IMT Institute for Advanced Studies, Lucca

Lucca, Italy

Modeling and Analysis of the Internet Topology

PhD Program in Computer Science and Engineering

XXVI Cycle

By

Giovanni Accongiagioco

2014

The dissertation of Giovanni Accongiagioco is approved.

Program Coordinator: Prof. Rocco De Nicola, IMT Institute for Advanced Studies, Lucca

Supervisor: Prof. Luciano Lenzini, University of Pisa

Supervisor: Ing. Enrico Gregori, IIT-CNR, Pisa

Tutor: Prof. Rocco De Nicola, IMT Institute for Advanced Studies, Lucca

The dissertation of Giovanni Accongiagioco has been reviewed by:

Prof. Rachid El-Azouzi, Univesité d'Avignon

Prof. Emanuele Giovannetti, Anglia Ruskin University

Prof. Avi Miron, Technion – Israel Institute of Technology

IMT Institute for Advanced Studies, Lucca

2014

a mia Madre, a mio Padre,
ai miei Nonni che non sono più qui,
a Michela.

Contents

List of Figures

List of Tables

List of Algorithms

Acknowledgements

The Acknowledgements section is divided in three parts: the first part lists the co-authored published and work in progress papers, identifying coauthors and funding projects. The second part contains "*academic*" acknowledgements, and is written in *English*, while the last part contains "*personal*" acknowledgements, and is written in *Italian*. The goal is both to ensure an appropriate writing style (i.e. maybe I have a better knowledge of the Italian language, although it does not necessarily mean that this makes me a better writer), and to give all the interested readers the possibility to understand.

This thesis is based on [AGL13; AGL14b; AGL14a], joint works with Enrico Gregori (IIT-CNR Pisa) and Luciano Lenzini (University of Pisa), and on [AAGL14b; AAGL14a], co-authored by Eitan Altman, INRIA, Sophia Antipolis, Enrico Gregori and Luciano Lenzini.

First and foremost, I would like to thank my supervisors Prof. Luciano Lenzini and Ing. Enrico Gregori for their support and guidance through all these years of pursuing my PhD. Besides my supervisors, I would also like to express my most sincere gratitude to Prof. Eitan Altman, who enlarged my academic vision during my visiting at the University of Avignon. His wide knowledge and different way of thinking has been of great value to me.

I would like to thank the reviewers, for improving my thesis by highlighting its strong and weak points, and for envisioning new working scenarios and paths of future research.

I am also grateful to Prof. Giovanni Stea for allowing me to successfully continue the work started during my master thesis. I would also like to thank Antonio Virdis and my former colleagues and labmates Matteo Maria Andreozzi, Daniele Migliorini, Generoso Pagano and Vincenzo Pii for their collaboration and for continuing our work together, which has led to the release of an LTE simulation framework.

I would like to thank Alexandre Reiffers for helping me moving my first steps in the world of Game Theory and for all the interesting and stimulating discussions. I would also like to thank Rodrigo, Yonathan, Adrian, Manjesh and all the other people in Avignon whom I had interesting discussions with. I would like to thank Eli Meirom and Prof. Ariel Orda for having provided me, during my short visit in Israel, a new path to explore for my upcoming works.

Finally, I would like to thank all the other people who I collaborated with or simply share thoughts, either at the IMT, at IIT/CNR or the Information Engineering Department: Luca, Francesca, Olga, Stefano, Massimo, Simone, Chiara, Alessandro, Valerio, Adriano, and my labmates.

Anzitutto voglio ringraziare i miei genitori, perché mi supportano ma soprattutto perché mi sopportano. Lasciare andare il proprio unico figlio non deve essere facile, ma voi continuate a sostenermi nel mio cammino senza (quasi) mai farmelo pesare. Grazie per tutte le cose che avete fatto per me, soprattutto per quelle che magari prima non ero in grado di capire, perché ancora oggi sono il mio primo stimolo per crescere domani.

Grazie ai miei amici di sempre, Luigi ma anche Tato e Fefo. Grazie per essere amici con cui poter condividere ogni cosa. Grazie soprattutto per essere come quei nodi, magari lenti, ma che più si mettono in tensione e più si stringono: anche quando la distanza mette il nostro rapporto in quiescenza, so

di avere in voi un rifugio sicuro nelle avversità della vita.

Grazie ai miei coinquilini vecchi e nuovi, perché grazie a voi mandare avanti una casa è stato sì un onere, ma anche un piacere: Gabriele, Marco, Roberto, Dimitri, Catia, Anna Chiara, Karim, e tutti gli altri, anche se hanno fatto un'apparizione più breve. Grazie ai miei amici "pisani" (anche se questa parola sarà quasi un insulto per voi tutti): Gabriele, Dimitri (anche se già inseriti tra i coinquilini), Pietro, Alessandro, Leonardo e Andrea, per avermi accompagnato in quest'avventura, attraverso vaneggiamenti, disquisizioni e missioni al limite dell'improbabilità.

Grazie a Fernando e Rita, per avermi accolto in casa vostra, e grazie a Benni perché mi insegna i segreti di una relazione in cui non si chiede (quasi) niente e si da tutto.

Grazie a P. Alfio, P. Francesco, P. Giovanni e P. Gianluca, perché poche vostre parole mi hanno insegnato più sulla vita di quanto avessi capito in oltre un quarto di secolo. Grazie per aver smontato tutti i miei preconcetti, i miei rancori, grazie per avermi mostrato quanto può essere bello vivere di un'acqua viva e dei frutti che porta a suo tempo. Grazie a Dio, per il suo perdono al di la del mio rifiuto, perché mi accompagna in terra arida come in quella fertile.

Dulcis in fundo, Grazie a Michela. Grazie perché la tua parola ed i tuoi gesti sono luce sul mio cammino. Grazie per ciò che sei per me, ma soprattutto grazie per ciò che sono io quando sono con te. Grazie perché mi permetti di amarti, grazie per il tuo sì. Grazie perché attraverso di te ho conosciuto il vero volto dell'amore, quell'amore che chiedo a Dio, per donarlo a te ogni giorno, per tutti i giorni della mia vita.

Thanks,
Grazie,
Giovanni

Vita

December 14, 1986	Born, San Giovanni in Fiore (CS), Italy
September 2005 - July 2008	Bachelor Degree in Computer Engineering Final mark: 110/110 cum laude University of Pisa, Italy
September 2008 - December 2010	Master Degree in Computer Engineering Final mark: 110/110 cum laude University of Pisa, Italy
January 2011 - February 2011	Research Assistant on the subject Internet Topology Modeling and Analysis IIT-CNR, Pisa, Italy
March 2011 - July 2014	PhD Student in CSE XXVI IMT Lucca, Italy
October 2012 - At Present	Active member of the European FP7-ICT Integrated Project 317672 CONGAS
January 2013 - July 2013	Visiting Student University of Avignon, France
March 2014 - At Present	Associate Researcher IIT-CNR, Pisa, Italy

Publications

1. G. Accongiagioco, M. M. Andreozzi, D. Migliorini, G. Stea, "Throughput-optimal Resource Allocation in LTE-Advanced with Distributed Antennas", *in Elsevier Computer Networks*, Volume 57, Issue 18, 24 December 2013, Pages 3997–4009

2. G. Accongiagioco, E. Altman, E. Gregori, L. Lenzini, "Peering vs Transit: a Game Theoretical model for Autonomous Systems connectivity", *in Proceedings of IFIP Networking 2014,* To Appear

3. G. Accongiagioco, E. Altman, E. Gregori, L. Lenzini, "A Game Theoretical study of Peering vs Transit in the Internet", *in NetSciCom 2014: 6th IEEE INFOCOM 2014 Workshop,* Pages 795-800

4. G. Accongiagioco, E. A. Meirom, E. Gregori, L. Lenzini, S. Mannor, A. Orda, "Agent-Based Model for Internet Hub Formation", *in European Conference on Complex Systems ECCS'14, 2014* (POSTER), To Appear

5. G. Accongiagioco, E. Gregori, L. Lenzini, "A Structure-Based Topology Generator for the Internet's Core", *4th InfQ Workshop*, Sorrento, Italy, June 2013, Italian workshop without proceedings, paper selected for journal extension

6. G. Accongiagioco, M. M. Andreozzi, D. Migliorini, G. Stea, "Throughput-optimal Resource Allocation in LTE-Advanced with Distributed Antennas", *3rd InfQ Workshop*, Lucca, Italy, July 2012, Italian workshop without proceedings, paper selected for journal extension

7. G. Accongiagioco, E. Gregori, L. Lenzini, "X-CENTRIC: an IXP-Centric Evolving NeTwoRk model for the Internet's Core", *submitted to IEEE Transactions on Networking, 2014*

8. G. Accongiagioco, E. Gregori, L. Lenzini, "S-BITE: a Structure-Based Internet Topology gEnerator", *submitted to Elsevier Computer Networks, 2014*

Patents

1. G. Accongiagioco, M. M. Andreozzi, M. Caretti, R. Fantini, D. Migliorini, D. Sabella, G. Stea "Scheduling Algorithm for Wireless Communication Networks"; joint work with TELECOM ITALIA S.p.A. Pub. No. WO/2013/097904, published 04/07/2013.

Presentations

1. G. Accongiagioco, "Traffic Measurement in the Internet AS-Level Environment using DNS data: Why and How", *at INRIA,* Sophia Antipolis, France, 2014

2. G. Accongiagioco, "Peering vs Transit: a Game Theoretical model for Autonomous Systems connectivity", *in IFIP Networking 2014,* Trondheim, Norway, 2014

3. G. Accongiagioco, "A Game Theoretical study of Peering vs Transit in the Internet", *in NetSciCom 2014,* Toronto, Canada, 2014

4. G. Accongiagioco, "X-CENTRIC: an IXP-Centric Evolving NeTwoRk model for the Internet's Core", *at IIT-CNR,* Pisa, 2014

5. G. Accongiagioco, "Peering vs Transit: a Game Theoretical model for Autonomous Systems connectivity", *at University of Technion,* Haifa, Israel, 2014

6. G. Accongiagioco, "A Structure-Based Topology Generator for the Internet's Core", *at 4th InfQ Workshop*, Sorrento, Italy, 2013

7. G. Accongiagioco, "Modeling and Analysis of the Internet's core via maximal cliques", *at University of Avignon,* Avignon, France, 2012

8. G. Accongiagioco, "Throughput-optimal Resource Allocation in LTE-Advanced with Distributed Antennas", *at 3rd InfQ Workshop*, Lucca, Italy, July 2012

Software

1. G. Accongiagioco, D. Migliorini, G. Pagano, V. Pii, A. Virdis *"SimuLTE™"*, simulation framework based on *OMNEST* for system level performance evaluation of *LTE* and LTE Advanced networks (*3GPP Release 8 and beyond*), released under the *GNU Lesser General Public License*.

Abstract

Over the last decades, the Internet has grown from a small-scale academic network linking few universities, to a large-scale complex system, reaching out to billions of people and enabling communications and computing worldwide. While the number of persons connected to the Internet 24/7 is growing exponentially, particularly through smartphones, most of them are unaware of the real Internet infrastructure. Recently, researchers have attempted to discover details about the characteristics of the Internet in order to create a model of its infrastructure, which could be exploited both to improve the performances and to identify and address possible weaknesses of the network. Despite several efforts in this direction, currently no model is known to represent the Internet effectively, especially due to the lack of understanding of the true driving forces behind the Internet evolution, and the excessively coarse granularity applied by the studies done to date. This thesis seeks to scientifically understand the driving forces lying behind the exponential growth of the "network of networks", through a fine-grained analysis of the available topological data. In the first instance, we show that the large-scale Internet can be broken down into a not-so-large core, which captures the most important structural properties of the Internet, and a periphery, representing the "tendrils" of the topology. The proposed decomposition technique is general and can be extended to other networks: multi-layer network analysis, a hot topic in the area of complex networks, can benefit from it as well. We point out the key role of Internet eXchange Points within the core of the Internet, and reveal through an economic analysis the emerging competition be-

tween them and the classical Network Service Providers. This analysis serves as ground truth to devise a network model able to represent the Internet's core effectively and potentially forecast its evolution. The model takes into account multi-layer interaction mechanisms, and can be further extended to the whole Internet, by devising simple attaching mechanisms for the periphery. Since the prediction properties of the model are based on data from real measurements, it is important to remove the biases introduced by the measuring infrastructure, in order to predict the future evolution of the real Internet topology. In the end, we show through a meaningful set of metrics, how the model is able to successfully capture the Internet's statistical and structural properties, outperforming existing topology generators in the literature.

Part I

Introductory Part

Chapter 1

Introduction

This chapter introduces the area of investigation, the motivation behind this work and its contributions. At the end it presents an overview of the thesis' structure.

1.1 Research Area

The Internet has become more and more part of the everyday life of billions of people. The spreading of smartphones, together with the deployment of high-speed wireless data networks such as LTE [STB09; AAMS13; AAMS12], allowed the pervasive growth of a huge variety of Internet-related services. Nowadays, wherever he is and with just few clicks, a person can perform a broad variety of tasks, such as sending an email, paying a bill or watching the last episode of some definitely unmissable tv show. Nevertheless, very little is known, by the majority of people, about what happens through their clicks, and how data flows through the network to and from their computer. In order to enlighten this aspect, we need to grasp the real meaning of the word "Internet", and its underlying physical infrastructure.

The Internet is a global system of interconnected computer networks. A computer network is a collection of computers (hosts) and dedicated

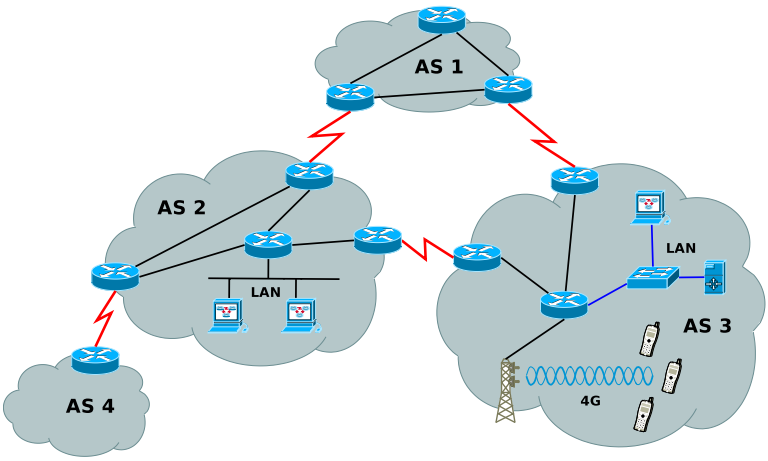


Figure 1.1.1: Basic internetwork

hardware components (NICs[1], switches, etc..) interconnected by communication channels that allow sharing of resources and information. A special device called *router* allows internetworking by forwarding packets between different computer networks. Two routers form an IP[2] [Pos81] connection to exchange data between them, and in such case are said to be adjacent. An Autonomous System (AS) [HB96] is a set of IP networks and routers, which operates under control of a single and well-defined administrative authority. Actually, the number of operators running inside an AS might be more than one, however from the perspective of an external observer, they must act as one, exhibiting a unique routing policy. Two ASes form a connection between them if they agree to exchange traffic and routing information through two or more directly connected routers. Figure 1.1.1 shows a small-scale example of how all these building blocks are linked to each other.

This "network of networks" is growing on a daily basis through the addition of new computer networks and new connection lines, however, since ASes are independently administered, this growth is not controlled

[1]NIC: Network Interface Card, used to connect a computer to a network
[2]IP: Internet Protocol, part of the TCP/IP protocol suite

by any central authority and thus the Internet can be considered as a self-organizing complex system. This phenomenon can result in major issues, since structure affects functioning, and a lack of knowledge of the former, can result in several unforeseen problems for the latter, like misfunctioning or even outages. Several researchers started to analyze the Internet in the hope of understanding its topological structure and evolutionary principles.

The first task carried out by researchers consists of measuring and mapping the Internet topology. This is accomplished by inferring a graph[3], which represents how the different entities interconnect to each other within the global Internet. Depending on the building block used, studies and ongoing projects focus on different level of abstractions, at least four:

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 [cai; Dim; Ipl; Por].

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 [GS06; Key10], by setting peculiar IP options in Traceroute probes [SBS08], by exploiting ad-hoc probes [GT00] and analyzing the IP ID values of the probes sent [BSS08; KYLC].

PoP-level: each node is represented by a collection of routers located in the same points of presence (PoPs), while edges are connections between pairs of PoPs. Data is typically gathered by applying reverse DNS[4] lookups [SMW02] or by looking for peculiar characteristics in the IP interface-level graph [FS08].

AS-level: each node is represented by an Autonomous System, while edges are connections between pairs of ASes, established via Border

[3]Graph: abstract representation of a set of objects where some object pairs are connected by links

[4]DNS: Domain Name System, translates domain names to numerical IP addresses

Gateway Protocol (BGP-4)[5] [RLH06]. Data is typically gathered via Route Collectors (see Section 2.2.1 of Chapter 2) or by applying IP-to-AS techniques to infer an AS-level topology from the IP interface-level graph [CSW01; MRWK03; MJR+04].

Once the Internet has been mapped through one of the above methodologies, the second task consists of analyzing and modeling the inferred topology. Indeed, a proper understanding of the Internet properties and the laws governing its growth, can result in a model able to foresee the network evolution, and tackle most of the problems shown above. The task of Internet modeling is made difficult by several problems [CHK+09; OW10], here we list the three identified *key problems*:

1. The Internet is exponentially growing and so is the number of its components. Moreover, being a "network of networks", the Internet is highly heterogeneous, so such components can be very different from each other, in sizes, purposes, policies, etc.

2. The network evolution is not dictated by a central authority, but is mainly the outcome of local economic and technical constraints.

3. Information about the structure and interconnections of an AS is not public, because publishing it might favor competitors, and current techniques used for data gathering are incomplete and often unreliable.

The first two problems are specific for the modeling part, while the last one is more related to the task of Internet mapping. Nevertheless, it is important to keep in mind this problem when analyzing data or validating models with data, because scientific discoveries might be biased by incorrect data and need to be rethought as our knowledge of it becomes more accurate.

Several studies have been made in the field of Internet modeling, at each of the level of abstractions previously shown (e.g. [ZLWX09]). Although it represents the Internet at a coarser level, the AS level of

[5] BGP: Exterior gateway protocol designed to propagate routing information between ASes

abstraction is probably the most interesting to study. On the one hand, it represents real-life interactions between organizations; on the other hand, available data tends to be more reliable than in the other cases. Despite of this, currently developed models (see Section 2.3 of Chapter 2) are rather crude, fail in properly considering such interactions as the driving forces behind the Internet evolution, and are becoming more and more questioned as data gets more accurate [RWM+11].

1.2 Contribution

The final goal of this thesis is a deeper understanding of the driving forces that are governing the growth of the Internet at the Autonomous System level. The path towards this goal is hindered by the three key problems which have been outlined in the previous section, and which are tackled throughout the work.

The first contribution of the thesis is a novel methodology to partition the Internet topology into two distinct blocks: the core and the periphery. The former is a small, densely connected component, which collects together the most important nodes of the network and captures the underlying community structure of the Internet. The latter is made up of most of the Internet nodes, and represents "tendrils" of the topology, outside of the central network backbone. This method uses one side of *key problem #1* against the other, as it tackles the large-scale nature of the Internet by exploiting its heterogeneity. By showing that most of the interesting, structural aspects of the Internet are captured by the core, we achieve a huge complexity reduction for subsequent analyses.

The second contribution is an economic analysis of the different interactions between the agents taking part into the core, using a game theoretical framework to tackle *key problem #2*. More specifically, we focus on the dichotomy between the peering[6] and transit[7] interconnection

[6] Peering: interconnection policy through which two ASes agree to mutually exchange traffic for free

[7] Transit: interconnection policy through which a customer AS pays is upstream provider to access the Internet

policies and point out the key role of Internet eXchange Points[8] and the emerging competition between them and Network Service Providers[9] (all these concepts will be detailedly introduced in Chapter 2).

The main achievement of the thesis combines the above results, which are used as ground truth to devise a novel model for the Internet's core. The model embodies an innovative dynamic which reflects the driving forces behind the evolution of the core, and is able to capture both statistical and structural characteristics of the Internet's core. Moreover, the model is versatile and capable of accurately predicting the evolution of the Internet's core as long as we are able to foresee the growth of the modeling parameters. As will be shown later, *key problem #3* becomes of crucial importance for this aspect: since the prediction properties are based on data from real measurements, it is important to remove the biases introduced by data incompleteness, in order to accurately predict the future evolution of the Internet topology.

In the last part, the model is extended to the whole Internet by devising simple attaching mechanisms for the periphery, and it is shown how it outperforms existing topology generators.

1.3 Motivation

In a continuously growing environment, modeling and understanding the Internet topology is a fundamental task from which many areas can benefit:

1. The performance of algorithms and protocols is often influenced by the underlying physical structure of the network. A better understanding of the Internet topology could help in the design of structure-aware algorithms, exploiting the topology knowledge for their advantage. Content Delivery Networks (CDN), for example, are distributed computer networks that spread contents (especially multimedia contents) to other users. Since these contents can ac-

[8] IXP: physical facility that allows ASes to interconnect directly
[9] NSP: high-level Internet Service Provider, part of the backbone

count for thousands of Gigabytes, a CDN exploits proxies[10] distributed around the globe to optimize data spreading. A better understanding of the Internet structure could help in the optimal placement of these proxies and boost the performances for final users (for example by reducing the hop distance).

2. Since networks are supposed to perform reasonably well even in the face of occasional malfunction of some components, fault-tolerant network design is a very important task. This issue is particularly difficult in the Internet environment, since there is no central authority overseeing its growth but it is instead the reflection of local efforts performed by individual ASes. A proper understanding of the Internet topology could therefore give a huge insight to engineers, helping them in the task of optimizing the Internet's fault tolerance.

3. As any other communication network, the Internet is subject to the spreading of infections by malicious programs that replicate and propagate themselves (worms[11]). An effective approach in facing this kind of epidemics is the containment strategy, obtained by placing firewalls[12] in strategic network points. It is obviously much easier to identify the optimal placement for these defense mechanisms if the structure of the network is accurately known.

4. Internet Traffic Engineering deals with the issue of performance evaluation and performance optimization of operational IP networks. A proper understanding of the network topology could help engineers in many tasks: reaching the highest level of capacity in the Internet backbone by optimizing link utilization, load balancing, management of SLAs[13] for granting and delivering the desired Quality of

[10] Proxy: server that acts as an intermediary for requests from clients seeking resources from other servers

[11] Worm: self-replicating malware computer program

[12] Firewall: device designed to permit or deny network transmissions based upon a set of rules

[13] SLA: Service Level Agreement, a service contract defining an agreement between a customer and one or more service providers

Service (QoS).

5. While new mechanisms (algorithms, protocols, firewalls, etc..) and services are deployed, it is important to ensure that current mechanisms and services will continue to operate. The ability to provide such guarantees can heavily rely on the knowledge of the underlying topology. For example, CDN services cannot be envisioned, made feasible, evaluated and optimized without intimate understanding of the topology.

To sum up, knowing the Internet structure is a very important point because structure affects functionality. Nevertheless, a point that is possibly even more important is that of foreseeing network growth. As a matter of fact, the Internet has an ever-changing topology. The ability to foresee network growth is the only way to predict if both currently operating and newly deployed mechanisms and services will continue to operate in the future.

1.4 Structure of the Thesis

The remainder of the thesis is structured as follows:

Chapter 2 describes the Internet structure and its evolution up to now, highlights the tools commonly used to analyze the topology and the state of the art in Internet topology modeling.

Chapter 3 illustrates our first contribution, a novel methodology to partition the Internet topology into core and periphery. Moreover, it highlights the growing importance of exchange points within the core. The Chapter is based on [AGL13].

Chapter 4 describes our second contribution: an economic analysis of the different interactions between the agents taking part into the core, using a game theoretical framework. This Chapter is based on [AAGL14b; AAGL14a].

Chapter 5 gives our evolving model for the Internet's core, analyzes its prediction capabilities and the biases introduced by the measuring infrastructure. The Chapter is based on [AGL14b].

Chapter 6 extends the core model to the whole Internet, and compares results obtained through it with existing topology generators. This Chapter is partly based on [AGL14a].

Chapter 7 concludes, summarizing important results obtained and giving an overview of the future works.

Chapter 2

Background

This chapter is split in three sections. In the first one we describe how the Internet evolved through time to reach its current structure. In the second section we illustrate how the network is being measured, and current techniques used for the analysis of the Internet topology. In the last section we describe the state of the art in Internet topology modeling, taking into account static, dynamic and agent-based models.

2.1 The Global Internet Ecosystem

The Internet ecosystem is made of tens of thousands Autonomous Systems, interconnected together in a complex and dynamic manner. In this section we illustrate how the Internet evolved through time to reach its current structure, and who are the entities involved in such evolution.

2.1.1 The birth of the Internet

The Internet originated in the late 1960s when the United States Defense Department developed ARPAnet (Advanced Research Projects Agency network), the first experimental network of computers and the progenitor of what was to become the global Internet. ARPAnet was the first successful experiment of a packet switching network, to which many

others followed (NPL, CYCLADES, Merit, etc.). With so many different networks, something was needed to unify them: in 1974, the TCP protocol was proposed [VCDS74], containing the first attested use of the term internet, as a shorthand for internetworking. At that time, an internet was any network using TCP/IP.

In 1986, the National Science Foundation (NSF) created the NSFnet backbone network to allow access to supercomputer sites in the U.S. from geographically spread organizations for research and education. It was around the time when ARPAnet was interlinked with NSFNET in the late 1980s, that the term "Internet" was first used as the name of the large and global TCP/IP network. By 1990, ARPAnet had been phased out, while NSFnet continued to grow, and more and more countries around the world connected to this Internet backbone [LCC+09].

The Autonomous Systems operating within the Internet at the NSFnet epoch can be distinguished in the following categories [Nor11]:

Transit ASes, that provide connectivity through themselves to other networks. These ASes typically correspond to large backbones and service providers, allowing Internet access to smaller ASes.

Stub/Multihomed ASes, that do not allow external traffic to pass through them. While stub ASes are connected to only one other AS, multihomed ASes can be connected to multiple of them, so as to increase fault-tolerance.

This paradigm produces as output a tree-like hierarchical structure for the Internet topology [PSV04], shown in Figure 2.1.1.

2.1.2 From NSFnet to the beginning of 21th century

Around 1992 the government-funded NSF determined that the Internet could and should be operated by the private sector, and devised a model composed of a set of competing commercial backbones. The progressive dismantlement of NSFnet gave rise to an early version of the commercial Internet model as we know it, with the emergence of economical relationships between ASes.

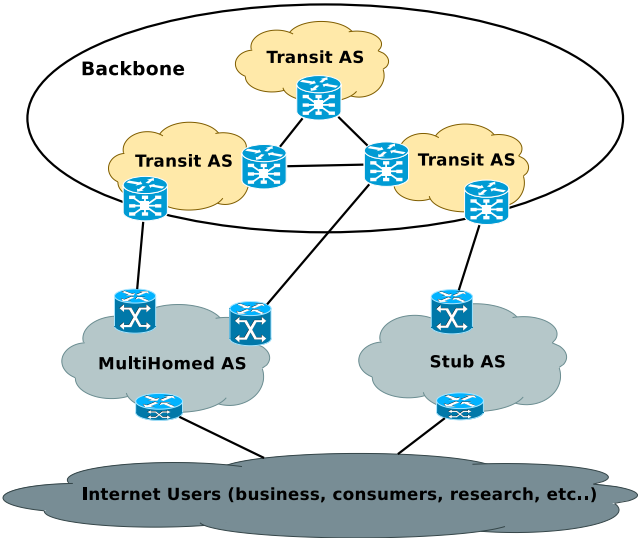


Figure 2.1.1: NSFnet-like hierarchical structure

The late twentieth-century Internet ecosystem was largely dominated by a hierarchical structure made of customer and provider Autonomous Systems. The customer AS pays its upstream provider for the traffic flowing on the link, both incoming and outgoing; in return, the provider grants to all its customer access to the entire Internet. This "customer-to-provider" relationship is also known as "transit": the upstream AS is called "transit provider", and the resulting interconnection between the ASes is known as transit link [Nor11].

The beginning of twenty-first century brought a new paradigm into the environment, since more and more ASes found it beneficial to establish "settlement-free" interconnections. This agreement envisages that neither party pays the other for accessing its customers. In this case the two ASes mutually agree to exchange traffic for free between them, and the only cost they incur is that of laying out the physical link. The relationship is also known as "peering", and the two ASes establishing the peering link are known as peers. Note however that, by exploiting a peering link, an AS cannot reach all the Internet routes, but just the customers of its peer

[GIL+11; Nor10a].

It is worth noting that there exists a third kind of relationship between ASes, denoted as "sibling-to-sibling" [Gao00]. This type of connection is typically applied between ASes owned by the same organization, therefore, from an economic standpoint, it is the same as peering. However, differently from peering, two sibling ASes may use each other to reach their respective providers (and therefore all the Internet routes). The focus of this study is the relationship between organizations, and this kind of connection does not capture this phenomenon. A possible way of dealing with sibling-to-sibling connections could be grouping the involved ASes in a single one, resulting in an AS that is under control of a single organization. Unfortunately, many times sibling tags are detected even if two ASes are not part of the same organization. This could happen if one AS is the customer of another AS but is also used as a transit backup, in case of BGP misconfigurations, and many others [MWA02]. Therefore, applying the grouping procedure described above could introduce more errors than benefits. Since the number of sibling connections in the measured Internet graph is very low, it is not expected to impact the obtained results, thus we are not dealing with sibling relationships in this thesis.

The primordial hierarchical structure, together with the emerging economical relationships, led to a decomposition of the Internet into tiers[1]. Although there is no authority that defines tiers of networks participating in the Internet, the most common definitions are the following [Nor11; Wikb]:

Tier-1 ASes, standing at the highest level in the hierarchy. They form an "elite club" of ASes able to reach the whole Internet without paying any other Autonomous System. To reach this goal, Tier-1 ASes are interconnected to each other through a full-mesh[2] of peering links. While this definition of Tier-1 can be separately applied to smaller Internet regions (e.g. Europe, America, and so on), we will always refer to the global Internet. In this context, the number of Tier-1

[1] Tier: level of a network

[2] Full-Mesh: interconnection such that each node is connected to any other node in the network

ASes is quite small (15-20) compared to the total number of ASes [Wikb].

Tier-2 ASes, which buy transit from higher-level ASes (typically Tier-1). Differently from Tier-1 ASes, they can engage in peering to reach a portion of the network, but they still have to purchase transit to reach the entire Internet. With a little abuse of terminology, in the following we equivalently refer to Tier-2 ASes as **Internet Service Providers (ISP)**. For Tier-3 ISPs see the keyword Access Provider.

Tier-3 ASes, which solely purchase transit from other networks in order to reach the Internet.

Depending on the belonging organization and the offered services, ASes can be further categorized into [pee]:

Network Service Providers (NSP): high-level transit providers [Nor10b], located in either Tier-1 or Tier-2. While all Tier-1 ASes can be considered NSPs, this is true for Tier-2 only if the AS sells transit to lower level ASes. Like Tier-1 ASes, NSPs typically span several geographic regions, and can be reached by smaller, regional ASes through their Points of Presence (PoPs).

Access Providers (AP): low-level regional providers, giving eyeballs (i.e. the end users) access to the Internet and its contents. They are typically located in Tier-3. An access provider AS wishing to grant Internet access to its users, needs to establish a link with a transit provider, and pay for the traffic flowing on this link. For the sake of conciseness, we use the term AP also to refer to all other kinds of Tier-3 ISPs (Mailbox Providers, Hosting Providers, etc.), since many APs provide these services themselves.

Content Providers (CP): responsible for the creation and management of contents, and the relationships with those who use, enhance, or support this content. Content providers are typically located in Tier-3, since network operations are outside the objectives of their mission. As shown in the next section, this is changing nowadays.

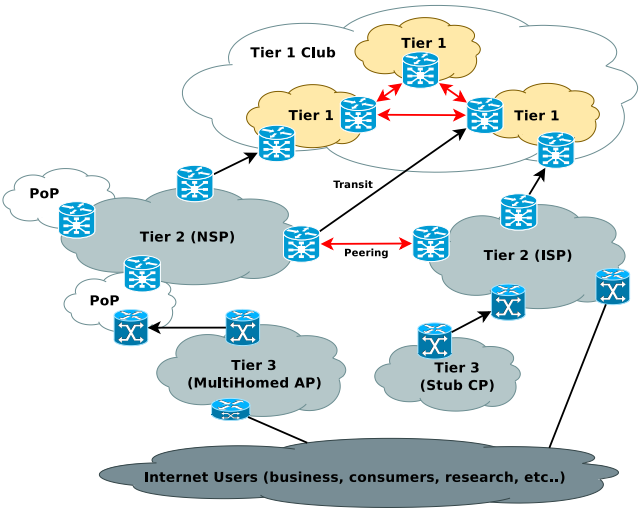


Figure 2.1.2: The 20th century Internet

Research Networks (EDU): used by universities and/or other research institutes.

The structure resulting from this new economic model, together with the above categorization, is illustrated in Figure 2.1.2.

2.1.3 Towards a Flat Structure

Recent years have shown an exponential growth of peering links, with its number approaching if not overtaking the number of transit links. The introduction of peering meshes, as reported in [DD10], is responsible for the evolution of the Internet from its previous, hierarchical structure, to a flatter one. The emergence of such phenomenon has to be sought after the birth of a new interconnection paradigm known as "public peering", as opposed to the classical one, referred to as "private peering". In particular, we have that [Nor11]:

Private Peering is the direct interconnection between the two parties establishing the settlement-free agreement. The involved ASes lay

out a point-to-point link to connect to each other. Tier-1 ASes use this kind of agreement.

Public Peering is, again, a settlement-free interconnection between the two parties, however in this case the involved ASes utilize a multi-party shared switch fabric. The public, physical facilities providing this technology are known as **Internet eXchange Points (IXPs)** [Eur].

The main benefit of establishing peering links, is that peer ASes can reach each other without passing through their respective upstream providers, thus saving transit money. However, since peers can only reach each other's customers, reaching multiple destinations requires the establishment of many peering links. For example, if hundreds of ASes want to exchange traffic between themselves without buying transit, they need to establish a full-mesh of peering links. Under private peering, each Autonomous System has to lay out physical interconnections one by one, therefore establishing the peering mesh requires it to build hundreds of links. This can result in high costs, possibly worse than establishing few transit links.

The growth of IXPs, in number and in size, made it easy to establish more and more peering meshes. In fact, when one AS joins an IXP, it can publicly peer with all (usually a subset) of the other ASes connected to the same IXP. This allows the Autonomous System to connect with hundreds of ASes at a cost given only by a single public peering link, plus the cost for joining the IXP. This last cost is related to hardware maintenance and shared by all the parties, as will be shown after. Therefore, the establishment of peering meshes through IXPs is an effective way of reaching many other participants without passing through upstream providers, thus saving lots of money. Moreover, this shortcut can also be an effective way of shortening the path between two ASes, thus increasing the overall network performances [Nor10c].

All these advantages are making the presence of IXPs a dominant factor inside the Internet's structure, transforming its past hierarchical structure into a flatter one [GILO11; ACF+12; AG10]. As shown in Figure

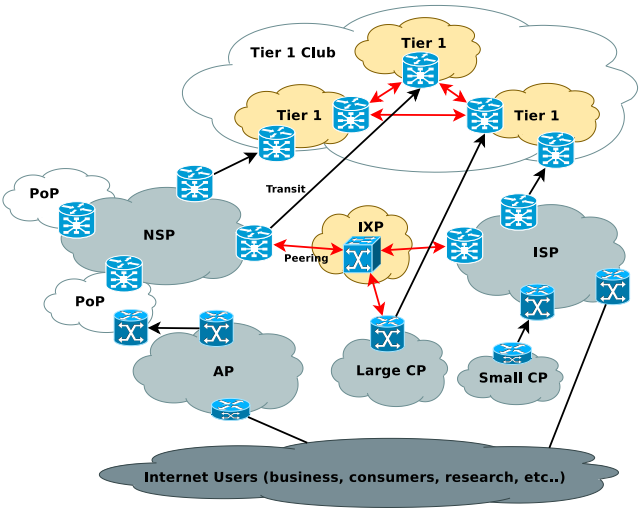


Figure 2.1.3: The 21th century Internet

2.1.3, large-scale Content Providers (e.g. Google) benefited from this architecture as well, since by establishing peering sessions over IXPs, they can effectively reduce the amount of traffic sent through costly transit connections [Nor11].

2.2 Internet as a Complex System

2.2.1 Measuring the Internet Topology

There are two main measurement techniques for gathering information about the Internet AS-level topology, based on active or passive probing.

The former, exploits a tool named traceroute to inject into the network special packets able to discover the path between two endpoints. Active probing is typically used when inferring router-level maps [GS06; Key10; SBS08; GT00; BSS08; KYLC], while there are several drawbacks when trying to obtain the AS-level topology, mainly due to dealiasing and router-to-AS mapping issues [CSW01; MRWK03; MJR+04]. Moreover, currently available datasets are incapable of giving an accurate view of

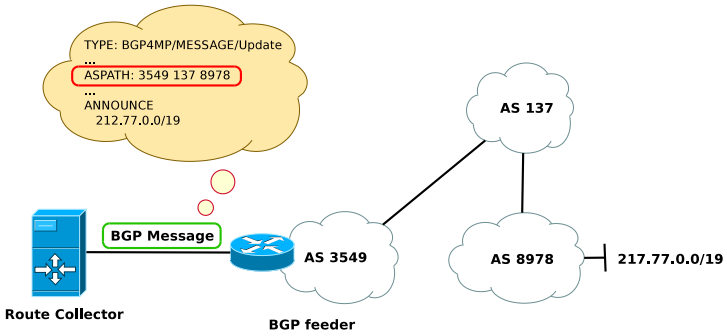


Figure 2.2.1: Measuring infrastructure

the topology evolution, something which is needed for this research work [cai; Dim; Ipl; Por].

The passive probing method consists in collecting inter-AS routing information (i.e. BGP messages) from the routers composing the Internet itself. The most common data sources exploited in this latter technique are provided by several research projects [rip; Rou; PCH; BGP] which deployed a set of route-collectors (RC) around the world aimed at collecting as much information as possible about the Internet routing. A Route Collector is a device that behaves like a BGP router, with the purpose of gathering BGP messages from cooperating ASes, refereed to as *feeders*. These projects make publicly available dumps of the BGP messages received by their RCs. The main source of information in BGP data about the AS-level topology is represented by the well-known mandatory AS_PATH attribute, which contains the sequence of ASes that the traffic crosses to reach some destination [RLH06]. Figure 2.2.1 shows a RC receiving a BGP message from the feeder AS 3549, highlighting the AS_PATH attribute contained in the message.

Since each feeder provides a representation of the network from its specific point of view, collected data has to be aggregated to obtain the complete Internet topology. The ISOLARIO dataset [Iso], created by a team of researchers at IIT-CNR, is generated using data collected by the RIS [rip], RouteViews [Rou], BGPmon [BGP] and PCH [PCH] route collector

projects. The data gathered is processed in order to obtain monthly-based AS-Level topology graphs [CGJ+02]. The routing information gathered by this methodology is considered reliable since it is collected directly from devices that are effectively participating in the inter-domain routing, and represents the result of the BGP decision process of each feeder. It is known that, due to the limited number of feeders, the inferred AS-level topology is incomplete [OW10]. Nevertheless it is a useful source of information, that can be used provided that one takes into account the incompleteness of available data when drawing any inference.

2.2.1.1 The Enriched Internet Graph

The dataset provided by ISOLARIO consists of several (one per month) AS-level Internet topology graphs, where nodes represent ASes, and edges are the BGP connections between them. Furthermore, the dataset is enriched through the following additional information:

Node Type: each node, as shown in Section 2.1.2, can be either a Network Service Provider, an Internet Service Provider, an Access Provider, a Content Provider (CP) or something else (Non-Profit, Educational, Research, etc..) [pee].

Node Peering Policy: each node has a specific peering inclination, or policy, indicating its willingness to participate in peering agreement. Each node publishes its peering policy, which can fall into one of the following categories: *open*, *selective* or *restrictive*. An "Open" peering inclination is a willingness to peer with any other AS in the ecosystem, while a "Selective" peering inclination reflects the existence of requirements for establishing peering sessions. Finally, a "Restrictive" peering inclination reflects a desire not to peer with anyone other than existing peers. As an example, Tier-1 ASes all have restrictive peering policies, because they peer with each other but refrain from doing so with any other AS [Nor10a].

IXP Participation Lists: each IXP provides through its website a publicly available list of ASes participating into the IXP. Therefore, for each

node, we can infer the IXPs it is participating into.

Link Type: as shown in Section 2.1.2, relationships between ASes can be either transit or peering. Unfortunately, this kind of information is typically hidden by network operators, due to its key role in economic agreements. Despite this, researchers have developed algorithms aimed at capturing such relationships [GIL+11; Gao00]. Thanks to this, the dataset can be enriched by "tagging" each edge in the graph with the appropriate relationship, either transit or peering.

Please note that while the first three piece of information are publicly available and collected on websites, such as [pee], the last one is the result of an a posteriori analysis, since network operators are typically reluctant to publicly reveal such information, and thus it is far less reliable. Nonetheless, when looking at the network from a high-level perspective, it is still a valid and useful piece of information.

2.2.2 Understanding the Internet Topology

With over forty-thousand nodes and hundreds of thousands links, it is definitely impossible to study the Internet AS-level topology via direct graphical representation. It is however feasible to obtain a statistical and structural description of it, using topological measures. Here we review the topological measures used for Internet topology analysis, typically borrowed from the framework of graph theory [PSV04; SLH06].

2.2.2.1 Graph Measures

Formally, we represent the AS-level Internet topology as a graph $G = (N, E)$. N is the set of nodes and $E \subseteq N \times N$ is the set of edges in the graph. As previously said, in the graph (or network) a node (or vertex) represents an Autonomous System, while edges (or links) are BGP connections between ASes: all these terms are used interchangeably through the thesis. We indicate with $|N|$ the number of nodes in the graph, and with $|E|$ its number of edges. Two nodes of the graph, say

$i, j \in N,\ i \neq j$, are said to be adjacent if they are linked by an edge, that is if $(i, j) \in E$. In this case the edge is said to be incident to the nodes it joins. If G is an undirected graph, than edges (i, j) are unordered pairs, meaning that (i, j) is identical to (j, i). If G is directed, than (i, j) are ordered pairs, and represent the "from i to j" relationship.

Unless otherwise stated, in the following we will always refer to undirected graphs, therefore, when we say that $(i, j) \in E$, we implicitly say that so does (j, i). As any undirected graph, G can be represented by means of an adjacency matrix A, such that:

$$a_{ij} = \begin{cases} 1 & if\ (i,j) \in E \\ 0 & otherwise \end{cases}$$

The number of node pairs in the graph can be written using Newton's Binomial coefficient as:

$$\left(\begin{array}{c} |N| \\ 2 \end{array} \right) = \frac{|N|(|N|-1)}{2}.$$

Please note that this is also the maximum possible number of edges in the graph, representing a full-mesh. The global network density d is given by the ratio of the number of edges in the network, and its maximum possible number, therefore:

$$d = \frac{|E|}{1/2 \cdot |N|(|N|-1)}.$$

A **clique** is a subset of nodes in a graph such that each node is adjacent to any other node. The clique is a full-mesh structure, since it has an edge for any node pair. Therefore, using the previous formula, a clique of k nodes has $k(k-1)/2$ edges.

The topological measures typically used in this field, can be divided in two groups, namely: local and global measures.

a) Local Measures A topological measure is local if it takes into account only node interconnections and interconnections between node neighbors. The set of neighbors of a node n can be formally expressed as:

$$N(n) = \{j \; s.t. \; (n, j) \in E\}$$

Therefore, given a node n, a local measure only considers pairs of edges $(i, j) \in E$ such that either i or j belong to $n \cup N(n)$. The most common local measures are:

Degree The degree k_n of the $n - th$ node of the graph is the total number of edges incident to it, therefore it is also equal to its number of neighbors $|N(n)|$. The interpretation of the degree is immediate, as it tells the propensity of a node to establish relationships with others. For directed graphs, it is possible to obtain two different degrees for each node: the *in-degree* is the number of links "terminating" in node n (from $N(n)$ to n), the *out-degree* is the number of links "originating" in node n (from n to $N(n)$).

Average neighbor degree The average degree of the neighbors of node n, written $k_{nn}(n)$, is formally defined as:

$$k_{nn}(n) = \frac{1}{k_n} \sum_j A_{nj} k_j = \frac{1}{k_n} \sum_{j \in N(n)} k_j.$$

A high (low) value of k_{nn} indicates that the neighbors of a node are high (low) degree nodes. Therefore, this measure can be interpreted as the attitude of nodes to establish connections with highly connected zones, i.e. hubs of the network.

Clustering The clustering coefficient c_n of the $n - th$ node of the graph is defined as the ratio of the number of edges between its neighbors to its maximum possible number, similarly to the network density. Formally, we write:

$$c_n = \frac{|E(n)|}{1/2 \cdot k_n(k_n - 1)}.$$

where $E(n) = \{(i, j) \; s.t. \; i, j \in N(n)\}$ is the set of edges between the neighbors of n, and k_n is its node degree. Please note that $c_n \in [0, 1]$: when $c_n = 0$, the topology is a star, and the neighbors of n are

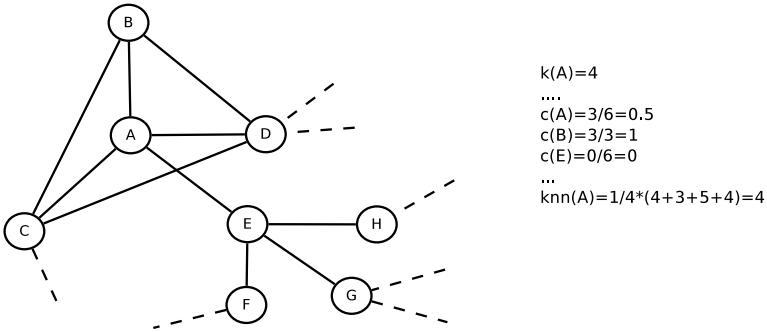


Figure 2.2.2: Local properties of a graph

all disconnected from each other; when $c_n = 1$, all the neighbors are interacting with each other in full-mesh, thus forming a clique. Therefore, this measure can be interpreted as the attitude of nodes to form densely connected groups, i.e. clusters.

Figure 2.2.2 illustrates an example of computation for the described local measures. Please note that $\{A, B, C, D\}$ is a full-mesh (the clustering of B is 1), while node E is the hub of a star topology (its clustering is 0).

b) Global Measures A topological measure is global if it takes into account all the possible nodes and interconnections of the graph.

Shortest Path Length The shortest path length (or geodesic) l_{ij} is equal to the number of edges forming the shortest path from node i to node j. The higher the shortest path length, the higher the number of hops between two nodes of the graph, therefore this measure can be immediately interpreted as indicator of the distance between any two nodes of the graph. Shortest paths are the "best" paths that can be used by two nodes to reach each other. Even if it is not always the case (due to economic and technical constraints [Gao00]), they are typically interpreted as the paths actually used by the nodes, therefore it is reasonable to think that a node crossed by many shortest paths is highly utilized, and therefore "central" in the network.

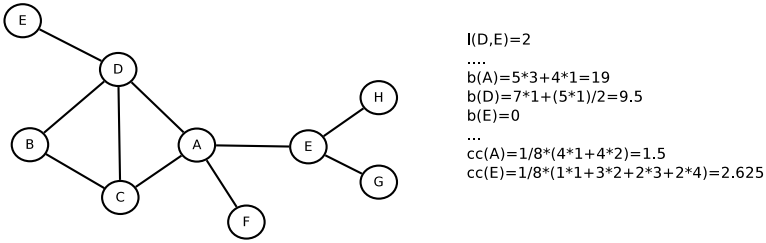


Figure 2.2.3: Global properties of a graph

Betweenness Centrality The betweenness centrality b_n of the $n-th$ node of the graph is the number of shortest paths passing through it among all possible ones. Since a high (low) value of betweenness, indicates that many (few) geodesics pass through a given node, this metric can be interpreted as the "centrality" of such node. Another betweenness measure is the edge betweenness, which counts the number of shortest paths passing through a given link.

Closeness Centrality Another centrality measure is the closeness centrality cc_n, which measures the average distance of of the $n-th$ node of the graph, to all the other nodes. Both betweenness and closeness can be normalized to represent values between zero and one.

Coreness [AHDBV06] The coreness of a node is defined as the highest k-core to which the node belongs. A k-core is a maximal connected subgraph in which all nodes have degree at least k, where maximal means that we cannot include another node to the subgraph without violating the minimum degree property (see Figure 2.2.4 in the next section). Coreness is a global measure similar to clustering, as it reveals the existence of densely connected zones in the network.

Figure 2.2.3 shows an example of computation for the described global measures. Note that the betweenness centrality of E is zero, as it is not crossed by any shortest path, however its closeness centrality is not infinity, as it is connected to the topology.

The topological measures just outlined, both local and global, can be used to statistically describe the graph at different levels of granularity.

Specifically, we use the following classification [RFT13]:

Coarse-grained Measures are those that concisely represent the graph with a unique scalar value. The average degree, k_{nn}, clustering coefficient, shortest path, betweenness and closeness, are all coarse-grained measures, obtained by averaging results obtained for a single node (or node pairs), to the whole graph. We use the notation $\langle \cdot \rangle$ to denote average values, e.g. $\langle k \rangle$ is the average degree of the network. Another coarse-grained measure typically used is the network diameter D, indicating the maximum geodesic considering all possible node pairs.

Medium-grained Measures are those that represent the graph using distributions of values. For example, the degree distribution $P(k)$ is the probability distribution of the degree measure over the whole network. The probability of observing a given degree k can be expressed as the ratio of the number of nodes with degree k, to the total number of nodes in the network. If the network has $|N|$ nodes and n_k of them have degree k, then we have:

$$P(k) = \frac{n_k}{|N|}.$$

We can apply the same reasoning to the shortest path length. In this case, $P(l)$ is the probability distribution of the geodesic measures over the whole network. If the network has $|N|$ nodes and n_l of them have geodesic l, we have:

$$P(l) = \frac{n_l}{1/2 \cdot |N|(|N|-1)}.$$

In the same way, we can define probability distributions for all the other local and global measures of the graph.

As will be shown later on, both coarse and medium-grained graph metrics have been extensively used for analyzing and modeling the Internet topology [PSV04], due to their ability to compactly represent its statistical characteristics. Nevertheless, they show several limitations, like the impossibility to study through them the existence of tiers and the interaction

between different network tiers. These aspects are crucial for a proper understanding of many real networks such as the Internet.

In order to tackle this problem, new **fine-grained measures** were introduced. The goal of this class of measures is capturing the structural characteristics of the network through the identification of their underlying community organization.

2.2.2.2 Community Structure

Most real networks typically contain parts in which the nodes are more tightly connected to each other than to the rest of the network. Although there is no unique definition, the sets of such nodes are usually called clusters or communities. They are a signature of the hierarchical nature of the system. Community detection has attracted much attention in recent years, and the literature on community detection is huge [For10]. Probably, the most famous method for community detection is based on modularity[3] maximization [New04]. However, this method has several drawbacks which bring to degeneracy for networks with highly hierarchical structure [GMC10], such as the Internet. Therefore, here we focus only on the most important methods which have been successfully applied to the Internet AS-level graph.

There are two main categories of community detection methods, providing either a partition or a cover of the network. A partition is a division of the graph in clusters, such that each node belongs to only one cluster, while a cover is a division of the graph into overlapping (or fuzzy) communities [For10].

Partition Methods divide the graph in clusters, such that each node belongs to only one cluster. The most important methods in this class are the k-core decomposition [CHK+07] and the k-dense method [GLO11b; SYK09]. The former relies on the previously defined concept of k-core, while the latter uses the concept of k-dense. A k-dense is a maximal connected subgraph in which each pair of

[3]Modularity: benefit function that measures the quality of a particular division of a network into communities

adjacent nodes has at least k common adjacent nodes. Both methods define nested communities, i.e. a k-core (k-dense) community is contained in a $(k-1)$-core ($(k-1)$-dense), and it is not possible for communities at the same k level to overlap.

Cover methods divide the graph into a set of overlapping communities. The most important methods in this class are the Clique Percolation Method (CPM) [PDFV05], the Greedy Clique Expansion (GCE) algorithm [LRMH10] and the agglomerativE hierarchicAl clusterinG based on maximaL cliquE (EAGLE) algorithm [SCCH09]. All of them detect communities starting from maximal cliques, and combine them to form communities using chains (CPM), fitness-based expansions (GCE) or clique-based modularity measure (EAGLE). Anyways, the basic building block for these methods are maximal cliques: a clique is maximal if it is not a subgraph of a larger clique, that is, it cannot be extended by including an adjacent vertex.

Figure 2.2.4 shows examples of k-cores, k-denses and maximal cliques according to above definitions. Please note how the granularity becomes finer as we go from k-core to cliques. Note also that while $\{A, B, C\}$, $\{A, B, D\}$, $\{A, C, D\}$, $\{B, C, D\}$ are cliques of size 3, they are not maximal since we can always add one more adjacent node and obtain a maximal clique of size 4.

2.2.2.3 State of the Art for the Internet Topology Analysis

The topological measures outlined in Section 2.2.2.1, both coarse and medium-grained, have been extensively used for analyzing the statistical characteristics of the Internet topology. Table 2.1 shows the behavior of coarse-grained topological measures for the Internet AS-level graph as a function of time.

Indeed, we observe that most of the average values do not show large fluctuations and seem to be more or less stable in time. The first feature that can be inferred through this table is the so called "***Small-World***" property, first observed in [WS98]. This characteristic, to be precise, refers to networks in which:

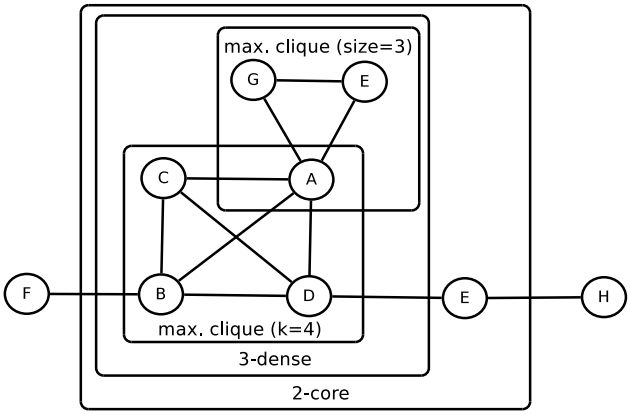


Figure 2.2.4: k-cores, k-denses and maximal cliques

Average Values	2006	2008	2011	2013
Nodes	22,634	27,896	37,278	44,900
Edges	65,268	85,185	122,799	179,770
Degree	5.77	6.11	6.58	8.00
Clustering	0.32	0.30	0.29	0.32
Knn	506.61	542.73	567.67	695.25
Shortest Path	3.71	3.72	3.76	3.72
Betweenness	30731.6	38003.8	51616.0	61141.9
Closeness	0.27	0.28	0.27	0.27
Coreness	3.00	3.17	3.40	4.15

Table 2.1: Internet average graph metrics

i. the average shortest path length $\langle l \rangle$ scales logarithmically (or slower) with the number of nodes, despite of the low global network density;

ii. the average clustering coefficient $\langle c \rangle$ is several orders of magnitude larger than the corresponding value for a graph where nodes are randomly interconnected (see next section).

Figure 2.2.5 shows the shortest path length frequencies observed for

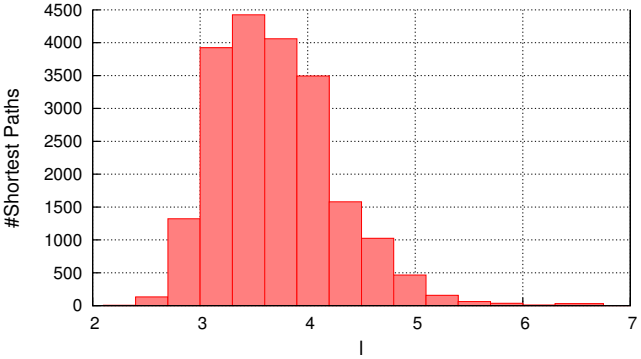


Figure 2.2.5: Internet shortest path frequencies - 2013

the Internet in April 2013. We observe that not only the average shortest path length is very small, even smaller than the logarithm of the number of nodes, but also the network diameter is pretty small, indicating that even in the worst-case scenario, each two nodes of the network can communicate in few steps. The small-world property can be interpreted as follows. The Internet has a low number of edges compared to its number of nodes; despite of this aspect, it is well connected, in the sense that two nodes of the network can reach each other in few hops, and there is a non-negligible presence of local clusters.

The second feature of the Internet topology, first observed in [FFF99], is the "***Scale-Free***" property. This property is based on the observation that some medium-grained measures, mainly the degree, are characterized by a heavy-tailed probability distribution, that can be reasonably approximated by power-law forms. In practice, heavy tail means that there is a non-negligible probability of getting very large values. A Gaussian distribution, for example, does not exhibit the heavy tail property, since the probability of obtaining large values becomes exponentially small as we leave the mean value. In this sense, heavy tail distributions typically represent wild as opposed to mild randomness. A power-law form, is a specific type of heavy-tailed distribution, whose density function follows the law:

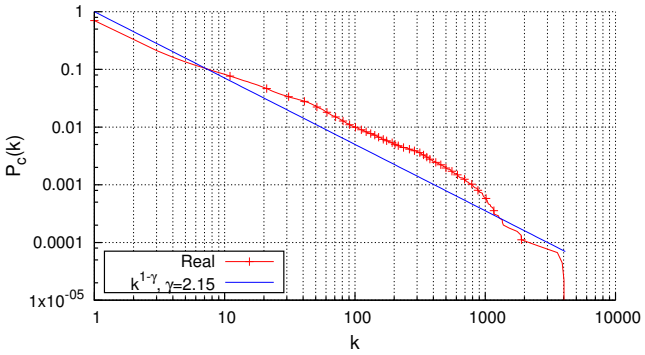


Figure 2.2.6: Internet degree ccdf - 2013

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

where $\gamma > 1$ is the power-law exponent. As typically done in the literature, we study the Complementary Cumulative Distribution Function (CCDF), which has the advantage of dealing with cumulative values, therefore it is considered much less noisy [PSV04]. We recall that the CCDF of a continuous random variable (e.g. the degree) can be expressed as the integral of its probability density function $P(k)$ as follows:

$$P_c(k) = 1 - P_{cdf}(k) = 1 - \int_{-\infty}^{k} P(k')dk' = \int_{k}^{\infty} P(k')dk'$$

where $P_{cdf}(k)$ is the cumulative distribution function (CDF). The CCDF is also known as the tail distribution, since its value at point k represents the probability that a random variable (e.g. the degree) has value greater than k. For power-law forms, it is easy to find that:

$$P_c(k) \sim k^{1-\gamma}$$

With a little abuse of terminology, in the following we will sometimes refer to the degree CCDF as "degree distribution". Figure 2.2.6 shows the

CCDF of the degree observed for the Internet in April 2013. The scale-free property can be observed by visualizing the plot in a log-log scale[4], where the power-law becomes a straight line, and the measured degree distribution approximately follows this line. Please note that while the exact definition of scale-free property refers to networks whose degree distribution follows, at least asymptotically, a power law, the term scale-free informally refers to the fact that the degree distribution maintains the same shape at all scales.

The small-world and the scale-free properties just outlined have been deeply investigated, and researchers exploited them in order to build models exhibiting such properties (see next section). However, several works have shown that they are incapable of accurately representing the most important properties of the Internet. First of all, we have to consider that more recent arguments made in [CGWJ02; WGJ+02], state that the power law behavior observed in [FFF99] is only evocative. This theory is also supported by the evidence that today, with data getting more accurate, the heavy-tailed degree distribution deviates substantially from an ideal power law. Second, but probably most important, these topological measures are inadequate to describe the hierarchical structure of the Internet [HRI+08], made of tiers and communities.

In order to understand the structural and functional properties of the Internet, it is fundamental to interpret the overall organization of the graph as the coexistence of different building blocks. Community detection methods have been applied to the Internet AS-level graph with the purpose of discovering such building blocks, or communities. There are many studies on the structural properties of the Internet's graph which partition the network into communities with techniques such as k-core decomposition [CHK+07] and the k-dense method [GLO11b; SYK09].

However, we believe that, within the Internet environment, each community shall identify dense subgraphs, and overlapping communities should be allowed. Take, for example, an AS participating into multiple, different, Internet eXchange Points. If we can represent each IXP with a community, it is obvious that such communities will overlap. Therefore,

[4]Log-Log: vertical and horizontal axis of the plot are logarithmically scaled

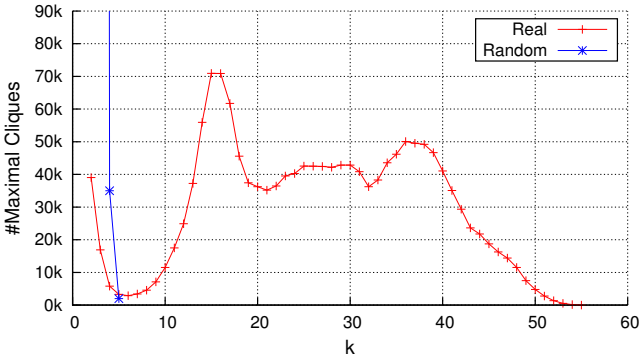


Figure 2.2.7: Internet maximal cliques distribution - 2013

cover methods better fit the topology under consideration. In this context, Internet-related works typically refer to the CPM method or to its variations [PDFV05; GLO11a; GLM12; LRMH10; SCCH09]. Anyways, as we have shown in the previous section, the most important cover methods detect communities starting from maximal cliques.

Cliques are a useful fine-grained tool to identify communities, because they identify full-mesh structures (e.g. peering meshes generated by IXPs) in the network, typically indicating the presence of high cohesion between the nodes. Figure 2.2.7 shows the maximal clique frequencies for the Internet topology as of April 2013. Each dot in the plot represents the number of maximal cliques of size k. With a little abuse of terminology, in the following we will also refer to this plot as "maximal clique distribution", even if it is not normalized. The high number of large cliques is an important, quantitative indicator of the underlying community structure of the Internet. In particular, the figure compares the clique distribution of the Internet topology, with that observed for a graph of the same size whose nodes are randomly interconnected, revealing how much the two structures differ. Although this distribution is highly sensitive to the addition/deletion of a small part of the network, we believe that a model aimed at representing the Internet realistically must take cliques into account, if it wants to properly grasp structural properties of the Internet.

2.3 The Art (and State) of Internet Modeling

The modeling of the Internet focuses, at the large-scale level, on the construction of graphs that reproduce the topological properties observed in the Internet AS-level map. Here we review the models known in the literature, highlighting their strengths and weaknesses. We divide the models in two main classes, depending on the used approach: complex networks based models, rooted in statistical physics, and strategic optimization based models, rooted in game theory [Wika].

2.3.1 Static and Dynamic Models

The simplest conceivable model for any complex network is that of static random graph, where nodes of a fixed size network are randomly connected with some probability. This model is characterized by a complete lack of knowledge of the principles guiding the creation of edges between nodes. The availability of the data and results shown in previous sections, led researchers to propose models better capturing the statistical characteristics of the Internet. Complex networks based models can be further divided in two classes: static models, which consider a graph of fixed size, and dynamic models, which focus more on the evolution, by dynamically adding new nodes to an initially small network [PSV04].

2.3.1.1 Random Graph Models

Erdős–Rényi The first theoretical model for random networks was proposed by Erdős and Rényi [ER60; Gil59]. This static model has been commonly used to describe complex networks with no apparent regularities, and was applied to the Internet as well. The random graph model assumes that the graph $G = (N, E)$ has a fixed set of $|N|$ different nodes and each of the possible $\frac{|N|(|N|-1)}{2}$ edges is present with probability p. In order for the model to be an accurate representation of the Internet, the first thing to do is to check if it reproduces the small-world and the scale-free properties observed for the measured topology. To this end, we can compute:

Degree: the average number of edges in the graph is given by:

$$\langle |E| \rangle = \frac{1}{2}|N|(|N|-1)p.$$

Since each edge contributes to the total network degree by 2, we can obtain the average degree as:

$$\langle k \rangle = \frac{2\langle |E| \rangle}{|N|} \simeq |N|p$$

In order to obtain the degree distribution, we observe that each node has degree k if it is connected to k other nodes and not connected to the $N-1-k$ others. Given that the wiring probability p is fixed and independent for each edge, the degree distribution is binomial:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}.$$

Clustering: for any node, the probability that two of its neighbors are connected to each other is given by the wiring probability, therefore the average clustering is p. Using above formulas we can rewrite the clustering as a function of the network size:

$$c = p = \frac{\langle k \rangle}{|N|}$$

note that at fixed values of the degree, as the network grows in size the clustering coefficient becomes increasingly smaller.

Shortest Path: Given the random position of edges and neglecting cycles, starting from node i we can reach, on average in one step, $\langle k \rangle$ nodes. Since each of these nodes is connected on average to other $\langle k \rangle$ nodes, we can reach in two steps $\langle k \rangle^2$ nodes, therefore in l steps, we can reach on average $\langle k \rangle^l$ nodes. In order to reach the whole graph in l steps we require that $\langle k \rangle^l = N$, therefore the average shortest path length is [New03]:

$$l = \frac{\log |N|}{\log \langle k \rangle}$$

note that this quantity grows logarithmically with respect to the size of the graph.

These results show how the random graph model exhibits one side of the small-world property, namely the "small" shortest path length, but fails to reproduce both the high clustering levels and the scale-free property observed for the Internet.

Watts-Strogratz The empirical observation of large and stationary clustering coefficient in real world networks, led to the construction of models that allow a tuning of this parameter. The Watts-Strogratz [WS98] model starts with a set of $|N|$ nodes placed on a ring, where each node is connected to its $2m$ nearest neighbors[5], half on the clockwise sense and the other half on the counterclockwise sense. For each node, each edge connected to a clockwise neighbor is preserved with probability $1-p$ and randomly rewired[6] with probability p. The following model exhibits the following properties [BW08]:

Degree: the degree distribution is binomial, similar to that of the Erdős-Rényi graph:

$$P(k) = \sum_{n=0}^{min(k-m,m)} (1-p)^n p^{m-n} \frac{(pm)^{k-m-n}}{(k-m-n)!} e^{-pm} \quad k \geq m$$

Clustering: the clustering coefficient takes advantage of the regular structure of the initial graph, and can be tuned by adjusting parameters p and m. It is possible to compute:

$$\langle c \rangle \simeq \frac{3m(m-1)}{2m(m-1)}(1-p)^3$$

note that, for $p \ll 1$, the clustering coefficient is finite and independent of the network size.

Shortest Path: the initial network of this model is similar to a regular grid, therefore it can be shown that the shortest path length grows linearly with the network size. Thanks to the shortcuts introduced

[5]The "near" property refers to the euclidean distance between the nodes
[6]Rewire: delete the link and generate a new link between different node pairs

by the rewiring steps, the average shortest path length reduces dramatically. In particular, for a sufficiently large rewiring probability $p \gg 1/|N|$ it can be shown that:

$$l \sim \log |N|$$

therefore the average shortest path length has values similar to that of random graphs.

These results show that, at least in the interval $1/|N| \ll p \ll 1$, the Watts-Strogratz model yields both sides of the small-world property, as it is able to produce high values of clustering. Unfortunately, even this model is unable to produce the scale-free degree distribution observed for the Internet and many other complex networks.

2.3.1.2 Evolving Network Models

Barabási-Albert Model As we said before, dynamic models focus more on modeling network growth rather than statically reproducing the graph. The first growing network model is due to Barabási and Albert [AB99; AB02], and is also known as BA-model. Following its introduction, many different models have been proposed, based on the same intuition. Basically, the network grows from an initial core of m_0 nodes to its final size $|N|$ by a sequence of time steps. Each time step $t = m_0, .., |N|$ a new node is added, and establishes connections with pre-existing nodes. This means that the "age" t of a network is equal to the number of vertices in the graph, and each vertex is uniquely identified by the time of its appearance. The system can be described by means of a "master equation" that dictates the time evolution of $p(k, s, t)$, the probability that node s has degree k at time t. In the large size limit, the stationary degree distribution can be obtained as:

$$P(k) = \lim_{t \to \infty} P(k, t)$$

where $P(k, t)$ is the degree distribution at time t, obtained by averaging $p(k, s, t)$ over all the nodes:

$$P(k,t) = \frac{1}{t}\sum_{s=0}^{t} p(k,s,t).$$

The Barabási-Albert model uses a mechanism known as "preferential attachment" (PA)[7]. According to this mechanism, at each time step t a new node appears in the network, and is linked to older nodes with a probability proportional to their degrees. Note that if $m_0 = 0$ then the total network degree is $2t$. The connectivity distribution of sites obeys to the master equation:

$$p(k,s,t+1) = \left[\frac{k-1}{2t}\right] p(k-1,s,t) + \left[1 - \frac{k}{2t}\right] p(k,s,t)$$

In [DM03] it is shown that, given the proper boundary conditions[8], the equation above can be rewritten, after transition to the continuous-time approximation, as:

$$P(k) + \frac{1}{2}\left[kP(k)\text{-}(k-1)P(k-1)\right] = 0$$

whose solution is:

$$P(k) = \frac{4}{k(k+1)(k+2)}$$

Interestingly enough, this distribution is scale-free. In particular it follows a power-law with exponent 3:

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

Starting from this basic model, many other models have been proposed by varying the attachment mechanism. Interestingly enough, the power-law degree distribution is retained only in the case of linear preferential attachment, and the addition of constants leads to a modification of the exponent. In [DMS00] the authors generalize the BA-model with the following assumptions:

[7] Preferential attachment is also known in literature as the rich-get-richer phenomenon
[8] i) the initial degree of nodes and ii) the initial network size

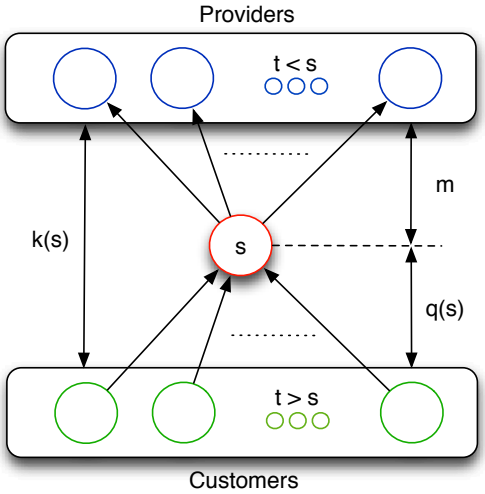


Figure 2.3.1: Generalized preferential attachment

i. Node s has an attractiveness proportional to its in-degree, and given by: $A_s = A^{(0)} + q_s$.

ii. At each time step t a new node appears and links to other m nodes with a probability proportional to their attractiveness.

iii. If we define $a = {}^{A^{(0)}}/_{m}$ and $m_0 = 0$, then the total network attractiveness at time t is $A_\Sigma = (m + A_0)t = (1 + a)m$.

Figure 2.3.1 shows the generalized PA mechanism.
The connectivity distribution $p(q, s, t)$ obeys to this master equation:

$$p(q, s, t+1) = \sum_{l=0}^{m} \mathcal{P}_s^{(m,l)} p(q - l, s, t)$$

where $\mathcal{P}_s^{(m,l)}$ is the probability that node s of starting in-degree $q - l$ receives exactly l links out of a total of m added in the time step (therefore obtaining in-degree q):

$$\mathcal{P}_s^{(ml)} = \binom{m}{l} \left[\frac{A_s}{A_\Sigma}\right]^l \left[1 - \frac{A_s}{A_\Sigma}\right]^{m-l}$$

Note that the probability that a node receives more than 1 link for each new node entering the network is vanishingly low for $t \to \infty$. Given the proper boundary conditions, after transition to the continuous-time approximation, the above equation can be solved and yields the following degree distribution:

$$P(q) \sim q^{-\gamma} \quad \gamma = 2 + a$$

As with previous models, in [KE02] and [CH03] the authors tackle the problem of measuring the average shortest path length and the clustering coefficient of the BA-model, with the following results:

Clustering The clustering coefficient can be computed, in the continuous approximation, by integrating the probability that any two nodes are connected over all the neighbors of a given node. The result states that:

$$c = \frac{m}{8|N|}(\log |N|)^2$$

Shortest Path The power law degree distribution implies that there are nodes in the network with a very large degree. These nodes act as hubs, therefore it is natural to expect a small shortest path length. It has been proven analytically that the average shortest path length of the Barabási-Albert network scales as:

$$l \sim \frac{\log |N|}{\log \log |N|}$$

The strength of this model is that of yielding, with simple dynamical rules, the scale-free property observed for the Internet topology. In particular, with an accurate choice of the power-law exponent $2 < \gamma < 3$, the model is able to reproduce the degree distribution observed in many real world networks. However, as far as the small-world property is concerned, the model is only able to reproduce the small shortest path length. In fact, these results reveal that the clustering coefficient decreases as the network size grows, which is an evident weakness of the model.

Newer Models Since both the Barabási-Albert model and the Watts-Strogratz model achieve good results with respect to some metrics but lack in others, more recent attempts in modeling the Internet focused on the extension of one or the other, in order to achieve the missing properties. The author of [New03] highlights a technique to generalize the random graph model by plugging inside it a specific degree distribution. This technique has been extended in [CZH07] to obtain a Watts-Strogratz model with tunable degree distribution. This technique is rather crude, as it simply reproduces observed properties without actually understanding the principles that lie behind them, therefore it is useless if we want to grasp the principles driving the growth of the Internet.

In [HK02], the BA-model is extended by the addition of a "triad formation step" in order to achieve higher and tunable clustering. The method is based upon the realization that in many real networks the probability of linking two nodes grows with the number of their common neighbors, as shown in [New01]: basically "friends of my friends tend to become my friends as well". An analytical formulation of this property and deriving consequences is described in [Vó3]. The realization of this phenomenon led to the construction of many other models based on the same considerations, like Community Guided Attachment [LKF05] and the Forest Fire Model [LKF05]. A new research branch intends to model the evolution of the Internet by taking advantage of existing economic models. The Wealth-based Internet Topology model (WIT) [WL10] captures the dynamics of open market competition between ASes by using a wealth function and forms new links by increasing the degree of ASes on the basis of their wealth. By using different wealth-exchange models, WIT can construct graphs with various degree distributions (among which the power law) and reproduce both the small-world effect and a high clustering.

While trying to reproduce coarse and medium-grained measures observed for the AS-level topology via reasonable considerations, all these models lack in explaining the driving forces responsible of the Internet's structure. As observed in Section 2.1.2, and further shown through the fine-grained measures of Section 2.2.2.3, the Internet has a hierarchical

structure made of tiers and communities [HRI+08]. Nevertheless, none of the proposed models generates this kind of structure. The importance of it can be understood through an enlightening story. Based on the preferential attachment mechanism, researchers claimed that the Internet displays a high degree of tolerance against random failures, but is extremely vulnerable to targeted attacks due to the presence of important hub nodes in the network [AJB00]. Reality is that Internet is extremely robust by design thanks to its multi-tiered structure and the ability to re-route traffic in case of failures [WAD09]. Moreover, a critical examination of the principles on which new ASes attach to old ones reveals that the mechanism is quite different from a preferential attachment [RWM+11], and the observed power law behavior is only evocative.

An interesting suggestion on the topic of modeling with communities comes from [LJKL09], where the authors try to obtain a community structure by exploiting the social similarity of nodes. The work however only hints at how this parameter should be set without giving a general definition. Moreover, the resulting community structure is tested using modularity measures that, as we have explained, poorly fit the data observed for the Internet's topology.

To sum up, empirical evidence shows that hierarchy and structure play a fundamental role in the shaping of the Internet, therefore, any realistic attempt to model this network has to deal with such features [PSV04; HRI+08]. This is a difficult but fundamental step if we want to avoid wrong inferences like the ones which have been made in the past.

2.3.2 Network Formation for the Internet

Static and dynamic models analyzed so far seek to model the network using approaches rooted in statistical physics. Network formation seeks to model how a network evolves by identifying mechanisms upon which nodes interact using strategic optimization. This kind of models where pioneered by Jackson and Wolinsky [JW96]. Basically, at the beginning of time, a network with a fixed number of nodes, also called agents or players in this context, is created. Every agent has a utility function,

which represents the benefits and costs of interconnecting with another player. Usually, forming or maintaining a link will have a cost, but having connections to other nodes will have benefits, such as the decreased distance to them. Each agent tries to maximize, in a myopic and selfish manner, its utility function. Since this utility function does not depend exclusively on himself, but also on the behavior of the other players, this task is classified as a game theoretical problem. The method seeks to understand, given some initial setting (parameters and utility functions), what kind of network structure will emerge as an equilibrium of this game.

Models based on this intuition where enhanced to take into account dynamic settings and evolution [CAAP06; Jac10]. An Internet model that finds its root in this kind of intuition is the heuristically optimized trade-off (HOT) model [FKP02], which shows how power-law forms can emerge as the result of trade-off mechanisms, i.e. the optimization of conflicting objectives for the different nodes.

Network formation models can be divided in two sub-classes, depending on their objectives and the tools used for the analysis: agent-based and game theory-based models.

Agent-Based Approach This approach typically relies on simulations to understand the dynamic of network formation and the resulting structure of the network. The main advantage of this method, is that it is possible to drastically reduce analytical considerations for the model, and once the initial setting and the utility function of each agent have been established, the outcome of the game can be found by simulating their behavior. Several agent-based computational models have been applied to the Internet, such as GENESIS [LDD12]. Despite of its interesting advantages, this approach has several drawbacks. First of all, in general it is not guaranteed that the simulation will converge, and it not known "where" it will converge: the simulation goes on in the hope of finding one of the possibly many equilibria. Second, even if it is possible to include more realistic considerations, this impacts on the analytical tractability of the model. As a matter of fact, analytical analysis can yield more interesting

results on the dynamic and the laws governing the interaction between the agents, something that is typically hidden by pure simulation.

Game Theoretical Approach The realm of network formation games (see [Jac10; Goy07]) investigates existence and properties of equilibria in a network created by rational players, each one with their costs and utility functions. Unfortunately these models can rarely be applied to study the Internet or any other real life network, due to the simplifications needed for mathematical tractability. An interesting advancement on this field comes from [MMO13], where it is shown that even very simple rules are able to provide properties which were measured on the topology and are not captured by notable models such as the BA-model. Despite of the impossibility to reproduce the whole Internet topology, this approach is very interesting since it makes it possible to deal with economic and technical constraints which are usually ignored by classical dynamic models. Thanks to this aspect, it is possible to model and capture important interaction mechanisms between the players that are typically hidden by other models, as we will do later (see Chapter 4).

2.3.3 Topology Generation

As already stressed in the introduction, one of the main goals for which a proper modeling of the Internet topology is important, is the ability to design topology-aware algorithms and routing protocols. To this end, it is fundamental that the graphs constructed by the models correctly reproduce the topological properties observed for the Internet. The computer science community has developed a series of *Internet Topology Generators*, which are based on the models previously described. Since the hierarchical nature of the Internet has been shown to play a fundamental importance, some of them try to embed structural properties based on this aspect. The most famous topology generators are:

ORBIS [MHK+07] is a topology generator which exploits the concept of dK degree distributions to construct the network. The 1K distribution corresponds to the degree distribution, while the 2K dis-

tribution is known as the joint degree distribution: $P(k_1, k_2)$. This distribution represents the probability that a randomly selected edge will be between nodes of degree k_1 and k_2. The maximum implemented mode is the 2K, which constructs the topology using as input the joint degree distribution. As the number d increases, the generated topology is more similar to the original one, however the input distribution size becomes impossible to manage.

GT-ITM [ZCD97] is the Georgia Tech Internetwork Topology Model. It is a hierarchical generator which uses the concepts of transit and stub so as to generate a topology made up of multiple levels, which can be either transit or stub domains. There are multiple attachment mechanisms (Random, Waxman, Exponential) dictating how to generate links within each domain, and it is possible to configure the number of transit-stub and stub-stub links. While this generator is unable to produce the scale-free Internet property, it is still interesting as it is one of the most configurable ones when it comes to reproducing hierarchical structures.

INET [WJ02] is an Autonomous System level topology generator developed by the University of Michigan. It uses a technique to combine the benefits of generalized random graphs, such as the small-world property, with a pre-built, tunable, degree distribution to also provide the scale-free property. Despite of this, it is unable to properly capture the hierarchical, community-structure of the Internet.

BRITE [MLMB01] is the Boston university Representative Internet Topology gEnerator. It is based on the Barabási-Albert model and, as such, it is able to reproduce the scale-free property.

The ORBIS generator is the most recent and complex, since it reconstructs the topology by exploiting a large amount of input data. The other three generators were presented around ten years ago, however, as reported by Haddai et Al. in their more recent survey [HRI+08], they are still the reference generators in the literature.

The main goal of the thesis, as already stressed, is developing a model which reflects the driving forces responsible for the structure of the Internet. The goodness of the model has to be validated, and we will do so by showing that it is able to both capture the statistical measures of the Internet topology, its community structure (in the form of maximal cliques) and outperforms the existing topology generators (see Chapter 6).

Part II

The Internet Core: Structure, Agents, Relationships

Chapter 3

The Twofold Internet Structure: Defining a Core

We tackle here the first problem stressed in the introduction: the large-scale heterogeneous nature of the Internet. Due to the high heterogeneity of the Internet topology, any model relying on a uniform class of nodes which all exhibit the same behavior is deemed to fail. Therefore, the first step that needs to be carried out is a deeper analysis of the topology, using concepts such as communities to identify and discover the different building blocks of the network.

We propose a technique that partitions the network topology into two distinct blocks: the **core**, which captures the underlying community structure of the Internet, and the **periphery**, representing the "tendrils" of the topology. The benefits of this innovative technique are twofold. First, it deals with the high heterogeneity of the Internet by highlighting a small yet well-structured core. This leads to a huge reduction in complexity and shows that the core of the large-scale Internet is not that large, and can further be broken down into a two-layer graph. Second, thanks to the simplifications introduced by the topology layering, it gives us the ability to deeply analyze the core, in order to discover statistical, structural and behavioral properties of its participating nodes. This second step will lay the groundwork for the subsequent chapters and for defining a proper

network model, first at the core level and then for the whole Internet [AGL13; AGL14a].

3.1 Introduction and Related Work

The Internet topology at the Autonomous System level, as stressed in Chapter 2, has been deeply analyzed in the past using several graph metrics to capture its statistical characteristics. The most important features observed as a result, are the small-world and scale-free properties. While these metrics have been widely used due to their ability to compactly represent statistical characteristics of the topology, they are unable to capture the structural characteristics of the network, such as the existence of tiers or the presence of communities.

Works dealing with the heterogeneity of the Internet using a community-based approach are the "*jellyfish*" and the "*medusa*" model [STF06; CHK+07]. Using different approaches, both models break down the full Internet map into different components exhibiting different behaviors, highlighting the existence of a central, strongly connected, component. The *jellyfish* model defines the core as a full-mesh, and then partitions the topology into layers by checking how far a node is w.r.t. the core. The *medusa* model uses instead the concept of k-core communities to define the "nucleus" of the network and separate it from the other components.

The methodology proposed here provides as well a partition of the Internet topology, dividing it into two main blocks. This is accomplished through two main steps. First of all we use real data to extract a small part of the topology which well represents what we call the "core", separating it from the rest of the network which we call the "periphery". Thanks to the extraction procedure, we show that although at first glance the Internet can be thought of as a large-scale complex system, its true nature is different, given that it is possible to extract a small core network exhibiting its most important structural properties. In the second step, the core is split into two layers, by observing that the emergence of cliques lies in the presence of two differently dense zones, and exploiting these different

densities.

The aims and tools used for this decomposition are quite different from the *medusa* and *jellyfish* methodologies which were described above. In particular, the main aims of that methodologies were to quantify the importance of nodes and checking if network models (e.g. the BA-model) produce the different levels highlighted for their model. Our purpose instead, is to understand which agents are responsible for the evolution of the Internet's core, and discover their statistical, structural and behavioral properties. In fact, with these in mind, it is possible to study the possible interactions between these agents and define mechanisms reflecting the driving forces of the Internet's growth, thus producing an accurate model for the Internet evolution. In addition, the main tool used for our partitioning are the maximal cliques, which have been shown to be more representative of the other techniques when applied to the Internet topology, due to the admitted overlaps (see Section 2.2.2.3 of Chapter 2).

Last, but not least, the power of our decomposition is also shown in [AGL13], where even a simple, data-driven analysis, is able to provide a fairly good model of the core. We skip that model here, since our main interest throughout the thesis is the proper understanding of the true driving forces behind the Internet's evolution.

The remainder of the chapter is organized as follows: in Section 3.2 we present our core definition, its extraction and decomposition techniques. Section 3.3 presents an in-depth analysis of the core, highlighting the importance of IXPs in its flat, community structure.

3.2 Topology Decomposition

In this section we provide our definition of the **core** of the Internet, in terms of the properties which it is assumed to fit. We then describe our extraction procedure, and decompose the core network into two layers by observing their different natures.

3.2.1 Core Extraction

Given our claims at the end of Section 2.2.2.3 in Chapter 2, here our aim is to identify the component of the Internet's topology that best represents its structure in the form of maximal cliques. We identify this component, a preferably small, densely connected subgraph, as the **core** of the Internet. Our core is thus made up of the minimum number of "important nodes" that grant a good fit for the maximal clique distribution of the graph of the whole Internet, according to the χ^2 test. The importance of a node is assigned through a weight, which depends on the number and size of the maximal cliques it participates in.

We extract the core of the Internet's measured graph using the dataset provided by ISOLARIO [Iso]. As already stressed, this dataset gathers data by several route collectors and processes it so as to obtain monthly-based AS-Level topology graphs. Our analysis covers almost eight years, from January 2006 to June 2013. Unless stated otherwise, for the sake of conciseness, we will refer to the time step of April 2013. However, the analysis was actually carried out for all of the time steps, and the results presented here are valid for the whole time span considered.

The extraction heuristic builds a graph $\hat{G}$ starting from an empty graph. It then makes it grow month by month in order to fit the maximal clique distribution of the graph of the Internet through the steps described in Algorithm 3.1. For each month m, the procedure starts by building a list L of nodes ordered by weight. We preserve the nodes of the previous graph $\hat{G}$ by putting them in the list with the highest possible weight (*lines 3-5*). Then, we observe the current (i.e. at month m) Internet graph and build a matrix $C_{num}[k][i]$ which reports the number of maximal cliques of size k for each node i. This matrix is used to assign a weight w to each node before inserting it into L (*lines 6-14*). We then perform a binary search on L to determine the minimum number of nodes h which provides a good fit for the maximal clique distribution, where the goodness of the fit is evaluated via Pearson's chi-squared test (*lines 18-28*). The test compares the maximal clique distribution of the Internet graph with the

Algorithm 3.1 Core Extraction

$\hat{G} = \emptyset$
for all $m \in$ Months **do**
  for all $i \in$ nodes($\hat{G}$) **do**
    insert(L, $< i, \infty >$)
  end for
  $G =$ get_topology(m) ▷ AS map/month m
  $C_{num} =$ get_maximal_cliques(G)
  $k_{max} =$ sizeof(C_{num}) ▷ Largest Clique Size
  for all $i \in$ nodes(G) **do**
    $w_i = \sum_{k=2}^{k_{max}} k * C_{num}[k][i]$ ▷ Weight
    if $i \notin L$ **then**
      insert(L, $< i, w_i >$)
    end if
  end for
  $C =$ maximal_clique_distribution(G)
  order_descending(L,w)
  $inf = 0; sup =$ sizeof(L)
  while $sup \neq inf$ **do**
    $h = inf + \frac{sup - inf}{2}$
    $G_h =$ tag_subgraph(G, get_subset(L,h))
    C_h= maximal_clique_distribution(G_h)
    $chi =$ chi_test(C, C_h, $k_{max}/2$, k_{max})
    if $chi >$ chi_limit($k_{max}/2$, p) **then**
      $sup = h$
    else
      $inf = h$
    end if
  end while
  $\hat{G} =$ tag_subgraph(G, get_subset(L,sup))
end for

one obtained by the subgraph induced by the h highest nodes in L, in the upper half of the distribution (the interval $[k_{max}/2, k_{max}]$) with the statistical significance parameter set to $p = 0.1$.

The test is performed on the upper part of the distribution because we believe that low values of k do not correspond to the network's core

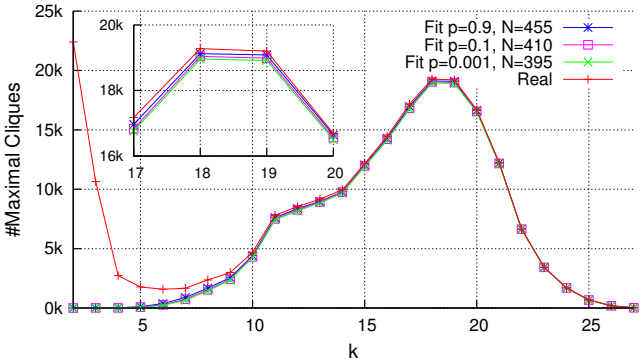


Figure 3.2.1: Maximal clique frequencies fit/Core size at various p - 2007

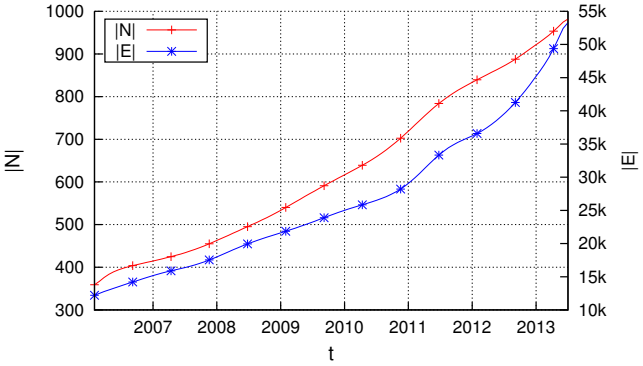


Figure 3.2.2: Core size: nodes and edges

but to less central parts, connected via Multi-Homed ASes and tree-like structures. The statistical significance parameter p influences the goodness of the fit. As $p \to 1$ the fit is more precise at the cost of an increasing core size, while for $p \to 0$ the core becomes smaller but the fit is less precise. Given the high variability of the maximal clique distribution, we are satisfied by a curve which exhibits the same trend, while an extremely precise fit is not needed. The experiments reported in Figure 3.2.1 show that a value of $p = 0.1$ gives a small core with a sufficiently precise fit.

The output $\hat{G} = (\hat{N}, \hat{E})$ is the core graph, which passed the χ^2 test

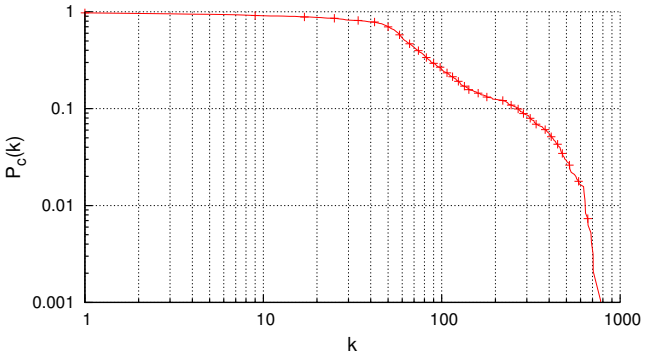


Figure 3.2.3: Degree ccdf for the Internet's core - 2013

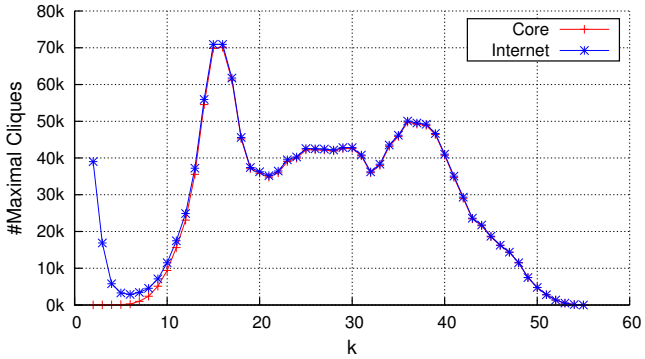


Figure 3.2.4: Maximal clique frequencies for the Internet's core - 2013

with the minimum value of h. The obtained core $\hat{G}$ is a small, relatively dense network. Figure 3.2.2 shows the evolution of the number of nodes and edges within the core.

As of April 2013, the core has only $|\hat{N}| = 956$ nodes and $|\hat{E}| = 50,440$ links, a heavy-tailed degree distribution (Figure 3.2.3), and a wide multimodal maximal clique distribution. Figure 3.2.4 shows how this distribution matches almost perfectly the one observed for the whole Internet. This behavior is stable throughout the whole considered time span.

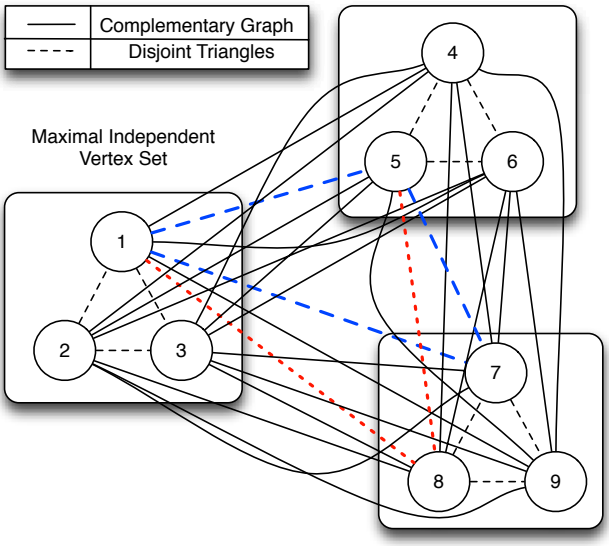


Figure 3.2.5: Graph with the highest number of maximal cliques

3.2.2 Emergence of Maximal Cliques

The extraction procedure and the results described in the previous section highlight that although the Internet is a large scale complex system, it is still possible to identify a very small, dense, central component that captures its structural properties very well in terms of communities, expressed through a high number of maximal cliques. Here we outline the driving forces behind the emergence of such a huge number of cliques.

In 1969 Moon and Moser [MM65] showed that a network of $|N|$ nodes has at most $3^{|N|/3}$ maximal cliques. This number can be achieved by complementing the graph which has $|N|/3$ disjoint triangles. More generally, a graph with a high number of maximal cliques can be obtained by partitioning the network into n sets and connecting each node of a set to all the other nodes in a different set. Each set is thus an independent vertex set (IVS), that is, a set of nodes where no two are adjacent. Figure 3.2.5 depicts a graph obtained by applying the Moon-Moser technique to a

network of 9 nodes, thus leading to $3^{9/3} = 27$ cliques. Basically each node of an IVS forms a clique with any couple of nodes taken from different IVSs; in Figure 3.2.5 node 1 forms the following cliques: $\{1, 4, 7\}$, $\{1, 4, 8\}$, $\{1, 4, 9\}$, $\{1, 5, 7\}$ (dashed blue), $\{1, 5, 8\}$ (dotted red), $\{1, 5, 9\}$, $\{1, 6, 7\}$, $\{1, 6, 8\}$, $\{1, 6, 9\}$.

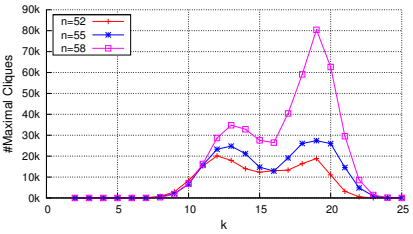
Clearly this network has a huge number of links resulting in an extremely high density. Now consider the following *two-layer random graph model*:

- There are n "*important*" nodes. Each node is connected to any other node with high probability q.
- The remaining $|N| - n$ nodes are partitioned into n IVSs, each of which has an *important* node and other l*ess-important* nodes.
- Each *less-important* node is connected with an *important* node of another IVS with probability $p \ll q$.

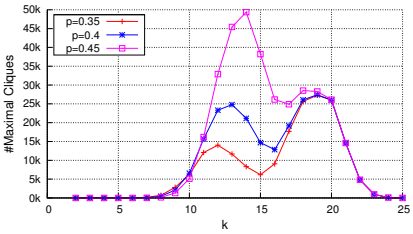
Basically, in this model we draw $|N|$ nodes, out of which n are *important*, and $\frac{n(n-1)}{2}q + (|N| - n)(n - 1)p$ edges: the first part of the sum is for *important* nodes, the other for *less-important* nodes.

This network is a subset of the Moon-Moser graph shown previously, where the network was separated into two components: a very dense one responsible for the generation of large cliques and a less dense component which generates smaller cliques. Figure 3.2.6 shows how this model generates a high number of cliques for various values of $|N|$, n, p, q. By changing the parameters we can generate all kinds of bimodal distributions: boosting the *important* nodes number n or their connection probability q results in a bigger right mode (Figures 3.2.6a and 3.2.6c), while doing the same for *less-important* ones strengthens the left mode (Figure 3.2.6b). As p increases, the left-hand mode can jump higher than the right-hand one, therefore in order to obtain coherent distributions with those observed for the Internet's topology, it is crucial that p and q remain far apart.

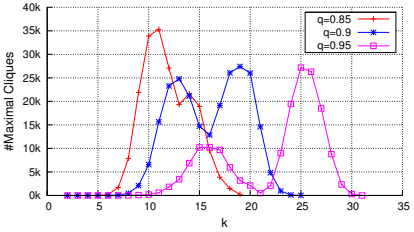
Our experiments show that in order to reproduce maximal clique distributions similar to that of Internet's topology, we need $n \approx 60 \div 70$



(a) Case 1: $|N| = 800, p = 0.4, q = 0.9$ - change n



(b) Case 2: $|N| = 800, n = 55, q = 0.9$ - change p



(c) Case 3: $|N| = 800, n = 55, p = 0.4$ - change q

Figure 3.2.6: Maximal clique distribution for 2-Layer Random Model

important nodes connected with a very high probability $q \approx 0.9 \div 0.95$ (these values depend on the actual year, of course).

3.2.3 Core Decomposition

The emergence of the high number of network conglomerates within the core seems to be tightly connected with the presence of zones with different densities inside it. As we have just shown, one possible reason for the high number of maximal cliques in the core is the coexistence of one very dense zone of the network and another less dense zone. This leads us to the analysis of the density d of the network. We recall from Chapter 2 that network density is defined as the ratio between its number of edges e and its maximum possible number: $d = 2e/(n(n-1))$, where n is the number of nodes.

Our main focus here is thus to analyze the density of the core just

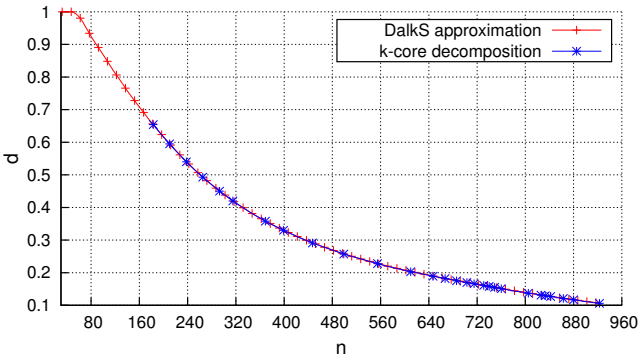


Figure 3.2.7: Network density as a function of n for the Core - 2013

extracted, in order to understand if such different dense zones exist. A well-known method for network decomposition is the k-core [CHK+07]; however, when applied to our core network this technique very soon stops yielding as result a large network of 183 nodes, corresponding to the 83-core, with density $d \approx 0.65$. Our goal instead is to extract a smaller and hopefully denser network. The problem of extracting the densest subgraph of a given size (DalkS, Densest at-least-k Subgraph [KBS09]) is known to be NP-Complete, however there exists a simple 2-approximation algorithm which solves it iteratively. At each step the method computes the degree of each node and removes the node with the minimum degree and all its links. Thanks to this procedure, the core can be decomposed by removing all its nodes one by one, until we obtain a clique, which of course has the maximum density $d = 1$. These techniques have been applied to the Internet, with different aims from ours in [WS04], where the authors searched for subgraphs of a specific density and highlighted the geographical locality of the obtained clusters.

Figure 3.2.7, which we refer to as the "**density curve**", shows the density of the graph $\hat{G}$ as the number of nodes decreases. The figure also shows the density of the various k-cores obtained via k-core decomposition. This pictures refers to the single time step of November 2012, however, the same behavior applies to all the other considered time steps. It is clear that the densities of the k-cores adhere to those obtained with

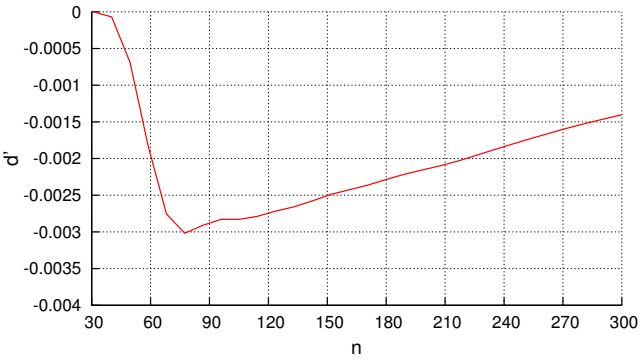


Figure 3.2.8: First derivative of network density for $n \in [30; 300]$ - 2013

the above method (up to the 83-core, composed of 183 nodes, where the k-core decomposition terminates).

Although there is no clear distinction between the two zones as conjectured in the previous section, Figure 3.2.7 highlights two main points: the existence of a very dense central zone made up of very few nodes, and the exponential decrease in density as the network size grows. Figure 3.2.8 shows the first derivative of network density, zoomed in the interval $[30; 300]$: the curve has a unique minimum for $\bar{n} \approx 70$, indicating an inflection point for the density curve. Consequently the exponential decrease is not a constant behavior, but something that happens only for $n > \bar{n}$, while for a smaller value of n, the density decrease is slower.

We believe that the different behavior of the density curve reflects a different nature of the network before and after $\bar{n}$. We thus separate the nodes of graph $\hat{G}$ into the following partitions:

The *Centrum* contains $N_1 = \bar{n}$ nodes obtained by applying DalkS to $\hat{G}$ up to the inflection point obtained at $\bar{d}$.

Layer-1 contains N_2 nodes with a direct connection to *centrum* nodes.

The second layer contains all the other nodes.

The second layer consists of very few nodes and has a negligible contribution to the metrics of interest, so this part can be discarded. Thanks to the

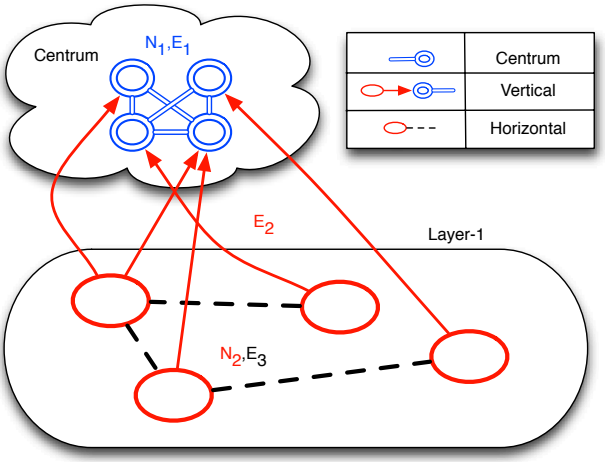


Figure 3.2.9: Core network decomposition

partitioning, we can now decompose the core into three networks:

The *centrum* network with N_1 nodes and E_1 edges, consists of all *centrum* nodes and their mutual connections.

The vertical network with $N_1 + N_2$ nodes and E_2 edges, consists of the *centrum* network plus all *layer-1* nodes and their vertical connections (that is, connections to the *centrum*).

The horizontal network with N_2 nodes and E_3 edges, consists of only *layer-1* nodes plus their horizontal connections (that is, connections to other *layer-1* nodes).

Figure 3.2.9 depicts our network. Interestingly, while the density curve changes at each time step, the inflection point $\bar{n}$ varies slowly from about 50 to 70 over the analyzed time span, therefore while the size of *layer-1* grows, the *centrum* is almost static or slowly growing. Moreover, the value of $\bar{d}$ at the inflection point is about $0.9 \div 0.95$. This value is in 1 to 1 correspondence with the connection probability of a random graph (see Section 3.2.2), and the values of $\bar{n}$ and $\bar{q}$ therein used to obtain maximal

clique distributions akin to that of Internet's topology, agree almost exactly with those just obtained.

3.3 Core Analysis

The decomposed network is further examined to better characterize its participants and their interconnections, and lay the groundwork for defining a core network model.

3.3.1 Structural Analysis

As shown in Figure 3.2.9, our core is a 2-layer graph made up of three different networks. We can now study the properties of these networks in terms of the degree and maximal clique distributions. The *centrum* is a network of very few well connected nodes. Figures 3.3.1 and 3.3.2 show that the main characteristic of this network is not the degree, which is similar for all nodes, but the formation of a small number of fairly large maximal cliques.

As shown in Figure 3.3.1, the *vertical* network has a very particular degree distribution, with a turning point at p_v ($p_v \approx 65$ as of November 2012). If we split the curve into two parts, for $k < p_v$ we have the distribution of the N_2 *layer-1* nodes, while for $k > p_v$ we have the distribution of the N_1 *centrum* nodes. Since the *layer-1* nodes are not connected to each other in the *vertical* network (these links have been placed, by construction, in the *horizontal* graph), the distribution for $k < p_v$ tells us exactly how many links each node should inject in the *vertical* network. Figure 3.3.3 zooms in on the first part and reveals that this distribution is approximately uniform $\mathcal{U}(1, p_v)$. As well as the *centrum* size $\bar{n}$, parameter p_v is also approximately constant, just slowly growing, during the analyzed time span.

The *horizontal* network (Figures 3.3.1 and 3.3.2) provides a small contribution in terms of large cliques. On the other hand, its degree distribution is similar to that of a scale-free network with a smooth cutoff, just like the networks obtained via classic preferential attachment modeling.

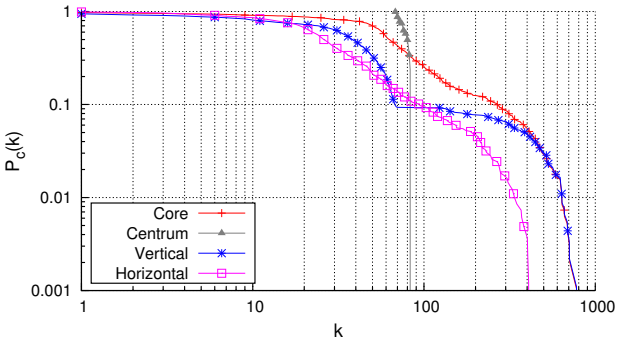


Figure 3.3.1: Degree ccdf for core decomposition - 2013

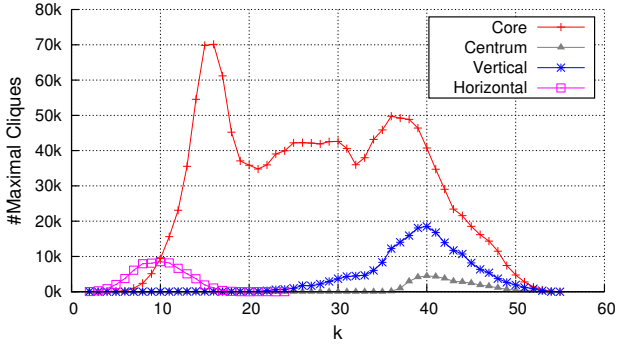


Figure 3.3.2: Maximal clique frequencies for core decomposition - 2013

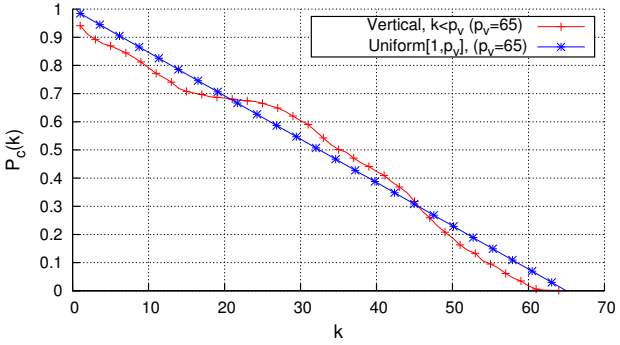


Figure 3.3.3: Real vs Uniform: vertical network degree for $k < p_v$ - 2012

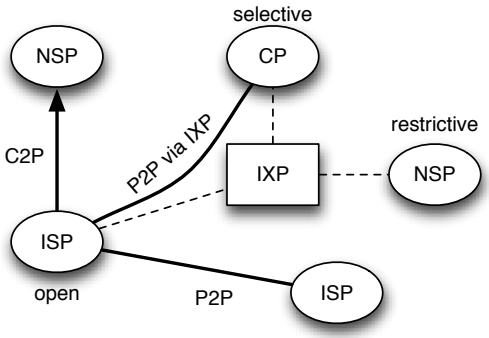


Figure 3.3.4: Types of nodes/links

3.3.2 Impact of IXPs within the Core

In order to better characterize the core, we exploited the enriched dataset shown in Section 2.2.1.1 of Chapter 2, and labeled nodes and links:

Links were labeled by adding the economic relationship between ASes. We recall that relationships can be either customer to provider (C2P, or transit), when the former pays the latter for the service, or peer to peer (P2P, or peering) when two ASes agree to exchange traffic for free.

Nodes were labeled according to the following data, publicly available from PeeringDB [pee]:

Node Type, either Network Service Provider (NSP), Internet Service Provider (ISP), Content Provider (CP) or Other (Non-Profit, Educational, Research, etc..);

Number of IXPs a node is attached to, which can be directly computed by exploiting IXP participation lists;

Node Peering Policy, either open, selective or restrictive, according to the willingness of the node to participate in peering sessions over an IXP.

Link/Network	Core	Centrum	Vertical	Horizontal
C2P	20.7%	2%	16.2%	28.5%
P2P	79.3%	98%	83.8%	71.5%

Table 3.1: Link types - 2012

Network	NSP	ISP	CP	Other
Centrum	70%	18.3%	0.1%	11.6%
Layer-1	37%	27.8%	16.1%	19.1%

Table 3.2: Node types - 2012

Network	$\overline{\mathbf{k_{Internet}}}$	$\overline{\mathbf{k_{P2P}}}$	$\overline{\#\mathbf{AttachedIXPs}}$
Centrum	525.5	444.6	8.35
Layer-1	112.5	55.8	3.48

Table 3.3: Node properties - 2012

Figure 3.3.4 summarizes the labels available. Our core network is predominantly made up of peering links (Table 3.1), especially the *centrum* which is nearly exclusively made up of P2P links. The vast majority of the nodes within the core are Network Service Providers (Table 3.2), which is expected given that these ASes provide backbone access to the Internet, and participate in IXPs. There are few Content Providers in the core, probably due to the incompleteness of the Internet graph, especially concerning peering links [RWM+11]. A comparison between the two layers of our decomposition highlights that *centrum* nodes typically have a higher degree (indicated with k in Table 3.3) and participate in more IXPs.

If we just consider the ASes which are NSPs, participate in four or more IXPs and have a peering degree that is higher or equal to 100, we find that 33 nodes of the *centrum* (55%) lie in this category, while only 28 nodes (3.8%) of *layer-1* nodes adhere to these kinds of specifications. A further analysis, reveals that *centrum* nodes falling into this category

typically have an Open/Selective peering policy, while all but one of the *layer-1* nodes have a Restrictive/Selective policy.

In summary, peering links are the glue that keeps the core well connected. The main difference between *centrum* and *layer-1* nodes is that while the former have a general tendency to establish well connected clusters by participating in many IXPs with a generally open peering policy, the latter (unless small ASes) tend to be more restrictive in establishing peering connections. Interestingly, Tier-1 nodes are located inside *layer-1* due to their restrictive peering policies (they do not peer at IXPs).

3.4 Conclusions

The decomposition technique enables us to shrink the large scale Internet, with over forty thousand nodes, to a relatively small network of only a hundred nodes. The benefits of this are twofold: on the one hand, while keeping the most interesting structural properties of the network intact, we are able to perform a very simple analysis of a two-layer graph. On the other, this analysis enlightens some important aspects related to the current structure of the Internet.

Although the core extraction and decomposition technique has been specifically applied to the Internet environment, it can be used, mutatis mutandis, on many other networks (e.g. social networks). Depending on the specific characteristics of the network under investigation, a core might emerge or not. The more regular the structure of the network is, the larger the core will be, otherwise (i.e. for heterogeneous networks such as the Internet) a smaller core will show up. Depending on the internal core density, multiple layers might emerge. The tool is thus fairly general, and can be used to gain insight on the internal structure of many different multi-layer networks.

The in-depth analysis carried out in Section 3.3.2, illustrates the key role of IXPs within the core, both in the *centrum*, where there is a high participation and a general openness, and in the layer-1, where nodes are still participating to IXPs, but with more restrictions. The core embodies perfectly the Internet flattening and the emergence of communities, two

intertwined phenomena leading the network growth.

The emerging key role of IXPs poses some questions: what is the relationship between IXPs and backbone service providers (NSPs), historically responsible for the Internet connectivity? And what is the relationship between other nodes, i.e. Internet Service Providers and Content Providers, and these two facilities (i.e. IXPs and NSPs) enabling them to communicate? Answering those questions will enable us to understand the mechanisms used by nodes to establish links between each other.

Chapter 4

Peering or Transit: IXPs vs NSPs

The second problem shown in the introduction, is tackled here. As we stressed, network evolution is not dictated by a central authority, but is mainly the outcome of local economic and technical constraints. Therefore, we have to understand the behavior and the decision of the Autonomous Systems living in the Internet ecosystem by taking them, rather than the network, as the main characters of the system, and study their reciprocal interactions. Here, we propose a model to analyze the decisions taken by ASes living in an Internet environment. In particular, due to the outcome of the analysis from the previous chapter, we focus on the core and on the key role of IXPs as an alternative to the old paradigm using NSP backbone providers [AAGL14b; AAGL14a].

4.1 Introduction and Related Work

The Internet, as stressed in the first chapter, consists of thousands of Autonomous Systems, independently administered networks that dynamically connect together to provide end-to-end reachability. Depending on their importance and the offered services, ASes can be sorted in different tiers and categories: content providers, access providers, transit providers

and so on [Nor11]. During the past century, Internet's structure was predominantly hierarchical, with customer ASes paying their providers to carry their traffic, and the latter providing default gateways to reach any requested destination. The pricing strategy for this transit policy is typically volume-based, metered using the 95th percentile traffic sampling technique (this allows customer ASes to burst, for a limited time period, beyond their committed base rate) [Nor10b].

Nowadays Internet, as already stated, is evolving from its previous structure and becoming flatter through the introduction of peering meshes, with ASes establishing bilateral agreements to exchange traffic between them for free. Peers must agree to each other's policy, which is used to avoid abuse of the peering relationship. Typical clauses include prohibition of using the peer as default gateway (therefore peers cannot be used to reach other Internet's routes) and traffic ratio balancing, meaning that the ratio between incoming and outgoing traffic over the link must not exceed some value (e.g. 2:1) [Nor10a].

The massive diffusion of peering meshes was mainly achieved thanks to the increasing deployment of Internet Exchange Points (IXPs) [GILO11]. The pricing strategy of an IXP, with respect to its customers, is typically flat. Each one of them pays a monthly-based fee, depending on the size (speed) of the port they are using and the cost of maintaining the equipment. Thanks to this mechanism, the IXP can share maintenance costs among all its participants [Nor10c]. It is worth noting that this pricing strategy doesn't allow standard cost function modelization (like in [ORS93]), since the addition of new participants potentially brings down the costs of an IXP customer.

ASes joining the Internet face the complex question of what is the best strategy for offering their services (e.g. traffic delivery) at the lowest possible cost. While the answer was easy in the past century, due to the existence of transit as the unique interconnection policy, today's answer is much more complex. In fact, peering policies and IXPs brought new variables to the problem, such as the fact that the outcome of an AS decision also depends on what other ASes, dealing with the same problem, do: due to this aspect, we believe that a game-theoretical analysis of the

problem would be highly insightful. Realistic modeling of the whole decision space of an AS is an extremely difficult task, therefore in our work we restrict our analysis to the problem of peering versus transit. As a matter of fact, a proper understanding of this problem is crucial if we really want to grasp the behavior of ASes living in the Internet ecosystem, and answer the two questions that arose in the last part of previous chapter.

In order to keep the problem analytically tractable we do not aim at modeling the whole network formation process. We rather study the interaction between ASes which connect to an existing network in order to serve some demands. In this context, a possible modelization of a network where access providers need to select a subset of content providers and fetch traffic from them in a cost-efficient manner is given in [JHA12]. However, the aim of this work is quite different from ours, as it concentrates on the economic analysis of neutral/non-neutral network features, without taking into account the difference between traffic and peering agreements. The models in [CRT00; N06] focus on the competition between backbone providers and on the conditions under which they agree or refuse to establish bilateral peering connections with each other. In [SS06], authors perform an interesting analysis on network pricing and analyze the economics of private Internet exchanges. However, as already explained in Section 2.1.3 of Chapter 2, private peering has different rules and costs compared with public peering. Nowadays Internet is largely dominated by public peering, occurring at IXPs, therefore in our work we concentrate on this last phenomenon, which allows us to give different insights on the present difference between transit and peering.

The work carried out in this chapter brings contributions both from a game theoretic perspective and an engineering perspective. First of all this is, to the best of our knowledge, a novel model to analyze the strategic choices of ASes living in an Internet environment with both technological and economic constraints. The modelization takes into account many realistic elements, which do not fall into standard frameworks, yet tries to keep the problem mathematically manageable. From a game-theoretic perspective, we prove that our game falls in a specific category for which

we both demonstrate the existence of equilibria and provide an algorithm for computing stable solutions. From an engineering perspective, the outcome of the analysis is highly insightful as it shows both the suboptimality of the decentralized solution and the emerging competition between the two facilities enabling either transit or peering connectivity: Network Service Providers and Internet Exchange Points.

In Section 4.2 we define the general model, in Section 4.3 we analyze a simplified "minimal complexity" model, derive the existence of equilibria and the algorithm to compute them, and in Section 4.4 we show inefficiencies of the decentralized solution. Section 4.5 extends convergence results to the general model, then Section 4.6 shows simulative results for both converging and non-converging cases.

4.2 General Scenario

In the following we describe the general scenario under investigation and derive the cost function.

4.2.1 Description

Using once again our enriched dataset (Section 2.2.1.1 of Chapter 2), we group ASes in their different categories. We concentrate on the following main categories present within the core of the Internet, and assign a specific role to each of them, so as to derive a game theoretic decision model:

Internet Service Provider (ISP)[1] this node gives to eyeballs (i.e. the end users) and lower tiers, access to the Internet and its contents. Each service provider has a traffic demand, hereafter demand, which represents the amount of traffic (uplink+downlink) that it handles.

[1] ISPs, NSPs and CPs are typically ASes. IXPs are not ASes, even if their infrastructure is under a single administrative control.

Content Provider (CP)[1] this node has physical access to the contents users are looking after, therefore an ISP with a demand for his specific content, has to connect to it in order to serve this demand.

Internet Exchange Point (IXP)[1] this is the facility that provides peering connection to all its participants. This means that all the nodes connected to a given IXP can potentially communicate with each other.

Network Service Provider (NSP)[1] this node is located at the highest hierarchical level of the network, meaning that each CP can be reached through it. ISPs can reach CPs by establishing a transit connection with an NSP. In particular, here NSPs can be either Tier-1 ASes, or high-level backbone service providers, whose main interest is to sell transit to lower tiers.

We consider a network with $i \in \{1, ..., I\}$ ISPs, $n \in \{1, ..., N\}$ CPs and $l \in \{1, ..., L\}$ transmit facilities (TF), that can be either the NSPs or the IXPs. Without loss of generality, we impose that TFs $j_1 \in \{1, ..., l_1\}$ are NSPs, while TFs $j_2 \in \{l_1 + 1, ..., L\}$ are IXPs.

The main difference between transmit facilities is that while links to NSPs are established through transit connections, links to IXPs are established through peering connections.

In a *transit connection*, or customer-to-provider (C2P) connection, the cost to the customer is a function of the amount of traffic that crosses the link (typically expressed as \$/Mbps).

In a *peering connection*, the price is generally flat and depends on the size of the port the customer buys. Moreover, when peering connections are maintained by an IXP, the costs are shared among all the participants.

Each ISP i has a demand for a CP n, which we indicate as ϕ_i^n. The players of our game are the ISPs, which need to decide how to split their demand among all possible transmit facilities. Figure 4.2.1 depicts our network scenario.

We indicate with $x_{i,l}^n$ the flow from ISP i to CP n via TF l. The strategy of ISP i is given by the vector $\mathbf{x_i} = (x_{i,1}^1, ..., x_{i,l}^n, ..., x_{i,L}^N) \in \mathbb{R}^{L \times N}$, while the strategy of all the other players is expressed as $\mathbf{x_{-i}} \in \mathbb{R}^{(I-1) \times L \times N}$. The

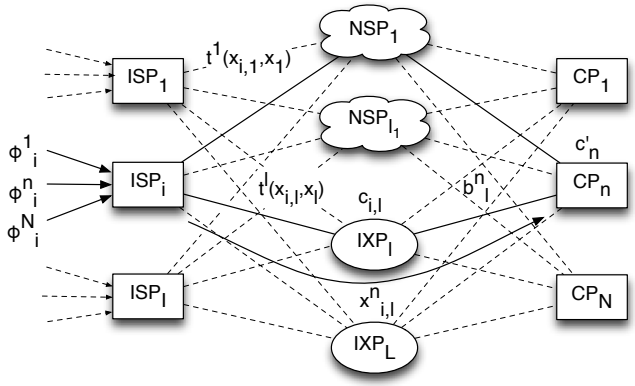


Figure 4.2.1: General model

goal of each player is to serve, at the minimum possible cost, his demand $\phi_{\mathbf{i}} = (\phi_i^1, ..., \phi_i^N) \in \mathbb{R}^N$ by splitting it into several flows $x_{i,l}^n$. Please note that we are not dealing with the issue of complete connectivity for the ISPs (in which case, it is sufficient to deploy a single transit connection): our aim is just to enable them to serve their specific demands.

We also indicate as:

- $x_{i,l} = \sum_{n=1}^{N} x_{i,l}^n$ the total flow from ISP i to TF l;
- $x_l = \sum_{i=1}^{I} \sum_{n=1}^{N} x_{i,l}^n$ the total flow at TF l.

Each player, say i, for each transmit facility (NSP or IXP) it connects to, incurs some costs:

TF usage cost this cost depends on the transmit facility used. If it is an NSP, then it is a function of $x_{i,l}$, the flow from the player to the NSP. Otherwise the TF is an IXP and the cost is shared among all the participants, therefore it also depends on the other players, in the form of x_l, the total flow at the IXP. Consequently, this cost can be written as a function:

$$t^l(x_{i,l}, x_l) \tag{4.2.1}$$

TF capacity cost each link between an ISP and a TF has a fixed capacity $c_{i,l}$; this means that we have a constraint of the form $x_{i,l} \leq c_{i,l}$. While we may introduce it in the problem "as is", this would make the model less manageable. Moreover, due to performance and congestion issues, network operators typically avoid reaching the capacity limit and keep a margin for traffic fluctuations. We can think of this performance degradation as a "virtual cost" for the ISP, and therefore model the constraint as a cost, that increases as the flow over the link approaches the capacity limit (as is typically done in the literature for M/G/1 Processor Sharing queues [ORS93]):

$$\frac{1}{c_{i,l} - x_{i,l}} x_{i,l} \tag{4.2.2}$$

We are aware that, in reality, network operators adjust this capacity when there is more demand for it, and the interconnection cost grows accordingly. However, this situation can be avoided as long as our working region is sufficiently far away from the saturation point. We will always assume that capacities are symmetric w.r.t. the players, therefore $c_{i,l} = c_l \; \forall i$. Typically the capacity of the NSP can be assumed to be much larger than that of IXPs: $c_{NSP} \gg c_{IXP}$ (see [Nor10b] and [Nor10c]).

CP reachability cost let's indicate with b_l^n the cost of transporting one unit of flow from TF l to CP n. This cost is not relevant from the player's perspective (it is paid by the CP), however it can be used to express the reachability of a given CP. In fact, while all the CPs are connected to the NSPs, an IXP can be connected only to a subset of CPs. This phenomenon can be expressed by putting:

$$b_l^n = \begin{cases} 0 & if\ l \leq l_1\ or\ IXP_l \rightarrow CP_n \\ \infty & otherwise \end{cases} \tag{4.2.3}$$

Thanks to all these considerations, the cost function for player i can be expressed as the sum of (4.2.1), (4.2.2) and (4.2.3):

$$C_i(\mathbf{x_i}, \mathbf{x_{-i}}) = \sum_{l=1}^{L} \left(t^l(x_{i,l}, x_l) + \frac{1}{c_l - x_{i,l}} x_{i,l} \right) + \sum_{l=1}^{L} \sum_{n=1}^{N} x_{i,l}^n b_l^n \tag{4.2.4}$$

In order to serve all the demands, each player i has to satisfy the flow constraint: for every CP, the total flow has to be equal to the demand ϕ_i^n. Therefore player i's best response $BR_i(\mathbf{x_{-i}})$ is obtained by minimizing cost function (4.2.4), subject to the flow constraints (4.2.5):

$$\begin{cases} BR_i(\mathbf{x_{-i}}) & = \arg\min_{\mathbf{x_i}} C^i(\mathbf{x_i}, \mathbf{x_{-i}}) \\ s.t. & \sum_l x_{i,l}^n = \phi_i^n \quad \forall n \end{cases} \tag{4.2.5}$$

The vector $\mathbf{x}^* = (\mathbf{x_1^*}, ..., \mathbf{x_I^*}) \in \mathbb{R}^{I \times L \times N}$ is an equilibrium of the game if and only if $\mathbf{x_i^*} \in BR_i(\mathbf{x_{-i}^*})\ \forall i$, that is, if the strategy of each player is a best response to the strategies of other players.

Throughout the chapter we will always refer to the description of Figure 4.2.1, however, mutatis mutandis, the results are still valid for scenarios where players are CPs or a mix of CPs and ISPs, as long as the demands are changed accordingly.

4.2.2 Transmit Facility Usage Cost

The TF usage cost is different between the NSPs and the IXPs. More specifically the NSP usage cost is linear in the amount of flow that each player sends to it [Nor10b]. Therefore we can write:

$$t^l(x_{i,l}, x_l) = a_l x_{i,l} \qquad l \leq l_1 \tag{4.2.6}$$

where $a_l,\ l \leq l_1$ represents the transit price of NSP l per unit of flow. We are aware that, due to economies of scale in the traffic delivery, transit costs are subadditive in reality. However, introducing this aspect would overcomplicate the model, hiding the truly interesting differences between transit and peering. Nevertheless, we are able to show that some of our results still hold for more generic transit cost functions (see below, Theorem 5).

For the IXP usage [Nor11], each player has to pay a share of the total cost of IXP maintenance. This share can be expressed as the ratio between the flow sent by player i on IXP l and the total flow crossing that IXP: $\frac{x_{i,l}}{x_l}$. Assume we can write the cost of maintenance of IXP l as a function h_l of

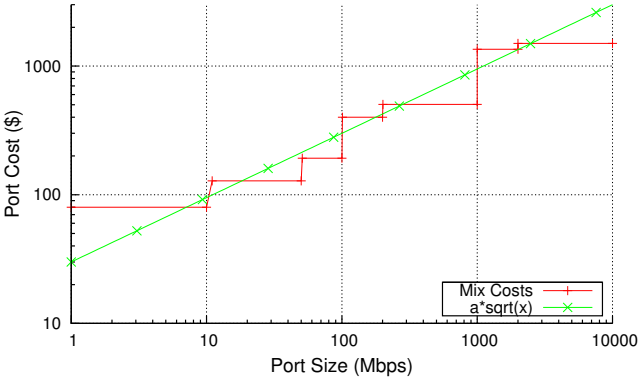


Figure 4.2.2: IXP port costs for MIX (log-log scale)

the total flow through the IXP, therefore the usage cost is:

$$t^l(x_{i,l}, x_l) = \frac{x_{i,l}}{x_l} h_l(x_l) \qquad l > l_1 \tag{4.2.7}$$

The cost of maintaining the equipment of an IXP is, in general, a non-linear function of several parameters. In order to keep the problem manageable, we will approximate this cost with that of a single port which handles x_l, the entire flow over the IXP. The cost of a port is a step-wise increasing function, as shown in Figure 4.2.2 for the MIX[2], an Italian IXP. This type of cost functions can be modeled (see [MDFL12]) by using a function like x^α with $\alpha \in [0.4; 0.7]$. For simplicity, we take $\alpha = 0.5$ as this value provides a fairly accurate fit (shown in Figure 4.2.2). Therefore, we express the maintenance cost as:

$$h_l(x_l) = a_l \sqrt{x_l} \tag{4.2.8}$$

where $a_l, l > l_1$ is a constant relating the total flow through IXP l with its maintenance cost. By putting together definitions (4.2.6), (4.2.7) and

[2]Milan IXP - public peering costs available online: http://www.mix-it.net

(4.2.8), the cost function (4.2.4) can be rewritten as:

$$C^i(\mathbf{x_i}, \mathbf{x_{-i}}) = \sum_{l=1}^{l_1} a_l x_{i,l} + \sum_{l=l_1+1}^{L} \left(\frac{a_l}{\sqrt{x_l}} x_{i,l} \right) + \\ + \sum_{l=1}^{L} \left(\frac{1}{c_l - x_{i,l}} x_{i,l} \right) + \sum_{l=1}^{L} \sum_{n=1}^{N} x_{i,l}^n b_l^n \tag{4.2.9}$$

Now, we define these new functions:

$$\begin{cases} f^l \left(\sum_i \sum_n x_{i,l}^n \right) = \begin{cases} a_1 & l \leq l_1 \\ \frac{a_l}{\sqrt{\sum_i \sum_n x_{i,l}^n}} & l > l_1 \end{cases} \\ g^l \left(\sum_n x_{i,l}^n \right) = \frac{1}{c_l - \sum_n x_{i,l}^n} \end{cases} \tag{4.2.10}$$

By using (4.2.10) and recalling that $x_{i,l} = \sum_n x_{i,l}^n$ and $x_l = \sum_i \sum_n x_{i,l}^n$, we can rewrite (4.2.9) as:

$$C^i(\mathbf{x_i}, \mathbf{x_{-i}}) = \sum_l \sum_n x_{i,l}^n \left[f^l \left(\sum_i \sum_n x_{i,l}^n \right) + g^l \left(\sum_n x_{i,l}^n \right) + b_l^n \right] \tag{4.2.11}$$

Equation (4.2.11) is the most general expression of the cost function for each player. Please note that (4.2.11) is in general a non-convex function of $x_{i,l}$, and therefore we cannot directly establish existence of pure equilibria. In particular it does not comply with the general assumptions used for link cost functions in the framework described in [ORS93]. Nevertheless, we cannot avoid dealing with functions of this shape if we want to properly grasp the difference between transit and peering strategies offered, respectively, by NSPs and IXPs.

4.3 Minimal Complexity Model (MCM)

In order to gain insights on the problem solution, in this section we analyze a simpler model, which we call Minimal Complexity Model (MCM), where $I, L, N = 2$. The MCM has two ISP players, two CPs (therefore each player only has two demands, namely ϕ_i^1 and ϕ_i^2) and two

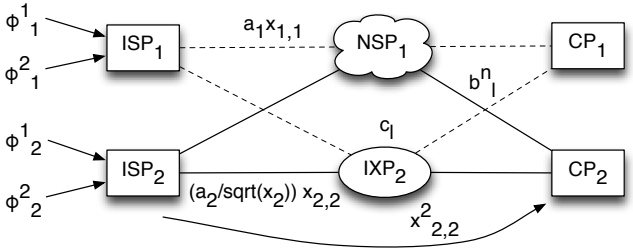


Figure 4.3.1: Minimal complexity model

transmit facilities available, either the NSP ($l = 1$) or the IXP ($l = 2$), as depicted in Figure 4.3.1. With some algebraic manipulations, explicitly shown in Appendix 4.8.1, we can rewrite cost function (4.2.11) for player 1 of the MCM as:

$$
\begin{aligned}
&C^1(x_1, x_2, y_1, y_2) = \\
&= \left(\phi_1^1 + \phi_1^2 - x_1 - x_2\right)\left(a_1 + \frac{1}{c_1 - (\phi_1^1 + \phi_1^2 - x_1 - x_2)}\right) + \\
&+ (x_1 + x_2)\left(\frac{a_2}{\sqrt{x_1 + x_2 + y_1 + y_2}} + \frac{1}{c_2 - (x_1 + x_2)}\right) + \\
&+ \left(\phi_1^1 - x_1\right) b_1^1 + \left(\phi_1^2 - x_2\right) b_1^2 + x_1 b_2^1 + x_2 b_2^2
\end{aligned}
\tag{4.3.1}
$$

where x_n is the flow sent from player 1 to CP n through the IXP and $\phi_1^n - x_n$ is, by constraint, the flow sent through the NSP. The same applies to y_n for player 2.

Assume now that the topology is fully connected, meaning that $b_l^n = 0 \ \forall l, n$. In this case, from the player's perspective, the cost does not depend on the facility used for a specific CP, but rather on the total amount of flow going through a specific TF, independently from the destination. Once again, with some algebraic manipulations (shown in Appendix 4.8.1) we rewrite cost function (4.3.1) for both players as:

$$
\begin{cases}
C^1(x, y) &= (\phi_1 - x)\left(a_1 + \frac{1}{c_1 - (\phi_1 - x)}\right) + x\left(\frac{a_2}{\sqrt{x+y}} + \frac{1}{c_2 - x}\right) \\
C^2(x, y) &= (\phi_2 - y)\left(a_1 + \frac{1}{c_1 - (\phi_2 - y)}\right) + y\left(\frac{a_2}{\sqrt{x+y}} + \frac{1}{c_2 - y}\right)
\end{cases}
\tag{4.3.2}
$$

where x is the cumulative flow of ISP_1 through the IXP and ϕ_1 is its cumulative demand. The same applies to y and ϕ_2 for ISP_2. The best response of player i is thus obtained by minimizing $C^i(x, y)$ defined in (4.3.2).

While simple, the MCM is interesting on its own as it provides a clear way to study the fundamental difference between transit and peering agreements, shedding light on the emerging competition between NSPs and large IXPs, first observed in [ACF+12].

4.3.1 Theoretical Results

Definition 1. Supermodular games [Yao95]

Consider a generic game G, where user's payoffs are given by an utility function $u : \mathbb{R}^k \to \mathbb{R}$. The game is said supermodular if the utility function is supermodular, that is:

$$u(x \vee y) + u(x \wedge y) \geq u(x) + u(y) \; \forall x, y \in \mathbb{R}^k$$

where $x \vee y$ denotes the componentwise maximum and $x \wedge y$ the componentwise minimum of x and y. If u is twice continuously differentiable, this property is given by the following condition:

$$\frac{\partial^2 u}{\partial z_i \partial z_j} \geq 0 \qquad \forall i \neq j$$

In our case we consider costs rather then utilities and minimization instead of maximization, therefore a game like ours is supermodular iff:

$$\frac{\partial^2 C(\mathbf{x})}{\partial x_i \partial x_j} \leq 0 \qquad \forall i \neq j \tag{4.3.3}$$

When the strategy space can be expressed using just two variables, as in the MCM case, we can write:

$$\frac{\partial^2 C(x, y)}{\partial x \partial y} \leq 0 \tag{4.3.4}$$

Theorem 1 of [AA03] proves the existence of equilibria for supermodular games, moreover it provides a way of computing them. The proof

is based on showing that **best response sequences are monotone** and therefore converge to a limit which is then shown to be a Nash Equilibrium Point (NEP). The monotonicity is a consequence of the "strategic complementarity" of the players: if one of them chooses a strategy x that decreases its own cost, this decision is beneficial for the cost of the other players too.

Here we relax the results on the existence of equilibria and convergence of best response sequences in supermodular games.

Definition 2. Symmetric supermodularity

We define as symmetric supermodular games, those for which (4.3.4) holds for all strategies $x = y$, meaning that the property holds along the symmetric axis.

Definition 3. Symmetric best response sequence

We call symmetric best response sequence a path (x_0, y_0), (x_1, y_1), ..., where $x_0 = y_0$ and $\forall i$, (x_i, y_i) satisfies $x_i = y_i$.

Theorem 4. *In symmetric supermodular games, pure equilibria exist and are given as the limit of symmetric best response sequences.*

Proof. Consider a sequence of best responses (x_0, y_0), (x_1, y_1), ... Due to symmetry we can choose this path to be a symmetric best response sequence. From definition 2 and by applying the same reasoning as in the original proof [AA03], we shall get monotone sequences whose limits are equilibria. □

Theorem 4 not only proves the existence of equilibria for symmetric supermodular games, but also gives an algorithm for computing them. Please note that, for this theorem to hold, the game does not need to satisfy (4.3.4) for all possible strategies, but just along the symmetric path. This result can be applied to our game thanks to Theorem 5 and Corollary 6.

Theorem 5. *The game defined in (4.3.2) is symmetric supermodular.*

Proof. Consider the cost function of our game. We compute the mixed second derivatives:

$$\begin{cases} \frac{\partial^2 C^1(x,y)}{\partial x \partial y} = -\frac{a_2}{2(x+y)^{\frac{3}{2}}} + \frac{3a_2 x}{4(x+y)^{\frac{5}{2}}} \\ \frac{\partial^2 C^2(x,y)}{\partial x \partial y} = -\frac{a_2}{2(x+y)^{\frac{3}{2}}} + \frac{3a_2 y}{4(x+y)^{\frac{5}{2}}} \end{cases} \tag{4.3.5}$$

Since we are interested only in their sign, we can multiply both derivatives in (4.3.5) by $(x+y)^{\frac{3}{2}}$, which is always positive as long as the flows are positive. Therefore we have:

$$\begin{cases} \operatorname{sgn}\left(\frac{\partial^2 C^1(x,y)}{\partial x \partial y}\right) = \operatorname{sgn}\left(\frac{a_2}{4} \cdot \frac{x-2y}{x+y}\right) \\ \operatorname{sgn}\left(\frac{\partial^2 C^2(x,y)}{\partial x \partial y}\right) = \operatorname{sgn}\left(\frac{a_2}{4} \cdot \frac{y-2x}{x+y}\right) \end{cases} \tag{4.3.6}$$

Consider now the symmetric axis, where $x = y$. With this equality both mixed second derivatives in (4.3.6) become negative, therefore due to condition (4.3.4) the game is symmetric supermodular. □

It is interesting to note that, as long as the transit cost function t^l of one ISP does not depend on the other ISP, the mixed second derivative (4.3.5) does not change. Therefore, symmetric supermodularity can be applied to game (4.3.2) even for more general transit cost functions (as outlined in Section 4.2.2). Please note that without the symmetric assumption, the game is neither supermodular, nor submodular, because we cannot say anything about the sign of the mixed derivatives.

Corollary 6. *The game defined in (4.3.2) has at least one pure equilibrium for symmetric demands, given as the limit of a symmetric best response sequence.*

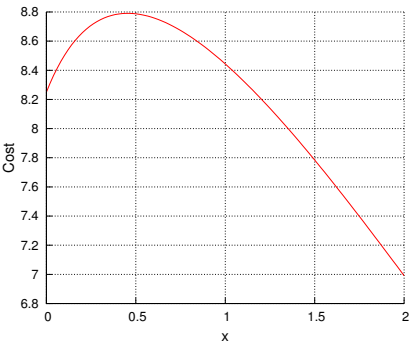
Proof. By hypothesis the demands satisfy $\phi_1 = \phi_2$. If we put this condition in system (4.3.2), the two players become symmetric, therefore we get this result combining Theorems 4 and 5. □

4.3.2 Cost Function Analysis

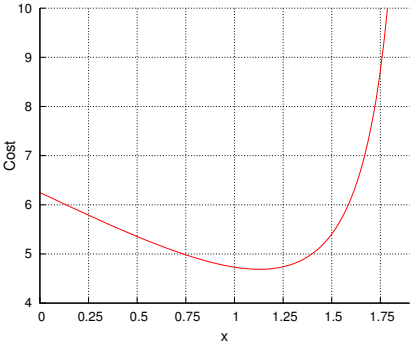
In order to gain insights on the outcome of the best response sequence algorithm, here we analyze the cost function. Consider the cost function of player 1 and suppose that the strategy y of player 2 is fixed, so:

$$C^1(x) = (\phi - x)\left(a_1 + \frac{1}{c_1 - (\phi - x)}\right) + x\left(\frac{a_2}{\sqrt{x+y}} + \frac{1}{c_2 - x}\right) \tag{4.3.7}$$

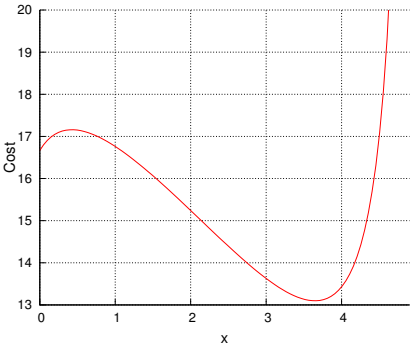
Lemma 7 and Theorem 8, whose proofs can be found in Appendix 4.8.2, tell us the shape of the cost function.



(a) Case 1: concave cost ($a_1 = 4$, $a_2 = 5$, $\phi = 2$, $c_1 = 10$, $c_2 = 5$, $y = 0.5$)



(b) Case 2: convex cost ($a_1 = 3$, $a_2 = 1$, $\phi = 2$, $c_1 = 10$, $c_2 = 2$, $y = 1.6$)



(c) Case 3: concave/convex cost ($a_1 = 4$, $a_2 = 5$, $\phi = 4$, $c_1 = 10$, $c_2 = 5$, $y = 0.5$)

Figure 4.3.2: Cost function

Lemma 7. *The second derivative of the cost function (4.3.7) is a monotonically increasing function.*

Theorem 8. *The cost function (4.3.7) can be either: always concave, always convex, or first concave and then convex.*

Figure 4.3.2 shows the possible cases. Please note that the shape depends both on the parameters and the strategy y of the other player: while for specific values of y the function might be convex as in Figure

Algorithm 4.1 Best Response Sequence

1: $\mathbf{startingpoint} = ...$ ▷ Initial strategies
2: $tolerance = ...$ ▷ NEP stationariety tolerance
3: $\mathbf{newx} = \mathbf{startingpoint}$
4: **repeat**
5: $\quad \mathbf{oldx} = \mathbf{newx}$ ▷ current step game strategies
6: $\quad$ **for** $i = 1; i < I; i++$ **do**
7: $\quad\quad \mathbf{x_{-i}} = \mathbf{oldx_{-i}}$ ▷ other players strategy
8: $\quad\quad \mathbf{x_i} = \arg\min_{\mathbf{x_i}} C^i(\mathbf{x_i}, \mathbf{x_{-i}})$ ▷ i strategy
9: $\quad\quad\quad s.t.\ \sum_l x_{i,l}^n = \phi_i^n \quad \forall n$
10: $\quad\quad \mathbf{newx_i} = \mathbf{x_i}$ ▷ next step game strategies
11: $\quad$ **end for**
12: **until** $||\mathbf{newx} - \mathbf{oldx}|| < tolerance$

4.3.2b, it can also be concave (Figure 4.3.2a) and in general is neither convex nor concave, as shown in Figure 4.3.2c.

4.3.3 The Simulator

The analysis of the cost function, performed in Section 4.3.2, suggests that in our game (4.3.2), even if the best response procedure converges to a NEP, there might be multiple equilibria, because of the presence of multiple local minima. This assumption can be verified via simulation, where we show a specific case in which the NEP reached can change, depending on the starting point of the algorithm.

We implemented in MATLAB [Mat] the general model (4.2.11) described in Section 4.2. Iteratively, each player performs its best response to the set of other players' strategies as shown in algorithm 4.1. If the simulation converges (this has not yet been proven for the general game (4.2.11)), the output $\mathbf{newx}$ is the NEP for the given input parameters, which are:

- The number of ISPs, TFs and CPs, respectively I, L, N.
- The cost function parameters a_l, c_l, b_l^n and demands ϕ_i^n.
- The $tolerance$ and the $\mathbf{startingpoint}$.

We can use the implemented algorithm on the MCM, which has fully connected topology and symmetric demands, by putting $I = 2$, $L = 2$, $N = 2$, $b_l^n = 0\ \forall l, n$ and $\phi_i^n = \phi^n$. Given the selected scenario and the best response sequence algorithm, Theorem 6 ensures the convergence of the simulation for whatever cost function coefficients. Moreover, thanks to the symmetric property, we can just investigate the strategy of player 1 (x), because player 2 will show exactly the same behavior.

We simulate the following parameters: $a_1 = 4$, $a_2 = 5$, $\phi^1 + \phi^2 = 4$, $c_1 = 10$, $c_2 = 5$. In this case, as previously shown in Figure 4.3.2c, the cost function could present multiple local minima, depending on the players' strategies. The simulation has multiple outcomes: if we start from the mean point ($x = 2$, $\phi - x = 2$) we end up in an equilibrium where traffic is split between the IXP and the NSP: $x^* = 3.64;\ \phi - x^* = 0.36$. The IXP is preferred because the usage cost is shared among the two players, however it is not used exclusively due to its smaller capacity not being able to serve all the traffic. With a bigger capacity, all the traffic would have been routed through the IXP. Otherwise, if we start from a strategy where the majority of traffic is routed through the NSP ($x = 0.4$, $\phi - x = 3.6$), we end up in an equilibrium where all the traffic flows through the NSP: $x^* = 0;\ \phi - x^* = 4$. This happens because when the IXP is routing a small amount of traffic, the flat port cost is too high to justify its use, therefore the players prefer the NSP. Once the NSP is serving all the traffic no player has an incentive to deviate, because he would pay the whole IXP cost by himself.

As we see, the outcome of the game is highly dependent on the starting point: the IXP is preferred only if it already has, at the beginning, a good amount of flow passing through him, otherwise all players will stick to the NSP. This result is consistent with reality, in fact, the necessary condition for an IXP to emerge is that it has a critical mass (represented by a fraction of the traffic/users in the Internet) which makes the value perceived by a potential participant greater than the cost he would incur in by joining the facility [Nor10c].

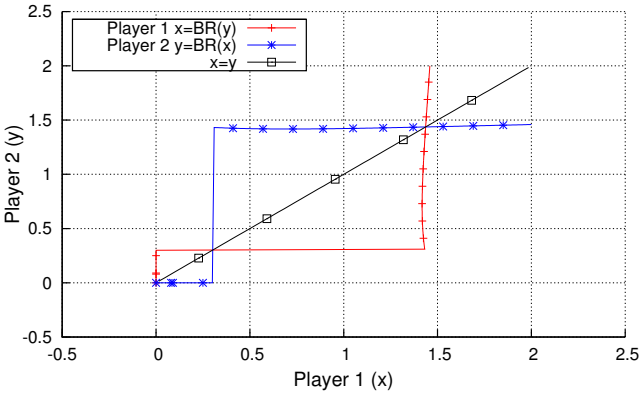


Figure 4.3.3: BRI - Case 1: $a_1 = 2$, $a_2 = 2$, $\phi = 2$, $c_1 = 10$, $c_2 = 3$

4.3.4 The best response behavior

With the purpose of understanding the number and position of NEPs, we draw the Best Response Intersection (BRI) picture. In this graph, shown in Figure 4.3.3, the line with tick marks represents the best response x of player 1 as a function of player 2's strategy y, while the line with cross marks does the exact opposite. The intersection points on the graph mean that both players are playing their best responses, therefore they are Nash Equilibrium Points. As we can see, there are three NEPs and, as expected due to the symmetric property, they are all on the symmetric axis [fCRVW04]:

Left Equilibrium is in $x = x_L^* = 0$ and corresponds to the scenario where all the traffic is routed through the NSP.

Right Equilibrium is for $x = x_M^* = 1.43$ and is the one where traffic is split between the IXP and the NSP.

Middle Equilibrium happens for $x = x_M^* = 0.31$. This is however a repulsive equilibrium, in fact, as soon as one of the two players deviate, they will never return to this point and reach instead one of the two others equilibria.

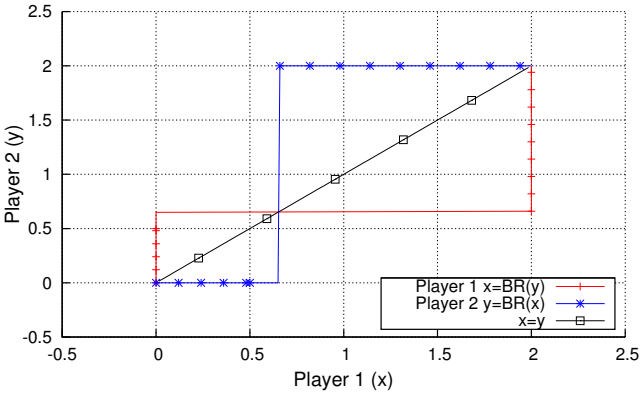


Figure 4.3.4: BRI - Case 2: $a_1 = 2$, $a_2 = 3$, $\phi = 2$, $c_1 = 10$, $c_2 = 4$

These three equilibria can be understood by observing Figure 4.3.2c: x_L^* and x_R^* are attractive, and correspond to the minima of the cost function, while x_M^* corresponds to the maximum of the cost function, and is thus repulsive. Of course, the last picture corresponds to the cost function for a specific strategy, therefore it cannot assert the position or the existence of equilibria, however it gives an insight on their meaning.

As we change the game parameters we observe that the shape of the best response is always the same, while the position of x_M^* and x_R^* changes. In particular, as shown in Figure 4.3.4, If the ratio a_2/a_1 increases (meaning that IXP cost w.r.t. NSP cost increases) then x_M^* gets nearer to x_R^*, making the left equilibrium is easier to reach. Moreover, we observe that if the capacity c_2 of the IXP is large enough, than in the right equilibrium all the traffic will flow through him.

4.4 Price of Anarchy, Stability and Fairness

4.4.1 Social Optimum

We now compare the performance of the distributed system, where each Service Provider acts on its own, with that of an ideal centralized system where decisions are took by some external entity. In this case the objective is to minimize the total cost of the two players, given by the summation

of the two costs in (4.3.2):

$$C(x,y) = \sum_i C^i(x,y) = (\phi_1 + \phi_2 - x - y)\, a_1 + \frac{\phi_1 - x}{c_1 - (\phi_1 - x)} + \\ + \frac{\phi_2 - y}{c_1 - (\phi_2 - y)} + \frac{x+y}{\sqrt{x+y}} a_2 + \frac{x}{c_2 - x} + \frac{y}{c_2 - y} \qquad (4.4.1)$$

Theorem 9 and Corollary 10, whose proofs can be found in Appendix 4.8.2, explain how to optimize this cost function.

Theorem 9. *The cost function (4.4.1) has a global minimum point. For symmetric demands this minimum point is attained at symmetric strategies, and it is either the left endpoint of the strategy space or the unique local minimum point of its convex part.*

Corollary 10. *The global minimum point of (4.4.1) is, for symmetric demands, either the left endpoint of the strategy space or the output of a standard algorithm for convex function optimization that starts from the right endpoint.*

The globally optimal solution to problem (4.4.1) can thus be computed by comparing the two candidate points highlighted in Corollary 10.

4.4.2 Alpha-Fair solution

Another metric for comparison comes from the theory of fairness. A unifying mathematical formulation, known as α-fairness [LKCS], says that given a set of users and utility functions $U_i(x)$, the α-fair solution to the problem of maximizing their utilities is given by:

$$\max_x \left(\sum_i \frac{U^i(x)^{(1-\alpha)} - 1}{1 - \alpha} \right)$$

For $\alpha = 0$, this is the same as maximizing the sum of the utilities, thus it gives the social optimum for the problem. The case $\alpha \to 1$ yields the proportional fair share assignment, however this solution is not feasible when we have to deal with cost function rather then utilities, and for $\alpha \to \infty$ it is equivalent to the max-min fairness. For $\alpha = 2$, the formula

gives us the "harmonic mean fair" solution, which we investigate here:

$$\max_x \left(\sum_i \frac{U^i(x)^{(1-\alpha)} - 1}{1-\alpha} \right) = \min_x \left(\sum_i -1 - \frac{1}{C^i(x)} \right)$$

The solution is computed numerically in the next subsection.

4.4.3 PoA/PoS/PoF Comparison

As is usually done in the literature [JHA12; LKCS], we define the Price of Anarchy (PoA) as the ratio between the worst decentralized solution (equilibrium) and the social optimum. Similarly, the Price of Stability (PoS) is defined as the ratio between the best equilibrium and the social optimum:

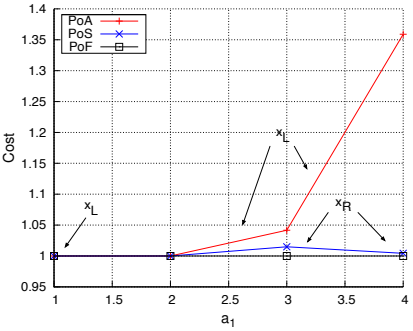
$$PoA = \frac{C\left(x^*_{worst}\right)}{C\left(x_{opt}\right)} \geq 1 \qquad PoS = \frac{C\left(x^*_{best}\right)}{C\left(x_{opt}\right)} \geq 1$$

In our case we have just two attractive equilibria, therefore the best and worst equilibria are either x^*_L or x^*_R. Following the same path, we define the price of fairness as the ratio between the fair and the optimal solution:

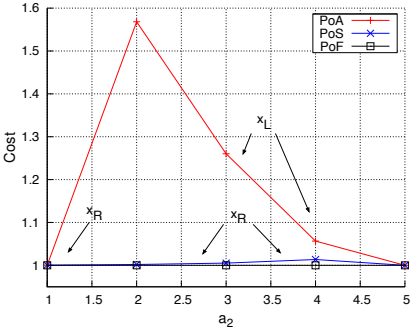
$$PoF = \frac{C\left(x_{fair}\right)}{C\left(x_{opt}\right)} \geq 1$$

Algorithm 4.1 has been extended to include numerical computation of the above-defined Prices of Anarchy, Stability and Fairness. We use as general configuration: $a_1 = 3$, $a_2 = 4$, $\phi = 2$, $c_1 = 10$, $c_2 = 3$ (except for Figure 4.4.1c, where c_2 increases as ϕ increases), and show PoA, PoS and PoF as parameters a_1, a_2 and ϕ change. Results are reported in Figure 4.4.1. As we see, it is always the case that $PoF = 1$, meaning that the harmonic mean fair solution is equal to the social optimum.

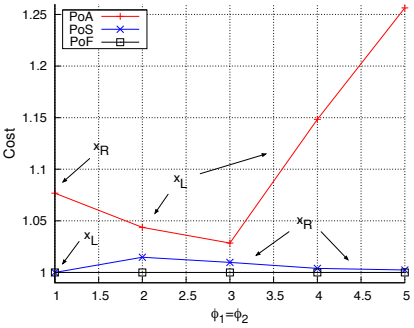
The PoA almost always corresponds to the left equilibrium. An exception to this is the case where there is a small amount of total traffic, shown in Figure 4.4.1c for $\phi = 1$: in this case the left equilibrium outperforms the right one, meaning that for small amounts of flow it is not convenient to share costs at the IXP. As ϕ increases, the advantages of sharing become



(a) PoA/PoS/PoF as a function of a_1



(b) PoA/PoS/PoF as a function of a_2



(c) PoA/PoS/PoF as a function of ϕ

Figure 4.4.1: Price of anarchy, stability and fairness

obvious. Figures 4.4.1a and 4.4.1b show that the PoA increases as a_1 increases and decreases as a_2 increases. An exception to this is the case $a_2 = 1$ of Figure 4.4.1b; with these parameters the cost function resembles that of Figure 4.3.2b, therefore we have only one equilibrium. The PoS is almost always very low, and it is always caused by the fact that the competition between ISPs reduces the amount of traffic through the IXP, thus reducing their opportunities to share costs.

4.5 Generalizations

While the results obtained by the MCM are interesting on their own, as they shed light on the competition between an Internet Exchange Point and a Network Service Provider, driven by the clear differences between transit and peering, one might argue that this topology is quite small and simplified to represent the Internet. Here we explicitly tackle this problem by both extending the convergence results to more general scenarios, and using simulation to cover other cases for which we couldn't obtain analytic results, thus analyzing a broad variety of general cases.

4.5.1 Extended Analytical Results

We consider a system with I ISPs, N CPs, and L TFs, with $l_1 = 1$, in a possibly disconnected topology. While having multiple IXPs is fundamental for understanding how players aggregate around exchange points, especially in presence of reachability constraints (IXPs may be connected to only a subset of CPs), this is not the case for NSPs, due to the fact that their cost is independent from other players' choice, and they are necessarily connected to all possible CPs. Therefore, without loss of generality for our problem, it is safe to consider only a single NSP for $l = 1$, and $L - 1$, for $l = 2..L$, IXPs as we do here. In the following we will always refer to this system, which is a single NSP version of Figure 4.2.1.

The cost function of Section 4.2 can thus be rewritten separating the NSP component from the IXPs, therefore:

$$C^i(\mathbf{x_i}, \mathbf{x_{-i}}) = \sum_n x^n_{i,1} \left(a_1 + \frac{1}{c_1 - \sum_n x^n_{i,1}} \right) + \tag{4.5.1}$$
$$+ \sum_{l \neq 1} \sum_n x^n_{i,l} \left(\frac{a_l}{\sqrt{\sum_i \sum_n x^n_{i,l}}} + \frac{1}{c_l - \sum_n x^n_{i,l}} + b^n_l \right)$$

where the CP reachability cost for the NSP has been removed since we know from (4.2.3) that $b^n_1 = 0 \ \forall n$.

Theorem 11 and Corollary 12, whose complete proofs can be found in Appendix 4.8.2, demonstrate existence of equilibria and convergence of

the symmetric best response algorithm for the general case just formalized.

Theorem 11. *The game defined in (4.5.1) is symmetric supermodular.*

Sketch of the Proof. The proof is based on showing that (4.3.3) holds along the symmetric axis for any possible combination of indexes:

$$\left.\frac{\partial^2 C^i(\mathbf{x_i}, \mathbf{x_{-i}})}{\partial x^{\bar{n}}_{j,\bar{l}} \partial x^n_{i,l}}\right|_{x^n_{i,l}=x^n_{j,l}} \leq 0 \qquad \forall i \neq j,\ \forall l, \bar{l}, n, \bar{n} \tag{4.5.2}$$

□

Corollary 12. *The game defined in (4.5.1) has at least one pure equilibrium for symmetric demands, given as the limit of a symmetric best response sequence.*

4.5.2 Subcases Analysis

After proving existence of equilibria for the general case, here we analyze some subcases in order to understand what kind of equilibria we should expect for specific scenarios. We have two main categories:

Fully Connected Topologies Suppose that we have a fully connected topology, meaning that $b^n_l = 0\ \forall l, n$. In such a case, we can take the summation over n and consider cumulative flows and demands:

$$\begin{cases} x_{i,l} = \sum_n x^n_{i,l} & \textit{cum. flow } ISP_i \rightarrow TF_l \\ \phi_i = \sum_n \phi^n_i & \textit{cum. demand } ISP_i \end{cases}$$

We can now substitute these two variables inside cost function (4.2.11), thus obtaining an equivalent problem where the strategy of each player is a vector $\mathbf{x_i} = (x_{i,1}, ..., x_{i,l}, ..., x_{i,L}) \in \mathbb{R}^L$. This means that, in fully connected topologies, our system is equivalent to another one where we only have a single CP, and each player has to serve a cumulative demand ϕ_i for that CP. This happens because there are no reachability constraints, therefore from a player's perspective the specific CP from which he has to fetch data is not relevant.

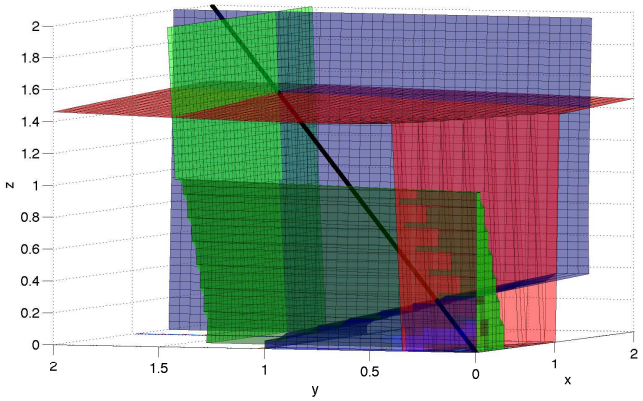


Figure 4.5.1: BRI for 3 players: $a_1 = 2$, $a_2 = 2$, $\phi = 2$, $c_1 = 10$, $c_2 = 3$

Symmetric IXPs Suppose that all the IXPs have the same costs, capacities and reachability matrix: $a_l = a_{IXP}$, $c_l = c_{IXP}$, $b_l^n = b_{IXP}^n \; \forall l \neq 1$. Due to their symmetry, there is an equilibrium where traffic is split equally among them [fCRVW04], therefore we might think of transforming this problem in an equivalent one having a single IXP with the same reachability matrix and transformed costs and capacities. Unfortunately, we were unable to perform this conversion due to the form of the cost function for the IXPs. In fact, as we see from (4.2.11) and (4.2.10), the non linear port cost h_l makes it quite different for players to share small traffic quantities rather then large ones.

The analysis of the two categories highlights that scenarios with multiple CPs can be highly simplified with fully connected topologies, while in the case of multiple IXPs, even if symmetric, the analysis can be quite difficult. To conclude this section, we analyze in more detail the simplified case where we have fully connected topology and just one IXP. Due to the fact that we can handle the multiple CPs as if there was just a single one, this scenario is quite similar to the MCM, except that we have a generic number I of players. However, thanks to Corollary 12, we now know that for symmetric demands $\phi_i = \phi \; \forall i$ we can compute equilibria as the limit of symmetric best response sequences. Just like it was done for the MCM, we can use algorithm 4.1 to compute the equilibria.

With the purpose of analyzing number and position of equilibria, for the case of $I = 3$ (it would be difficult to represent more dimensions), we can draw the Best Response Intersection (BRI) graph, shown in Figure 4.5.1. Just like in the MCM, the picture shows three equilibrium points, obtained by the intersection of the three surfaces representing the players' best responses. As we see from the straight line crossing all such points, the three equilibria are symmetric, with the leftmost (traffic split between NSP and IXP) and the rightmost (all flows through the NSP) being the stable ones.

4.6 Simulations

In this section, we use our MATLAB implementation to test the behavior of the system. As a matter of fact, thanks to Theorem 11, we have proven convergence for symmetric scenarios, showing that symmetric supermodularity holds for the general scenario of Section 4.2. We do not have convergence guarantees for asymmetric topologies. However, we know that if the best response sequence algorithm converges, then it converges to an equilibrium [AA03]. Therefore we can use our simulator both to study our system in general cases, and to assess convergence for specific cases.

Simulations have been performed using the symmetric best response algorithm 4.1: iteratively, each player performs its best response to the set of other players' strategies. If the simulation converges, the output is the equilibrium for the given input parameters, which are:

- the number of ISPs, TFs and CPs, respectively I, L, N;
- the cost function parameters a_l, c_l, b_l^n and demands ϕ_i^n.

Moreover, it is important to set a **startingpoint**, because, as we saw, on startup IXPs need a critical mass, represented by a share of the total traffic in the system, in order to be able to attract players.

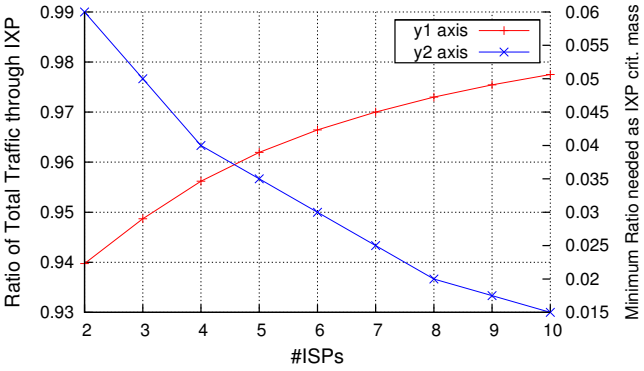


Figure 4.6.1: Traffic ratios and equilibrium breakpoint as I grows

4.6.1 Growing Number of ISPs/IXPs

We start with showing the behavior of system for symmetric cases, for which convergence has been proven, and checking what happens as the number of agents in the system grows. The base configuration used is $I = 2$, $L = 2$, $N = 2$, $b_l^n = 0\ \forall l, n$ and $\phi_i^n = \phi^n = 2$, and all tests have been performed with fully connected topology and symmetric demands. The cost coefficients used are: $a_1 = 1$, $a_2 = 1.5$, $c_1 = 10$, $c_2 = 6$. As long as flows and capacities are properly balanced, the existence of multiple CPs does not seem to affect the results of the simulations, therefore here we check what happens when we have either more ISPs or IXPs.

Generic Number of ISPs When the number I of players increase, the benefits of joining an IXP increases as well, due to the fact that costs are shared among multiple participants: in fact, as shown in Figure 4.6.1 on the y1 axis, the fraction of traffic flowing through the IXP at the equilibrium increases with I. We recall from Section 4.3.3 that IXPs need a critical mass to be used, which in our case corresponds to a fraction of the total traffic in the system. Very interestingly, the y2 axis of Figure 4.6.1 shows that this fraction decreases as the number of player grows.

Generic Number of IXPs In order to have an interesting case study as L grows, we test a scenario where the global IXP capacity does not change,

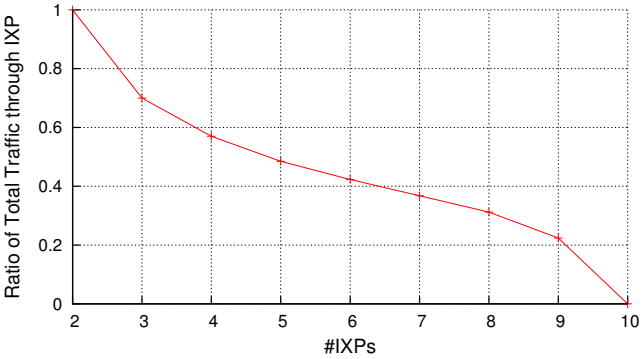


Figure 4.6.2: Traffic ratio as L grows

therefore $c_l = c_2/(L-1)\ \forall l \neq 1$. This means that instead of having one "large" IXP with a high capacity, we have multiple IXPs with less capacity. In order to have meaningful capacities for the small IXPs, we increased global flows and capacities to: $c_1 = 50$, $c_2 = 25$, $\phi_i^n = \phi^n = 10$. As shown in Figure 4.6.2 the fact that IXPs only offer small ports is detrimental for the players, and after a certain point they will all stick to the NSP.

Simulations show that while the IXP critical mass decreases with a larger player base, this effect is counterbalanced by the fact that the critical mass increases with the number of IXPs. Therefore, the results found in the two player case still hold in more realistic scenarios: IXPs need a critical mass to emerge even in scenarios with more ISPs and IXPs, otherwise we still end up in an equilibrium with dominant NSP connectivity.

4.6.2 Flow Path Analysis

In this section we show results of simulations regarding the path followed by traffic flows. The simulations were performed both for symmetric scenarios, for which convergence has been proven in the general case, and asymmetric scenarios, for which we have no proofs. In fact, as we'll see later on, in this last case it is possible for players to never reach an equilibrium.

4.6.2.1 Symmetric Case

We simulate a scenario with $I = 10$ ISPs, $L = 4$ TFs ($L-1$ symmetric IXPs) and $N = 4$ CPs. The connectivity matrix is:

$$b_l^n = \begin{cases} \infty & (l,n) = (2,2) \vee (l,n) = (4,1) \vee (l,n) = (4,4) \\ 0 & \textit{otherwise} \end{cases}$$

The cost parameter a_l has been chosen in order to be similar to present reality [Nor10b], therefore we choose $a_1 = a_{NSP} = 3$ and $\forall l \neq 1$, $a_l = a_{IXP} = 30$ as seen in the fit performed for the MIX (Figure 4.2.2). All users have symmetric flows $\phi_i^n = 12.5\ \forall i, n$ and their capacities to the facilities are $c_1 = c_{NSP} = 100$ and $\forall l \neq 1$, $c_l = c_{IXP} = 20$, so that $c_{NSP} \gg c_{IXP}$.

As already happened in the MCM, depending on the **startingpoint** we notice the existence of multiple equilibria. In fact, if the initial condition is such that one or more IXPs are underutilized, than at equilibrium those IXPs will not be used. This phenomenon corroborates the outcomes of the MCM, showing that indeed even in general scenarios the competition between NSPs and IXPs, and even between IXP themselves, strongly emerges. Differently from the MCM, in this case we observe more than two stable equilibria, since any combination with one or many unused IXPs can be an equilibrium.

Suppose now that the **startingpoint** is such that flows are split equally among the facilities, so that all IXPs have the critical mass to attract players. Figure 4.6.3 shows the scatterplot at equilibrium. In this plot, each dot represents the flow quantity that each user sends on a given path (that is, to a fixed CP through a given IXP). Due to symmetry, we observe that all users will behave symmetrically on the same path, and this is exactly what happens in Figure 4.6.3. There is generally a low utilization of the NSP, which rises a little bit for those CPs with a worse reachability matrix ($CP_1,\ CP_4$).

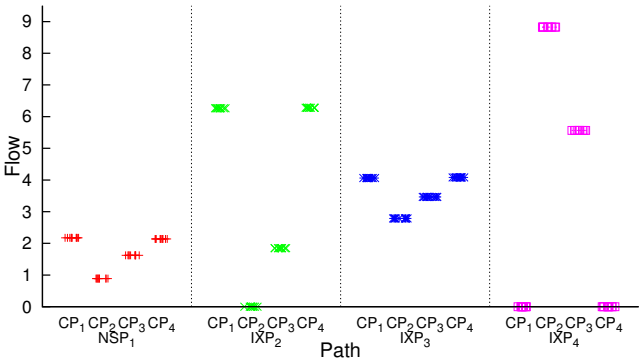


Figure 4.6.3: Symmetric case flows scatterplot: $\phi_i^n = 12.5$

4.6.2.2 Asymmetric Case

We now show the impact of asymmetric players' demands. In this case, convergence of the best response sequence is not guaranteed by Corollary 12, however, we know that if the simulation converges we certainly reach an equilibrium.

We simulate a scenario with exactly the same parameters as in the symmetric case, except that now the demands grow linearly from $\phi_1^n = 10 \ \forall n$ to $\phi_I^n = 15 \ \forall n$. The average demand is still 12.5, but now the demand of the last player is 1.5 times that of the first one. The scatterplot at equilibrium is shown in Figure 4.6.4. Very interestingly, even if demands are asymmetric, paths of flow tend to be almost symmetric for the IXPs, while they spread apart for the NSP. This happens because the benefits of sharing costs at the exchange points is bigger when the traffic ratio is approximately the same between participants, therefore players tend to "symmetrize" around the IXPs. Due to the fact that flows around the IXPs are more or less symmetric, players will send the traffic residual through the NSP, which will see highly asymmetric patterns.

In previous case, the asymmetry in players' demands was not very pronounced. Let's now see what happens when the demands go from $\phi_1^n = 6.5 \ \forall n$ to $\phi_I^n = 18.5 \ \forall n$, meaning that last player demand is nearly three times that of player one. Again, Figure 4.6.5 shows the scatterplot at equilibrium. Due to the heavily unbalanced demands, the symmetric

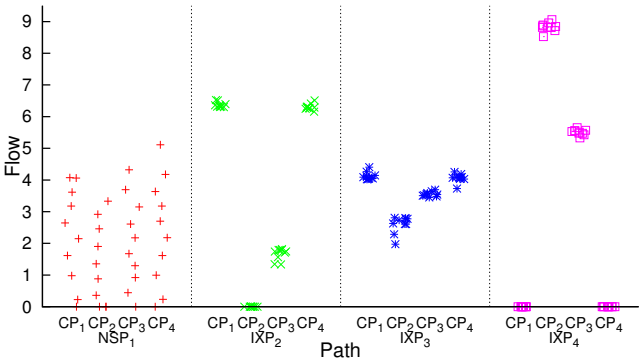


Figure 4.6.4: Asymmetric case flows scatterplot: $\phi_i^n = 10 \rightarrow 15$

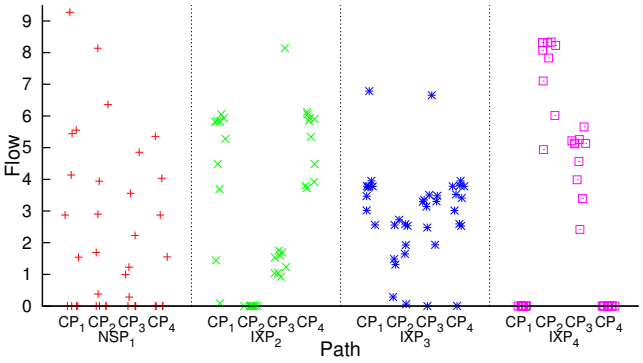


Figure 4.6.5: Asymmetric case flows scatterplot: $\phi_i^n = 6.5 \rightarrow 18.5$

patterns around the IXPs are still present, but much less pronounced. While in the previous case equilibrium was driven by the simple rule of "symmetric behavior", in this case the outcome is more difficult to predict. In general, due to the asymmetry, cost benefits of players' for using exchange points decrease, therefore we observe, on average, a higher quantity of flow going through the NSP.

The phenomena emerged through this analysis can provide some preliminary insight on how to devise optimal policies to handle peering traffic at IXPs. More specifically, the "symmetric behavior" rule highlights that it is beneficial to balance traffic as much as possible, therefore IXP owners should create few classes of traffic (namely, few different port

sizes), and participants should try to aggregate traffic on these ports, since unbalanced flows must be handled at NSPs and bring to suboptimality. While simulations have been carried out with a limited network size due to computational constraints, the conclusions are fairly general, therefore we expect similar results to hold for larger-scale scenarios.

4.6.3 Non Convergence

Even quite simple scenarios for which we cannot apply Theorem 11, might lead to a situation where players' behavior oscillates, never reaching an equilibrium. Consider a system with two symmetric IXPs and an asymmetric starting point s.t. a group of players send more traffic to one of the IXPs and less to the other, while the other group of players do the opposite. Due to the asymmetric assumption we cannot apply Corollary 12, and simulation shows that this scenario might never reach an equilibrium. This happens when players enter a never-ending oscillation between the first and the second group, as detailedly shown in the following example.

Example 13. Non-Convergence.

Consider a system with fully connected topology, $I = 16$ ISPs, $L = 3$ TFs (one NSP and two symmetric IXPs) and $N = 1$ CP. Once again we use cost parameters $a_{NSP} = 3$ and $a_{IXP} = 30$. Capacities are $c_{NSP} = 200$ and $c_{IXP} = 70$ and we even take symmetric demands $\phi_i^n = 50\ \forall i, n$. According to Corollary 12, equilibrium can be reached following a sequence of symmetric best responses. Instead, we set an asymmetric starting point:

$$\begin{cases} x_{i,l} = 24.9 & if\ (l = 1 \wedge i \leq 8) \vee (l = 2 \wedge i > 8) \\ x_{i,l} = 25.1 & if\ (l = 2 \wedge i \leq 8) \vee (l = 1 \wedge i > 8) \\ x_{i,l} = 0 & otherwise \end{cases}$$

Simulation shows that players never reach an equilibrium, and go through a never-ending oscillation between two points:

$$\begin{cases} x_{i,l}^- = 0\ x_{i,l}^+ = 50 & if\ (l = 1 \wedge i \leq 8) \vee (l = 2 \wedge i > 8) \\ x_{i,l}^- = 50\ x_{i,l}^+ = 0 & if\ (l = 2 \wedge i \leq 8) \vee (l = 1 \wedge i > 8) \\ x_{i,l}^- = x_{i,l}^+ = 0 & otherwise \end{cases}$$

This happens because, on each iteration, each player of the first group sees the second group of players on a different IXP, and finds it beneficial to deviate on that TF. The same happens for the players of the second group, which in turn deviate altogether to the IXP of the first group. After the deviation, situation is reversed, therefore the two groups keep deviating all the time, never reaching an equilibrium. Please note that using a symmetric starting point would immediately lead to an equilibrium where $x_{i,l} = 25$ if l is an IXP, and zero otherwise.

4.7 Conclusions

The proposed model gives insight into the economy of different types of Autonomous Systems and the driving forces behind the decisions they make when joining the Internet. The peculiar pricing strategies of players doesn't allow standard modelization, however, by exploiting peculiar properties of the game, we are able to prove analytically the existence of multiple equilibria, and provide an algorithm to compute the stable ones. From a game theoretic perspective, while the theory on supermodularity is well-developed, we relaxed this concept and introduced the new category of Symmetric Supermodular games. Thanks to this we were able to prove existence of equilibria and convergence of best response sequences in our game. This is the first case, to the best of our knowledge, where results on supermodularity are applied even if the property does not hold for the game in general, by showing that it holds along the symmetric path. From an engineering perspective, the outcome of the analysis is highly insightful as it shows different interesting aspects. First of all, we observe the suboptimality of the decentralized solution, originated by the non-cooperative behavior of the ASes, by showing the existence of a Price of Anarchy and Stability. Second, we have shown that also for asymmetric cases the system often reaches an equilibrium. Such equilibrium suggests that players tend to "symmetrize" their traffic as much as possible with respect to the peering exchange points, and send their asymmetric traffic quota via the transit service providers. This observation can provide insights on how to devise optimal policies to

handle peering traffic at IXPs. Last, but probably most important, we highlight the growing competition between IXPs, providing customers the ability to lay out peering connections, and NSPs, high-level providers selling transit connections, even for large-scale scenarios.

The developed model is a simple yet effective way of analyzing how Autonomous Systems manage their traffic, and what are the possible ways to optimize their costs for handling such traffic. Therefore, it is reasonable to think of simulating the model using as input realistic traffic demands expected in the Internet environment. Unfortunately, traffic data between ASes over the Internet is undisclosed, and the few studies related to this subject rely on private data [ACF+12]. Nevertheless, we are currently trying to evaluate traffic patterns using available DNS data, so as to indirectly assess the fraction of traffic flowing between different ASes.

The growing importance of IXPs is a key aspect in the flattening, community structure of the Internet. The novel aspect introduced by IXPs, is that its customers, even when acting in a myopic and selfish manner, can find benefits in being together on IXPs. These benefits increase for, but are not limited to, customers with symmetric traffic levels, and hence of similar importance in the hierarchy. Therefore, an IXP acts as an attractor, and its participants typically become very well connected between themselves to exploit all the benefits of this facility. This interaction paradigm will be deepened in the next chapter, where we show how IXP participants become well "correlated" between themselves, and exploit this phenomenon as the ground truth for the core network model.

4.8 Appendix

4.8.1 MCM Cost Function Derivation

Consider the model in Figure 4.3.1. Given that we have $I, L, N = 2$, where $l = 1$ is the NSP and $l = 2$ is the NSP, we can explicitly rewrite the general cost function (4.2.11) for the two players as in (4.8.1).

$$\begin{cases} C^1(\mathbf{x_i}, \mathbf{x_{-i}}) = & \left(x^1_{1,1} + x^2_{1,1}\right)\left(a_1 + \frac{1}{c_1 - (x^1_{1,1} + x^2_{1,1})}\right) + \left(x^1_{1,2} + x^2_{1,2}\right) \cdot \\ & \cdot \left(\frac{a_2}{\sqrt{x^1_{1,2} + x^2_{1,2} + x^1_{2,2} + x^2_{2,2}}} + \frac{1}{c_2 - (x^1_{1,2} + x^2_{1,2})}\right) + \\ & + x^1_{1,1} b^1_1 + x^2_{1,1} b^2_1 + x^1_{1,2} b^1_2 + x^2_{1,2} b^2_2 \\ C^2(\mathbf{x_i}, \mathbf{x_{-i}}) = & \left(x^1_{2,1} + x^2_{2,1}\right)\left(a_1 + \frac{1}{c_1 - (x^1_{2,1} + x^2_{2,1})}\right) + \left(x^1_{2,2} + x^2_{2,2}\right) \cdot \\ & \cdot \left(\frac{a_2}{\sqrt{x^1_{1,2} + x^2_{1,2} + x^1_{2,2} + x^2_{2,2}}} + \frac{1}{c_2 - (x^1_{2,2} + x^2_{2,2})}\right) + \\ & + x^1_{2,1} b^1_1 + x^2_{2,1} b^2_1 + x^1_{2,2} b^1_2 + x^2_{2,2} b^2_2 \end{cases} \tag{4.8.1}$$

Constraints shown in (4.2.5) can also be rewritten as:

$$\begin{cases} x^1_{1,1} + x^1_{1,2} = \phi^1_1 \\ x^2_{1,1} + x^2_{1,2} = \phi^2_1 \end{cases} \qquad \begin{cases} x^1_{2,1} + x^1_{2,2} = \phi^1_2 \\ x^2_{2,1} + x^2_{2,2} = \phi^2_2 \end{cases} \tag{4.8.2}$$

For the sake of readability, we apply the following variable renaming to our problem:

$$\begin{cases} x_1 = x^1_{1,2} \\ x_2 = x^2_{1,2} \end{cases} \qquad \begin{cases} y_1 = x^1_{2,2} \\ y_2 = x^2_{2,2} \end{cases} \tag{4.8.3}$$

Meaning that x_n is the amount of traffic sent from player 1 to CP n through the IXP, while y_n is the analogous for player 2. Due to constraints (4.8.2) we have that:

$$\begin{cases} x^1_{1,1} = \phi^1_1 - x_1 \\ x^2_{1,1} = \phi^2_1 - x_2 \end{cases} \qquad \begin{cases} x^1_{2,1} = \phi^1_2 - y_1 \\ x^2_{2,1} = \phi^2_2 - y_2 \end{cases} \tag{4.8.4}$$

By substituting the variables defined in (4.8.3) and the ones obtained in (4.8.4), we get system (4.8.5), which gives the cost functions of the two players for the MCM model.

$$\begin{cases} C^1(x_1,x_2,y_1,y_2) = & \left(\phi_1^1+\phi_1^2-x_1-x_2\right)\left(a_1+\frac{1}{c_1-\left(\phi_1^1+\phi_1^2-x_1-x_2\right)}\right)+ \\ & +\left(x_1+x_2\right)\left(\frac{a_2}{\sqrt{x_1+x_2+y_1+y_2}}+\frac{1}{c_2-\left(x_1+x_2\right)}\right)+ \\ & +\left(\phi_1^1-x_1\right)b_1^1+\left(\phi_1^2-x_2\right)b_1^2+x_1b_2^1+x_2b_2^2 \\ C^2(x_1,x_2,y_1,y_2) = & \left(\phi_2^1+\phi_2^2-y_1-y_2\right)\left(a_1+\frac{1}{c_1-\left(\phi_2^1+\phi_2^2-y_1-y_2\right)}\right)+ \\ & +\left(y_1+y_2\right)\left(\frac{a_2}{\sqrt{x_1+x_2+y_1+y_2}}+\frac{1}{c_2-\left(y_1+y_2\right)}\right)+ \\ & +\left(\phi_2^1-y_1\right)b_1^1+\left(\phi_2^2-y_2\right)b_1^2+y_1b_2^1+y_2b_2^2 \end{cases} \tag{4.8.5}$$

We further know that the topology is fully connected, meaning that $b_l^n = 0\ \forall l, n$. In this case, from the player's perspective, the cost does not depend on the facility used for a specific CP, but rather on the total amount of flow going through a specific TF, independently from the destination content provider. Therefore we can consider these cumulative flows and cumulative demands as new variables of our problem:

$$\begin{cases} x = x_1 + x_2 & \textit{flow } ISP_1 \rightarrow IXP \\ y = y_1 + y_2 & \textit{flow } ISP_2 \rightarrow IXP \end{cases} \tag{4.8.6}$$

and

$$\begin{cases} \phi_1 = \phi_1^1 + \phi_1^2 & \textit{cum. demand } ISP_1 \\ \phi_2 = \phi_2^1 + \phi_2^2 & \textit{cum. demand } ISP_2 \end{cases} \tag{4.8.7}$$

By plugging (4.8.6) and (4.8.7) in system (4.8.5), we obtain the final form of the cost functions for the two players, shown in (4.3.2).

4.8.2 Proofs

Proof of Lemma 7. The second derivative of function (4.3.7) w.r.t. x is:

$$\frac{\partial^2 C^1(x)}{\partial x^2} = \frac{2c_1}{\left(c_1-(\phi-x)\right)^3} + \frac{2c_2}{\left(c_2-x\right)^3} + \frac{3a_2x}{4\left(x+y\right)^{\frac{5}{2}}} - \frac{a_2}{\left(x+y\right)^{\frac{3}{2}}} \tag{4.8.8}$$

We have that $c_1 \gg c_2, \phi$, therefore the first term is negligible when trying to check the sign of this derivative. Given that $x, y > 0$, $c_2 > x$ and $x \leq \phi$,

the second term is an always positive increasing function. For a fixed value of y, the summation of the third and forth term is an always negative increasing function. Therefore (4.8.8) is a monotonically increasing function, as it is the summation of two increasing functions. □

Proof of Theorem 8. According to Lemma 7 the second derivative of (4.3.7) is monotonically increasing. Therefore we can only have three cases:

$$\begin{cases} \frac{\partial^2 C^1(x)}{\partial x^2} < 0 \ \forall x & \Rightarrow C^1 \text{ always concave} \\ \frac{\partial^2 C^1(x)}{\partial x^2} > 0 \ \forall x & \Rightarrow C^1 \text{ always convex} \\ \exists! \bar{x} \ s.t. \ \left.\frac{\partial^2 C^1(x)}{\partial x^2}\right|_{x=\bar{x}} = 0 & \Rightarrow C^1 \text{ concave in } [0;\bar{x}] \\ & \quad \text{and convex in } [\bar{x};c_2] \end{cases}$$

□

Proof of Theorem 9. The strategy space is $x \in [0, \min(c_2, \phi_1)[$ and $y \in [0, \min(c_2, \phi_1)[$. Within this set, C is continuous, therefore, by Weierstrass' theorem, it has a global minimum, which might be either a stationary point or one of the interval endpoints. In order to find all the stationary points, we study the function gradient:

$$\begin{cases} \frac{\partial C(x,y)}{\partial x} = -a_1 - \frac{c_1}{(c_1-(\phi_1-x))^2} + \frac{a_2}{2\sqrt{x+y}} + \frac{c_2}{(c_2-x)^2} \\ \frac{\partial C(x,y)}{\partial y} = -a_1 - \frac{c_1}{(c_1-(\phi_2-y))^2} + \frac{a_2}{2\sqrt{x+y}} + \frac{c_2}{(c_2-y)^2} \end{cases} \tag{4.8.9}$$

By hypothesis we have symmetric demands: $\phi_1 = \phi_2 = \phi$. If we sum the two equations in (4.8.9) we have that:

$$-\frac{c_1}{(c_1-(\phi-x))^2} + \frac{c_2}{(c_2-x)^2} = -\frac{c_1}{(c_1-(\phi-y))^2} + \frac{c_2}{(c_2-y)^2}$$

which clearly is true only for symmetric strategies, therefore we must have $x = y$.

The capacity of the NSP is typically $c_1 \gg \phi, x$ therefore we can simplify $\frac{c_1}{(c_1-(\phi-x))^2} \approx \frac{1}{c_1}$. In order to find the stationary points we need to find the roots of equation:

$$-a_1 - \frac{1}{c_1} + \frac{a_2}{2\sqrt{2x}} + \frac{c_2}{(c_2-x)^2} \tag{4.8.10}$$

Unfortunately, this is a fifth-degree polynomial, therefore we don't have an explicit solution. Consider now the derivative of equation (4.8.10):

$$-\frac{a_2\sqrt{2}}{8x\sqrt{x}}+\frac{2c_2}{(c_2-x)^3} \tag{4.8.11}$$

This is a monotonically increasing function (as it is the sum of two increasing functions), that goes to $-\infty$ for $x \to 0$ and to $+\infty$ for $x \to c_2$, therefore it has one root. As a consequence, we know that eq. (4.8.10), representing the first derivative of C, is convex and has limits $+\infty$ for $x \to 0, c_2$. Therefore, it has a unique minimum point corresponding to the root of equation (4.8.11).

Given the form of its first derivative (4.8.10), the cost function (4.4.1) is concave in its first part (where the derivative decreases) and convex in the second part (where the derivative increases). The points of minimum of the concave part are its two endpoints, while the convex part has a unique minimum point. The right endpoint of the concave part is inside the convex part (C is continuous), therefore the convex minimum is an improvement over it. Henceforth, the global minimum point is either the minimum point of the convex part, or the left endpoint of the concave part, which is also the left endpoint for $C(x, y)$. □

Proof of Corollary 10. Given such an algorithm, we execute it on the function giving as initial point the right endpoint of the strategy set, where we know that the function is convex. The output of the algorithm is the local minimum of the convex part, therefore according to theorem 10 the global minimum point is either this point or the left endpoint. □

Proof of Theorem 11. First of all, we can use constraint (4.2.5) to reduce the number of variables of our system. In fact if we perform the summation over n on left and right member, and separate the NSP component from the IXPs, we obtain:

$$\sum_n x_{i,1}^n + \sum_{l\neq 1}\sum_n x_{i,l}^n = \sum_n \phi_i^n \tag{4.8.12}$$

Now we substitute $\sum_n x_{i,1}^n$ taken from (4.8.12) inside (4.5.1) and rear-

range terms, so as to obtain:

$$C^i(\mathbf{x_i}, \mathbf{x_{-i}}) = \sum_{l\neq 1}\sum_n x^n_{i,l}\left(\frac{a_l}{\sqrt{\sum_i\sum_n x^n_{i,l}}} + \frac{1}{c_l - \sum_n x^n_{i,l}}\right) + \\ + \left(\sum_n \phi^n_i - \sum_{l\neq 1}\sum_n x^n_{i,l}\right)\left(a_1 + \frac{1}{c_1 - \left(\sum_n \phi^n_i - \sum_{l\neq 1}\sum_n x^n_{i,l}\right)}\right) + \\ + \sum_{l\neq 1}\sum_n x^n_{i,l} b^n_l \tag{4.8.13}$$

where all the flow variables have $l \neq 1$. Now we need to compute the mixed second derivatives of equation (4.8.13). We observe that the second term, referring to the NSP, has no mixed components, due to the fact that the cost does not depend on other players' choice, therefore this term becomes zero in the computation. The same happens with the last term, therefore we have:

$$\frac{\partial^2 C^i(\mathbf{x_i}, \mathbf{x_{-i}})}{\partial x^{\bar{n}}_{j,\bar{l}} \partial x^n_{i,l}} = \frac{\partial}{\partial x^{\bar{n}}_{j,\bar{l}}}\left[\left(\frac{a_l}{\sqrt{\sum_i\sum_n x^n_{i,l}}} + \frac{1}{c_l - \sum_n x^n_{i,l}}\right) + \right. \\ \left. + \sum_n x^n_{i,l}\left(-\frac{a_l}{2\left(\sum_i\sum_n x^n_{i,l}\right)^{3/2}} + \frac{1}{\left(c_l - \sum_n x^n_{i,l}\right)^2}\right)\right] \\ \forall i \neq j,\ \forall l, \bar{l}, n, \bar{n} \tag{4.8.14}$$

Following the same reasoning previously done, we observe that the second and forth term in (4.8.14) do not depend on $x^{\bar{n}}_{j,\bar{l}}$, therefore their contribution in the final derivative is zero. Moreover, we observe that first and third term only have flow variables with index l, therefore for any $\bar{l} \neq l$ the whole derivative becomes zero:

$$\frac{\partial^2 C^i(\mathbf{x_i}, \mathbf{x_{-i}})}{\partial x^{\bar{n}}_{j,\bar{l}} \partial x^n_{i,l}} = 0 \qquad \forall i \neq j,\ \forall l \neq \bar{l},\ \forall n, \bar{n} \tag{4.8.15}$$

while in the other case we have:

$$\frac{\partial^2 C^i(\mathbf{x_i}, \mathbf{x_{-i}})}{\partial x^{\bar{n}}_{j,l} \partial x^n_{i,l}} = -\frac{a_l}{2\left(\sum_i \sum_n x^n_{i,l}\right)^{3/2}} + \frac{3a_l \sum_n x^n_{i,l}}{\left(\sum_i \sum_n x^n_{i,l}\right)^{5/2}} \quad \forall i \neq j,\ \forall l, n, \bar{n} \tag{4.8.16}$$

Please note that, regardless of the chosen $n, \bar{n}$, the derivatives are all the same. In order to prove symmetric supermodularity, we have to show that property (4.3.3) holds for both (4.8.15) and (4.8.16). While in the first case this is trivial, for the second one we multiply (4.8.16) by the positive quantity $\left(\sum_i \sum_n x^n_{i,l}\right)^{3/2}$, thus obtaining that:

$$\operatorname{sgn}\left(\frac{\partial^2 C^i(\mathbf{x_i}, \mathbf{x_{-i}})}{\partial x^{\bar{n}}_{j,l} \partial x^n_{i,l}}\right) = \operatorname{sgn}\left(\frac{a_l}{4} \cdot \frac{\sum_n x^n_{i,l} - 2\sum_{j \neq i} \sum_n x^n_{j,l}}{\sum_i \sum_n x^n_{i,l}}\right) \tag{4.8.17}$$

Along the symmetric axis we have that $x^n_{i,l} = x^n_{j,l}\ \forall i \neq j,\ \forall l, n$, meaning that each couple (i, j) of players send, to a fixed CP n through a given IXP l, the same quantity of flow. With this condition, equation (4.8.17) is always negative, therefore (4.5.1) is symmetric supermodular. □

Proof of Corollary 12. By hypothesis the demands satisfy $\phi^n_i = \phi^n_j\ \forall i \neq j,\ \forall n$. Therefore ISPs keep playing along the symmetric axis [fCRVW04], and we obtain this result by combining Theorems 4 and 11. □

Part III

A Novel Model for the Internet Topology

Chapter 5

The evolving Internet's Core: an IXP-centric network model

In this chapter, we propose an IXP-centric network model aimed at accurately reproducing the evolving core of the Internet at the Autonomous System (AS) level. The proposed model, called *X-CENTRIC* (iXp-Centric Evolving NeTwoRk model for the Internet's Core), exploits the results from previous chapters, that tackled two of the most important problems faced when modeling the Internet. Therefore, X-CENTRIC focuses on the Internet's ***core***, as it is the most interesting and structured part of the network, and uses an innovative dynamic which reflects the driving forces behind the evolution of the Internet, with the emerging flattening of its structure through the massive diffusion of IXPs.

Indeed, the game-theoretic model shown in the previous chapter highlights the role of IXPs as "concentrators", since the benefits of joining these facilities specifically emerge when there are many well connected participants. The natural evolution of this mechanism is the formation and growth of dense peering meshes around the IXPs. This consideration serves as the basis for the deployment of the dynamic rules of X-CENTRIC, as detailedly explained in Section 5.2. These dynamic rules are embedded

in X-CENTRIC through a small set of physically meaningful parameters, thanks to which the model has the potential to forecast the evolution of the core network. Results show that X-CENTRIC is able to both capture the statistical characteristics of the Internet's core and its community structure, measured in the form of maximal cliques [AGL14b].

At the end of the chapter, we highlight the crucial importance of the third key problem stressed in the thesis introduction. Indeed, the prediction properties are based on data from real measurements, therefore it is important to remove the biases introduced by data incompleteness, in order to accurately predict the future evolution of the Internet topology. We conclude that the model is versatile and capable of accurately predicting the evolution of the Internet's core as long as we are able to foresee the growth of the modeling parameters, which can be done under a stable measuring infrastructure.

5.1 Introduction and Related Work

Current efforts to analyze the Internet topology focus on the driving forces behind its evolution. Many studies have focused on the analysis of the Internet's topology evolution [DD11], highlighting the flattening phenomenon [DD10; LWY12] and the impact of IXPs [GILO11; AG10], however none have been able to devise a model which reproduces the evolving Internet topology under such driving forces. Moreover, studies in this field typically do not take into account the measuring infrastructure used to infer the Internet topology [GIL+12], and the possible biases introduced by the fact that this infrastructure changes over time.

The main research contribution of this chapter is X-CENTRIC, a novel IXP-centric evolving model for the Internet's core, whose ground truths are the structural properties of the topology and the awareness of the IXPs impact on its flattening (see Chapters 3 and 4). X-CENTRIC makes use of a well-defined service model, able to outline the true driving forces behind link creation and the Internet's evolution. Using a set of correlation parameters, we highlight the importance of IXPs within the Internet's core as responsible for its community structure, and define a

mechanism through which exchange point customers become attached. The successful validation carried out, using even sensitive metrics such as maximal cliques, highlights how fundamental our mechanism is if we want to grasp the true driving forces behind the evolution of the Internet's core.

X-CENTRIC reproduces the statistical characteristics of the Internet's core, such as the heavy-tailed degree distribution and its community structure (in the form of maximal cliques). In addition, it also captures the evolution of the network by looking at the growth of the modeling parameters, possibly making predictions on the future evolution of the core.

Unlike growing network models, X-CENTRIC can deal with aggregated events. Specifically, in growing network models, such as the BA-model, the time evolution is measured with respect to the number of nodes added to the graph. X-CENTRIC takes a different approach, and builds different graphs in different time steps: time is measured with respect to the evolution of modeling parameters at the different steps. The difference between the two approaches can be easily understood by an analogy with physics: we study the evolution of the temperature, pressure, etc.. (modeling parameters) of a room, rather than the interaction of its molecules (nodes). The strength of our method lies in the fact that modeling parameters are both a high level representation of the system and a way of reproducing the driving forces behind the node interactions.

Using X-CENTRIC and existing topological data, we highlighted a critical aspect of the measurement infrastructure: with the existing topological data it is impossible to study the long-term evolution of the Internet, as the measuring infrastructure is too variable over time and generates long-term fluctuations that hide the real evolution of the topology. However, the insights provide directions for future research in the field of Internet measuring and data gathering, and should be considered when trying to apply evolutionary network models to such data.

X-CENTRIC is thus versatile and capable of accurately predicting the evolution of the Internet's core as long as we are able to foresee the growth of the modeling parameters. Under a stable measuring infrastructure,

with knowledge of such parameters, the model would be able to predict the future evolution of the Internet.

The remainder of this chapter is organized as follows: in Section 5.2 we recall the core structure presented in Chapter 3 and illustrate our innovative dynamic, based on considerations from Chapter 4. In Section 5.3 we devise the X-CENTRIC model and compare its results with measured data, then Section 5.4 analyzes the evolution of the core, compares it with the measuring infrastructure, and highlights the predictive properties of our model.

5.2 Core Analysis and Correlations

As shown in Figure 3.2.9 of Chapter 3, the core of the Internet is a two-layer graph made up of three different networks. Here we analyze the three networks taking into account: i) the structural and behavioral properties of the nodes (Chapter 3); ii) the role of IXPs as concentrators (Chapter 4).

The rational behind our analysis is the awareness that the core is made up of many peering links, and its nodes participate in many IXPs. Thus, links within the core are mainly links between subsets of customers of an IXP. Internet eXchange Points act as concentrators between participants of the same IXP, which therefore form a community. We believe that, due to their role, IXPs naturally generate "correlation" properties in the three networks, where customers of the same IXP have a high correlation between them, and a low correlation with nodes participating in other IXPs. A proper model of the Internet's core should thus take into account strategies that avoid the destruction of such correlations, since they are an important indicator of the community structure (i.e. the massive presence of maximal cliques) in the core network.

There are two main questions that enable us to build a core model, and that we need to answer in our analysis: *how many links* should each node draw? And what kind of *attachment mechanism* should it follow? The answers depend on the reference network, and the mechanisms built upon those answers should be able to provide the correlations mentioned above.

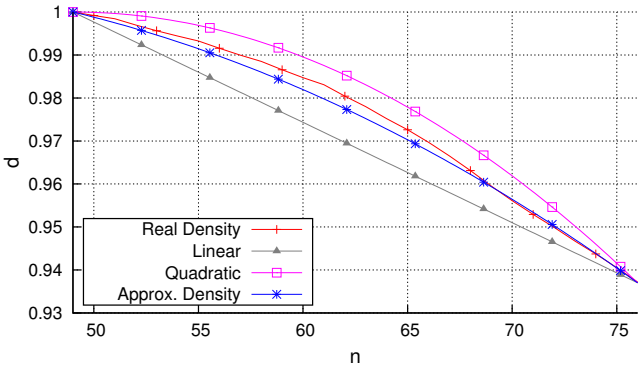


Figure 5.2.1: Density curve approximation - 2013

5.2.1 Centrum Network

The *centrum* is a network of very few nodes (at most 80 in 2013), which participate in many IXPs and, as such, are very well connected. This small network has a very high density, therefore it makes no sense to look for the existence of correlated/non-correlated nodes. The number of links that each node should draw is well captured by the first part of the density curve (before the inflection point), shown in Figure 5.2.1. A further analysis reveals that this curve can be easily modeled using just three parameters: the size of the largest clique $\bar{k}$, the size of the *centrum* $\bar{n}$, and the *centrum* density $\bar{d}$. In fact, as shown in Figure 5.2.1, the actual density curve is well approximated by a curve which averages a linear and a quadratic interpolation between points $(\bar{k}, 1)$ and $(\bar{n}, \bar{d})$. The approximated density curve can be used to derive the number of links drawn per node.

As far as the attachment mechanism is concerned, it is possible to show that the "directed preferential attachment" mechanism [AGL13] is able to reproduce the high number of maximal cliques observed. Specifically, the node has a picture of the directed *centrum* network, where each edge goes from the newer connecting node to the older connected node, and the preferences are [1 : 2]. This means that for each node, an outgoing link adds a preference of 1, while an incoming link adds a preference of 2. In any case, the small network size makes the problem of deriving a good

attachment mechanism far less crucial compared with the same problem applied to the other networks.

5.2.2 Vertical Network

The *vertical* network has a very particular degree distribution, shown in Figure 3.3.3 of Chapter 3. This distribution is in 1 to 1 correspondence with the number of vertical links towards the centrum. Therefore, from it we can infer that the number of links that each node should inject to the *centrum* is uniformly distributed in $\mathcal{U}(1, p_v)$.

Vertical links reflect the willingness of a node to connect to the *centrum*. We believe that the main mechanism through which a *layer-1* node lays out connections towards the *centrum* is related to its willingness to join an IXP. As shown in Chapter 3, *centrum* nodes participate in many IXPs with an Open Peering policy. Therefore, when a *layer-1* node joins the IXP, it will connect with many *centrum* participants. In addition, since these participants are on the IXP, they are also very well connected with each other, therefore the addition of a new node will contribute to form new large cliques.

This assumption can be verified with real data. $X(i)$ indicates the set of *centrum* nodes to which *layer-1* node i is connected, and $E_{X(i)}$ the set of edges between such nodes. Therefore, $c_{X(i)} = 2|E_{X(i)}|/(|X(i)|(|X(i)|-1))$ indicates the density of $X(i)$. If $c_{X(i)} \to 1$ then $X(i)$ tends to a clique, while for $c_{X(i)} \to 0$ the nodes of $X(i)$ tend to an independent vertex set. An example of computation for $c_{X(i)}$ is shown in Figure 5.2.2. We measured $c_{X(i)}$ $\forall i \in N_2$, that is, for all *layer-1* nodes. Figure 5.2.3 compares the distribution of $c_{X(i)}$ obtained with real data, with the one that would have been obtained if each *layer-1* node had chosen $X(i)$ using a classic preferential attachment (PA). The huge distance between the two distributions validates our thesis: the driving force behind the choice of nodes is far from random, and derives instead from their willingness to join an IXP, thus forming cliques with other participants.

We studied the distribution in Figure 5.2.3, which we call a "*vertical correlation distribution*", over time and noticed that its structure comes

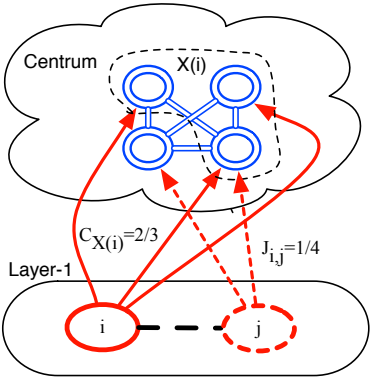


Figure 5.2.2: Correlation indexes computation

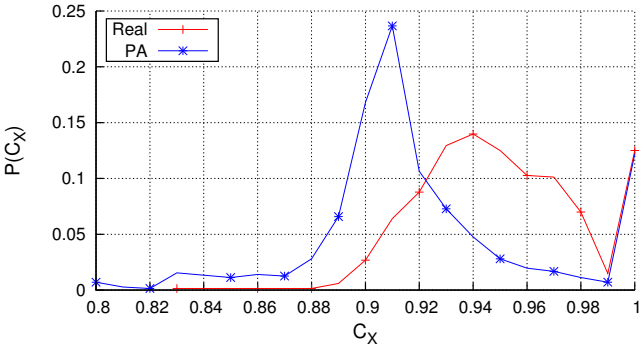


Figure 5.2.3: Vertical correlation metric - Real vs PA - 2012

from the superposition of a Gaussian distribution with small variance, and a mirrored power law distribution. We believe that this behavior derives from the two possible rules that a node can use when joining an IXP: manually configuring its peers or joining the Route Server [Eur]. In a manual configuration, the node will typically lay out a small number of links. Such connections are also fine-tuned by the administrator, therefore there is a high chance that peers are already part of the same community, thus they are already peering with each other. This phenomenon is reflected in Figure 5.2.4, where the vertical correlation distribution for

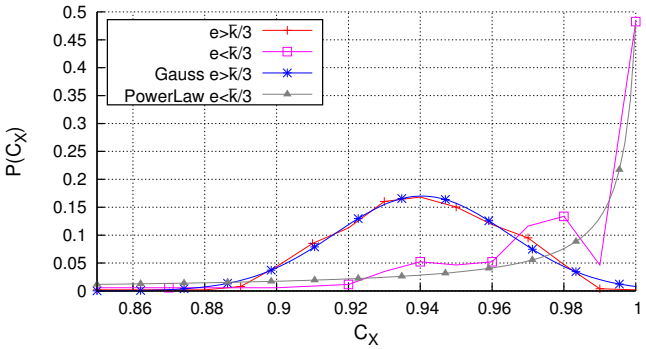


Figure 5.2.4: Vertical correlation metric - Real vs Approximation - 2012

nodes injecting few links (square markers curve), is well fitted by a power law with an average close to 1 (triangle markers curve). If the node uses the Route Server, it will automatically peer with all the participants in the IXP, therefore the number of connections laid out will be typically high. Peers will not necessarily be part of the same community (there is no fine-tuning), therefore the correlation will be lower than in the previous case. Figure 5.2.4 highlights the vertical correlation distribution for nodes injecting many links (cross markers curve), well fitted by a Gaussian with small variance (star markers curve).

To sum up, our attachment mechanism for the vertical network still uses a preferential attachment to ensure a heavy-tailed degree distribution, however we need a second constraint to ensure that the vertical correlation follows the distribution highlighted in Figure 5.2.4.

5.2.3 Horizontal Network

We found that the number of links laid out by nodes in the *horizontal* network is far less crucial than in the previous networks. In order to keep the model simple, we use a uniform distribution $\mathcal{U}(1, p_h)$ for the horizontal network as well. Parameter p_h is easily calculated so that when the network formation is complete, the total number of links is $|\hat{E}|$, the same as the measured core.

As shown in Chapter 3, although their participation is less pronounced

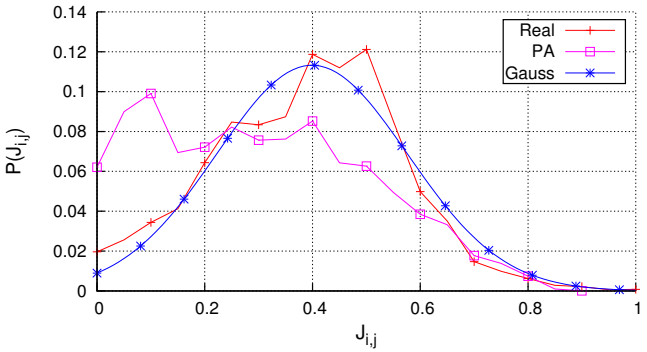


Figure 5.2.5: Horizontal correlation metric - Real vs PA vs Gauss - 2012

than that of *centrum* nodes, *layer-1* nodes participate in IXPs as well, however their peering policy is typically Selective or Restrictive. As such, we believe that the main mechanism through which a *layer-1* node lays out connections towards other *layer-1* nodes is related to a configuration step, through which two *layer-1* nodes already on the same IXP, decide to create a direct link with each other. Since these nodes are already on the same IXP, this means that they share a considerable number of neighbors, therefore this link creation will also heavily contribute to the formation of cliques.

As with the vertical network, we verified this assumption with real data. For each edge $(i, j) \in E_3$, in the horizontal network, $X(i)$ and $X(j)$ indicate the sets of *centrum* nodes to which nodes i and j are connected, respectively. We compute the Jaccard Similarity between these two sets $J_{i,j} = |X(i) \cap X(j)|/|X(i) \cup X(j)|$. If $J_{i,j} \to 1$ then i and j have the same *centrum* neighbors, while for $J_{i,j} \to 0$ the nodes have totally different *centrum* neighbors. An example of computation for $J_{i,j}$ is shown in Figure 5.2.2. Figure 5.2.5 compares the distribution of $J_{i,j}$ obtained with real data, with the distribution that would have been obtained if each *layer-1* node had chosen its horizontal neighbor using classic preferential attachment. As in the previous case, the huge distance observed between the two distributions validates our thesis: even this choice is far from random, and can be pictured as a configuration step through which *layer-1* nodes on the same IXP create a new direct link, thus contributing to the clique

formation process.

We studied the distribution in Figure 5.2.5, which we call a "*horizontal correlation distribution*", over time and noticed that its structure is always a Gaussian distribution. To sum up, our attachment mechanism for the horizontal network still uses preferential attachment to ensure a heavy-tailed degree distribution, however we need a second constraint to ensure that the horizontal correlation follows the distribution highlighted in Figure 5.2.5.

5.3 X-CENTRIC

The two-layer core structure so far obtained, along with the new dynamic rules highlighted in Section 5.2, enable us to build a novel model for the Internet's core, based both on the observed structural properties and on behavioral rules for the nodes verified with real data.

In this section we propose X-CENTRIC, a model exploiting our findings aimed at accurately reproducing the Internet's core, capturing both its statistical characteristics, such as the heavy-tailed degree distribution, and its community structure, measured in the form of maximal cliques.

5.3.1 Core Modeling

Thanks to the partitioning of the core, it is possible to generate it in three different steps, each of which builds a separate part of the network.

X-CENTRIC uses the following parameters, which were highlighted in the previous section and are summarized here:

$$\begin{cases} |N|,\ |E| & \textit{core parameters} \\ \bar{n},\ \bar{k},\ \bar{d} & \textit{centrum parameters} \\ \mathcal{U}\left(1, p_v\right) & \textit{vertical distribution} \\ \mathcal{N}_v\left(\mu_v, \sigma_v^2\right) & \textit{vertical correlation} \\ \mathcal{N}_h\left(\mu_h, \sigma_h^2\right) & \textit{horizontal correlation} \end{cases}$$

and builds the network using a three-step algorithm.

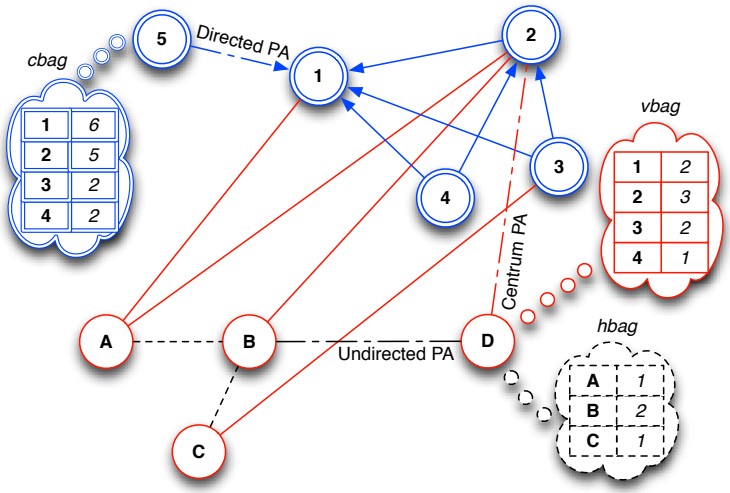


Figure 5.3.1: Core attachment mechanisms

5.3.1.1 Centrum Network construction (alg. 5.1)

We start with a single node and for each new incoming node $i < \bar{n}$ we calculate how many links to add by exploiting the interpolated density curve (*line 2*). The new links are drawn according to the mechanism shown in Figure 5.3.1 for node 5, referred to as "Directed preferential attachment" (*lines 8-13*), where preferences are implemented through a bag of nodes as shown in [CT12]. Basically the node has a picture of the directed *centrum* network, where each edge goes from the newer connecting node to the older connected one, and preferences are $[1 : 2]$, meaning that for each node an outgoing link adds a preference of 1, while an incoming link adds a preference of 2 (e.g. node 2 with 2 in-links and an out-link has preference 5). This choice emulates a $[0 : 1]$ preference mechanism, which would have been too detrimental for new nodes, and is needed to enforce PA in a region where a similar degree of all nodes would have neglected its use. Thanks to the fact that connected nodes have a higher preference than connecting ones, we have that the more a node leaves the center, the less the probability of a new node connecting to it.

Algorithm 5.1 Centrum Construction

```
1: function DENSITY(i)
2:     return 1 + (i−k̄)(d̄−1)(i+n̄−2k̄) / 2(n̄−k̄)²
3: end function
4:
5: add_to_bag(cbag, 0)
6: for i = 1; i < n̄; i++ do
7:     e = n_edges(i, DENSITY(i))                ▷ Interpolated density
8:     while e ≠ 0 do
9:         n = cbag[uniform(0, sizeof(cbag))]
10:        add_edge(i, n)
11:        add_to_bag(cbag, i, n, n)             ▷ Directed PA i → n
12:        e = e − 1
13:    end while
14: end for
```

5.3.1.2 Vertical Network construction (alg. 5.2)

Starting from the *centrum* network, we add new nodes. Each incoming node $\bar{n} \leq i < |N|$ adds e links towards the *centrum*, where e is extracted from $\mathcal{U}(1, p_v)$ (*line 6*). Note that in this step we do not add links between *layer-1* nodes.

The node extracts a number d from the vertical correlation distribution $\mathcal{N}_v\left(\mu_v, \sigma_v^2\right)$ (*line 7*) and then chooses a set X of *centrum* nodes using the preferential attachment mechanism (*lines 10-17*), shown in Figure 5.3.1 for node D. It then computes c_X and checks whether $|c_X - d| < \epsilon$, meaning that c_X is near d (*lines 18-19*). If the condition is true, X is accepted and the node creates the links (*lines 20-23*), otherwise it picks a new set X and repeats the procedure. Thanks to this simple mechanism, we are able to both use preferential attachment and ensure that the resulting vertical correlation distribution is akin to $\mathcal{N}_v\left(\mu_v, \sigma_v^2\right)$.

Although the node knows the vertical links between layer-1 nodes and the *centrum*, it is unaware of the internal structure of the *centrum*. This means that at the beginning all *centrum* nodes have an equal preference of 1 (*lines 1-3*), which changes with the injection of new edges. In order to inject just vertical links, we compute preferences only for *centrum* nodes,

Algorithm 5.2 Vertical Construction

1: **for** $i = 0; i < \bar{n}; i++$ **do**
2: add_to_bag($vbag, i$) ▷ Equal starting preferences
3: **end for**
4:
5: **for** $i = \bar{n}; i < |N|; i++$ **do**
6: $e = \mathcal{U}(1, p_v)$
7: $d = \mathcal{N}_v(\mu_v, \sigma_v^2)$ ▷ Requested vertical correlation
8: **repeat**
9: $tbag = vbag, te = e$
10: **while** $te \neq 0$ **do** ▷ Select X with Centrum PA
11: $n = tbag$[uniform(0, sizeof($tbag$))]
12: **if** edge_not_exists(i, n) **then**
13: add_to_bag($tbag$, n)
14: insert(X,n)
15: $te = te - 1$
16: **end if**
17: **end while**
18: c_X = vert_correlation(X) ▷ Test correlation
19: **until** $|c_X - d| < \epsilon$
20: **for all** $n \in X$ **do** ▷ Add X neighbors to i
21: add_edge(i,n)
22: add_to_bag($vbag$, n) ▷ Centrum PA $\rightarrow n$
23: **end for**
24: **end for**

adding a preference of 1 for each undirected incident edge (e.g. node 2 has preference 3 due to its starting preference and the connections with A and B).

5.3.1.3 Horizontal Network construction (alg. 5.3)

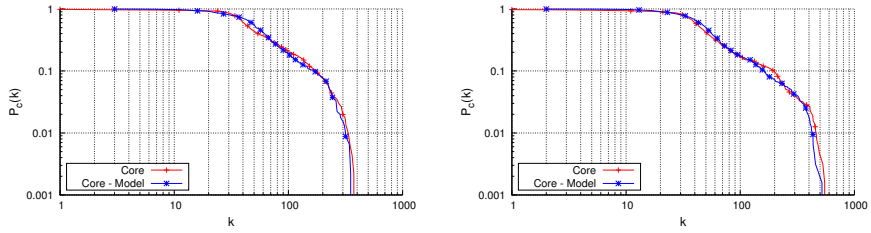
Starting from the previously obtained network, we iterate again over the nodes $\bar{n} \leq i < |N|$. Each node adds e links towards other *layer-1* nodes, where e is extracted from $\mathcal{U}(1, p_h)$ (*line* 6). Given that we already generated the vertical network, it is easy to compute p_h knowing both the current number of nodes and edges, and its total number $|N|$, $|E|$.

Algorithm 5.3 Horizontal Construction

1: **for** $i = \bar{n}; i < |N|; i++$ **do**
2: add_to_bag($hbag$, i) ▷ Equal starting preferences
3: **end for**
4:
5: **for** $i = \bar{n}; i < |N|; i++$ **do**
6: $e = \mathcal{U}(1, p_h)$
7: **while** $e \neq 0$ **do**
8: X_i=get_neighbors(i)
9: $d = \mathcal{N}_h\left(\mu_h, \sigma_h^2\right)$ ▷ Requested horiz. correlation
10: **repeat**
11: $n = hbag$[uniform(0, sizeof($hbag$))]
12: **if** edge_not_exists(i, n) **then**
13: X_n = get_neighbors(n)
14: $J(i, n)$ = horiz_correlation(X_i,X_n)
15: **end if**
16: **until** $|J(i, n) - d| < \epsilon$
17: add_edge(i, n)
18: add_to_bag($hbag$, n, i) ▷ Undirected PA $i - n$
19: $e = e - 1$
20: **end while**
21: **end for**

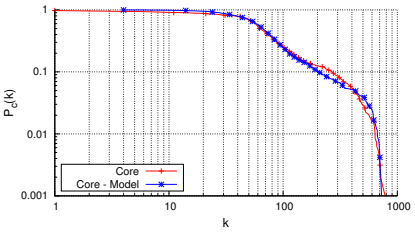
Similarly to the previous step, node i extracts a number d from the horizontal correlation distribution $\mathcal{N}_h\left(\mu_h, \sigma_h^2\right)$ (*line* 9), then chooses a *layer-1* node n using the preferential attachment mechanism (*line 11*), shown in Figure 5.3.1 for node D. Afterwards, it computes $J_{i,n}$ and checks whether $|J_{i,n} - d| < \epsilon$, meaning that $J_{i,n}$ is near d (*lines 14,16*). If the condition is true, n is accepted and the node creates the links (*lines 17-19*), otherwise it picks another node n and repeats the procedure. Thanks to this mechanism, we are able to both use preferential attachment and ensure that the resulting horizontal correlation distribution is akin to $\mathcal{N}_h\left(\mu_h, \sigma_h^2\right)$.

The preferences are separated: while step 2 uses preferences to the *centrum*, step 3 refers to only *layer-1* nodes (e.g. node B has preference 2 due to connections with nodes A and C).



(a) 2007

(b) 2010



(c) 2013

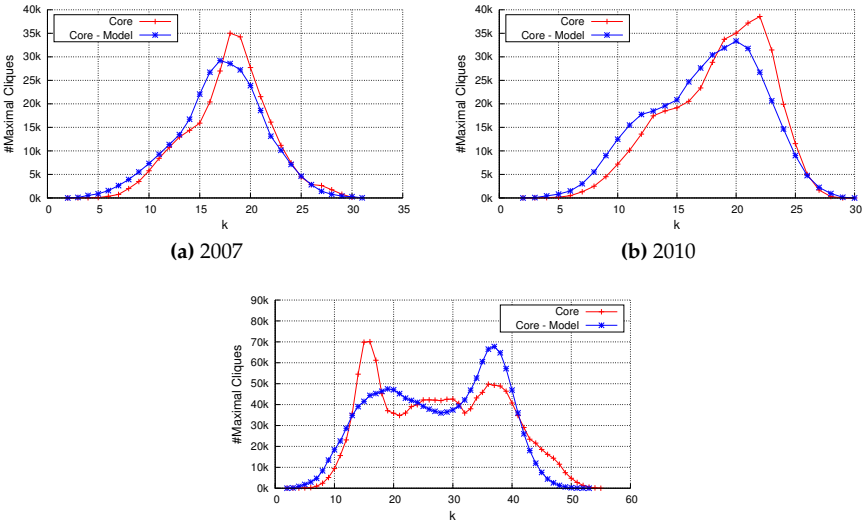
Figure 5.3.2: Core network: degree distribution comparison

5.3.2 Results

We compared the extracted core $\hat{G}$ with the results obtained by X-CENTRIC. Input parameters were extracted according to the procedure outlined in Section 5.2.

Figure 5.3.2 compares the degree distribution of the core $\hat{G}$ with the one obtained from X-CENTRIC, while Figure 5.3.3 compares their maximal clique distribution. X-CENTRIC provides a very good fit for the degree distribution, matching it throughout the whole analyzed time span, as shown in Figures 5.3.2a, 5.3.2b and 5.3.2c. The fit for the maximal clique distribution is fairly good, the model is able to match the evolving structure of the core despite of the evident change in number and size of the maximal cliques (see Figures 5.3.3a, 5.3.3b and 5.3.3c). We know that this fit is not perfect, however this is not the main concern, given the high variability of this distribution observed on real data (see [RWM+11]).

Finally, Table 5.1 compares the core and X-CENTRIC taking into account other standard graph metrics, and shows how our algorithm is able



(a) 2007

(b) 2010

(c) 2013

Figure 5.3.3: Core network: maximal clique distribution comparison

to match the statistical characteristics of the Internet.

5.4 Core Evolution

Here we carry out the analysis of the Internet's core evolution by observing the growth of the parameters used by X-CENTRIC, listed at the beginning of Section 5.3.1 and extracted from the measured data using the guidelines in Section 5.2.

5.4.1 Parameters Evolution

The ability to devise a model with a few scalar parameters, makes it possible to analyze and predict the future evolution of the Internet. In fact, if we can understand the time evolution of such parameters, and predict their future value, then X-CENTRIC is able to construct a topology matching what the future Internet would be.

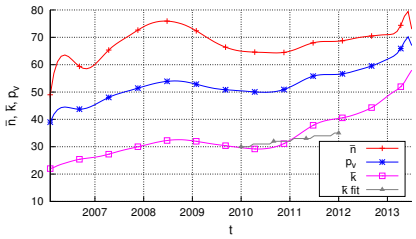
	2006		2008		2010	
Avg. Values	**Real**	**Model**	**Real**	**Model**	**Real**	**Model**
Degree	66.54	65.32	77.05	76.20	76.85	76.30
Clustering	0.642	0.622	0.628	0.620	0.631	0.644
Knn	156.07	139.54	175.97	161.12	246.13	211.03
Shortest Path	1.843	1.839	1.872	1.844	1.925	1.891
Betweenness	164.01	163.32	202.82	196.15	293.01	281.90
Closeness	0.548	0.548	0.541	0.547	0.525	0.530
Coreness	37.5	36.38	43.28	42.06	42.59	41.66

	2012		2013	
Avg. Values	**Real**	**Model**	**Real**	**Model**
Degree	93.80	91.63	105.52	104.64
Clustering	0.663	0.678	0.665	0.669
Knn	313.35	319.90	330.77	330.86
Shortest Path	1.984	1.921	1.978	1.910
Betweenness	443.93	409.63	467.18	428.39
Closeness	0.515	0.521	0.521	0.530
Coreness	51.38	49.03	58.37	56.26

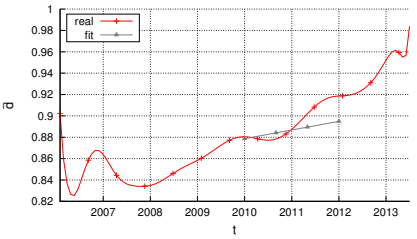
Table 5.1: Real vs X-CENTRIC: average metrics comparison

Therefore, here we analyze the time evolution of such parameters, shown in Figures 3.2.2 (from Chapter 3) and 5.4.1. In general, all the parameters are growing with time. The number of nodes $|N|$ within the core grows linearly, while the number of links $|E|$ grows quadratically (Figure 3.2.2), which keeps the core well connected and highly dense. Figure 5.4.1a reveals that the *centrum* is getting bigger, and so is its largest clique. Like the size of the *centrum*, the number of vertical links towards the *centrum* is also increasing. In fact, Figure 5.4.1a highlights that these three parameters are strictly correlated: the *centrum* size $\bar{n}$, the size of the largest clique $\bar{k}$ and the maximum number of vertical links towards the *centrum* p_v grow similarly, just shifted one from the other.

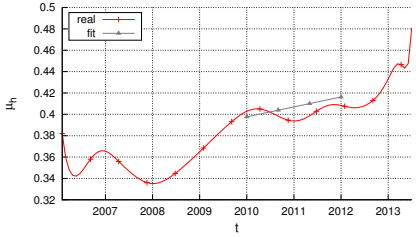
The centrum density $\bar{d}$ and the average correlation values μ_h and μ_v



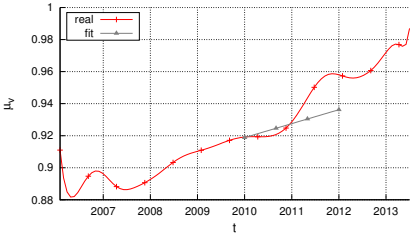
(a) $\bar{n}$, $\bar{k}$, p_v



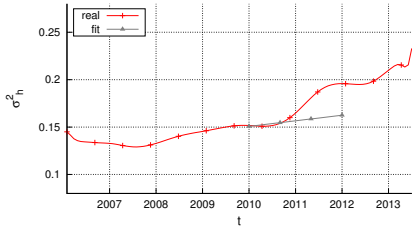
(b) $\bar{d}$



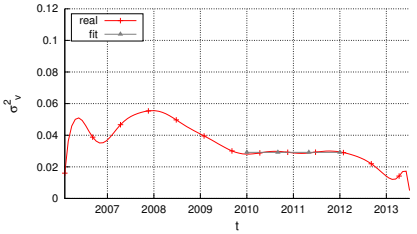
(c) μ_h



(d) μ_v



(e) σ_h^2



(f) σ_v^2

Figure 5.4.1: Evolution of core parameters

are also increasing, as shown in Figures 5.4.1b, 5.4.1c and 5.4.1d. The growth in correlation values over time seems to indicate the increasing presence of IXPs within the core. On the one hand, the growth of μ_v indicates that the presence of *centrum* nodes on exchange points is increasing. On the other, the growth of μ_h reflects the fact that more and more horizontal links are due to IXPs. While these parameters are all growing over time, their specific trend is difficult to understand and model. The centrum density $\bar{d}$ grows linearly during some time slots, and seems to be more stable in others (Figure 5.4.1b), as if there were different attitudes

in different "epochs". This behavior is clearly confirmed in the evolution of the average correlation values μ_h and μ_v. The vertical correlation μ_v (Figure 5.4.1d) grows during 2008-2009 and 2011-2012, but it is almost stable in the other time slots. A similar trend is exhibited by μ_h (Figure 5.4.1c) although the growing time slots are slightly different. As shown in Figures 5.4.1e and 5.4.1f, the variances are almost stable or linearly growing. In either case, their value is very small and their evolution is not expected to have a significant impact on the model.

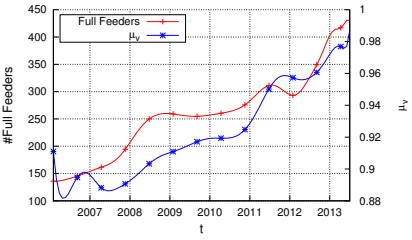
To sum up, while it is difficult to model the trends of the parameters, they appear to have a sort of semi-stationary behavior: we observe some time-slots, or epochs, where they are almost stable, and other ones where they grow. Initially this behavior seems to be directly due to the Internet itself, but further analysis reveals that it is more related to the perception of the Internet that we get through its measuring infrastructure.

5.4.2 Measuring Infrastructure

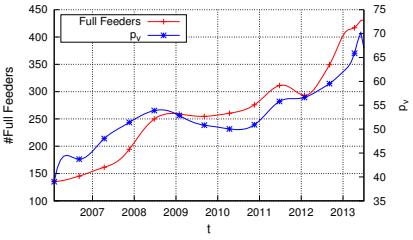
As previously mentioned, the dataset used in our modeling and analysis is provided by Isolario [Iso], which infers the AS-level topology graph using data coming from several cooperating ASes, indicate as *feeders*. Since each feeder provides a representation of the network from its own specific point of view, the data collected are aggregated to obtain the complete Internet topology [GLM12].

As the number of feeders increases, the number of links observed in the Internet topology increases as well. Therefore, it is difficult to understand if the growing network topology is an effect of the evolution in the network itself, or is due to the increased visibility of the Internet topology obtained through the new feeders. Our aim here is to understand how much the measuring infrastructure influences our viewpoint of the Internet's evolution.

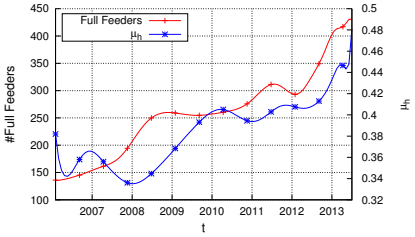
As shown in [GLM12], each feeder provides a different contribution to route collectors. To better quantify the total contribution of feeders, the authors of [GLM12] subdivide them in three different categories: minor, partial and full feeders. Several cooperating ASes are minor feeders and



(a) Feeders vs μ_v



(b) Feeders vs p_v



(c) Feeders vs μ_h

Figure 5.4.2: Evolution of core parameters vs number of full feeders

do not provide any relevant information to the route collectors, while a few are full feeders and are responsible for most of the inferred Internet topology. In addition, feeders can change over time, as new ASes begin cooperating in route collector projects, while other ASes terminate their cooperation.

Figure 5.4.2 shows the number of full feeders over time, highlighting how the Internet's measuring infrastructure is far from stable. More specifically, Figures 5.4.2a, 5.4.2b and 5.4.2c compare the evolution of the measuring infrastructure with that of the main parameters of X-CENTRIC. As shown in Figure 5.4.2a, the trend of parameter μ_v is tightly coupled with the growing number of full feeders: as the number of feeders increases, μ_v grows as well, while it is almost constant with a stable number of feeders. Figure 5.4.2b reinforces this claim, showing that it is also true for p_v. Since p_v itself is coupled with other centrum parameters ($\bar{n}$, $\bar{k}$, as previously observed in Figure 5.4.1a), also those parameters are indirectly coupled with the evolution of the measuring infrastructure. Unlike

previous parameters, the correlation between μ_h and the number of full feeders, shown in Figure 5.4.2c, is more difficult to grasp. In this case, there seems to be a time-shift between the two trends, as if the impact of new feeders on μ_h is captured a few months after their establishment.

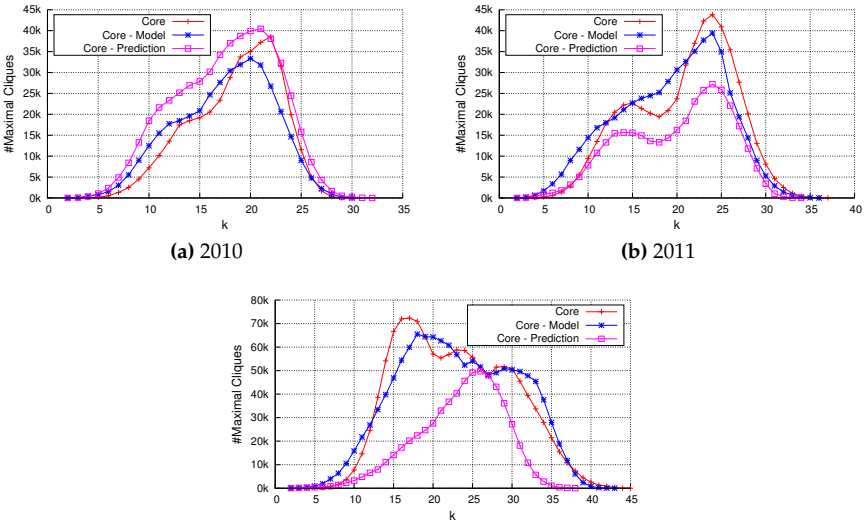
Since the evolution of core parameters is heavily biased by the measuring infrastructure, one possible approach would be to try to find a sufficiently stable time-slot and analyze the evolving core of the Internet within that time slot. As shown in Figure 5.4.2, the number of feeders seems sufficiently stable between the second half of 2008 and the beginning of 2011. Unfortunately, a further analysis highlights that even in this time-slot, the measuring infrastructure changes considerably: out of the 260 average number of full feeders, only 155 are stable, while there are 250 other feeders that appear/disappear during the time span. A second, more drastic approach would be to extract a subpart of the topology obtained considering only those feeders that are steadily present along the analyzed time span. Unfortunately, there are few feeders that are stable for a sufficiently long period of time, and considering only them would misrepresent the Internet.

We believe that the insights arising from this analysis shed light on directions for future research in the field of Internet measuring and data gathering, and should be kept in mind when trying to apply evolutionary network models to such data.

5.4.3 Predictive Properties

Despite its negative results, this analysis does not invalidate the main results for X-CENTRIC and its ability to potentially predict the future evolution of the Internet under a stable measuring infrastructure. In this section, we illustrate the prediction properties of the model.

We take as input data the evolution of parameters observed in the 2006-2009 time span and perform a linear fit of their observed trend (Figure 5.4.1). We suppose that future data are unknown, and use these fits to "predict" the value of the parameters for the 2010-2012 time span. The fitted parameters are used as input for X-CENTRIC, so as to generate a



(a) 2010 **(b)** 2011

(c) 2012

Figure 5.4.3: Maximal clique distribution for core prediction

prediction on the evolution of the Internet's core.

While matching the statistical description of the Internet topology is relatively easy, matching its structure in terms of cliques is far more difficult, therefore we focus on this task. Figure 5.4.3 compares the predicted evolution of the maximal clique distribution with its actual evolution (both from measured data and from our model). As we see, while the short and medium-term predictions are quite good (Figures 5.4.3a and 5.4.3b), the long-term prediction is far from reality (Figure 5.4.3c). By looking at the growth of parameters, it is immediately clear that such behavior must be sought after the sudden change in their trend, especially for some of them. For example, for $\bar{d}$ and μ_v (Figures 5.4.1b, 5.4.1d), the linear fits are quite different from the actual trend, and the effect of the wrong estimation of their values is promptly reflected in a bad long-term prediction of topological properties. As observed in the previous section, this sudden difference in the parameters' trend is caused by the changing measuring infrastructure.

We conclude that X-CENTRIC is capable of accurately predicting the

evolution of the Internet's core as long as we are able to predict the growth in the parameters observed. If we had complete knowledge of such parameters, the model would be able to reproduce the topology even in the event of drastic changes to the measuring infrastructure.

5.5 Conclusions

We have proposed X-CENTRIC, an innovative model able to outline the true driving forces behind link creation and the evolution of the Internet's core. X-CENTRIC captures both the statistical characteristics of the Internet's core, such as its heavy-tailed degree distribution, and its community structure, measured in the form of maximal cliques, over a time span of almost eight years. X-CENTRIC also has the potential to forecast the evolution of the network, which can be represented by a set of few well-defined parameters. To the best of our knowledge this is the first evolving model that achieves both these goals.

X-CENTRIC is versatile and capable of accurately reproducing the core, and even predicting its evolution as long as we are able to predict the growth of its parameters. The analysis carried out reveals that, to date, our perception on the evolution of the Internet is tightly connected with that of its measuring infrastructure, which fluctuates over time and biases possible conclusions. Therefore, despite the potential predictive properties of the model, it cannot be fully exploited to this end until the measuring infrastructure is sufficiently stable.

In conclusion, X-CENTRIC is able to generate an evolving topology which well represents the core of the Internet, and provides great insights that could be exploited for future research. Since the model uses data derived from real measurements, it is important to remove the biases introduced by the measuring infrastructure, in order to observe the true nature of the evolving network parameters, understand their real trend and predict the future evolution of the Internet's topology.

X-CENTRIC is able to only reproduce the core of the Internet. In the last part of the thesis, we show how to add the periphery, so as to obtain a model for the whole Internet. We anticipate that adding

the periphery is a far easier task than modeling the core. In fact, this periphery is represented by a sparse graph, with nodes not participating in communities and exhibiting no correlation properties. Therefore, we expect that properly corrected BA-like models are sufficient for describing link creation outside of the core.

Chapter 6

Modeling and Validation for the Internet Topology

Thanks to the model proposed in the previous chapter, we have been able to accurately reproduce the evolving core of the Internet at the Autonomous System level. The model is based on several insights regarding both the structure and the dynamical rules responsible for the interconnection of ASes, thus highlighting the driving forces behind the evolution of the core network. The last compulsory step to obtain a topology generator for the whole Internet is the addition of peripheral nodes to the core: here we propose a simple dynamic for node attachment with this purpose [AGL14a].

The generator, called X-CENTR-ITE (X-CENTRIC Internet Topology gEnerator), is validated against the measured Internet graph, and it is shown how it outperforms the reference generators in the literature when looking at both coarse and medium-grained metrics such as the degree distribution, and fine-grained metrics, quantitatively captured, in our case, by the maximal cliques distribution. To the best of our knowledge, this is the first model able to match, with good accuracy and for such a long time span, the statistical and structural metrics of the Internet.

6.1 Introduction and Related Work

We use as ground truth for this work, the X-CENTRIC model proposed in the previous chapter. In order to add the periphery to the obtained core, we first exploit existing data so as to analyze the behavior of peripheral nodes and define the criteria in order to attach them to the previously obtained core.

In Section 6.2 we complete the previously obtained model for the Internet's core by adding the periphery. Section 6.3 presents the final results, comparing our model with the four reference generators, described in Section 2.3.3 of Chapter 2: ORBIS [MHK+07], GT-ITM [ZCD97], INET [WJ02] and BRITE [MLMB01]. These generators were tuned using available data, so that they could yield their best possible results.

6.2 Adding the Periphery

The X-CENTRIC model shown in the previous chapter is a first important step for the development of our topology generator. In this section we analyze the periphery, i.e. the partition outside the core, in order to define the rules followed by peripheral nodes joining the Internet. Our final result is obtained by putting all the pieces together, so as to obtain a topology generator for the whole Internet.

6.2.1 Periphery Analysis

Following a similar path to the one used for the core, we try to understand how to model the periphery by once again analyzing the dataset provided by Isolario [Iso]. The analysis, as usual, reports results with specific reference to the time step of April 2013. While some of the numbers are changing in time, as will be shown in the next section, the general behavior is the same throughout the whole analyzed time span of eight years.

We recall some notations from previous chapters: $G = (N, E)$ is the graph of the whole Internet and $\hat{G} = (\hat{N}, \hat{E})$ is the graph of the Internet's

core, where $\hat{N}$ and $\hat{E}$ are the core nodes and links, which we indicate here as c-nodes and c-links. Furthermore, we define as $\tilde{N} = N \backslash \hat{N}$ and $\tilde{E} = E \backslash \hat{E}$ all the Internet nodes and links lying outside the core, in the periphery, which we indicate as p-nodes and p-links.

In order to include p-nodes and p-links in our topology generator, we need to define the behavior of the peripheral nodes joining the Internet. More specifically, we need to answer two questions:

1. How many *p-links* does each p-node draw when entering the graph?
2. What kind of *attachment mechanism* should each p-node follow?

According to the analysis performed throughout the thesis, the most complex, structured, and interesting part of the Internet is the core, while the periphery is made up of a large number of nodes mainly organized into trees, with a negligible contribution to the community structure of the Internet. A straightforward approach to modeling this part of the Internet would be to use the standard Barabási-Albert method. In this case, the answer to the above questions would simply be to draw a constant number of p-links for each p-node, using classic preferential attachment. Our analysis reveals that this approach, without proper corrections, is wrong, both from an engineering standpoint and when looking at raw data.

Question 1 The first step to answer Question 1 lies in understanding the hierarchical structure of p-nodes. Once we have a core, in fact, we can hierarchically organize the peripheral nodes, according to their "*distance*" from the core. More specifically, for each p-node, we define its distance from the core as the shortest path length between the peripheral node and any of the c-nodes. Figure 6.2.1 represents the layering obtained with this definition, where a p-node lies in layer x if it has distance x from the core. Figure 6.2.1 highlights that the majority of links are up-directed, connecting different layers, while only 16% of the links connect nodes in the same layer, which therefore lie in the same hierarchical level. In order to "break the tie" between them, we hierarchically order nodes from the same layer so that the higher the degree, the higher the node level.

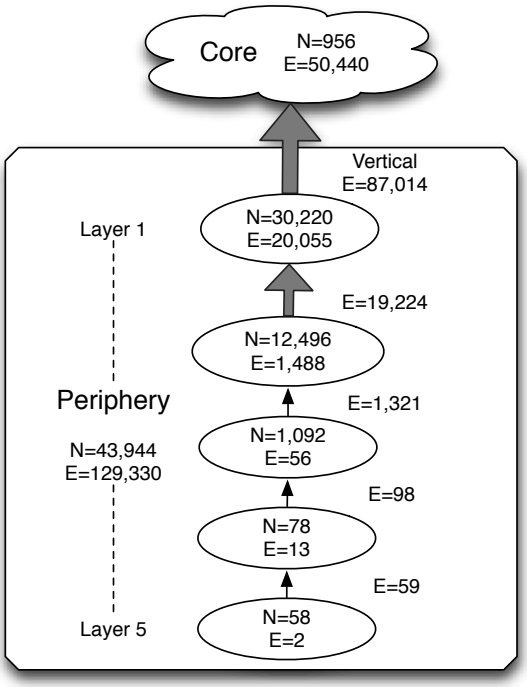


Figure 6.2.1: Periphery layering as of April 2013

Therefore, if nodes A and B are in the same layer and have degrees k_A and k_B with $k_A > k_B$ respectively, the up-directed link goes from node B to node A. This "weak" hierarchical definition is very powerful, because now each p-node only has up-directed links, which can be mapped in a 1:1 correspondence with the p-links that the node draws in the graph, thus answering our first question. We know that the tiebreak rule is not rigorous, however the number of links for which it is applied is quite low and does not affect the results too much. Figure 6.2.2 shows the CCDF of the up-directed degree distribution for our dataset, where the degree k_{up} of each node is equal to the sum of its up-directed links. The distribution is heavy-tailed, and can be roughly approximated by a power law P_1, with exponent $\gamma_1 = 2.2$ and a sharp cut-off at $k_1 = 60$. Although the real distribution is not exactly a power-law, the simple approximation we

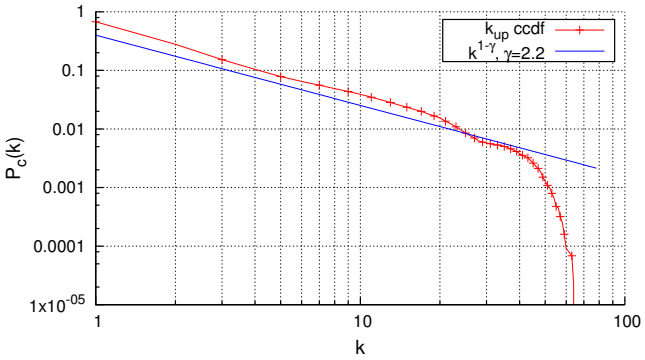


Figure 6.2.2: k_{up} degree ccdf - 2013

make here yields good results, as will be shown later.

Question 2 Given the nature of the peripheral network and the lack of community structure, it is clear that the preferential attachment mechanism could be used. When p-nodes are added to the graph, we already have an existing core network, therefore we need to set some initial preference for the c-nodes, in order for the p-nodes to attach themselves to these c-nodes. In the classic PA definition, each node's preference is equal to its existing degree. However, as we saw in the analysis in Section 3.3.2 from Chapter 2, while the core is mainly made up of peering links, the periphery is dominated by customer-to-provider links. From an engineering standpoint these two phenomena are completely different, therefore we cannot use degrees obtained though peering links to set up preferences for customer-to-provider links. For example, within the core Tier-1 ASes and big Network Service Providers, attracting many customer p-nodes, coexist with other nodes (Content Providers, etc.) which are mainly interested in keeping peering relationships with core nodes, and care much less about customer connectivity. Therefore the high degree of some core nodes, typically obtained by joining many IXPs, is a bad indicator of their willingness to attract customers (and therefore have a high preference for them), thus we have to use a different criteria to set preferences. Determining the correct preferences for core nodes entails exploiting our dataset in order to understand what happens to the core *after* peripheral nodes have

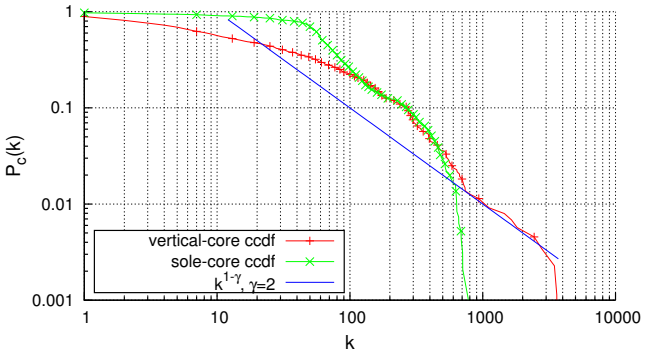


Figure 6.2.3: Vertical/Sole core degree ccdf - 2013

been added. Figure 6.2.3 shows the "vertical-core degree distribution", which is the degree distribution of c-nodes obtained by only considering the p-links, coming from the periphery (mainly C2P links, represented by the bold arrow in Figure 6.2.1). This distribution, as shown in Figure 6.2.3, is very different from the "sole-core degree distribution", which is the degree distribution of c-nodes obtained by only considering c-links (mainly P2P links). Due to the fact that we need to lay out the vertical, customer-to-provider links between the periphery and the core, it is much more appropriate to extract preferences from a distribution which approximately fits the vertical-core degree distribution, rather than the sole-core distribution. Figure 6.2.3 shows that a good fit is obtained from a power law P_2 with exponent $\gamma_2 = 2$ and cut-off $k_2 = 4000$.

6.2.2 P-nodes Addition Algorithm

By answering questions 1 and 2 we now have all the ingredients to understand what happens when a peripheral node joins the graph. The p-nodes addition algorithm takes as input the core network as generated in Section 5.3.1 of previous chapter, and builds as output the whole Internet topology by adding the periphery.

It is well known that preferential attachment by itself is unable to generate the "*Small-World*" effect, as it yields a small diameter but is unable to produce high values for the clustering coefficient. In order to

Algorithm 6.1 P-node Addition

```
1:  for n = 0; n < |N|; n++ do                          ▷ Set Preferences
2:      x =(int)get_plaw_number(k2,γ2);                  ▷ P2
3:      add_to_bag(bag, x, n);
4:  end for
5:
6:  for n = |N|; n < tot_nodes; n++ do                  ▷ Attach
7:      e =(int)get_plaw_number(k1,γ1);                  ▷ P1
8:      while e ≠ 0 do
9:          m = bag[uniform(0, sizeof(bag))]
10:         add_edge(m, n)
11:         add_to_bag(bag, m, n)
12:         e − −
13:         if uniform(0,1) < p then                     ▷ Triangle
14:             h = select_neighbor(m);
15:             add_edge(h, n)
16:             add_to_bag(bag, h, n)
17:             e − −
18:         end if
19:     end while
20: end for
```

solve this problem, we exploit an idea that is often used in the literature (see [LKF05; New01]) and add a further step to the attachment mechanism. After drawing a link to a c-node, the peripheral node tries to link with one of its neighbors. This technique ensures the formation of triangles, which are quite common in the Internet's graph (e.g. multi-homed ASes), and produces high clustering values.

The algorithm needs two distributions as further input, which we both approximate with power laws: the up-directed link distribution P_1, with exponent γ_1 and cut-off k_1, and the preference distribution P_2, with parameters γ_2 and k_2.

Algorithm 6.1 performs the following steps:

- assign a preference to each of the c-nodes, extracted from distribution $P_2(\gamma_2, k_2)$;
- for each new p-node, assign a number of p-links to be drawn, ex-

tracted from $P_1(\gamma_1, k_1)$;

- for each link to be drawn:
  - select a node m from the graph according to its preference and connect to it;
  - with probability p choose a neighbor of m and connect to it;
  - update preferences for the inserted p-node.

The algorithm terminates when the generated graph has size $|N|$, same as the Internet's graph.

6.3 Results

6.3.1 X-CENTR-ITE

The X-CENTR-ITE topology generator is obtained by the serial execution of the three-step core generation algorithm (X-CENTRIC) presented in the previous chapter, and the p-node addition algorithm just shown. While we already carried out a time analysis for the core and its parameters, here we have to check the behavior of the new parameters introduced. A temporal analysis of the distributions P_1 and P_2 reveals that:

- the up-directed link distribution P_1 is approximately constant until 2011, then the cutoff grows from 40 up to 60;
- the preference distribution P_2 is time-dependent in both the exponent γ_2 and cutoff k_2. The exponent γ_2 is slowly increasing over time from 1.6 to 2, while the cutoff is approximately constant until 2011, then almost doubles.

This apparently strange behavior, as already stressed in Section 5.4.2 of Chapter 5 can be attributed to the changing measuring infrastructure. In the next subsection we test our generator showing how a fine-tuning of these parameters yields good results. As a matter of fact, the generated topologies match graph metrics measured for the Internet, thus confirming the robustness of our method through the analyzed eight-year period.

We recall that the maximal clique distribution is virtually unaffected by the addition of p-nodes (except for small cliques), therefore this new step is irrelevant for this metric. Furthermore, we compare our results with that of the reference generators in the literature.

6.3.2 Comparison with existing Topology Generators

We compare the measured Internet graph with the results obtained with X-CENTR-ITE and those obtained using four Internet Topology Generators: ORBIS, GT-ITM, INET and BRITE. Parameters for X-CENTR-ITE were chosen as previously described, while for the other generators:

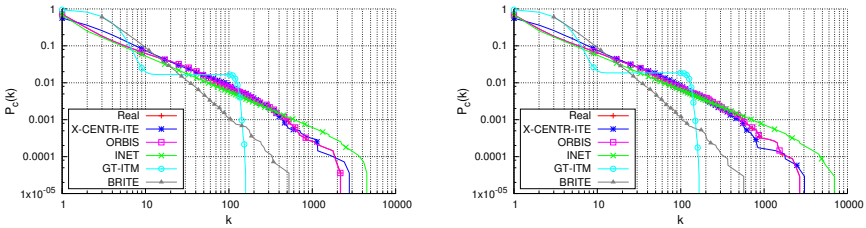
ORBIS [MHK+07] is a topology generator which exploits the concept of dK degree distributions to construct the network. We extracted the 2K distribution from the measured Internet topology, and directly used it as input for the generator, so that the generator could yield its best possible results.

GT-ITM [ZCD97] is a hierarchical generator, which uses the concepts of "transit" and "stub" so as to generate a multi-tiered topology. The available parameters enable us to reproduce the network partitioning into core/periphery, and to properly configure their sizes and the edges between them. Therefore we used them to reproduce the observed Internet structure as accurately as possible.

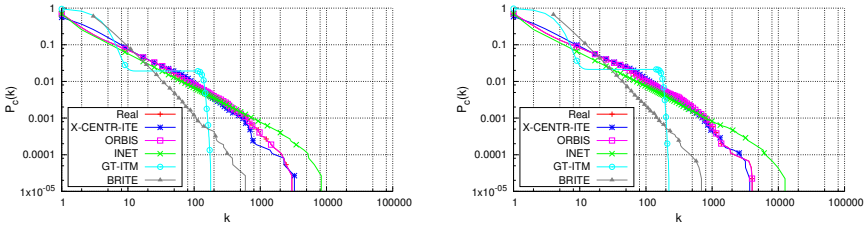
INET [WJ02] is another hierarchical topology generator, using a pre-built degree distribution to provide a good fit for this metric. This generator can be configured through a small set of parameters related to topological metrics, while it is not possible to configure the network structure.

BRITE [MLMB01] uses the AS-Level Barabási-Albert model, and can be configured by modifying the parameters already highlighted for the generalized BA-model presented in Section 2.3 of Chapter 2.

Figure 6.3.1 compares the degree distribution, while Figure 6.3.2 compares the maximal clique distribution. GT-ITM is unable to produce



(a) 2008 **(b)** 2010

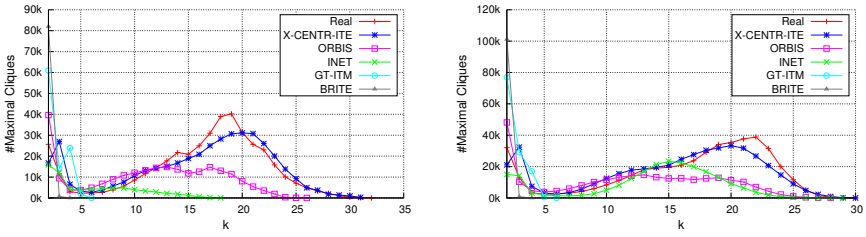


(c) 2011 **(d)** 2013

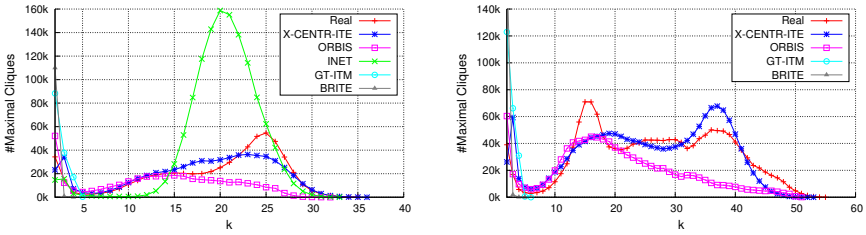
Figure 6.3.1: Degree distribution comparison

scale-free degree distributions, therefore this metric is always wrong with such generator. The BRITE generator is able to reproduce scale-free distributions, however it is not possible to configure a distribution for the up-directed links. Therefore, all the produced distributions (Figure 6.3.1a to 6.3.1d) have a correct shape but wrong starting and ending point. The INET generated topologies provide a fairly good fit for the degree distribution, especially in the first year (Figure 6.3.1a), while in the last year (Figure 6.3.1d) the degree can assume very large values that are not observed in the actual topology. Of course, by using as input the 2K degree distribution, ORBIS matches perfectly the degree distribution. The X-CENTR-ITE model is able to produce very good fits for the degree distribution, if appropriately tuned. In particular, to generate above results we used $\gamma_1 = 2.2$, while other parameters are growing so as to match the increasing average degree of the Internet: $k_1 = 40 \rightarrow 60$, $k_2 = 2000 \rightarrow 4000$ and $\gamma_2 = 1.6 \rightarrow 2$.

As for the maximal cliques (Figure 6.3.2), BRITE produces tree-like graphs, therefore the number of maximal cliques produced is negligible.



(a) 2008 **(b)** 2010



(c) 2011 **(d)** 2013

Figure 6.3.2: Maximal clique distribution comparison

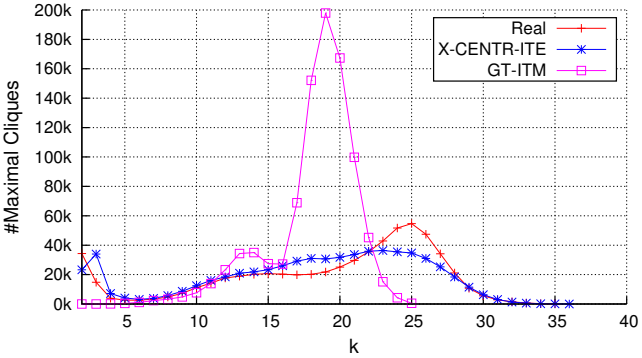


Figure 6.3.3: Real vs GT-ITM vs X-CENTR-ITE - 2011

Even GT-ITM, when used to generate the whole Internet, is unable to build large cliques. However, it is also possible to apply GT-ITM to just build the core, by setting the appropriate parameters. In this case we obtain better results, shown in Figure 6.3.3. Even if it is able to generate a bimodal clique distribution more similar to that of the Internet, GT-ITM

2006	Real	X-CNTR-ITE	ORBIS	INET	GT-ITM	BRITE
Degree	5.77	6.05	5.77	5.56	6.02	6.00
Clustering	0.32	0.30	0.13	0.39	0.034	0.003
Knn	506.61	464.65	503.86	1156.73	27.29	23.42
Shortest Path	3.71	3.57	3.44	3.05	4.36	4.53
Betweenness	30731.6	29087	27505	23163.3	37979.2	39944.9
Closeness	0.27	0.28	0.29	0.33	0.23	0.22
Coreness	3.00	3.09	2.96	2.81	3.58	3.00
2008	**Real**	**X-CNTR-ITE**	**ORBIS**	**INET**	**GT-ITM**	**BRITE**
Degree	6.11	6.15	6.11	6.18	6.29	6.00
Clustering	0.30	0.29	0.12	0.49	0.03	0.002
Knn	542.73	534.12	539.49	1802.88	30.51	23.86
Shortest Path	3.72	3.60	3.48	2.96	4.35	4.60
Betweenness	38003.8	36245.9	34440	27339.7	46832.1	50172.6
Closeness	0.28	0.28	0.29	0.34	0.23	0.22
Coreness	3.17	3.29	3.14	3.15	3.79	3.00
2010	**Real**	**X-CNTR-ITE**	**ORBIS**	**INET**	**GT-ITM**	**BRITE**
Degree	6.58	6.55	6.21	7.34	6.91	6.00
Clustering	0.29	0.27	0.11	0.61	0.03	0.002
Knn	567.67	540.26	529.56	3698.65	41.94	24.54
Shortest Path	3.76	3.71	3.52	2.83	4.18	4.69
Betweenness	51616.0	50578.8	43136.4	34054.8	59619.6	68800.1
Closeness	0.27	0.27	0.29	0.36	0.24	0.21
Coreness	3.40	3.39	3.18	3.84	4.18	3.00
2013	**Real**	**X-CNTR-ITE**	**ORBIS**	**INET**	**GT-ITM**	**BRITE**
Degree	8.00	8.05	8.01	8.25	8.43	8.00
Clustering	0.32	0.31	0.13	0.65	0.03	0.002
Knn	695.25	724.20	692.25	6317.73	70.58	30.00
Shortest Path	3.72	3.67	3.45	2.72	3.83	4.33
Betweenness	61141.9	59842.1	54928.0	38725.6	63708.3	74865
Closeness	0.27	0.28	0.29	0.37	0.26	0.23
Coreness	4.15	4.20	4.11	4.45	5.07	4.00

Table 6.1: Real vs Models: average metrics comparison

is still way far from the actual maximal clique distribution. Moreover, this tool is still incapable of capturing the scale-free degree distribution of the Internet. The INET generator is able to produce larger cliques,

however the fit is almost always wrong. In particular, the number of cliques produced in 2008 is very low (Figure 6.3.2a), and while in 2010 it would seem that the distribution is similar to what observed for the Internet (Figure 6.3.2b), our further analysis reveals that this conjecture is wrong. In fact, the maximal clique distribution for INET is bell-shaped, centered around a single value, while the distribution for the Internet is wider, and with more modes. Furthermore, after 2010 the distribution for INET starts growing uncontrollably: as shown in Figure 6.3.2c, the number of cliques is way too high, while in 2013 we were even unable to perform the computation, due to memory limits, using the algorithm in [CT12] on a computer with 128-GB RAM. Finally, ORBIS is capable of fitting a little better the left mode of the clique distribution, however it is totally unable to capture the existence of large cliques belonging to the right mode. X-CENTR-ITE provides a fairly good fit for the maximal cliques, generating a multimodal maximal clique distribution akin to that of the Internet throughout the whole analyzed time span. This is an important quantitative indicator of the presence of a community structure in the model itself.

Even when compared to generators using a far more complex input such as the joint degree distribution, X-CENTR-ITE provides an almost perfect fit for the degree, and outperforms other models when looking at the maximal cliques. Finally, Table 6.1 illustrates how our model is able to match the graph metrics of the Internet's topology through time, outperforming previous generators.

To sum up, although classical models produce topologies that fit well one metric or the other, none of them provides a good fit for all the observed metrics. These results would seem to confirm that our topology generator successfully addresses the problem of both representing the Internet's structure, in the form of maximal cliques, and matching statistical metrics such as the degree distribution, clustering coefficient, shortest path length, and so on. To the best of our knowledge, this is the first model that has achieved both these goals.

Part IV

Closing Part

Chapter 7

Conclusions

The Internet is a complex system that evolved during the past few decades from a small research network to a worldwide multipurpose network. A proper understanding of the characteristics and key factors driving the evolution of the Internet and its structure is a fundamental task from which many areas can benefit: deployment of CDNs, topology-aware routing algorithms, prediction of outages and traffic engineering. Being driven by a mix of technological and economic constraints, the characteristics of the Internet are hard to be revealed. Despite several attempts by researchers, much is still unknown and a proper model for the evolving Internet topology has yet to be given. In this thesis we have addressed three key problems related to the Internet modeling, since we believe that any model overlooking them is deemed to fail.

The first key problem is the large-scale heterogeneous nature of the Internet. Due to the high heterogeneity of the Internet topology, a model cannot rely on a uniform class of nodes which all exhibit the same behavior. Therefore, we carried out an in-depth analysis of the topology, using the concept of maximal clique to identify and discover the different building blocks of the network. This methodology enabled us to partition the Internet topology into a small core and a periphery. On the one hand, the core is a small, very dense network which well represents the most interesting structural properties of the network. On the other, the periphery

a sparse network of many nodes loosely connected between each other, representing the "tendrils" of the topology. Please note that while this technique has been applied specifically to the Internet topology, it is fairly general and can be applied to any complex network, thus revealing the existence of heterogeneous zones in the graph.

The second key problem is that network evolution is not dictated by a central authority, but is mainly the outcome of local economic and technical constraints. This problem is intertwined with the previous one, since heterogeneous nodes will have heterogeneous local goals. We highlighted the peering/transit dichotomy as the outcome of the emerging importance of Internet eXchange Points within the core, which are becoming the main antagonists of backbone Network Service Providers (e.g. Tier-1 Networks). Our analysis covers several sides of this aspect, revealing how good IXP policies enforce cooperation even in an environment where agents act in a myopic and selfish manner. This analysis enabled us to think of novel IXP-centric interaction mechanisms that were never conjectured in previous Internet topology models.

The proposed model for the Internet's evolving core, X-CENTRIC, uses as ground truth both the structural and behavioral patterns previously outlined. The model is based on data from real measurements, so it has to deal with the third key problem: techniques for data gathering are incomplete and often unreliable. In this sense, we have shown the robustness of the model through a long time span, highlighting how changes of the measuring infrastructure are correctly reflected in similar changes to the modeling parameters. In addition, X-CENTRIC is able to correctly predict the evolution of the core, in terms of the measured metrics, in temporal windows where the measuring infrastructure is stable.

Once the important structural and behavioral patterns within the core have been captured, the task of adding the periphery becomes relatively easy, since this part is still characterized by simple transit mechanics that are well-captured by the preferential attachment mechanism. In the last part of the thesis we devised a topology generator exploiting the previously obtained core model together with a simple mechanism for

attaching the periphery. The generator is able to successfully capture the Internet's structure, measured in the form of maximal cliques. As a matter of fact, this metric captures very well the emerging role of IXPs in the Internet topology. In addition, the model takes into account all the other statistical graph metrics, among which the heavy-tailed degree distribution, and outperforms existing topology generators in the literature.

The ability to represent the Internet as a multi-layer network, study the interactions between these layers, and deploy a model which takes into account such interactions, give to the thesis added value, since the analysis of multi-layer networks is a hot topic of several on-going European and worldwide projects [CON; MUR].

7.1 Future Work

Revealing the basic driving forces of the Internet is a first important step for modeling its growth. In this sense, the outcomes of this thesis are of utmost importance. Nevertheless, many aspects of the inferred model, such as the network partitioning, the dual layered core, and the different correlations are "given from above" through the analysis of existing data, rather than being the result of some kind of direct interaction between the nodes.

The first possible enhancement for the X-CENTRIC model, is trying to obtain the multi-layered structure of core and periphery as the result of a direct interaction between nodes, optimizing a tangled fitness function which might spontaneously lead to the formation of different tiers, hopefully similar to those measured for the real Internet. Alternatively, it might be advantageous to utilize concepts brought from evolutionary game theory. These tools utilize mix strategies and might be used to obviate the need to artificially separate the core into multiple layers.

Another important aspect, is deepening the economic analysis carried out in Chapter 4. First of all, the problem of peering vs transit can be enhanced using hierarchical Stackelberg models. In this case, we can set up a scenario where in one level ISPs compete for the best transit, and on

the second level they try to form coalitions at IXPs. Moreover, while we currently examine a static situation, both transit and peering costs can be renegotiated over time. In order to keep into account this phenomenon, we could add as players the NSPs and IXPs, and then analyze a dynamic game where their strategies are intertwined with that of ISPs. Finally, in the model we only focus on the strategic complementarity that ISPs have by sharing costs at IXPs. Actually, it has been shown that, especially in asymmetric cases, peering might be detrimental, due to effects of business stealing [AG09]. This can induce a trade-off between the benefits of sharing the costs of connectivity and the costs of losing final demand, or having to charge a lower price for it, that would be interesting to analyze.

Furthermore, a very important point to analyze is network formation around IXPs. While we studied the contribution of IXPs in fixed topologies, it is interesting to understand the dynamic that brings customers to attach to IXPs, so as to devise a model where nodes directly attach to exchange point customers without a superimposed correlation metric. A preliminary work on an agent-based model that we are developing to study network formation around IXPs can be found in [AMG+14].

Finally, we believe that the different network models should be compared using an impartial, standard framework.

In conclusion, the dependence of modeling from accurate data is indisputable. In the case of X-CENTRIC, removing the biases introduced by the measuring infrastructure would enable us to predict the future evolution of the Internet topology, therefore we believe that enhancements in the field of data gathering are much needed.

References

[AA03] E. Altman and Z. Altman. S-modular games and power control in wireless networks. *IEEE Trans. on Automatic Control*, 48, 2003. 82, 83, 96

[AAGL14a] Giovanni Accongiagioco, Eitan Altman, Enrico Gregori, and Luciano Lenzini. A Game Theoretical study of Peering vs Transit in the Internet. In *NetSciCom 2014 INFOCOM Workshop - To Appear*, 2014. xv, 10, 71

[AAGL14b] Giovanni Accongiagioco, Eitan Altman, Enrico Gregori, and Luciano Lenzini. Peering vs Transit: a Game Theoretical Model for Autonomous Systems Connectivity. In *IFIP Networking 2014 - To Appear*, 2014. xv, 10, 71

[AAMS12] Giovanni Accongiagioco, Matteo Andreozzi, Daniele Migliorini, and Giovanni Stea. Throughput-optimal resource allocation in LTE-Advanced with distributed antennas. In *3th Workshop of Italian group on Quantitative Methods in Informatics (InfQ)*, pages 1–8, 2012. 3

[AAMS13] Giovanni Accongiagioco, Matteo Andreozzi, Daniele Migliorini, and Giovanni Stea. Throughput-optimal resource allocation in lte-advanced with distributed antennas. *Computer Networks*, 57(18):3997 – 4009, 2013. 3

[AB99] Réka Albert and Albert-László Barabási. Emergence of Scaling in Random Networks. *Science*, 286(509), 1999. 39

[AB02] Reka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 2002. 39

[ACF+12] Bernhard Ager, Nikolaos Chatzis, Anja Feldmann, Nadi Sarrar, Steve Uhlig, and Walter Willinger. Anatomy of a large european IXP. In *Proc. of the 2012 ACM SIGCOMM Conference*, pages 163–174, 2012. 19, 82, 104

[AG09] D'Ignazio Alessio and Emanuele Giovannetti. Asymmetry and discrimination in Internet peering: evidence from the LINX. *International Journal of Industrial Organization*, 27(3):441–448, 2009. 156

[AG10] Mohammad Zubair Ahmad, Ahmad and Ratan Guha, Guha. Impact of Internet Exchange Points on Internet Topology Evolution. In *Proceedings of the 2010 IEEE 35th Conference on Local Computer Networks*, LCN '10, pages 332–335, Washington, DC, USA, 2010. IEEE Computer Society. 19, 114

[AGL13] Giovanni Accongiagioco, Enrico Gregori, and Luciano Lenzini. A Structure-Based Topology Generator for the Internet' s Core. In *4th Workshop of Italian group on Quantitative Methods in Informatics (InfQ)*, pages 1–8, 2013. xv, 10, 52, 53, 117

[AGL14a] Giovanni Accongiagioco, Enrico Gregori, and Luciano Lenzini. S-bite: a structure-based internet topology generator. *Submitted to Elsevier Computer Networks*, 2014. xv, 11, 52, 137

[AGL14b] Giovanni Accongiagioco, Enrico Gregori, and Luciano Lenzini. X-centric: an ixp-centric evolving network model for the internet' s core. *Submitted to IEEE Transactions on Networking*, 2014. xv, 11, 114

[AHDBV06] José Alvarez-Hamelin, Luca Dall'Asta, Alain Barrat, and Alessandro Vespignani. k-Core Decomposition: A Tool for the Visualization of Large Scale Networks. *Advances in Neural Information Processing Systems*, 18, 2006. 27

[AJB00] Reka Albert, Hawoong Jeong, and Albert-Laszlo Barabasi. The Internet's Achilles' heel: Error and attack tolerance of complex networks. *Nature*, 406:200–0, 2000. 44

[AMG+14] Giovanni Accongiagioco, Eli A. Meirom, Enrico Gregori, Luciano Lenzini, Shie Mannor, and Ariel Orda. Agent-Based Model for Internet Hub Formation, 2014. Poster at European Conference on Complex Systems ECCS'14, Lucca, Italy. 156

[BGP] BGPmon. http://bgpmon.netsec.colostate.edu. 21

[BSS08] 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. 5, 20

[BW08] Alain Barrat and Martin Weigt. On the properties of small-world network models. *Eur. Phys.*, 13:1–19, 2008. 38

[CAAP06] Hyunseok Chang, Ann Arbor, Park Ave, and Florham Park. To Peer or not to Peer : Modeling the Evolution of the Internet ' s AS-level Topology. *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, 2006. 45

[cai] CAIDA: The Cooperative Association for Internet Data Analysis. 5, 21

[CGJ+02] Hyunseok Chang, Ramesh Govindan, Sugih Jamin, Shenker Scott J., and Walter Willinger. On Inferring AS-Level Connectivity from BGP Routing Tables. *Technical Report UM-CSE-TR-454-02*, 2002. 22

[CGWJ02] Qian Chen, Ramesh Govindan, Walter Willinger, and Sugih Jamin. The origin of power laws in Internet topologies revisited. *INFOCOM. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, pages 608–617, 2002. 34

[CH03] Reuven Cohen and Shlomo Havlin. Scale-Free Networks Are Ultrasmall. *Physical Review Letters*, 90(5):5–8, February 2003. 42

[CHK+07] 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*, 104(27):11150–4, July 2007. 29, 34, 52, 61

[CHK+09] Kimberly Claffy, Young Hyun, Ken Keys, Marina Fomenkov, and Krioukov Dmitri. Internet Mapping : from Art to Science. *CAIDA Research Report*, 2009. 6

[CON] CONGAS EU Project - Dynamics and COevolution in Multi Level Strategic iNteraction GAmeS. http://www.congas-project.eu. 155

[CRT00] Jacques Cremer, Patrick Rey, and Jean Tirole. Connectivity in the Commercial Internet. *Journal of Industrial Economics*, 48(4):433–72, December 2000. 73

[CSW01] 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. 6, 20

[CT12] Gábor Csárdi and Nepusz Tamás. *iGraph Reference Manual*. 2012. 123, 149

[CZH07] Yan W. Chen, Lu F. Zhang, and Jian P. Huang. The Watts–Strogatz network model developed by including degree distribution: theory and computer simulation. *Journal of Physics A: Mathematical and Theoretical*, 40(29):8237–8246, July 2007. 43

[DD10] Amogh Dhamdhere and Constantine Dovrolis. The Internet is flat: modeling the transition from a transit hierarchy to a peering mesh. In *Proceedings of the 6th International COnference*, Co-NEXT. ACM, 2010. 18, 114

[DD11] Amogh Dhamdhere and Constantine Dovrolis. Twelve Years in the Evolution of the Internet Ecosystem. *IEEE/ACM Trans. Netw.*, 19(5):1420–1433, October 2011. 114

[Dim] DIMES project. http://www.netdimes.org. 5, 21

[DM03] Sergey N. Dorogovtsev and Jose Fernando F. Mendes. *Evolution of Networks: from Biological Nets to the Internet and WWW*. Oxford University Press, USA, 2003. 40

[DMS00] Sergey N. Dorogovtsev, Jose Fernando F. Mendes, and Alexander N. Samukhin. Structure of Growing Networks: Exact Solution of the Barabási–Albert's Model. *Arxiv preprint cond-mat/0004434*, 2000. 40

[ER60] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst Hung. Acas. Sci*, 5:17–60, 1960. 36

[Eur] European Internet Exchange Association. https://www.euro-ix.net/services. 19, 119

[fCRVW04] Shih fen Cheng, Daniel M. Reeves, Yevgeniy Vorobeychik, and Michael P. Wellman. Notes on equilibria in symmetric games. In *Proceedings of the 6th International Workshop GTDT*, pages 71–78, 2004. 88, 95, 110

[FFF99] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On Power-Law Relationships of the Internet Topology. *SIGCOMM Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, 1999. 32, 34

[FKP02] Alex Fabrikant, Elias Koutsoupias, and Christos H. Papadimitriou. Heuristically Optimized Trade-Offs: A New Paradigm for Power Laws in the Internet. In *Proceedings of the 29th ICALP '02*. Springer-Verlag, 2002. 45

[For10] Santo Fortunato. Community detection in graphs. *Physics Report*, 486(3-5):75–174, 2010. 29

[FS08] Dima Feldman and Yuval Shavitt. Automatic large scale generation of internet pop level maps. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pages 1–6. IEEE, 2008. 5

[Gao00] L. Gao. On inferring autonomous system relationships in the Internet. In *Global Telecommunications Conference, 2000. GLOBECOM'00. IEEE*, volume 1, 2000. 16, 23, 26

[Gil59] Edgar N. Gilbert. Random Graphs. *Annals of Mathematical Statistics*, 30:1141–1144, 1959. 36

[GIL+11] Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, and Luca Sani. BGP and Inter-AS Economic Relationships. *IFIP International Federation For Information Processing*, pages 54–67, 2011. 16, 23

[GIL+12] 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 2012 ACM conference on Internet measurement conference*, IMC '12, pages 253–264, New York, NY, USA, 2012. ACM. 114

[GILO11] 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, January 2011. 19, 72, 114

[GLM12] Enrico Gregori, Luciano Lenzini, and Simone Mainardi. Parallel k -Clique Community Detection on Large-Scale Networks. *IEEE transactions on parallel and distributed systems*, pages 1–11, 2012. 35, 131

[GLO11a] Enrico Gregori, Luciano Lenzini, and Chiara Orsini. k-clique Communities in the Internet AS-level Topology Graph. *2011 31st International Conference on Distributed Computing Systems Workshops*, pages 134–139, June 2011. 35

[GLO11b] Enrico Gregori, Luciano Lenzini, and Chiara Orsini. k-dense communities in the internet AS-level topology. *2011 Third International Conference on Communication Systems and Networks (COMSNETS 2011)*, pages 1–10, January 2011. 29, 34

[GMC10] B.H. Good, Y.A. De Montjoye, and A. Clauset. Performance of modularity maximization in practical contexts. *Physical Review E*, 81(4):046106, 2010. 29

[Goy07] S. Goyal. *Connections: an introduction to the economics of networks*. Princeton University Press, July 2007. 46

[GS06] 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. 5, 20

[GT00] 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. 5, 20

[HB96] J. Hawkinson and T. Bates. Guidelines for creation, selection, and registration of an autonomous system (as). RFC 1930 (Best Current Practice), March 1996. Updated by RFC 6996. 4

[HK02] Petter Holme and Byung-Ji Kim. Growing scale-free networks with tunable clustering. *Physical Review E*, 65(2):026107, 2002. 43

[HRI+08] Hamed Haddai, Miguel Rio, Gianluca Iannaccone, Andrew Moore, and Richard Mortier. Network Topologies: Inference, Modeling and Generation. *IEEE Communication Surveys*, 10(2):48–69, 2008. 34, 44, 47

[Ipl] University of Washington iPlane project. http://iplane.cs.washington.edu. 5, 21

[Iso] IIT-CNR Isolario project. http://www.isolario.it. 21, 54, 131, 138

[Jac10] M.O. Jackson. *Social and Economic Networks*. Princeton Univ.Press, 2010. 45, 46

[JHA12] T Jiménez, Y Hayel, and E Altman. Competition in access to content. *in Proceedings of IFIP Networking*, pages 211–222, 2012. 73, 91

[JW96] Matthew O. Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44 – 74, 1996. 44

[KBS09] Samir Khuller and Barna Barna Sah. On finding dense subgraphs. *In ICALP '09*, pages 597 – 608, 2009. 61

[KE02] Konstantin Klemm and Victor M. Eguiluz. Growing scale-free networks with small-world behavior. *Physical Review E*, 65(5):057102, 2002. 42

[Key10] Ken Keys. Internet-scale IP Alias Resolution Techniques. *ACM SIGCOMM Computer Communication Review*, 40(1):50–55, 2010. 5, 20

[KYLC] 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*. 5, 20

[LCC+09] 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. 14

[LDD12] Aemen Lodhi, Amogh Dhamdhere, and Constantine Dovrolis. GENESIS: An agent-based model of interdomain network formation, traffic flow and economics. *2012 Proceedings IEEE INFOCOM*, pages 1197–1205, March 2012. 45

[LJKL09] Yixiao Li, Xiaogang Jin, Fansheng Kong, and Jiming Li. Linking via Social Similarity : The Emergence of Community Structure in Scale-free Network. *Web Society. SWS '09. 1st IEEE Symposium on*, 2009. 44

[LKCS] Tian Lan, David Kao, Mung Chiang, and Ashutosh Sabharwal. An axiomatic theory of fairness in network resource allocation. INFOCOM '10. 90, 91

[LKF05] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over Time : Densification Laws , Shrinking. *Physical Review*, 2005. 43, 143

[LRMH10] Conrad Lee, Fergal Reid, Aaron Mcdaid, and Neil Hurley. Detecting Highly Overlapping Community Structure by Greedy Clique Expansion. *Workshop on Social Network Mining and Analysis*, 10, 2010. 30, 35

[LWY12] Miao Li, Hui Wang, and Jiahai Yang. Flattening and preferential attachment in the internet evolution. In *APNOMS*, pages 1–8. IEEE, 2012. 114

[Mat] Mathworks. Matlab - http://www.mathworks.it/products/matlab/. 86

[MDFL12] Murtaza Motiwala, Amogh Dhamdhere, Nick Feamster, and Anukool Lakhina. Towards a cost model for network traffic. *ACM SIGCOMM Comput. Commun. Rev.*, 42(1):54–60, January 2012. 79

[MHK+07] Priya Mahadevan, Calvin Hubble, Dmitri Krioukov, Bradley Huffaker, and Amin Vahdat. ORBIS: Rescaling Degree Correlations to Generate Annotated Internet Topologies. *SIGCOMM Comput. Commun. Rev.*, 37(4):325–336, August 2007. 46, 138, 145

[MJR+04] 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. 6, 20

[MLMB01] Alberto Medina, Anukool Lakhina, Ibrahim Matta, and John Byers. BRITE: Universal Topology Generation from a User's Perspective. Technical report, BU, 2001. 47, 138, 145

[MM65] J. W. Moon and L. Moser. On cliques in graphs. *Israel Journal of Mathematics*, 3(1):23 – 28, 1965. 58

[MMO13] Eli A. Meirom, Shie Mannor, and Ariel Orda. Formation Games and the Internet Structure. *CoRR*, abs/1307.4102, 2013. 46

[MRWK03] 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. 6, 20

[MUR] US MURI Project - Multi-Layer and Multi-resolution Networks ofInteracting Agents in Adversarial Environments. https://wiki.engr.illinois.edu/display/MuriHMAN/. 155

[MWA02] Ratul Mahajan, David Wetherall, and Tom Anderson. Understanding BGP Misconfiguration. *SIGCOMM Comput. Commun. Rev.*, 32(4):3–16, August 2002. 16

[N06] Economides N. The Economics of the Internet Backbone. *Handbook of Telecommunications Economics*, Majumar, S., Vogelsang, I., Cave, M. (Eds.):374–412, 2006. 73

[New01] Mark E. J. Newman. Clustering and preferential attachment in growing networks. *Physics Review E*, 64:1–13, 2001. 43, 143

[New03] Mark E. J. Newman. Random graphs as models of networks. *Handbook of Graphs and Networks: From the Genome to the Internet*, (1), 2003. 37, 43

[New04] Mark E. J. Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6):066133, 2004. 29

[Nor10a] William B. Norton. A Study of 28 Peering Policies. Technical report, Dr. Peering White Paper, 2010. 16, 22, 72

[Nor10b] William B. Norton. Internet Transit Prices - Historical and Projected. Technical report, Dr. Peering White Paper, 2010. 17, 72, 77, 78, 99

[Nor10c] William B. Norton. The Art of Peering - The IX Playbook. Technical report, Dr. Peering White Paper, 2010. 19, 72, 77, 87

[Nor11] William B. Norton. *The 2013 Internet Peering Playbook*. 2011. 14, 15, 16, 18, 20, 72, 78

[ORS93] Ariel Orda, Raphael Rom, and Nahum Shimkin. Competitive routing in multiuser communication networks. *IEEE/ACM Transactions on Networking*, 1(5):510–521, October 1993. 72, 77, 80

[OW10] Ricardo Oliveira and Walter Willinger. The (In)Completeness of the Observed Internet AS-level Structure. *IEEE/ACM Transactions on Networking*, 18(1):109–122, February 2010. 6, 22

[PCH] Packet Clearing House. http://www.pch.net. 21

[PDFV05] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–8, June 2005. 30, 35

[pee] PeeringDB - http://www.peeringdb.com. 17, 22, 23, 66

[Por] University of Pisa Portolan project. http://portolan.iet.unipi.it. 5, 21

[Pos81] Jon Postel. RFC 791 Internet Protocol - DARPA Inernet Programm, Protocol Specification, September 1981. 4

[PSV04] Romualdo Pastor-Satorras and Alessandro Vespignani. *Evolution and Structure of the Internet: a Statistical physics approach*. Cambridge Univ. Pr., 2004. 14, 23, 28, 33, 36, 44

[RFT13] Ryan A. Rossi, Sonia Fahmy, and Nilothpal Talukder. A Multi-Level Approach for Evaluating Internet Topology Generators. In *IFIP Networking*, pages 1–9, 2013. 28

[rip] Routing Information Service (RIS) at RIPE Network Coordination Centre. 21

[RLH06] Yakov Rekhter, Tony Li, and Susan Hares. RFC 4271 - A Border Gateway Protocol 4 (BGP-4), 2006. 6, 21

[Rou] University of Oregon Route Views Project. http://www.routeviews.org. 21

[RWM+11] 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. 7, 44, 67, 127

[SBS08] 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. 5, 20

[SCCH09] Huawei Shen, Xueqi Cheng, Kai Cai, and Mao-Bin Hu. Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications*, 388(8):1706–1712, April 2009. 30, 35

[SLH06] Lung-de Shyu, Seng-yong Lau, and Polly Huang. On the Search of Internet AS-level Topology Invariants. *Global Telecommunications Conference, 2006. GLOBECOM '06. IEEE*, 2006. 23

[SMW02] 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. 5

[SS06] Srinivas Shakkottai and R. Srikant. Economics of Network Pricing with Multiple ISPs. *IEEE/ACM Trans. Netw.*, 14(6), December 2006. 73

[STB09] S. Sesia, I. Toufik, and M. Baker. *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley InterScience online books. Wiley, 2009. 3

[STF06] 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. 52

[SYK09] Kazumi Saito, Takeshi Yamada, and Kazuhiro Kazama. The k-Dense Method to Extract Communities. *Studies in Computational Intelligence*, 165:243–257, 2009. 29, 34

[VÓ3] Alexei Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67(5):1–15, May 2003. 43

[VCDS74] Vincent Vinton Cerf, Yogen Dalal, and Carl Sunshine. RFC 675 - SPECIFICATION OF INTERNET TRANSMISSION CONTROL PROGRAM, 1974. 14

[WAD09] 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. 44

[WGJ+02] Walter Willinger, Ramesh Govindan, Sugih Jamin, Vern Paxson, and Scott Shenker. Scaling phenomena in the Internet: critically examining criticality. *Proceedings of the National Academy of Sciences of the United States of America*, 99 Suppl 1:2573–80, February 2002. 34

[Wika] Wikipedia - Network Formation. http://en.wikipedia.org/wiki/Network_formation. 36

[Wikb] Wikipedia - Tier 1 network. http://en.wikipedia.org/wiki/Tier_1_network. 16, 17

[WJ02] Jared Winick and Sugih Jamin. Inet-3.0: Internet Topology Generator. Technical report, University of Michigan, 2002. 47, 138, 145

[WL10] Xiaoming Wang and Dmitri Loguinov. Understanding and Modeling the Internet Topology : Economics and Evolution Perspective. *IEEE Transactions on Networking*, 18(1):257–270, 2010. 43

[WS98] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–2, June 1998. 30, 38

[WS04] Avishai Wool and Gonen Sagie. A clustering approach for exploring the Internet structure. *Proceedings of 23rd IEEE Convention of Electrical and Electronics Engineers in Israel*, pages 149 – 152, 2004. 61

[Yao95] David D. Yao. S-modular games, with queueing applications. *Queueing Systems*, 21(3-4):449–475, September 1995. 82

[ZCD97] Ellen W. Zegura, Kenneth Calvert, and M. Jeff Donahoo. A Quantitative Comparison of Graph-based Models for Internet Topology. *IEEE/ACM Transactions on Networking*, December 1997. 47, 138, 145

[ZLWX09] Cheng Zhang, Yanheng Liu, Jian Wang, and Minhui Xia. Modeling Router-Level Internet Topology. *2009 International Workshop on Chaos-Fractals Theories and Applications*, pages 331–335, November 2009. 6