which there is at least one active route that includes the related `ASPATH` that is announced from a router in `SOURCE` and reaches at least one subnet in `DESTINATION`.

To deal with the new input, it is enhanced part of the first step of the original algorithm in order to obtain only information concerning a specific region R, as shown in Figure 4.6. The economic tagging of each AS connection that make up the considered Geographic AS Path is initially performed using exactly the same methodology proposed in Section 3.2 (line 4–9) exploiting the presence of provider-free ASes in the AS path. Note that the initial tagging phase needs to be performed on the full AS paths, since the presence of a provider-free AS need to affect every connection of the AS path, irrespectively of any geographic information. Once the classic tagging phase is completed, it is introduced an additional discarding phase (line 10), in which all the economic tags related to AS connections not geolocated in R are removed from the partial results. As a consequence of the new version of the first step, the remaining steps of the original algorithm are fed only with economic tags related to AS connections established in R and, thus, the resulting economic tagged topology is related to R.

4.5 Economic analysis

The application of the enhanced economic tagging algorithm to the sets of geographic AS paths allows deeper insights of each regional ecosystem which reveal the real nature of the regional differences that were only deduced from the undirected analysis of the Internet. Table 4.4 and Table 4.6 show the results obtained by applying the enhanced economic algorithm with the most conservative scenario – i.e. $N_{MAG} = 1$ – listing the number of economic tags inferred for each region. The following inferences still hold even for other N_{MAG} values, whose results can be found in [4].

The most relevant characteristic highlighted by the economic analysis is the large proliferation of potential p2p connections in the European ecosystem, representing the 58.63% of the total. This feature is in contrast with the peering behaviors of other regions, where the amount of p2c connections is larger (around 70% of the total) than the amount of p2p connections. Together with the conclusions drawn about the geographic undirected graphs, this allows to understand the real nature of the flat structure of the European Internet ecosystem. This joint analysis shows that Europe

Continent	**p2c**	**p2p**	**s2s**
AF	1,417 (64.73 %)	751 (34.30 %)	21 (0.97 %)
AP	15,031 (71.87 %)	5,681 (27.16 %)	202 (0.97 %)
EU	38,815 (40.94 %)	55,594 (58.63 %)	397 (0.41 %)
LA	5,999 (43.88 %)	7,575 (55.41 %)	95 (0.69 %)
NA	34,030 (73.24 %)	12,114 (26.07 %)	318 (0.68%)
World	89,138 (52.43 %)	79,498 (46.76 %)	1,375 (0.81 %)

Table 4.4: Geographic economic tag distribution (continental level)

is rich of small/medium transit providers that, in addition to offering transit to end-users and stub ASes, tend to establish settlement-free p2p connections among them. The proliferation of these small/medium providers is also the reason for the development in Europe of largely-populated IXPs which in turn facilitated the establishment of settlement-free relationships among ASes, helping to create the large amount of p2p connections just described. This is confirmed by the large percentage of p2p connections found in Germany and Great Britain, that are respectively hosting the DE-CIX and the LINX.

Another interesting analysis is to investigate how the economic relationships change when the focus is moved from the worldwide scenario to narrower geographic scopes. Inter-regional providers may indeed decide to establish regional p2p connections – exchanging only routes of regional customers – in those regions where their pervasiveness is similar, while they may decide to establish a p2c agreement elsewhere. This is especially true for large transit ASes. In Table 4.5 are summarized the tag changes with respect to the World scenario, focusing especially on the change from global transit (p2c, c2p, s2s) to continental peering (p2p). Although the number of these connections may look not relevant at a first glance, it should be considered that these tags are referred to AS connections that compose the core of the region. This is highlighted by the large number of tag changes that involve only non-stub ASes

Continent	**Tag changes**	**Transit to peering**	**Among multi-regional ASes**	**Among non-stub ASes**
AF	184 (8.58 %)	155 (7.22 %)	157 (7.32 %)	154 (7.18 %)
AP	580 (2.78 %)	396 (1.90 %)	394 (1.89 %)	472 (2.26 %)
EU	1,400 (1.47 %)	893 (0.94 %)	391 (0.41 %)	1,057 (1.11 %)
LA	435 (3.20 %)	335 (2.46 %)	156 (1.14 %)	385 (2.83 %)
NA	620 (1.33 %)	457 (0.98 %)	426 (0.91 %)	514 (1.10 %)

Table 4.5: Economic relationship changes (continental level)

Country	P2C	P2P	S2S
US	32,594 (73.21 %)	11,626 (26.11%)	298 (0.68 %)
RU	7,562 (54.60 %)	6,196 (44.73 %)	91 (0.67 %)
UK	5,036 (38.29 %)	8,073 (61.39 %)	41 (0.32 %)
DE	3,455 (36.18 %)	6,070 (63.57 %)	23 (0.25 %)
CA	2,693 (65.50 %)	1,391 (33.83 %)	27 (0.67 %)
FR	2,066 (31.13 %)	4,560 (68.72 %)	9 (0.15 %)
IT	1,213 (53.60 %)	1,039 (45.91 %)	11 (0.49 %)
JP	1,434 (64.13 %)	788 (35.24 %)	14 (0.63 %)

Table 4.6: Geographic economic tag distribution (country level)

and that involve only multi-regional ASes (Table 4.5). It is interesting also to note that the percentage of tag changes of the North American and European region is very similar and is lower with respect to the other continents, highlighting the importance of these two ecosystems in the definition of the BGP economic relationships. A similar analysis is reported in Table 4.7, where are summarized the number of economic tags changes comparing the national and the global inferences. The number of economic tags changes in every country scenario is slightly larger with respect to the changes experienced by the related continent, with the exception of the U.S.A., confirming its centrality in the Internet ecosystem.

Once again, it must be noted though that a part of the p2p connections of every economic topology is caused by the incompleteness of BGP data, as already highlighted in the previous chapter. Due to the small number of feeders actually announcing their full routing table to the route collectors (see Chapter 5 for more details) and due to the presence of BGP decision processes on ASBRs, only a small subset of the

Country	Tag changes	Transit to peering	Among multi-regional ASes	Among non-stub ASes
US	641 (1.44 %)	446 (1.00 %)	434 (0.97 %)	530 (1.19 %)
RU	378 (2.73 %)	238 (1.71 %)	83 (0.59 %)	248 (1.79%)
UK	456 (3.49 %)	367 (2.81 %)	401 (3.07 %)	406 (3.11 %)
DE	382 (4.03 %)	320 (3.37 %)	311 (3.28 %)	348 (3.67 %)
CA	247 (6.10 %)	230 (5.68 %)	238 (5.88 %)	212 (5.24 %)
FR	505 (7.71 %)	440 (6.72 %)	373 (5.69 %)	475 (7.25 %)
IT	117 (5.23 %)	109 (4.87 %)	110 (4.92 %)	101 (4.51 %)
JP	154 (7.00 %)	136 (6.18 %)	138 (6.27 %)	131 (5.95 %)

Table 4.7: Economic relationship changes (country level)

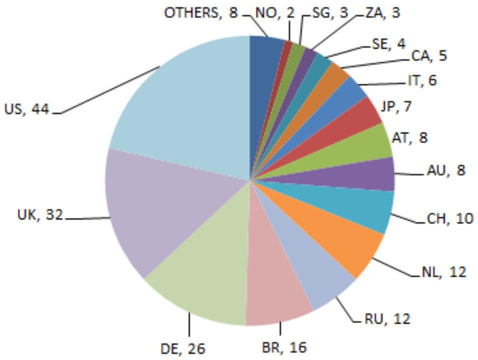


Figure 4.7: Geographic distribution of route collector feeders (country level)

possible AS paths are actually collected. In these conditions, the algorithm described in Section 3.2 may not record any AS path including a provider-free AS when deciding the nature of a connection `(A B)`. This means that the algorithm is unable to prove the existence of a transit connection between `A` and `B`, and thus may conclude that the economic relationship among them is p2p, i.e. `A` and `B` announce to each other only their networks and the networks announced to them by their customers. Obviously, the narrower the geographic focus is, the larger is the impact of the incompleteness on the economic topologies inferred, leading the topologies of the countries that host at least a feeder to be more complete than others, and the inferences made to be more reliable. Note that, the more the Internet is pervasive in a country, the higher is the probability that the same country is hosting at least one feeder, as can be seen in Fig. 4.7, enforcing the quality of topologies and economic inferences made. Thus the topologies and economic inferences made for the G8 members can be considered more reliable than those made for other countries, since they host more than the 70% of the feedeers of the route collectors. Likewise, the European (91 feeders) and North American (49) topologies and inferences are more reliable than those related to Asia Pacific (20), Latin America (16) and Africa (3). Note that, despite the Europe is hosting a larger number of feeders than North America, the number of p2p connections is still much larger in Europe, thus the rich set of European p2p relationships cannot be imputed only to lack of information, highlighting the different peering behaviors in the European and North American Internet regions.

5

BGP data incompleteness

In the last decade many studies have used the Internet AS-level topology to perform several analyses, from discovering its graph properties to assessing its impact on the effectiveness of worm-containment strategies. Yet, the BGP data typically used to reveal the topologies are far from being complete. Some issues about the extent of the incompleteness of BGP data were initially raised in [56], but only several years later was the first attempt to quantify such incompleteness performed. In [28], the authors compared the topology inferred from BGP data and the topology inferred from data contained in the regional IRRs, highlighting that about 40% of connections were missing in BGP-derived topologies. This topic was then to some extent neglected, but was recently re-evaluated in [30] and [87]. In [30] is provided an analysis of the BGP data obtained from each AS participating in RouteViews [10] and RIS [9], and their contribution was found to be heavily redundant. A completely different approach was applied in [87]. The AS-level topology inferred via BGP data was compared with a ground truth made up of proprietary router configurations of two major ISPs (a Tier-1 and a Tier-2 ISP), of two research networks (Abilene and GEANT) and several content providers. Their results were significant, but they are not reproducible and are heavily biased by the ground truth selected. Nevertheless, this work can still be considered a milestone in the analysis of the incompleteness of BGP data. Its most relevant result was that economic relationships established between ASes strongly affect the information that can be revealed by each monitor. The main rationale behind this lies in the economic-driven nature of the inter-domain routing [52] already introduced in Chapter 3 and on the different inter-AS economic relationships applied by ASes. In particular, basing on their definition, p2p connections are hidden from any of the providers or peers of a given AS. Basing on that, in [87] was claimed that most of the connections that are missing are p2p connections, and that monitors should be placed in the periphery of the Internet to discover them. A similar conclusion about the invisibility of p2p connections was also drawn in [64], which proposed a tool based on traceroute probes that is able to discover missing p2p connections established on IXPs. Despite this strong evidence of the incompleteness of BGP data gathered by route collectors, only a few works have addressed the issue of how new BGP monitors should be located to minimize this lack of information. In [101], the authors extended a model based on techniques developed in biological research for estimating the size of populations to work on the Internet AS graph. Their results showed that thousands of connections are missing, and the authors estimated that 700 route monitors would be able to see almost the totality of the connections. However, their heuristic only marginally took into account the existence of inter-AS economic relationships, thus the optimal number of route monitors found was seriously underestimated, as will be highlighted in Section 5.3.

In this chapter, BGP data currently gathered by the most known BGP route collector projects (BGPmon [1], PCH [7], RIS [9] and RouteViews [10]) is analysed, highlighting and explaining the causes of its incompleteness. It will be shown that the current view of the Internet is extremely *narrow* – due to the low number of ASes that are actively feeding the route collectors – and *biased* – due to the nature of the feeding ASes, which is mostly managed by worldwide ISPs. This top-down view does not allow the route collector infrastructure to discover a large set of p2p connections that may be established among ASes that are part of the lower part of the Internet hierarchy, as already highlighted in [33, 35, 64] and [87]. In addition to this classic analysis is also provided an innovative metric, named *p2c-distance*, which takes into account the presence of BGP decision processes and BGP export policies crossed by BGP UPDATE messages before reaching a route collector and provides a better understanding of the level of completeness of the data gathered. Unlike other approaches, this allow to analyse and quantify the level of incompleteness of BGP data by relying only on the route collector infrastructure and without exploiting any private data [87]. Then, in order to overcome the large amount of incompleteness highlighted, it is formulated a Minimum Set Cover problem that exploits the inter-AS p2c-distance within a generic AS-level topology to select the minimum number of ASes that should provide full routing information to the route collector infrastructure. Even though this kind of problems have been proved to be NP-complete [54], this methodology exploits the graph properties reducing the size of the graph in which the solution has to be searched by a) leveraging on the extreme low densities of the covering matrices, b) applying classic mathematical reduction techniques to these matrices [37, 55, 83, 94, 95] and c) using an exhaustive search on the remaining uncovered components of the original covering matrix. Furthermore, it is formulated an additional Maximum Coverage (MC) problem that allow to understand how much the situation can be enhaced with a limited number of new feeders. To solve it, we used a classic greedy algorithm considering only the coverage of each element in the Minimum Set Cover solution. It can be proved that this approach is a $(1 - \frac{1}{e}) \approx 0.632$ approximation for Maximum Coverage [65]. Finally, the solutions obtained are analysed and compared by applying the proposed methodologies on the global topology and five regional AS-level topologies of the Internet, highlighting the impact that the geographical peculiarities of the Internet have on the selection of the optimal set of ASes. As already highlighted in Chapter 4, geography plays an important role in routing decisions inside worldwide ASes and, consequently, also on the establishment of local economic inter-AS relationships and on the related BGP export policies. Typically the routing decisions are carried out at a continental level. Thus, to have full coverage of a given world-wide AS and its neighbors, multiple BGP connections with the route collectors need to be set up, one for each continent where the ASes are located. Furthermore, it is also provided an analysis of the current status of the route collector regional infrastructure by identifying how many ASes in the optimal solution are currently connected to them, by working out how many new ASes should be connected in each region, and by providing their typical characteristics.

5.1 The dark side of BGP-based measurements

Without any doubt, BGP data is the best data to infer the Internet AS-level topology, since the AS-level information is directly contained in the `AS_PATH` BGP attribute and no further heuristics have to be applied (see Chapter 2). BGP route collectors receive BGP routing information from cooperating ASes – i.e. *feeders* – to which they establish a BGP session. Thanks to these routing data it is possible to re-create the dynamics of the inter-domain routing as seen from a customer of the BGP feeders. In this section it is analysed the incompleteness of BGP data from a perspective different from the current state of the art. First, it is outlined the amount of data collected thanks to the route collector projects and is analysed the contribution of each of their BGP feeders. Finally, it is investigated the impact that the geographical distribution of the feeders of the route collectors has on the ability to discover regional Internet properties.

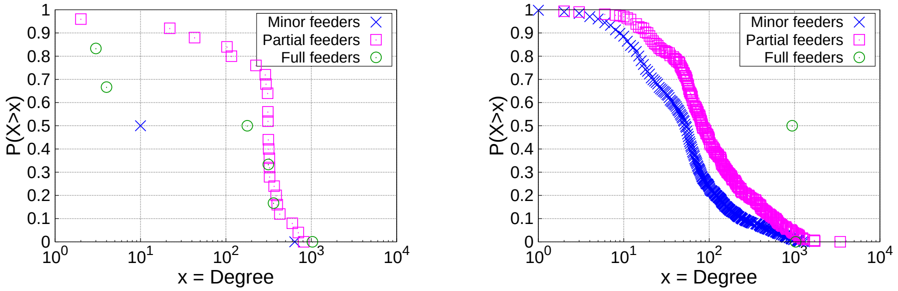
5.1.1 BGP feeder contribution analysis

Despite the well-known aims of the route collector projects, several of their BGP feeders do not provide any relevant contribution [33]. To better quantify the total contribution of BGP feeders, they can be subdivided on the basis of the amount of IPv4 and IPv6 space[1] that each of them advertised to the route collectors (Table 5.1): *minor feeders* announce an IPv4 space smaller than that of a single /8 IPv4 subnet or an IPv6 space smaller than that of a single /32 IPv6 subnet – that are typically the minimum IPv6 allocation done by regional IRRs[2], *full feeders* announce an IPv4 (IPv6)

---

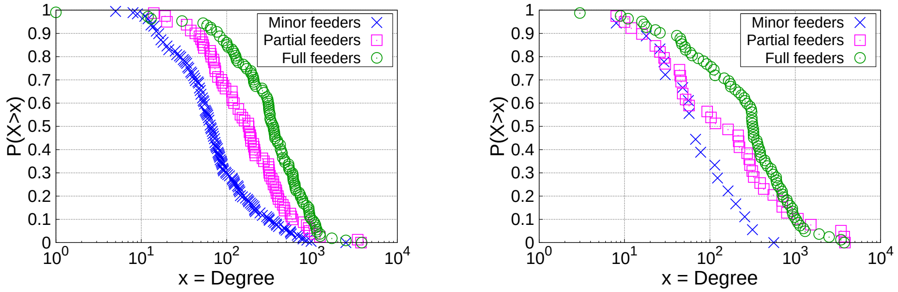
[1] The IPv4 (IPv6) space is computed considering only *not overlapping* subnets, i.e. those subnets that are not included in any other subnet

[2] `http://www.ripe.net/internet-coordination/press-centre/understanding-ip-addressing`



(a) BGPmon

(b) PCH



(c) RIS

(d) RouteViews

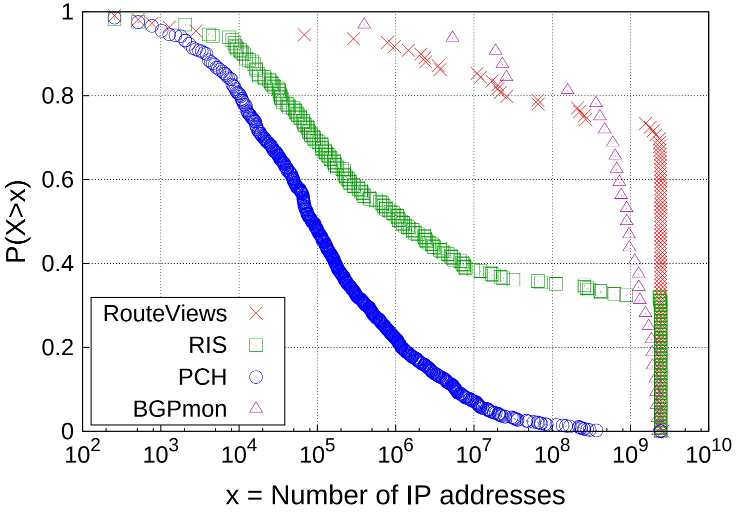
Figure 5.1: Route collector CCDF of the node degree of feeders

space closer to the full Internet IPv4 (IPv6) space currently advertised[3], while *partial feeders* include those ASes in between. For sake of readiness, in this work the full Internet IPv4 (IPv6) space is considered to be represented by the maximum number of IPv4 (IPv6) addresses announced by an AS, and an AS is considered to announce an IPv4 (IPv6) space closer to that if it announces more than the 75% of its value. Following this subdivision, the number of IPv4 full feeders is 5 for BGPmon, 2 for PCH, 96 for RIS and 78 for RouteViews, while the number of IPv6 full feeders is 1 for BGPmon, 1 for PCH, 57 for RIS and 29 for RouteViews. Together they make up a set of 152 IPv4 full feeders – i.e. 13.95% of the total number of IPv4 feeders – and 76 IPv6 full feeders – i.e. 16.67% of the total number of IPv6 feeders.

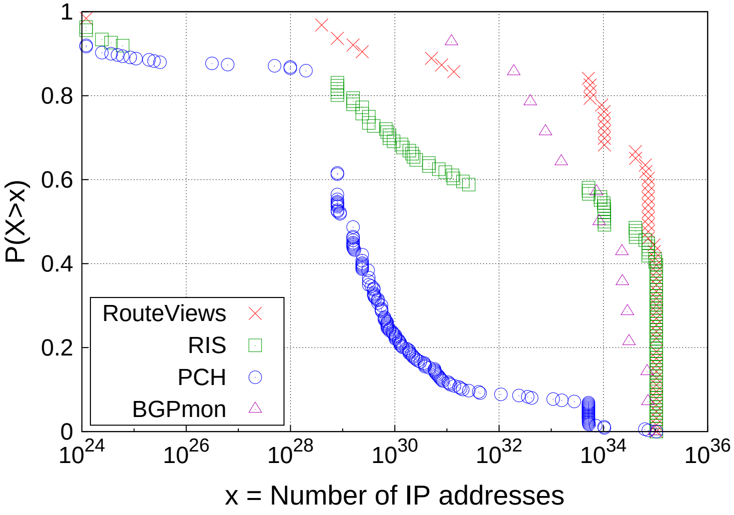
Figure 5.2a and Figure 5.2b show respectively the amount of non overlapping IPv4 and IPv6 addresses announced from the BGP feeders of each project. As can be seen from the height of the vertical tail of the CCDFs – that represents the percentage of feeders announcing to the route collectors a full routing table – the full feeders

---

[3] More information about the current IPv4 and IPv6 space advertised can be found at [8]



(a) IPv4 space



(b) IPv6 space

Figure 5.2: CCDF of the amount of IPvX space from each AS per project

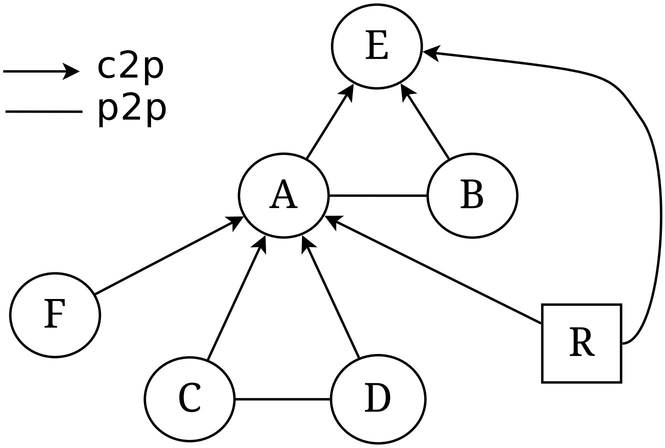


Figure 5.3: Connectivity scenario I

represents the clear majority only in RouteViews project, while in PCH the percentage is close to zero. Interestingly, only 63 ASes announce to the route collectors both the IPv4 and the IPv6 full routing table. This behavior can be just partially explained by the nature of the feeders themselves. Indeed, there still exists a large number of ASes on the Internet that do not introduced IPv6 on their networks, thus their router are unable to announce any IPv6 route to the route collector infrastructure. However, only 11 ASes out of the 89 that announce only a IPv4 full routing table do not originate any BGP message containing IPv6 prefixes, meaning that the largest amount of these AS networks are IPv6-capable. The most believable motivation for this behavior is that – for any technical or commercial reason – some of the feeders are just interested to announce their IPv4 (IPv6) reachability, and limit the amount of information concerning the IPv6 (IPv4) reachability. This can be proved by the presence of several IPv4 (IPv6) full feeders in the minor/partial set of IPv6 (IPv4) feeders. In detail, 37 out of 89 IPv4 full feeders appear either as IPv6 minor or partial feeders, while 8 out of 13 IPv6 full feeders appear either as IPv4 minor or partial feeders. However, given the low number of the ASes that are either part of the IPv4 (152) or the IPv6 (76) set of full feeders, in this work the set of overall full feeders is considered to be as composed by the union of the two sets.

A feature that can be used to gain a deeper insight into these different classes of BGP feeders is their node degree distribution (Fig. 5.1), which is computed on the union of the AS-level topologies inferred from BGPmon, PCH, RIS and RouteViews datasets. In each project the full feeder set is mainly composed of ASes that have developed a large number of BGP connections, which is a typical behavior of tran-

Route collector	***# of feeders***					
	full		*partial*		*minor*	
	IPv4	IPv6	IPv4	IPv6	IPv4	IPv6
BGPmon						
netsec.colostate	5	1	25	13	2	0
PCH						
ams.pch.net	0	0	14	97	269	16
atl.pch.net	0	0	2	14	45	3
ber.pch.net	0	0	3	6	13	2
bur.pch.net	0	0	7	8	21	2
cai.pch.net	0	0	0	0	2	0
cdg.pch.net	0	0	3	11	21	5
cpt.pch.net	0	0	0	0	12	1
dac.pch.net	0	0	0	0	13	0
dub.pch.net	0	0	0	3	13	2
equinix-paris.pch.net	0	0	3	15	35	5
eze.pch.net	0	0	0	1	2	1
fra.pch.net	0	0	15	74	166	10
gnd.pch.net	0	0	0	0	1	0
hkg.pch.net	0	0	6	12	29	3
iad.pch.net	0	0	12	27	68	4
icn.pch.net	0	0	1	1	3	1
jax.pch.net	0	0	0	0	3	1
jpix.pch.net	0	0	3	7	13	2
ktm.pch.net	0	0	0	3	17	1
lax.pch.net	0	0	5	3	19	1
lga.pch.net	0	0	3	11	19	1
lhr.pch.net	0	0	15	56	175	10
lonap.pch.net	0	0	1	16	33	5
mia.pch.net	0	0	4	10	32	4
mnl.pch.net	0	0	0	0	4	0
mpm.pch.net	0	0	0	0	12	1
muc.pch.net	0	0	2	0	10	1
nl-ix.pch.net	0	0	1	12	53	5
nrt.pch.net	0	0	2	3	8	0
nyiix.pch.net	0	0	6	18	69	5
ord.pch.net	1	0	4	19	30	2
paix-sea.pch.net	0	0	1	4	13	1
pao.pch.net	1	1	12	19	53	5
per.pch.net	0	0	0	4	7	1
sea.pch.net	0	0	5	17	71	0
sfinx.pch.net	0	0	1	17	34	6
sgw.pch.net	0	0	8	16	33	7
sna.pch.net	0	0	9	18	69	4
syd.pch.net	0	0	1	5	14	2
tie-ny.pch.net	0	0	1	12	23	3
tll.pch.net	0	0	0	3	3	0
tmp.pch.net	0	0	0	5	8	1
trn.pch.net	0	0	2	0	19	1
vie.pch.net	0	0	6	21	55	3
waw.pch.net	0	0	3	7	11	0
wlg.pch.net	0	0	0	4	8	1
yow.pch.net	0	0	0	1	3	1
yyz.pch.net	0	0	1	4	30	3
zrh.pch.net	0	0	2	19	37	4
RIS						
rrc00	22	11	0	0	1	1
rrc01	13	7	1	5	44	2
rrc03	8	6	4	17	58	2
rrc04	5	1	0	4	7	0
rrc05	6	2	3	10	31	0
rrc06	1	1	0	0	4	1
rrc07	2	3	2	3	10	1
rrc10	4	4	5	8	12	0
rrc11	6	2	3	7	13	3
rrc12	16	20	2	17	32	1
rrc13	9	2	0	1	9	0
rrc14	6	4	3	5	6	0
rrc15	10	5	1	3	2	0
RouteViews						
route-views2	27	0	5	0	0	0
route-views4	15	5	1	7	0	0
route-views6	0	9	0	5	0	0
route-views.eqix	10	6	1	4	4	0
route-views.isc	11	2	0	1	3	0
route-views.jinx	1	1	1	5	6	0
route-views.kixp	0	0	0	1	1	0
route-views.linx	19	5	2	8	3	2
route-views.saopaulo	12	6	3	2	0	0
route-views.sydney	4	2	3	4	0	0
route-views.telxatl	4	0	0	4	0	0
route-views.wide	2	1	0	1	2	0

Table 5.1: Route collector feeder details (January 2013)

sit ISPs. To confirm this, the nature of these ASes was analysed by browsing their websites and parsing their entries in the regional IRRs, and the vast majority of full feeders were proven to be large ISPs. Furthermore, from this analysis it was found that 11 out of the 16 ASes listed as being provider-free in [13] are currently feeding the route collectors.

Since the vast majority of full feeders are large ISPs, the view of the Internet (at the AS-level) extracted from these projects is likely to represent more the Internet viewed by some of the most important ISPs in the world rather than the real Internet. A view of the Internet from the top of the AS hierarchy is not able to discover a large number of connections. In fact, due to BGP export policies, a route collector connected with

ASes that are part of the top of the hierarchy is not able to reveal all the p2p connections that are established at the lower levels. On the other hand, the lower in the hierarchy the BGP feeder is located, the greater the chance to gathering information about an AS path involving a previously hidden p2p connection. Consider for example Fig. 5.3. In this case, if the route collector `R` is connected to AS `E` located at the top of the economic hierarchy, it cannot reveal either the p2p connection between `A` and `B`, or the p2p connection between `C` and `D`. On the other hand, if `R` is connected to `A`, it can reveal the p2p connection between `A` and `B`, but not between `C` and `D`. Moreover, it is fundamental that the route collectors establish a c2p relationship with their feeders. Otherwise, even if the route collector is connected to `A`, the connection `(A,B)` will not be revealed.

A real-world example of the importance of obtaining full routing tables from BGP feeders located in the lowest part of the Internet hierarchy is represented by PCH. This data source is potentially extremely useful for discovering hidden AS connections, since its route collectors are deployed on 49 different IXPs and connected to 1,276 ASes, about three times the total number of BGP feeders of RIS (432) and RouteViews (172). In addition, many of its BGP feeders have small node degree value (Fig. 5.1), which is a rough indication of their location at the bottom of the Internet hierarchy. Nevertheless, as shown in Table 5.2, the number of AS connections detected by BGP data gathered by PCH and not discovered by RouteViews and RIS is extremely low, since 86,364 connection out of the total 92,282 connections discovered – i.e. the 93.59% of the total number of connections of PCH – are already revealed by the other route collector projects. This happens because PCH mainly establishes only p2p connections with its BGP feeders except with its two providers, i.e. its route collectors obtain only routes concerning prefixes owned directly by their BGP feeders or announced by their customers. Consequently, it is likely that almost every connection found by PCH represents a p2c (c2p) economic agreement. Thus the issue of p2p connection discovery has not been solved even though it currently represents the largest set of hidden connections [64, 35, 18], greatly limiting the topology discovery potentiality of its route collectors.

A deeper insight into the amount of information provided by each BGP feeder can be found by analysing the difference between the direct node degree and the inner node degree (Fig. 5.4). The *direct* node degree of a BGP feeder `X` is defined as the cardinality of the set of its neighbors that are discovered using only BGP data directly announced by `X` to a route collector, and the *inner* node degree of `X` is defined as the cardinality of the set of its neighbors that are discovered using BGP data announced by every BGP feeder but `X`. A similar approach was proposed in [39], but with a different purpose. It is thus possible to differentiate between two different classes of behavior of BGP feeders: a) ASes that announce just a partial view of the Internet (degree difference < 0), like those ASes that consider the route collectors as peers

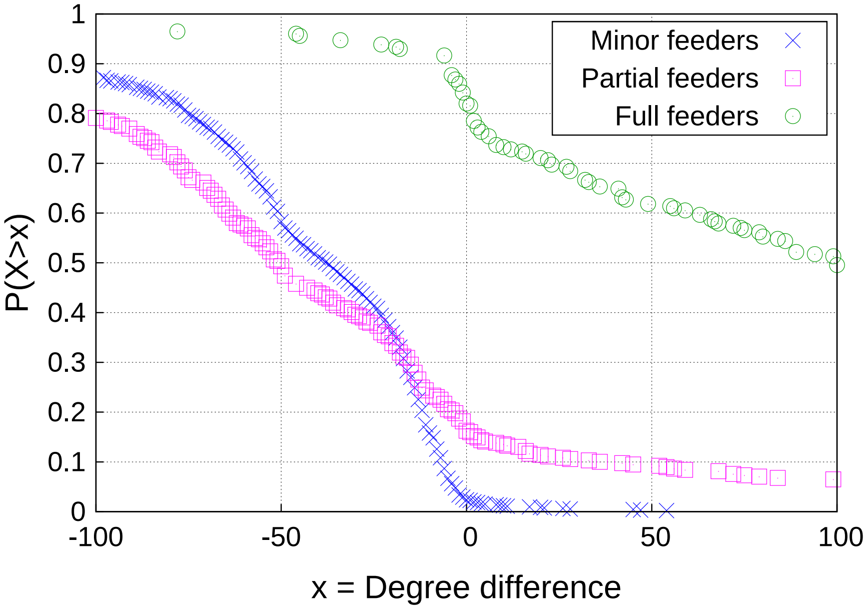


Figure 5.4: CCDF of the degree difference of BGP feeders

and not as customers, and b) ASes that contribute with connections not contained in any AS path announced by other BGP feeders (degree difference > 0), such as p2p and p2c connections that are hidden from the other BGP feeders due to the effect of BGP export policies crossed during the propagation of the routing information.

The first class is typical of minor feeders, whose connectivity is mostly discovered via other feeders (Fig. 5.4). On the other hand, the second class is typical of full feeders, that typically introduce previously undiscovered AS connections involving them. However, it can noted how some of the full feeders partially hide their connectivity despite advertising their full routing table to a route collector project. This phenomenon has been recorded in about 20% of the full feeders (see negative values of degree difference in Fig. 5.4) and is caused by the BGP decision process on the feeder side. Depending on the policies established among ASes and on technical decisions, some direct connections may not be announced to the route collector. For example, as highlighted above, some ASes may decide to announce only their IPv4 (IPv6) full routing table. However, for the same economic and technical reasons it is possible that the same direct connections hidden from the route collector are announced to other neighbors, propagated on the Internet, and finally detected by the

	BGPmon	*PCH*	*RIS*	*RouteViews*
# of nodes	44,182	43,803	44,016	44,006
# of edges	103,402	92,282	158,541	135,501
# of common edges	86,364			

Table 5.2: Topology characteristics

route collector from another feeder as a side effect. Some other slight exceptions are related to minor and partial feeders with a positive degree difference (less than 5%). These feeders are likely to be located in the bottom part of the Internet hierarchy and some of their p2c connections may result as hidden to the route collector infrastructure because of the cross effect of their multi-homed nature and of multiple BGP decision process crossed by UPDATE messages before reaching the route collector infrastructure.

This last phenomenon highlights that the presence of multiple BGP decision processes along the AS path may limit the completeness of AS-level topology collected, since each BGP ASBR selects and announces only the best route per-destination [97] to their neighbors. In summary, the information that a BGP feeder announces to the route collector is the result of its BGP decision process which, in turn, is fed only with routes that are the result of the BGP decision processes of its neighboring ASes, and so on. Each BGP decision process, from an AS-level measurement perspective, is a route filter which can potentially reduce the AS-level connectivity information received from each route collector. As a consequence, the higher the distance of an AS from the BGP route collectors, the higher the number of BGP decision processes crossed and, thus, the probability that one or more of them will filter out some AS connections.

5.1.2 Geographical coverage

The incompleteness of BGP data is even stronger if analysed from a geographical perspective. Table 5.3 details the total number of BGP feeders as well as the number of them that supply the full routing table to any of route collector projects. To perform this analysis we geolocated the IP address of each BGP feeder using the Maxmind GeoIPLite database [6] and considering the world being subdivided into five macro-regions: Africa, Asia-Pacific (i.e. Asia and Oceania), Europe, Latin America (the Caribbean, Central America, Mexico and South America) and North America, as done in Chapter 4. Table 5.3 highlights that most full feeders are located either in Europe or in North America. Interestingly, in Africa only a couple of feeder is found even though Africa hosts two route collector of RouteViews and three of PCH. This means that every inference about the African part of the Internet is mostly obtained through views located in different regions. Thus, some relevant characteristics of the African

Region	**Feeders**		**Full feeders**	
	IPv4	*IPv6*	*IPv4*	*IPv6*
Africa	19	6	3	2
Asia-Pacific	104	59	22	18
Europe	699	310	93	72
Latin America	21	40	16	17
North America	369	105	60	32
World	1,090	456	152	76

Table 5.3: Geolocation of BGP feeders

Internet may be hidden from the current route collectors, e.g. the largest amount of African p2p connectivity. This is not only a problem regarding Africa, in fact the number of feeders in other regions is low as well, if compared with the total number of ASes of the Internet.

5.2 A novel methodology to deal with BGP data incompleteness

Given the large amount of incompleteness of data collected via route collectors, the first step to infer an Internet AS-level topology closer to reality is to introduce a larger number of new BGP feeders that announce their full routing tables. One of the biggest obstacles is the vast number of ASes that make up the Internet. Obtaining routing information from each of them would require the participation of thousands administrators in a project that may not be appealing for many of them, and represents a practically unfeasible task. However, the vast majority of ASes that currently hide part of their connectivity are mostly ASes that offer IP transit to other ASes and ASes that are connected to IXPs [18, 41, 58, 64]. Although there are far fewer of them than the total number of ASes, these ASes contribute substantially to provide Internet connectivity. In fact, both world-wide ISPs and small/medium ISPs are part of this set of ASes.

In this section, it is proposed a methodology to identify the minimal set of BGP feeders that would be needed to gather as much BGP data as possible *about* the ASes whose connectivity plays a major role in their economic market – i.e. the non-stub ASes – thus minimizing the number of hidden connections present in the core of the Internet and, at the same time, minimizing the impact of BGP filters, such as BGP export policies and BGP decision processes. The very same methodology also provides an additional ranking list of candidate feeders to allow any route collector project to identify which ASes are the best candidates to maximize their coverage with a limited amount of resources.

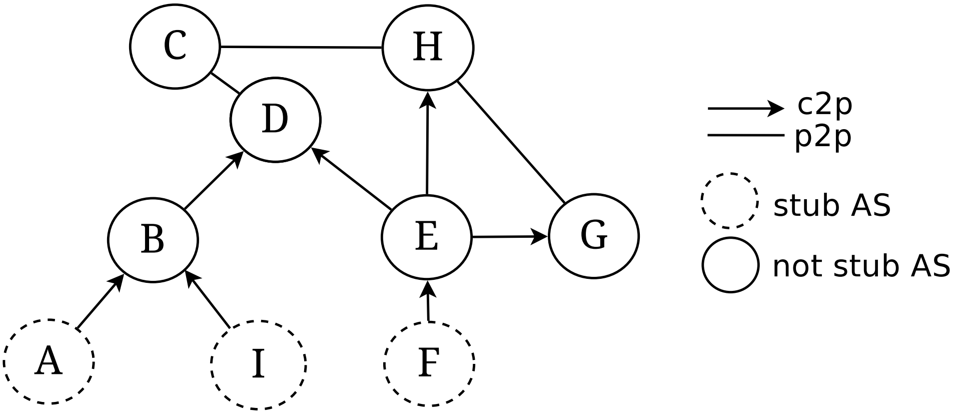


Figure 5.5: Connectivity scenario II

5.2.1 A new metric: p2c distance

As already highlighted, a major role in filtering is played by inter-AS economic relationships due to their influence on the amount of routing information that each AS announces to its neighbors. Given the BGP export policies related to each economic relationship [52], a necessary – but not sufficient – condition for a route collector to reveal the full connectivity of a given AS is that there exists at least one AS path that is made up exclusively of p2c connections from that AS to the route collector. This because only customers in a p2c connection are able to obtain routes towards every Internet destination. Moreover, it is preferable that routing information arrives at a route collector having crossed the lowest number of p2c connections as possible, in order to limit the filtering effects of BGP decision processes.

On the basis of the two conditions hypothesized above, it is defined a new metric that is able to capture the level of completeness of data gathered by the current set of route collectors. The *p2c-distance* of one AS `X` from another AS `Y` is defined as the minimum number of consecutive p2c connections that connect `X` to `Y` in the considered economic topology or, likewise, the minimum number of consecutive c2p connections that connect `Y` to `X`. This metric quantifies at the same time the distance and the amount of transit connections crossed by any BGP UPDATE message to reach a route collector, and can be exploited to reveal which part of the Internet is well-monitored and which part still represent a dark zone. Note that this metric still relies on an inference made on the current Internet AS-level topology, but the p2c and c2p connections have been proved to be extremely accurate [33, 35, 64, 87].

To better understand how this metric works, consider the connectivity scenario depicted in Fig. 5.3. In this example, the route collector `R` has a p2c-distance of 1 from AS `A` and `E`, while the p2c-distance from `B`, `C`, `D` and `F` is not defined. This means that `R` has the possibility to reveal every p2p connection established by `A`, `B` and `E`. On the other hand, it also means that `R` is not able to reveal the p2p connectivity of `C`, `D` and `F` in any way, thus `R` will not reveal the connection `(C, D)` in any AS path. Nevertheless, `R` can discover the p2c (c2p) connectivity of each AS of the scenario.

5.2.2 BGP feeder selection

Given the definition of p2c-distance, a complete view of the Internet can only be obtained by connecting a route collector to each stub AS, as already concluded in [87]. Stub ASes are ASes typically managed by local access providers – which provide connectivity to end users but not to other ASes – and organizations that do not have the Internet transit as part of their core business (e.g. banks and car manufacturers), and appear in BGP data as the originating element, i.e. the right-most element in every`AS PATH` attribute that involves them. Due to the nature of their related organizations, these ASes tend to be customers in the economic relationships established with other ASes, representing a perfect starting point to minimize the p2c-distance of every AS that make up the Internet. However, since p2c connections are already discovered by route collectors connected to the top of the hierarchy [87], most BGP data collected from a hypothetical route collector infrastructure connected to each of them would be redundant. Moreover, since it is not possible to infer a priori which stub AS is actually interested in establishing p2p connections, it is impossible to reduce the number of new feeders required to obtain full Internet AS-level connectivity. This means that, based on January 2013 data, it is required a BGP connection with each of the 35,967 stub ASes (out of 44,389 total ASes). This makes this approach practically unfeasible. A good trade-off solution between the possibility to discover hidden p2p connections and the feasibility of obtaining such data is however represented by those ASes that are actually interested in deploying p2p connections to improve the quality of their services, i.e. non-stub ASes [41]. The lack of interest of stub ASes in establishing p2p connections is highlighted by the low percentage of them (7%) that participates in at least one IXP[4], where ASes typically interconnect with settlement-free p2p connections to reduce the amount of their traffic directed to their providers (see [15, 18, 25, 58] and [64] for more details on IXPs).

In this work, the objective is to select new BGP feeders such that each non-stub AS has a finite and bounded p2c distance from the route collector infrastructure to

---

[4] The set of ASes that participate to at least one IXP was collected by downloading and parsing the participant list web page of 198 IXPs that were found to be active January 15th, 2013. This kind of data has been collected once per month since February 2012 and is publicly available at [4].

minimize the effects of BGP filters and, consequently, increasing the possibility to reveal the hidden p2p connectivity of the actual core of the Internet. As will be shown later on, stub ASes will still play a key role in the solution of the problem, even though the analysis do not actually aim to discover their connectivity. However, in Section 5.3 is also shown an additional scenario that consider as part of the objective function also those stub ASes that are found to participate on (at least) one IXP.

5.2.3 Identifying the feeders

The first part of the methodology is dedicated to the discovery of the minimum number of full feeders required to obtain a complete[5] view of the Internet core. The identification of the feeders can be modeled as a MSC problem that can be described through the following Integer Linear Programming (ILP) formulation:

$$\text{Minimize} \qquad \left(\sum_{AS_i \in \mathcal{U}} x_{AS_i} \right) \tag{5.1}$$

subject to

$$\sum_{AS_j : AS_n \in S^{(d)}_{AS_j}} x_{AS_j} \geq 1 \qquad \forall AS_n \in \mathcal{N} \tag{5.2}$$

$$x_{AS_i} \in \{0, 1\}, \qquad \forall AS_i \in \mathcal{U} \tag{5.3}$$

where $\mathcal{U} = \{AS_1, AS_2, \ldots, AS_n\}$ is the set of all ASes composing the Internet, $\mathcal{N} \subset \mathcal{U}$ is the set of non-stub ASes and $S^{(d)}_{AS_i}$ represents the *covering set* of AS_i at fixed p2c-distance d, i.e. the set of ASes in $\mathcal{N}$ that have a p2c-distance value of at most d from AS_i. The goal of this MSC problem is, given d, to obtain the minimum number of elements of $\mathcal{S}^{(d)}$ such that their union is $\mathcal{N}$ or, in other words, to select the minimum number of feeders from $\mathcal{U}$ such that the p2c-distance of any non-stub AS from at least one of them is at most d. The parameter $d \geq 0$ defines the maximum number of BGP decision processes[6] that `UPDATE` messages generated by each non-stub AS will traverse before reaching a potential feeder and, thus, indicates the number of filters encountered that can cause loss of information. Note that AS_i belongs to $S^{(d)}_{AS_i}$ for any fixed d. Furthermore, x_{AS_i} is 1 if $S^{(d)}_{AS_i}$ is part of the final solution, 0 otherwise. In other words, the problem aims to minimize the number of ASes (5.1) required to cover each non-stub AS (5.2) and, thus, there is a non-zero probability to discover its complete connectivity from a route collector. Note also that imposing $d = 0$ implies that the solution is composed of the entire set of non-stub

---

[5] Note that in this methodology it is assumed that every AS in the Internet is able to provide a full routing table to the route collectors.

[6] The number of BGP decision processes encountered before reaching a route collector is $d + 1$, since feeders introduce an additional BGP decision process before announcing BGP data to the route collectors.

ASes. The larger the value of d, the heavier the filtering effects introduced by BGP decision processes but the smaller the number of required BGP feeders and, thus, the number of required BGP connections. To better understand the problem consider the scenario depicted in Fig. 5.5. In this example, $\mathcal{U} = \{A, B, C, D, E, F, G, H, I\}$ and $\mathcal{N} = \{B, C, D, E, G, H\}$. Thus, $S_A^{(1)} = S_I^{(1)} = \{B\}$, $S_B^{(1)} = \{B, D\}$, $S_C^{(1)} = \{C\}$, $S_D^{(1)} = \{D\}$, $S_E^{(1)} = \{E, D, G, H\}$, $S_F^{(1)} = \{E\}$, $S_G^{(1)} = \{G\}$, $S_H^{(1)} = \{H\}$. One of the optimal solutions to cover every non-stub AS is $\{S_B^{(1)}, S_C^{(1)}, S_E^{(1)}\}$, i.e. ASes `B`, `C` and `E` should be selected as new feeders.

5.2.4 Solving our MSC problem

MSC problems are known to be NP-hard [65], however that does not mean that it is always practically impossible to solve them [114]. Several techniques to reduce the problem into smaller problems were developed in the last century in different research fields, and can be particularly effective depending on the nature of the problem. In our case, the iterative application of the concepts of *essentiality* and *dominance* described in the Quine-McCluskey procedure ([83, 94, 95]) on the covering matrix related to the described MSC problem were found to be particularly effective. The *covering matrix* of a given MSC problem is the Boolean matrix in which each row represents a covering element and each column represents an element that has to be covered. A generic element (i, j) of that matrix thus contains 1 if the element placed at row i covers the element placed at column j, 0 otherwise. In our case, each row of the covering matrix represents an AS in $\mathcal{U}$ and each column represents an AS in $\mathcal{N}$, thus its size is $|\mathcal{U}| \times |\mathcal{N}|$, and each element (i, j) of the matrix contains 1 if AS i covers AS j, i.e. $S_{AS_i}^{(d)}$ contains AS j. The size of this matrix can be then reduced by applying the following techniques:

Essentiality. *A row of the covering matrix M is defined as essential iff it is the only row covering a given element. Consequently, an AS is essential iff its covering set contains a non-stub AS covered only by the AS itself. More formally, AS$_x \in \mathcal{U}$ is essential iff* $\exists AS_n \in S_{AS_x}^{(d)} : AS_n \in (S_{AS_x}^{(d)} \setminus \bigcup_{y \neq x} S_{AS_y}^{(d)})$

Dominance. *A row of the covering matrix M is defined as dominated by another row iff every element covered by the considered row are also covered by the dominating row. Consequently, AS_x dominates AS_y iff the non-stub ASes covered by AS_x are covered also by AS_y. Formally, given $x, y \in rows(M)$, x dominates y iff $S_{AS_y}^{(d)} \subseteq S_{AS_x}^{(d)}$, and $rows(M)$ is the set of rows of matrix M.*

Using essentiality and dominance it is possible to devise a technique to retrieve an optimal solution for our MSC problem. In detail, it is proposed a procedure that consists of the following four phases: *phase a*) Selection of essential covering sets,

```
1  **Input:** Covering matrix $M$ of size $|\mathcal{U}| \times |\mathcal{N}|$
2  ------------------------------------------
3    $O =$ null matrix of size $|\mathcal{U}| \times |\mathcal{N}|$
4    **while** $M \neq O$
5      $O = M$
6      **foreach** row $r \in M$
7        **if** $AS_r$ is essential
8          **insert** $AS_r$ in $\mathcal{P}$
9      **delete** from $M$ each column $c$ covered by $AS_r$
10     **foreach** row $r \in M$
11       **foreach** row $s \in M \wedge s \neq r$
12         **if** $AS_r$ dominates $AS_q$
13           **delete** $AS_q$ from $M$
14           record <$AS_r, AS_q$> in $\mathcal{T}$
15         **else if** $AS_q$ dominates $AS_r$
16           **delete** $AS_r$ from $M$
17           record <$AS_q, AS_r$> in $\mathcal{T}$
18           **break**
19
20   **if** $M$ is not empty
21     partition $M$ in a block matrix
22     **foreach** block $b_i \in M$
23       compute $\mathcal{P}_{b_i}$ applying exhaustive search on $b_i$
24       add $\mathcal{P}_{b_{i_0}}$ to $\mathcal{P}$
25       **insert** $\mathcal{P}_{b_{i_{j>0}}}$ in $\mathcal{I}$
26
27   **insert** $\mathcal{P}$ in $\mathcal{I}$
28   **foreach** <dominating $AS_r$, dominated $AS_q$> in $\mathcal{T}$
29     **if** $S_{\mathcal{P}} = S_{\mathcal{P}-AS_r+AS_q}$
30       **insert** $AS_q$ in $\mathcal{I}$
31   ------------------------------------------
32   **Output:** solution set $\mathcal{P}$, set of element in a possible optimal solution $\mathcal{I}$
```

Figure 5.6: MSC reduction procedure

phase b) Deletion of dominated covering sets, *phase c*) Exhaustive search on the remaining covering matrix, and *phase d*) Selection of ASes that are part of at least one optimal solution. Details about the procedure are depicted in Fig. 5.6.

The aim of phase *a*) of the procedure is to find *essential* ASes, that are required to be part of the final solution (lines 6–9). Indeed, whenever an AS `X` is found to be essential, this means that `X` is the only AS to cover at least one non-stub that is required to be covered, and has to be part of solution. Every AS contained in $S_X^{(d)}$ then can considered covered (line 9), reducing the columns of the covering matrix. Then, in phase *b*), the procedure tries to find if any AS is dominated by another AS (lines 10–18). In this case, the row corresponding to the dominated AS is deleted from the covering matrix. Note that an AS is considered to be dominated by another AS even if the two ASes have the same covering set. In this case, one of the two ASes is randomly chosen to continue the procedure. Every dominated AS is maintained in a separated set together with the related dominating AS in order to check their presence in at least an optimal solution during the final phase. Essentiality and dominance foster each other, since the former reduces the columns of the matrix on the basis of the available rows, whereas the latter reduces the rows of the covering matrix on the basis of the available columns. For this reason phases *a*) and *b*) are iteratively applied until a cyclic core is found [37] or, in other words, the size of the covering matrix cannot be reduced anymore (line 4).

In our case, essentiality and dominance have been particularly effective due to the economic nature of the Internet and the way the covering matrices are populated. The Internet is composed mainly by organization whose economic market is not driven by Internet traffic transit services [59] and which are typically located in a single country [60]. Moreover, the national Internet markets are typically closed [85] and, as a consequence, the number of providers present in each country is limited. The (relatively) poor heterogeneity of choices that an AS located in a single country have in choosing its providers make several ASes to have a common set of providers, and thus several rows in the covering matrix to be similar, often leading to dominating/dominated pair of rows. This large similarity among rows and the low density of the p2c-distance covering matrices allow us to reduce the problem greatly, often finding an optimal solution by just applying essentiality and dominance iteratively. Consider for example the scenarios related to January 2013. In this scenario, the original covering matrices sizes were $44,389 \times 8,422$, and the matrix densities[7] were 0.00029 ($d = 1$), 0.00325 ($d = 2$) and 0.01982 ($d = 3$). The reduction technique just depicted shrunk the covering matrices respectively to a cyclic core of 70×66 ($d = 1$), 3×4 ($d = 2$) and to an empty cyclic core ($d = 3$).

If a non-empty cyclic core is found, the problem could be solved via exhaustive search or by applying a branch and bound technique. Back to the scenario of January 2013, this would mean to compute $\sum_{k=1}^{66} \binom{70}{k}$ comparisons. However, it is still possible to reduce the problem (phase *c*) by exploiting the Internet regionality characteristic ([60, 85]) to partition the remaining covering matrices in diagonal blocks (line 21) by permuting opportunely their rows and columns using standard mathematical techniques. This allows to divide the remaining covering problem in disjoint and smaller subproblems that can be solved independently [86] via exhaustive search (lines 22–24), computing for each block b_i the set of every possible solution $\mathcal{P}_{b_i}$. The

---

[7] In this work is defined as *matrix density* the ratio of non-zero entries to total number of entries in a matrix.

final optimal solution is then represented by the union of the solution found so far and one of the solutions found per each block. Consider again the January 2013 scenario depicted above and focus on $d = 1$, that showed the largest cyclic core. In that case, it was possible to decompose the cyclic core in 16 sub-matrices – each composed by elements located in the same country – whose maximum size was 10×8, that can be solved with at maximum $\sum_{k=1}^{8} \binom{10}{k} = 1,012$ operations via exhaustive search. Given the results obtained, there was no need to complicate even more the procedure with additional reducing steps, even if it would still be possible to reduce the size of cyclic cores furthermore by using Gimpel's technique [55].

Finally, in phase *d*) is checked if any of the dominated elements recorded during the procedure could be part of at least one optimal solution (lines 28–30) using the following lemma and corollary:

Lemma 1. *An element AS_q that is dominated by an element AS_r – with $AS_r \in \mathcal{P}$ and not part of the cyclic core – is part of at least one optimal solution only if the solution $\mathcal{P}^*$ obtained by swapping AS_q with AS_r in $\mathcal{P}$ is still an optimal solution, i.e. $|S_{\mathcal{P}}| = |S_{\mathcal{P}^*}|$, where $\mathcal{P}^* = \mathcal{P} - AS_r + AS_q$.*

Proof. By hypothesis, if $|S_{\mathcal{P}}| \neq |S_{\mathcal{P}^*}|$, the space of elements covered uniquely by AS_r in $\mathcal{P}$ that made AS_r as essential in the MSC procedure – i.e. $S_{\mathcal{P}} \setminus S_{\mathcal{P}-AS_r}$ – cannot be fully covered by AS_q. Now, consider the solution $\mathcal{P}'$ obtained by forcing AS_q in the initial solution set and by applying the MSC procedure once again. Since the MSC procedure is devised to retrieve an optimal solution ([83, 95]), it is also able to retrieve the solution with the minimal cardinality that involves AS_q. Since AS_q was found to be dominated by AS_r in the original MSC procedure, this means that the space of elements $S_{AS_q} \setminus S_{AS_r}$ was found to be covered by essential elements. Thus, even forcing AS_q in solution these elements are still going to be selected as essentials by the MSC procedure, since S_{AS_q} do not include the elements that characterize their essentiality. In other words, the space of elements $S_{AS_q} \setminus S_{AS_r}$ is covered redundantly and do not justify the presence of AS_q in an optimal solution. The remaining covering space $S_{AS_q} \cap S_{AS_r}$, by hypothesis, covers only partially the space of elements $S_{\mathcal{P}} \setminus S_{\mathcal{P}-AS_r}$, that has to be covered by an additional AS, i.e. AS_r itself or one of the elements that were found to be dominated during the original MSC procedure. Thus, the cardinality of $\mathcal{P}'$ has to be larger than $\mathcal{P}$, since there is an element (AS_q) that do not cover uniquely any space of elements in $\mathcal{N}$. □

Corollary 1. *An element AS_q that is dominated by an element AS_r – with AS_r part of block b_i of the cyclic core – is part of at least one optimal solution only if there exists at least a solution $\mathcal{P}_{b_{i_j}}$ found via exhaustive search on b_i where the solution $\mathcal{P}^*_{b_{i_j}}$ obtained by swapping AS_q from $\mathcal{P}_{b_{i_j}}$ with its dominating element AS_r is still*

```
Input: Covering matrix $M$ of size $|\mathcal{I}| \times |\mathcal{N}|$
--------------------------------------------------
$\mathcal{C} = \emptyset$
while ($\mathcal{C} \neq \mathcal{N}$)
  $AS_r$ = row covering the largest number of columns
  append $AS_r$ to $\mathcal{R}$
  foreach row $s \in M$
    if $S_{AS_q} = S_{AS_r}$
      insert $AS_q$ in $\mathcal{R}$ as alternative to $AS_r$
  delete from $M$ each column covered by $AS_r$
  $\mathcal{C} = \mathcal{C} \cup S_{AS_r}$
--------------------------------------------------
Output: Ranking list $\mathcal{R}$
```

Figure 5.7: Greedy heuristic

an optimal solution, i.e. $|S_{\mathcal{P}_{b_{i_j}}}| = |S^*_{\mathcal{P}_{b_{i_j}}}|$*, where* $\mathcal{P}^*_{b_{i_j}} = \mathcal{P}_{b_{i_j}} - AS_r + AS_q$ *and* $0 \leq j < |\mathcal{P}_{b_i}|$.

As result of this procedure, the algorithm provides: i) a set $\mathcal{P}$ of ASes composed of the set of ASes that were inserted into the solution during phase a) and of one of the solution per each block found $\mathcal{P}_{b_i}$, and ii) a set $\mathcal{I}$ of ASes containing every AS that can be part of at least one optimal solution.

5.2.5 Ranking the candidates

The solution of the MSC problem by itself provides only a quantification of the number of feeders required to obtain an optimal coverage of the Internet core, but does not help much in understanding how much the coverage would be improved with just a limited set of feeders. This because the aim of the MSC problem is to completely cover the set of non-stub ASes, and at each step are selected the ASes due to the *essentiality* concept described earlier, instead of the ASes that would maximize the partial coverage. The results obtained from the MSC problem have thus a theoretical importance – since they allow to quantify the current coverage of the route collectors and to identify which ASes should join them – but at the same time have a rather low practical importance – since it is almost impossible to connect each feeder found to a route collector. In this perspective, the second part of the methodology consists in a procedure to rank each AS found to be part of at least one MSC problem optimal solution, i.e. belonging to $\mathcal{I}$, to provide to any route collector project an ordered list of ASes that should join them in order to maximize their coverage with limited resources.

This result can be obtained by solving via greedy algorithm [36] a particular MC problem restricted on the elements found to be part of $\mathcal{I}$, i.e. the rows of the covering matrix of this MC are the rows corresponding to the ASes belonging to set $\mathcal{I}$. The MC problem can be described through the following ILP formulation for every given $k > 0$:

$$\text{Maximize} \qquad \left(\sum_{AS_j \in \mathcal{N}} y_{AS_j} \right) \tag{5.4}$$

subject to

$$\sum_{AS_i \in \mathcal{I}} x_{AS_i} \leq k \tag{5.5}$$

$$\sum_{AS_i \in \mathcal{I} \wedge AS_j \in S_{AS_i}} x_{AS_i} \geq y_{AS_j}, \qquad \forall AS_j \in \mathcal{N} \tag{5.6}$$

$$y_{AS_j} \in \{0, 1\}, \qquad \forall AS_j \in \mathcal{N} \tag{5.7}$$

$$x_{AS_i} \in \{0, 1\}, \qquad \forall AS_i \in \mathcal{U} \tag{5.8}$$

where $\mathcal{U}$, $\mathcal{N}$, $\mathcal{I}$ and $S^{(d)}_{AS_i}$ represent respectively the set of ASes, the set of non-stub ASes, the set of ASes part of at least one MSC optimal solution and the covering set of AS_i, x_{AS_i} is 1 if $S^{(d)}_{AS_i}$ is part of the MC problem solution, 0 otherwise, and y_{AS_j} is 1 if AS_j is part of the final coverage, 0 otherwise.

An *exact* solution $\mathcal{E}_k$ of this MC problem is the set of k ASes (5.5) that covers the largest set of non-stubs (5.4) chosen from the output of the MSC problem (5.6). However, it must be noted that it is not possible to use exact solutions to retrieve the desired AS ranking list, since $\mathcal{E}_k$ may differ from $\mathcal{E}_{k-1}$ by more than one element, and an AS that is part of $\mathcal{E}_{k-1}$ may be no longer part of $\mathcal{E}_k$.

Instead of looking for exact solutions, it is possible to exploit the greedy heuristic and the concept of dominance described above to obtain an approximate solution $\mathcal{G}_k$ which can be interpreted as the first k ASes that should be added to the route collectors on the basis of their potential contribution to the coverage. In detail, for a given k, the greedy heuristic consists in k steps, selecting at each step the AS which covers the maximum number of non-stubs currently uncovered. This means that $\mathcal{G}_k$ is obtained adding to $\mathcal{G}_{k-1}$ the AS selected by the heuristic at step k. It must be also noted that this approach is proved to be *at least* a $(1 - \frac{1}{e}) \approx 0.632$ approximation of $\mathcal{E}_k$ [65].

Fig. 5.7 depicts the pseudo-code of the greedy algorithm contextualized into our framework. Given the set $\mathcal{I}$ of ASes, the heuristic selects at each step the AS that covers the largest number of non-stubs and appends it into the ranking list $\mathcal{R}$ (lines 5-6). Then, it searches for alternatives of the selected AS, i.e. ASes covering the same

	# ASes					
	AF	*AP*	*EU*	*LA*	*NA*	*W*
ASes	770	6,576	17,657	2,490	16,032	41,127
Edges	1,980	19,829	77,465	8,175	43,331	144,475
Non-stub ASes	229	1,589	3,697	659	2,531	7,282
p2c	1,380	14,868	39,812	4,844	33,701	93,898
p2p	533	4,737	37,225	3,231	9,430	49,251
s2s	31	158	359	54	323	1,256

Table 5.4: Main characteristics of AS-level topologies

set of non-stub ASes currently covered by the selected AS (line 9). Finally, the remaining covering sets are updated by deleting from them the non-stub ASes covered by the selected AS (line 10). Instead of stopping when a given bound k is reached, as described in the classical MC formulation, the proposed algorithm stops when *all* the non-stubs ASes becomes covered (line 4). Note also that, each time an AS is added to the ranking list, it is possible to keep track of the percentage of non-stubs covered, thus the ranking list can also be used to identify which ASes should become feeders in order to cover a given percentage of non-stub ASes.

5.3 Towards an ideal route collector infrastructure

Finding a list of ASes that should become BGP feeders exploiting the methodology just illustrated entails computing the p2c-distances between ASes on a suitable economic-tagged AS-level Internet topology. A good starting point would seem to be the classic global AS-level topology of the Internet tagged according to one of the economic tagging algorithms proposed in the literature. However, as shown in [61], this could lead to misleading and incomplete results. An AS connection present in the global topology may hide multiple connections between the same two ASes but located in different geographic regions, each potentially regulated by different economic relationships. Applying our methodology to this coarse-level representation of

p2c-distance	*# not stub ASes*					
	AF	*AP*	*EU*	*LA*	*NA*	*W*
1	3	24	92	20	55	165
2	14	94	268	67	114	625
3	12	84	248	41	63	589
$>$ *3*	263	1,372	3,267	642	2,294	7,056

Table 5.5: Regional distribution of p2c-distances from RCs

the Internet may thus lead to an underestimation of the correct number of BGP feeders required to obtain a complete view of the Internet core. To illustrate this it is shown both the results obtained by applying our methodology to the global topology of the Internet, referred to as *World (W)*, and to five regional topologies: *Africa (AF)*, *Asia-Pacific (AP)*, *Europe (EU)*, *Latin America (LA)*, *North America (NA)*.

It will be firstly shown the impact of the geography in BGP feeder selection, highlighting that the analysis of the Internet from a global point of view underestimates the number of BGP feeders required. Then, it will be analysed the candidate feeders selected by our methodology, identifying their particular characteristics. Finally, it will be compared the coverage of the current BGP feeders with the ideal set drawn by our the results of the MSC methodology, and it will be shown how much the coverage of the route collector infrastructure would improve if new feeder would be wisely chosen via MC problem.

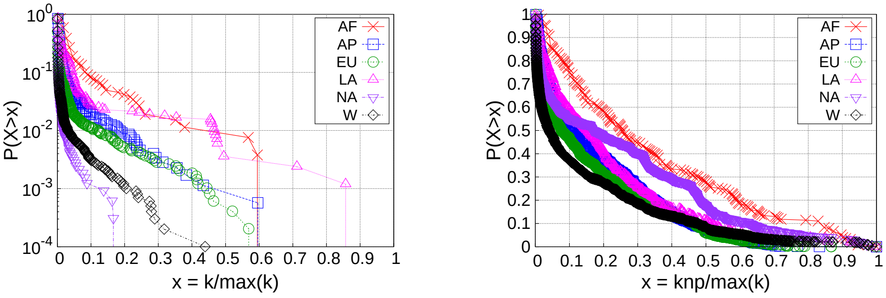
5.3.1 Global vs regional analysis

The global and the regional topologies have been inferred using the methodologies described in Section 3.2 and Section 4.4 with the most conservative time parameter, i.e. $N_{MAG} = 1$. Results for the remaining topologies can be found at [4]. The main characteristics of these topologies are reported in Table 5.4.

Firstly, it will be computed the p2c-distances for each of the available topologies, as required by the MSC procedure. Note that these values can also be used to highlight which zones of the Internet are poorly captured by the route collector infrastructure, thus providing further proof of the incompleteness of the current collected topologies. To confirm this, it will be analysed the p2c-distances of each non-stub AS from the route collector infrastructure by considering only routing information obtained via full feeders (see Table 5.5). Note that it will be considered to be ∞ the p2c-distance of ASes that cannot reach any route collector using only p2c connections. Most ASes are currently either too far from the route collector infrastructure or cannot be reached

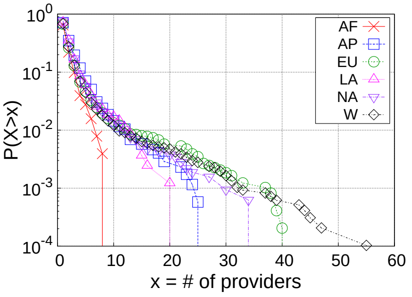
Region	$\lvert\mathcal{P}\rvert$ ($\lvert\mathcal{I}\rvert$)		
	$d = 1$	$d = 2$	$d = 3$
AF	171 (264)	157 (247)	156 (247)
AP	883 (1,757)	792 (1,582)	775 (1,530)
EU	2,285 (4,896)	2,082 (4,376)	2,049 (4,321)
LA	463 (837)	416 (735)	409 (735)
NA	1,544 (3,233)	1,439 (3,031)	1,421 (2,899)
W	4,653 (9,952)	4,083 (8,755)	3,958 (8,489)

Table 5.6: MSC procedure results



(a) Normalized node degree

(b) Normalized average provider degree



(c) Number of providers

Figure 5.8: CCDF of node properties of candidate BGP feeders

by any route collector via c2p connections alone, thus potentially representing hideouts for AS connectivity which need further investigation.

Once the p2c-distances had been calculated, it will be solved the related MSC problem concerning each of the economic topologies available. The results are summarized in Table 5.6, where are shown the cardinality of the solution set $\mathcal{P}$ and, in round brackets, the cardinality of the set of ASes that can be part of a solution ($\mathcal{I}$) for each topology. In each geographic scenario the number of BGP feeders required is significantly smaller than the number of non-stub ASes (Tables 5.4 and 5.6). More importantly, the sum of BGP feeders required by regional scenarios is higher than the number of those required by the *World* scenario. This result was expected since the complete capture of the connectivity of an AS with a large geographic range may entail deploying multiple BGP feeders around the world. Inter-regional ASes typically follow a regional principle to route their traffic, in order to maximize their performance and minimize latency [59, 85]. To do this, they tend to subdivide their ASes into different routing areas by exploiting the features of Interior Gateway Protocols (IGPs)

Region	$\lvert\mathcal{I}\rvert$ (% out of the total ASes)	
	On IXPs	*Stubs*
AF	38 (14.39 %)	92 (34.84 %)
AP	342 (19.46 %)	842 (47.94 %)
EU	1,033 (21.09 %)	2,602 (53.14 %)
LA	186 (22.22 %)	353 (42.17 %)
NA	264 (8.16 %)	1,713 (53.24 %)
W	1,731 (17.39 %)	5,247 (52.72 %)

Table 5.7: Characteristics of candidate BGP feeders

such as OSPF and IS-IS and set up connections that can only be exploited in regional traffic routing. A total of 1,016 of 7,903 non-stub ASes are present in more than one single geographical topology and thus may fit this description.

5.3.2 Candidate feeder analysis

Focus now on the characteristics of the elements of the set of candidate BGP feeders found by applying the monitor placement algorithm with parameter $d = 1$, which represents the best trade-off between AS-level connectivity discovery and the number of BGP feeders required. With $d = 1$, the positioning algorithm finds the set of BGP feeders required to obtain BGP routing information filtered by at most two BGP decision processes from each non-stub AS: the source AS and the BGP feeder itself. The results obtained with other values of d are available at [4].

Table 5.7 and Figure 5.8 show the most relevant characteristics of these elements. Figure 5.8 depicts $a)$ the degree distribution, $b)$ the *average provider degree* (k_{np}) distribution, where k_{np} is computed for each AS as the average degree of its providers, and $c)$ the number of providers of the candidate feeders. Note that graphics $a)$ and $b)$ have been normalized with the maximum value of the node degree (k) found in the

Region	$n = \lvert\mathcal{P}\rvert - \lvert\mathcal{F}\rvert$		
	$d = 1$	$d = 2$	$d = 3$
AF	169	156	155
AP	869	786	770
EU	2,251	2,057	2,025
LA	452	407	401
NA	1,526	1,425	1,408
W	4,600	4,403	3,930

Table 5.8: Additional feeders required in each region

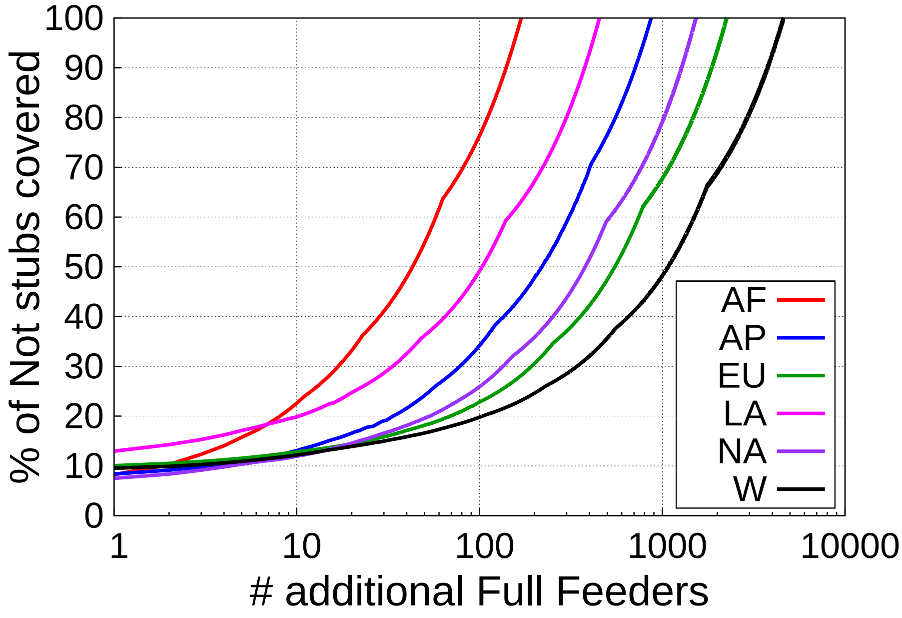


Figure 5.9: MC greedy algorithm results $(d = 1)$

related region, in order to allow full-scope analysis and to trace the characteristics of a typical candidate BGP feeder. The most frequent classes of ASes found to be candidate are: a) stub ASes (see Table 5.7), b) ASes that have set up a small number of BGP connections (see Fig. 5.8a), and c) ASes that have chosen a rather small number (see Fig. 5.8c) of small-medium ISPs (see low-medium values of the normalized k_{np} in Fig. 5.8b) to be their providers.

Only a small percentage of these ASes are present on at least one IXP (see Table 5.7), and this implies that their typical interconnectivity behavior is to not establish public peering with other ASes. Thus, it can concluded that the typical AS that should become a BGP feeder of the current route collector infrastructure is a small multihomed AS, that has set up multiple connections with different regional providers to guarantee route diversity and increase the reliability of its reachability. This is not surprising since these ASes are likely to be located at the bottom of the hierarchy and, thanks to multi-homing can cover several non-stub ASes at once.

		# of current BGP feeders $\in \mathcal{I}$ (% out of the total BGP feeders)					
		AF	*AP*	*EU*	*LA*	*NA*	*W*
Full feeders	IPv4	2 (66.67 %)	11 (50.00 %)	37 (39.78 %)	10 (62.50 %)	16 (26.67 %)	44 (28.94 %)
	IPv6	2 (100 %)	9 (50.00 %)	27 (37.50 %)	9 (52.94 %)	9 (28.13 %)	25 (32.89 %)
Partial feeders	IPv4	4 (66.67 %)	17 (40.48 %)	68 (29.69 %)	2 (66.67 %)	22 (23.91 %)	19 (32.75 %)
	IPv6	2 (50.00 %)	14 (37.84 %)	61 (29.61 %)	2 (10.53 %)	15 (23.81 %)	81 (24.69 %)
Minor feeders	IPv4	4 (33.33 %)	17 (34.81 %)	130 (34.48 %)	1 (50.00 %)	67 (30.88 %)	270 (30.68 %)
	IPv6	0 (– – %)	2 (50.00 %)	10 (31.25 %)	1 (25.00 %)	2 (20.00 %)	19 (36.53 %)

Table 5.9: Number of current feeders included in the set of elements candidated to be part of at least an optimal solution

5.3.3 Current status of the route collector infrastructure

Analyse now how many of the current BGP feeders are present in the ideal set of candidates. Their distribution for each region is shown in Table 5.9, as well as the percentage of the total number of feeders in the region that fall into that category. The main result is that only a small percentage of the current full feeders are actually part of an optimal solution in any of the topologies analysed. This is a direct consequence of their position in the Internet hierarchy. These ASes are not likely to have a large number of providers, thus their contribution is limited. Nevertheless, there are few other classes of current BGP feeders in the set of candidates, thus highlighting that, in terms of p2c-distance, only a few of them are placed in an optimal position. It is also interesting to understand how many BGP feeders should be added to the current route collector infrastructure in order to improve the quality of data. To determine these values, it can be used the methodology illustrated in Section 5.2.3 considering the current set of full feeders $\mathcal{F}$ as part of the initial set of solution $\mathcal{P}$, and considering the number of additional BGP feeders as $n = |\mathcal{P}| - |\mathcal{F}|$. Results per each geographic region are reported in Table 5.8. A comparison of the number of additional ASes required and the number of non-stub ASes (see Table 5.4) reveals that the methodology covered every non-stub AS with a number of new BGP feeders which is about 50-60% of the number of non-stub ASes in each region.

Region	**Current number of full feeders**	**# Not stub ASes covered (percentage)**	
		Current status	*Doubling Full Feeders*
AF	3	17 (5.82%)	36 (12.33%)
AP	24	118 (7.50%)	279 (17.73%)
EU	96	360 (9.29%)	870 (22.45%)
LA	23	87 (11.28%)	200 (25.94%)
NA	62	169 (6.69%)	536 (21.20%)
W	6	384 (5.27%)	2,137 (29.35%)

Table 5.10: Coverage improvements by doubling the number of full feeders

Finally, a priority for each AS found to be part of $\mathcal{P}$ can be identified by analysing the results of the related MC problem. Fig. 5.9 shows the percentage of non stub ASes covered in each region by introducing in the set of full feeders the ASes selected by each step k of the greedy algorithm. As can be seen, if new full feeders were chosen accordingly to the results of the MC problem, then it would be possible to capture a *more complete* AS-level view of the Internet with a limited amount of new elements and, consequently, with limited costs. For example, as can be seen from Table 5.8, just by doubling the number of full feeders in each region, it is possible to double also the coverage of the non stub ASes. Thus our methodology would seem to be a very useful tool to drive the growth of the route collector infrastructure. The complete priority list of ASes per region can be found at [4].

It must be stressed that our methodology extracts the optimal solution for the input data provided, thus, if a new BGP feeder is introduced, the solution may no longer be optimal. In any case, it still represents an upper bound to the number of additional full feeders needed. In fact, the introduction of new data may add previously hidden connections and may lead the tagging algorithm exploited to infer a higher number of p2c connections. These new connections may change the p2c-distance of several ASes that might be reached by exploiting a lower number of feeders. However, it would still be possible to apply the methodology once again on the new data to obtain a new optimal value.

6

Conclusions

The Internet is a complex system that evolved during the last two decades from a mere local research network to become a worldwide multipurpose network, often believed to be one of the greatest masterpieces of the human kind. Being mostly economic driven, its characteristics are extremely hard to be revealed. Researchers tried to shed some light during the years about that by exploiting route collectors, which currently represents the most reliable source of information about the inter-AS infrastructure of the Internet. However, they faced two major problems: Internet routing data publicly available is *incomplete*, i.e. only a small set of ASes agree to cooperate and participate to a route collector project, and *biased*, i.e. those ASes that agree to cooperate are mostly worldwide ISPs that use these projects as a gratuitous opportunity to advertise their reachability. This because the feeders which contributes with their full routing tables to the route collector projects are typically large ASes such as provider-free and worldwide ISPs, which makes the current vision of route collectors to not capture any of the p2p connections established by small or medium-sized ASes. Studies on Internet topology structure must be fully aware of the high level of data incompleteness, since a topological analysis of the Internet as viewed from these monitors is like analysing a roadmap of a given country where the highways are known, but most of the secondary roads are not shown! To quote Sir Arthur Conan Doyle, *“It is a capital mistake to theorize before you have all the evidence. It biases the judgment"*.

A possible solution to deal with such incompleteness is to increase the number of feeders. In this perspective, in this thesis is described a systematic methodology to *i*) infer the minimum number of feeders required to maximize the amount of AS-level routing data collected, and *ii*) retrieve a ranking list of candidate feeders, to understand the improvements that can be obtained with a limited amount of resources. Furthermore, it was found that multi-homed ASes with a small-medium size are the most useful contributors in a topology discovery perspective. The author is fully aware that it is extremely hard to convince the identified ASes to join a route collector, but it is his strong belief that the greatest problem is that ASes are not stimulated enough to join, since no direct service is offered in change of their voluntary participation. Thus, it is not hard to understand that currently route collectors are used mostly by large ISPs, that see in them a free-of-charge opportunity to advertise their network reachability. It might be a good idea to create services based on the real-time analysis of the inter-domain routing from different points of views in return of full routing tables, following the *do ut des* principle. Such services would be valuable for many ASes, ranging from local ISPs to CDNs, and would encourage their participation. In this perspective, the ranking list of candidate feeders provided by our methodology is extremely useful, since it allows to identify an initial limited set of potential users of these services. Another alternative to improve the amount of routing data available is to exploit tools based on active probes. Some of the traceroute-based projects developed so far [30, 11, 3, 81] are indeed able to bypass the reluctance in disclosing the routing information of AS owners by placing agents directly on user applications and, thus, obtaining data that would not be collected otherwise.

Despite the large incompleteness of data, another problem of the Internet analyses performed so far concerns the characteristics that are being analysed. It has been widely proved that the Internet cannot be studied as a mere undirected graph, and several efforts have been done during the years to delineate a proper methodology to enhance the undirected graph with a proper label indicating the routing choices applied by ASes [40, 42, 52, 59, 67, 87, 109] and to analyse the Internet at a geographic-level, in order to highlight characteristics that would be lost in a global analysis [48, 61, 73, 82, 85]. Much efforts still need to be done both in the inter-AS economic relationship inferences and in geographic inferences.

This thesis propose to enhance the state of the art by providing two methodologies to exploit BGP data to discover which economic relationships are established between couples of ASes working directly on BGP raw data. It was shown that BGP data can contain several spurious entries that are clearly in contrast with the valley-free rule introduced in [52]. These routes are potentially sources of error for any economic-driven AS-level analyses. It was thus provided two different approaches to deal with them. The first is an algorithm keen of these events and able to handle them using the lifespan of each AS Path. To infer inter-AS relationships, this algorithm relies on a priori knowledge of a list of provider-free ASes in order to understand whether an AS is transiting traffic for another AS. Using the concept of *two-way validation*, it is also able to assign a level of reliability to each inferred tagged connection from the Internet topology. The second algorithm proposed uses a new methodology to detect the effects of BGP path exploration phases, in order to purge BGP data from spurious events that would lead to wrong inferences about the Internet economic ecosystem. The tagging algorithm is still based on the valley-free principle and on the a priori knowledge of a small set of provider-free ASes in order to infer inter-AS relationships.

The most relevant result is that about one every two different AS paths collected by these projects is found only in spurious routes. This large amount of spuriousness has been proved to impact on the inferences drawn about the 8% of connections of the Internet AS-level topology.

Concerning geographic aspects, in this thesis it is proposed an innovative tagging algorithm able to geolocate AS connections starting from BGP data. This algorithm allowed to infer regional AS-level topologies, that have analyzed both from an undirected graph and economic perspective. From this analysis, it was found that the study of the Internet at a global level fails to display several characteristics that, on the other hand, play a fundamental role in the regional Internet connectivity. In particular were found strong evidence of structural differences between the European and the North American Internet topologies, that reflect different historical developments of the Internet in those regions. The same methodology proposed can be applied at other regional levels, such as countries. However, it must be noted that some studies [104, 92] have found that the more the geographic scope is narrow, the lower the precision of the IP geolocation database used is.

Every result introduced in this thesis can be found on the Isolario website [4].

6.1 Future works

There is plenty of room for improvements for the techniques proposed in this thesis. For example, the geographic tagging algorithm proposed assumes that two ASes establish a BGP session in each region where they are co-located and, consequently, the connection between these ASes may be not correctly geolocated. Moreover, this coarse-grained assumption does not allow to distinguish between regional and worldwide peer-to-peer connections. However, the most important step to do in the next future is to increase the quantity of data collected by route collectors to obtain the real structure of the Internet. In this thesis was proposed an useful methodology to select those ASes that are the most useful in enhancing the quality of topology information data. This kind of data is extremely hard to obtain, but it also true that the greatest problem in data gathering is that ASes are not stimulated enough to join any of the current projects, since no direct service is offered by the projects in change of the voluntary participation of ASes. Thus, it makes sense to imagine that currently the route collector projects are used by several large ISPs to promote their reachability. In particular, it might be a good idea to create services based on the real-time analysis of the inter-domain routing from different points of view in return of full routing tables, following the *do ut des* principle. Such services would be valuable for many ASes, ranging from local ISPs to CDNs, and thus encourage them to participate. Otherwise, it might be useful to exploit alternative tools to improve the amount of data available.

Specifically, some of the traceroute-based projects [3, 30, 45, 81] are able to bypass the reluctance in disclosing the routing information of AS owners by placing agents directly on user applications and, thus, obtaining data that would not be collected otherwise.

References

1. BGPmon. http://bgpmon.netsec.colostate.edu.
2. CAIDA UCSD Archipelago project. http://www.caida.org/projects/ark/.
3. DIMES project. http://www.netdimes.org.
4. IIT-CNR Isolario project. http://www.isolario.it.
5. Internet Assigned Numbers Authority. http://www.iana.org.
6. Maxmind GeoIPLite Database. http://www.maxmind.com/app/geoip_country.
7. Packet Clearing House. http://www.pch.net.
8. Potaroo - IPv4 Address Report. http://www.potaroo.net/tools/ipv4.
9. RIPE NCC Routing Information Service. http://www.ripe.net/data-tools/stats/ris/routing-information-service.
10. University of Oregon Route Views Project. http://www.routeviews.org.
11. University of Pisa Portolan project. http://portolan.iet.unipi.it.
12. University of Washington iPlane project. http://iplane.cs.washington.edu.
13. Wikipedia - Tier 1 network. http://en.wikipedia.org/wiki/Tier_1_network.
14. Dimitris Achlioptas, Aaron Clauset, David Kempe, and Cristopher Moore. On the bias of traceroute sampling: or, power-law degree distributions in regular graphs. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC '05)*, pages 694–703, 2005.
15. Bernhard Ager, Nikolaos Chatzis, Anja Feldmann, Nadi Sarrar, Steve Uhlig, and Walter Willinger. Anatomy of a Large European IXP. In *Proceedings of the 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '12)*, pages 163–174, 2012.
16. Réka Albert and Albert-László Barabási. Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.
17. Réka Albert, Hawoong Jeong, and Albert-László Barabási. The Internet's Achilles' Heel: Error and Attack Tolerance of Complex Networks. *Nature*, 406:378–382, 2000.
18. Brice Augustin, Balachander Krishnamurthy, and Walter Willinger. IXPs: mapped? In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement (IMC '09)*, pages 336–349, 2009.
19. Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *Science*, 286:509–512, 1999.
20. Adam Bender, Rob Sherwood, and Neil Spring. Fixing Ally's Growing Pains with Velocity Modeling. In *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement (IMC '08)*, pages 337–342, 2008.
21. Larry Blunk, Manish Karir, and Craig Labovitz. RFC 6396 - Multi-Threaded Routing Toolkit (MRT) Routing Information Export Format, 2011.

22. Béla Bollobás. *Random Graphs.* Academic press, London, UK, 1985.
23. Yuri Breitbart, Minos Garofalakis, Ben Jai, Cliff Martin, Rajeev Rastogi, and Avi Silberschatz. Topology Discovery in Heterogeneous IP Networks: the NetInventory System. *IEEE/ACM Transactions on Networking*, 12(3):401–414, 2004.
24. Xue Cai, John Heidemann, Balachander Krishnamurthy, and Walter Willinger. Towards an AS-to-Organization Map. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement (IMC '10)*, pages 199–205, 2010.
25. Juan Camilo Cardona Restrepo and Rade Stanojevic. A History of an Internet Exchange Point. *ACM SIGCOMM Computer Communication Review*, 42(2):58–64, 2012.
26. Shai Carmi, Shlomo Havlin, Scott Kirkpatrick, Yuval Shavitt, and Eran Shir. A Model of Internet Topology using k-shell Decomposition. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 104(27):11150–11154, 2007.
27. Hyunseok Chang. *An Economic-Based Empirical Approach to Modeling the Internet's Inter-Domain Topology and Traffic Matrix.* 2006.
28. Hyunseok Chang, Ramesh Govindan, Sugih Jamin, Scott Shenker, and Walter Willinger. Towards Capturing Representative AS-level Internet Topologies. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 44(6):737–755, 2004.
29. Hyunseok Chang, Jamin Sugih, and Walter Willinger. Inferring AS-level Internet Topology from Router-Level Path Traces. In *Proceedings of Workshop on Scalability and Traffic Control in IP Networks (ITCOM '01)*, pages 196–207, 2001.
30. Kai Chen, David R. Choffnes, Rahul Potharaju, Yan Chen, Fabian E. Bustamante, Dan Pei, and Yao 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*, pages 217–228, 2009.
31. Pei-chun Cheng, Beichuan Zhang, Daniel Massey, and Lixia Zhang. Identifying BGP Routing Table Transfers. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 55(3):213–218, 2011.
32. Pei-chun Cheng, Xin Zhao, Beichuan Zhang, and Lixia Zhang. Longitudinal Study of BGP Monitor Session Failures. *ACM SIGCOMM Computer Communication Review*, 40(2):34–42, 2010.
33. Kimberley C. Claffy. Border Gateway Protocol (BGP) and Traceroute Data Workshop Report. *ACM SIGCOMM Computer Communication Review*, 42(3):28–31, 2012.
34. David Clark. The Design Philosophy of the DARPA Internet Protocols. In *Proceedings of the 1988 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '88)*, pages 106–114, 1988.
35. Rami Cohen and Danny Raz. The Internet Dark Matter - on the Missing Links in the AS Connectivity Map. In *Proceedings of the 25th IEEE International Conference on Computer Communications (INFOCOM '06)*, pages 1–12, 2006.
36. Thomas H. Cormen, Clifford Stein, Ronald L. Rivest, and Charles E. Leiserson. *Introduction to Algorithms.* McGraw-Hill Higher Education, 2nd edition, 2001.
37. Olivier Coudert. Two-level Logic Minimization: an Overview. *Integration of the VLSI Journal*, 17(2):97–140, 1994.
38. Shivani Deshpande, Marina Thottan, Tin Kam Ho, and Biplab Sikdar. An Online Mechanism for BGP Instability Detection and Analysis. *IEEE/ACM Transaction on Computers*, 58(11):1470–1484, 2009.
39. Amogh Dhamdhere, Himalatha Cherukuru, Constantine Dovrolis, and KC Claffy. Measuring The Evolution of Internet Peering Agreements. In *Proceedings of the 11th International IFIP TC-6 Conference on Networking (NETWORKING '12)*, volume 2, pages 136–148, 2012.

40. Giuseppe Di Battista, Thomas Erlebach, Alexander Hall, Maurizio Patrignani, Maurizio Pizzonia, and Thomas Schank. Computing the Types of the Relationships Between Autonomous Systems. *IEEE/ACM Transactions on Networking*, 15(2):267–280, 2007.
41. Xenofontas Dimitropoulos, Dmitri Krioukov, Marina Fomenkov, Bradley Toward Topology Dualism: Improving the Accuracy of AS Annotations for RoutersHuffaker, Young Hyun, Kimberley C. Claffy, and George Riley. AS Relationships: Inference and Validation. *ACM SIGCOMM Computer Communication Review*, 37(1):29–40, 2007.
42. Xenofontas Dimitropoulos, Dmitri Krioukov, Bradley Huffaker, KC Claffy, and George Riley. Inferring AS Relationships: Dead End or Lively Beginning? In *Proceedings of the 4th International Conference on Experimental and Efficient Algorithms (WEA '05)*, pages 113–125, 2005.
43. Benoit Donnet and Timur Friedman. Internet Topology Discovery: a Survey. *IEEE Communications Survey and Tutorials*, 9(4):56–69, 2007.
44. Alex Fabrikant, Umar Syed, and Jennifer Rexford. There's something about MRAI: Timing diversity can exponentially worsen BGP convergence. In *Proceedings of the 30th IEEE International Conference on Computer Communications (INFOCOM '11)*, pages 2966–2974, 2011.
45. Adriano Faggiani, Enrico Gregori, Luciano Lenzini, Simone Mainardi, and Alessio Vecchio. On the Feasibility of Measuring the Internet Through Smartphone-based Crowdsourcing. In *Proceedings of the 8th International Workshop on Wireless Network Measurements (WINMEE '12)*, pages 1–6, 2012.
46. Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On Power-law Relationships of the Internet Topology. In *Proceedings of the 1999 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '99)*, pages 251–262, 1999.
47. Dima Feldman and Yuval Shavitt. Automatic Large Scale Generation of Internet PoP Level Maps. In *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '08)*, pages 2426–2431, 2008.
48. Dima Feldman, Yuval Shavitt, and Noa Zilberman. A Structural Approach for PoP Geo-Location. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 56(3):1029–1040, 2012.
49. Anja Feldmann, Olaf Maennel, Z. Morley Mao, Arthur Berger, and Bruce Maggs. Locating Internet Routing Instabilities. In *Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '04)*, pages 205–218, 2004.
50. Ashley Flavel, Olaf Maennely, Belinda Chiera, Matthew Roughan, and Nigel Bean. Clean-BGP: Verifying the Consistency of BGP Data. In *Proceedings of the 2008 ICNP Workshop on Internet Network Management (INM '08)*, pages 1–6, 2008.
51. Ashley Flavel, Matthew Roughan, Nigel Bean, and Olaf Maennel. Modeling BGP Table Fluctuations. In *Proceedings of the 20th International Teletraffic Conference on Managing Traffic Performance in Converged Networks (ITC20 '07)*, pages 141–153, 2007.
52. Lixin Gao. On Inferring Autonomous System Relationships in the Internet. *IEEE/ACM Transactions on Networking*, 9(6):733–738, 2001.
53. Lixin Gao and Feng Wang. The Extent of AS Path Inflation by Routing Policies. In *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '02)*, volume 3, pages 2180–2184, 2002.
54. Michael R. Garey and David S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1990.
55. James F. Gimpel. A Reduction Technique for Prime Implicant Tables. *IEEE Transactions on Electronic Computers*, 14(4):535–541, 1965.
56. Ramesh Govindan and Anoop Reddy. An Analysis of Internet Inter-Domain Topology and Route Stability. In *Proceedings of the 16th IEEE International Conference on Computer Communications (INFOCOM '97)*, pages 850–857, 1997.

57. Ramesh Govindan and Hongsuda Tangmunarunkit. Heuristics for Internet Map Discovery. In *Proceedings of the 19th IEEE International Conference on Computer Communications (INFOCOM '00)*, volume 3, pages 1371–1380, 2000.
58. Enrico Gregori, Alessandro Improta, Luciano Lenzini, and Chiara Orsini. The Impact of IXPs on the AS-level Topology Structure of the Internet. *Computer Communications*, 34(1):68–82, 2010.
59. Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, and Luca Sani. BGP and Inter-AS Economic Relationships. In *Proceedings of the 10th International IFIP TC-6 Conference on Networking (NETWORKING '11)*, volume 2, pages 54–67, 2011.
60. Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, and Luca Sani. Inferring Geography from BGP Raw Data. In *Proceedings of the 4th IEEE International Workshop on Network Science for Communication Networks (NETSCICOM '12)*, pages 208–213, 2012.
61. Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, and Luca Sani. On the Incompleteness of the AS-level Graph: a Novel Methodology for BGP Route Collector Placement. In *Proceedings of the 12th ACM SIGCOMM Conference on Internet Measurement (IMC '12)*, pages 253–264, 2012.
62. Mehmet H. Gunes and Kamil Sarac. Analytical IP Alias Resolution. In *Proceedings of the IEEE International Conference on Communications (ICC '06)*, volume 1, pages 459–464, 2006.
63. John Hawkinson and Tony Bates. RFC 1930 - Guidelines for Creation, Selection, and Registration of an Autonomous System (AS), 1996.
64. Yihua He, Georgos Siganos, Michalis Faloutsos, and Srikanth V. Krishnamurthy. Lord of the Links: A Framework for Discovering Missing Links in the Internet Topology. *IEEE/ACM Transactions on Networking*, 17(2):391–404, 2009.
65. Dorit S. Hochbaum. *Approximation Algorithms for NP-hard Problems*. PWS Publishing Co., Boston, MA, USA, 1997.
66. Bradley Huffaker, Amogh Dhamdhere, Marina Fomenkov, and Kimberley C. Claffy. Toward Topology Dualism: Improving the Accuracy of AS Annotations for Routers. In *Proceedings of the 11th International Conference on Passive and Active Measurement (PAM '10)*, pages 101–110, 2010.
67. Benjamin Hummel and Sven Kosub. Acyclic Type-of-Relationship Problems on the Internet: an Experimental Analysis. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement (IMC '07)*, pages 221–226, 2007.
68. Geoff Huston. *ISP Survival Guide: Strategies for Running a Competitive ISP*. John Wiley & Sons, Inc., New York, NY, USA, 1999.
69. Geoff Huston, Mattia Rossi, and Grenville Armitage. A Technique for Reducing BGP Update Announcements Through Path Exploration Damping. *IEEE Journal on Selected Areas in Communications*, 28(8):1271–1286, 2010.
70. Ken Keys. Internet-scale IP Alias Resolution Techniques. *ACM SIGCOMM Computer Communication Review*, 40(1):50–55, 2010.
71. Ken Keys, Hyun Young, Matthew Luckie, and Kimberley C. Claffy. Internet-Scale IPv4 Alias Resolution with MIDAR. *to appear in IEEE/ACM Transactions on Networking*.
72. Craig Labovitz, Abha Ahuja, Abhijit Bose, and Farnam Jahanian. Delayed Internet Routing Convergence. *IEEE/ACM Transactions on Networking*, 9(3):293–306, 2001.
73. Sandor Laki, Peter Matray, Peter Haga, Tamas Sebok, Istvan Csabai, and Gabor Vattay. Spotter: A Model Based Active Geolocation Service. In *Proceedings of the 30th IEEE International Conference on Computer Communications (INFOCOM '11)*, pages 3173–3181, 2011.
74. Ron Larson and Elizabeth Farber. *Elementary Statistics: Picturing the World*. Prentice Hall College Div, 5th edition, 2004.

75. Barry M. Leiner, Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Larry G. Roberts, and Stephen Wolff. A Brief History of the Internet. *ACM SIGCOMM Computer Communication Review*, 39(5):22–31, 2009.
76. Mike Little. Goals and Functional Requirements for Inter-Autonomous System Routing, 1989.
77. Kirk Lougheed and Yakov Rekhter. RFC 1105 - Border Gateway Protocol (BGP), 1989.
78. Ratul Mahajan, David Wetherall, and Tom Anderson. Understanding BGP Misconfiguration. In *Proceedings of the 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '02)*, pages 3–16, 2002.
79. Zhuoqing M. Mao, David Johnson, Jennifer Rexford, Jia Wang, and Randy H. Katz. Scalable and Accurate Identification of AS-level Forwarding Paths. In *Proceedings of the 23rd IEEE International Conference on Computer Communications (INFOCOM '04)*, volume 3, pages 1605–1615, 2004.
80. Zhuoqing M. Mao, Jennifer Rexford, Jia Wang, and Randy H. Katz. Towards an Accurate AS-level Traceroute Tool. In *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '03)*, pages 365–378, 2003.
81. Pietro Marchetta, Pascal Mérindol, Benoit Donnet, Antonio Pescapé, and Jean-Jacques Pansiot. Topology Discovery at the Router Level: a New Hybrid Tool Targeting ISP Networks. *IEEE Journal on Selected Areas in Communications*, 29(6), 2011.
82. Peter Matray, Peter Haga, Sandor Laki, Istvan Csabai, and Gabor Vattay. On the Network Geography of the Internet. In *Proceedings of the 30th IEEE International Conference on Computer Communications (INFOCOM '11)*, pages 126–130, 2011.
83. Edward J. McCluskey. Minimization of Boolean Functions. *Bell System Technical Journal*, 35(6):1417–1444, 1956.
84. Wolfgang Mühlbauer, Anja Feldmann, Olaf Maennel, Matthew Roughan, and Steve Uhlig. Building an as-topology model that captures route diversity. In *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '06)*, pages 195–206, 2006.
85. William B. Norton. *The Internet Peering Playbook: Connecting to the Core of the Internet*. DrPeering Press, Palo Alto, CA, USA, 2011.
86. T. Sasao O. Coudert. *Two-Level Logic Minimization, Logic Synthesis and Verification*. Kluwer Academic Publishers, 2001.
87. Ricardo Oliveira, Dan Pei, Walter Willinger, Beichuan Zhang, and Lixia Zhang. The (In)Completeness of the Observed Internet AS-level Structure. *IEEE/ACM Transactions on Networking*, 18(1):109–122, 2010.
88. Ricardo Oliveira, Beichuan Zhang, Dan Pei, Rafit Izhak-Ratzin, and Lixia Zhang. Quantifying Path Exploration in the Internet. In *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement (IMC '06)*, pages 269–282, 2006.
89. Jean-Jacques Pansiot and Dominique Grad. On Routes and Multicast Trees in the Internet. *ACM SIGCOMM Computer Communication Review*, 28(1):41–50, 1998.
90. Dan Pei, Beichuan Zhang, Daniel Massey, and Lixia Zhang. An Analysis of Convergence Delay in Path Vector Routing Protocols. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 50(3):398–421, 2006.
91. Larry L. Peterson and Bruce S. Davie. *Computer Networks: A Systems Approach, 3rd Edition*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003.
92. Ingmar Poese, Steve Uhlig, Mohamed A. Kaafar, Benoit Donnet, and Bamba Gueye. IP Geolocation Databases: Unreliable? *ACM SIGCOMM Computer Communication Review*, 41(2):53–56, 2011.
93. Jon Postel. RFC 791 - Internet Protocol, 1981.
94. Willard V. Quine. A Way to Simplify Truth Functions. *The American Mathematical Monthly*, 62(9):627–631, 1955.

95. Willard V. Quine. On Cores and Prime Implicants of Truth Functions. *The American Mathematical Monthly*, 66(9):755–760, 1959.
96. Yakov Rekhter and Tony Li. RFC 1654 - A Border Gateway Protocol 4 (BGP-4), 1994.
97. Yakov Rekhter, Tony Li, and Susan Hares. RFC 4271 - A Border Gateway Protocol 4 (BGP-4), 2006.
98. Jennifer Rexford, Jia Wang, Zhen Xiao, and Yin Zhang. BGP routing stability of popular destinations. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement (IMW '02)*, pages 197–202, 2002.
99. Matei Ripeanu, Adriana Iamnitchi, and Ian Foster. Mapping the Gnutella Network. *IEEE Internet Computing*, 6(1):50–57, 2002.
100. Eric C. Rosen. RFC 827 - Exterior Gateway Protocol (EGP), 1982.
101. Matthew Roughan, Simon Jonathan Tuke, and Olaf Maennel. Bigfoot, Sasquatch, the Yeti and Other Missing Links: What We Don't Know About the AS Graph. In *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement (IMC '08)*, pages 325–330, 2008.
102. Matthew Roughan, Walter Willinger, Olaf Maennel, Debbie Perouli, and Randy Bush. 10 Lessons from 10 Years of Measuring and Modeling the Internet's Autonomous Systems. *IEEE Journal on Selected Areas in Communications*, 29(9):1810–1821, 2011.
103. Amit Sahoo, Krishna Kant, and Prasant Mohapatra. Characterization of BGP Recovery Time under Large-Scale Failures. In *Proceedings of the IEEE International Conference on Communications (ICC '06)*, volume 2, pages 949–954, 2006.
104. Yuval Shavitt and Noa Zilberman. A Geolocation Databases Study. *IEEE Journal on Selected Areas in Communications*, 29(10):2044–2056, 2011.
105. Rob Sherwood, Adam Bender, and Neil Spring. Discarte: a Disjunctive Internet Cartographer. In *Proceedings of the 2008 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '08)*, pages 303–314, 2008.
106. Georgos Siganos, Sudhir Leslie Tauro, and Michalis Faloutsos. Jellyfish: a Conceptual Model for the AS Internet Topology. *Journal of Communications and Networks*, 8(3):339–350, 2006.
107. Neil Spring, Ratul Mahajan, and David Wetherall. Measuring ISP Topologies with Rocketfuel. In *Proceedings of the 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '02)*, pages 133–145, 2002.
108. Daniel Stutzbach, Reza Rejaie, and Subhabrata Sen. Characterizing Unstructured Overlay Topologies in Modern P2P File-sharing Systems. *IEEE/ACM Transactions on Networking*, 16(2):267–280, 2008.
109. Lakshminarayanan Subramanian, Sharad Agarwal, Jennifer Rexford, and Randy H. Katz. Characterizing the Internet Hierarchy From Multiple Vantage Points. In *Proceedings of the 21st IEEE International Conference on Computer Communications (INFOCOM '02)*, volume 2, pages 618–627, 2002.
110. Quaizar Vohra and Enke Chen. RFC 4893 - BGP Support for Four-octet AS Number Space, 2007.
111. Matthias Wählisch, Thomas C. Schmidt, Markus Brün, and Thomas Häberlen. Exposing a Nation-centric View on the German Internet — a Change in Perspective on AS-Level. In *Proceedings of the 13th International Conference on Passive and Active Measurement (PAM '12)*, pages 200–210, 2012.
112. Duncan J. Watts and Steven H. Strogatz. Collective Dynamics of 'small-world' Networks. *Nature*, 393(6684):440–442, 1998.
113. Walter Willinger, David Alderson, and John C. Doyle. Mathematics and the Internet: A Source of Enormous Confusion and Great Potential. *Notices of the American Mathematical Society*, 56(5):586–599, 2009.

114. Gerhard J. Woeginger. Exact Algorithms for NP-Hard Problems: A Survey. *Combinatorial Optimization - Eureka, You Shrink! - Lecture Notes in Computer Science*, 2570:185–207, 2003.
115. Kuai Xu, Zhenhai Duan, Zhi-Li Zhang, and Jaideep Chandrashekar. On Properties of Internet Exchange points and Their Impact on AS Topology and Relationship. In *Proceedings of the 3rd International IFIP TC-6 Conference on Networking (NETWORKING '04)*, pages 284–295, 2004.
116. Soon-Hyung Yook, Hawoong Jeong, and Albert-László Barabási. Modeling the Internet's Large-scale Topology. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 99(21):13382–13386, 2002.
117. Beichuan Zhang, Raymond Liu, Daniel Massey, and Lixia Zhang. Collecting the Internet AS-level Topology. *ACM SIGCOMM Computer Communication Review*, 35(1):53–61, 2005.
118. Yu Zhang, Ricardo Oliveira, Yangyang Wang, Shen Su, Baobao Zhang, Jun Bi, Hongli Zhang, and Lixia Zhang. A Framework to Quantify the Pitfalls of Using Traceroute in AS-Level Topology Measurement. *IEEE Journal on Selected Areas in Communications*, 29(9):1822–1836, 2011.
119. Yu Zhang, Ricardo Oliveira, Hongli Zhang, and Lixia Zhang. Quantifying the Pitfalls of Traceroute in AS Connectivity Inference. In *Proceedings of the 11th International Conference on Passive and Active Measurement (PAM '10)*, pages 91–100, 2010.
120. Shi Zhou, Guo-Qiang Zhang, and Guo-Qing Zhang. Chinese Internet AS-Level Topology. *IET Communications*, 1(2):209–214, 2006.

Acknowledgments

Foremost, I would like to thank my advisors Prof. Luciano Lenzini and Enrico Gregori for their support and their patience. Their guidance helped me in all these years of pursuiting my PhD.

Besides my advisor, I would also like to express my most sincere gratitude to Walter Willinger and Balachander Krishnamurthy, that opened my mind to new working scenarios while I was visiting them at the AT&T Labs. Their wide knowledge and different way of thinking have been of great value for me. In particular, without having attended the talk "Lies, Damned Lies, and Statistics" that Walter did in Pisa during the last few months of my master thesis, I do not think that I would achieved any of the results provided in this thesis.

I am also grateful to Prof. Massimo Pappalardo and Prof. Paolo Nobili for for their very useful advice provided during the formulation and solution of the minimization/maximization problems.

I would warmly thank Isabel Cabrera, Melva Jackson, David Lenicek, Howard Karloff, Flip Korn and David Kormann, for they encouragement and support during my (several) comic adventures in the United States. Without their help, everything would have been much harder for me!

I would like to thank also my fellow labmates both in IIT/CNR and in the Information Engineering Department: Valerio Arnaboldi, Carmine Benedetto, Luca Bisti, Cristiano Bonato Both, Rafael Kunst, Adriano Faggiani, Pietro G. Giardina, Massimiliano La Gala, Valerio Luconi, Giovanni Minutiello, Maddalena Nurchis, Chiara Orsini, Fabio Pezzoni, Alessandro Pischedda, Dario Puliero, Marco Tiloca, Carlo Vallati and Valerio Lorenzo, for the stimulating discussions, and for all the fun we have had in the last years.

A special thank goes to Luca Sani and Lorenzo Rossi, for being my friends even before colleagues. The sleepless nights we were working together before deadlines, they discussions we had about everything and the fun we had together will never be forgotten.

Last but not least, I would like to thank my family and my closest friends Alice, Chiara, Claudia, Federico, Iolanda and Nicola. These years have been extremely hard for me, and I doubt I would have accomplished any of the work described in this thesis without the strong support and encouragement of any of them.