Dissertação de Mestrado

An information-theoretic approach of network structure and dynamics

de **Cristopher Gabriel de Sousa Freitas**

orientado por
Prof. Dr. André Luiz Lins de Aquino

Universidade Federal de Alagoas
Instituto de Computação
Maceió, Alagoas
26 de Outubro de 2020

UNIVERSIDADE FEDERAL DE ALAGOAS
Instituto de Computação

AN INFORMATION-THEORETIC APPROACH OF NETWORK STRUCTURE AND DYNAMICS

Dissertação de Mestrado submetida ao Instituto de Computação da Universidade Federal de Alagoas como requisito parcial para a obtenção do grau de Mestre em Modelagem Computacional do Conhecimento.

Cristopher Gabriel de Sousa Freitas

Advisor: Prof. Dr. André Luiz Lins de Aquino

Banca avaliadora:

Fabiane da Silva Queiroz	Prof. Dr., UFAL
Rian Gabriel Santos Pinheiro	Prof. Dr., UFAL

Maceió, Alagoas
26 de Outubro de 2020

Catalogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 – 1767

F866i Freitas, Cristopher Gabriel de Sousa.
An information-theoretic approach of network structure and dynamics / Cristopher Gabriel de Sousa Freitas. – 2020.
44 f. : il.

Orientador: André Luiz Lins de Aquino.
Dissertação (mestrado em Modelagem Computacional do Conhecimento) – Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2020.

Texto em inglês.
Bibliografia: f. 39-44.

1. Redes de computadores. 2. Teoria da informação. 3. Redes complexas. I. Título.

CDU: 004.7

Folha de Aprovação

CRISTOPHER GABRIEL DE SOUSA FREITAS

An information-theoretic approach of network structure and dynamics

Dissertação submetida ao corpo docente do Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas e aprovada em 26 de OUTUBRO de 2020.

Prof. Dr. ANDRE LUIZ LINS DE AQUINO

Instituto de Computação - UFAL

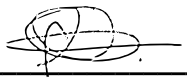
Orientador

Banca Examinadora:

Professor Dr. RIAN GABRIEL SANTOS PINHEIRO

Instituto de Computação - UFAL

Examinador interno



Professora Dra. FABIANE DA SILVA QUEIROZ

Centro de Ciências Agrárias - UFAL

Examinador externo

Happiness can be found, even in the darkest of times, if one only remembers to turn on the light. – Albus Dumbledore

Abstract

Modern networks are facing the most challenges in recent years. Life is changing, and the Internet became one of the essential services such as power, healthcare, banking. The Internet has opened global information, and nowadays, a stable and efficient network is a global right and requirement. Legacy networks inherited many concepts and infrastructure from circuit-switching networks, with inflexible design relying most on hardware. Another critical aspect of networking is debugging it, which requires specialized personnel with properly designed software. For these reasons – and currently increasing demand – Internet Service Providers are dealing with challenges that require innovation and scientific strategies. This scenario led to new paradigms such as software-defined networks (SDNs) to address most of the current issues by bringing the network logic from hardware design to software development and virtualization.

The emergence of SDNs has drawn the attention of network scientists and engineers to new roads. The Internet is the most extensive distributed system, and it relies upon distributed devices communicating efficiently through protocols. A computer network is a sophisticated collection of network devices and end-systems. SDNs allow a protocol-agnostic, centralized view, and flexible control of the network, favoring the development of new strategies that are not achievable into legacy IP networks. In this work, we study two main aspects of computer networks: structure and dynamics. By understanding the network characteristics and modeling its dynamical processes, we can uncover how network structure and dynamics affect its robustness. To achieve this understanding, we propose the usage of information-theory quantifiers for network characterization. For network topology, we introduce the Fisher Information Measure for quantifying the network characteristics, using alongside the Network Entropy in a bi-dimensional representation that allows us to identify if a network is closer to a *random*, *small-world* or *scale-free* topology.

We evaluated the traffic time-series using the Normalized Permutation Entropy and the Statistical Complexity for network traffic. We observe the traffic generation models based on heavy-tail distributions can not reproduce the actual traffic dynamics. We believe this understanding through information-theory quantifiers can help develop fault-management solutions and network automation. Instead of focusing on the massive amount of data available for networks, we can observe how the quantifiers describe the network behavior.

Keywords: Software-defined Networks, Information Theory, Complex Networks.

Resumo

As redes modernas estão enfrentando os maiores desafios nos últimos anos. A vida está mudando e a Internet se tornou um dos serviços essenciais, como energia, saúde, serviços bancários. A Internet abriu informações globais e, hoje em dia, uma rede estável e eficiente é um direito e um requisito global. As redes legadas herdaram muitos conceitos e infraestrutura das redes de comutação de circuitos, com um design inflexível que depende principalmente do hardware. Outro aspecto crítico da rede é a depuração, o que requer pessoal especializado com software projetado adequadamente. Por essas razões - e atual aumento de demanda - os Provedores de Serviços de Internet estão lidando com desafios que exigem inovação e estratégias científicas. Esse cenário levou a novos paradigmas, como redes definidas por software (SDNs) para resolver a maioria dos problemas atuais, trazendo a lógica de rede do *design* de hardware ao desenvolvimento de software e virtualização.

O surgimento de SDNs atraiu a atenção de cientistas e engenheiros de rede para novos caminhos. A Internet é o sistema distribuído mais extenso e depende de dispositivos distribuídos que se comunicam com eficiência por meio de protocolos. Uma rede de computadores é uma coleção sofisticada de dispositivos de rede e sistemas finais. As SDNs permitem uma visão centralizada e agnóstica de protocolo e controle flexível da rede, favorecendo o desenvolvimento de novas estratégias que não são alcançáveis em redes IP legadas. Neste trabalho, estudamos dois aspectos principais das redes de computadores: estrutura e dinâmica. Ao compreender as características da rede e ao modelar seus processos dinâmicos, podemos descobrir como a estrutura e a dinâmica da rede afetam sua robustez. Para alcançar esse entendimento, propomos o uso de quantificadores da teoria da informação para caracterização de redes. Para topologia de rede, apresentamos a Medida de Informação de Fisher para quantificar as características da rede, usando juntamente com a Entropia da Rede em uma representação bidimensional que nos permite identificar se uma rede está mais próxima de uma topologia *aleatória*, *small-world* ou *scale-free*.

Avaliamos as séries temporais de tráfego usando a Entropia de Permutação Normalizada e a Complexidade Estatística para tráfego de rede. Observamos que os modelos de geração de tráfego baseados em distribuições de cauda pesada não podem reproduzir a dinâmica real do tráfego. Acreditamos que esse entendimento por meio de quantificadores da teoria da informação pode ajudar a desenvolver soluções de gerenciamento de falhas e automação de rede. Em vez de focar na enorme quantidade de dados disponíveis para as redes, podemos observar como os quantificadores descrevem o comportamento da rede

Palavras-chave: Redes definidas por software, Teoria da Informação, Redes Complexas.

List of Figures

Contents

Publications

During the first year of the Master's degree, many subjects were under research and review. First, it was necessary a comprehensive study of Software-defined Networking, which comes in handy when developing and validating the techniques studied in a quasi-real scenario. Secondly, as many problems arise, we have used the knowledge from several research fields such as Complex Networks, Information Theory, and Machine Learning. Here it is a list of papers produced directly related to this proposal (chronological order):

1. Characterizing complex networks using Information Theory planes. Expanded abstract presented at: MEDYFINOL, Dec 2018, Santiago, Chile. XX Conference on Nonequilibrium Statistical Mechanics and Nonlinear Physics, 2018.

2. An intelligent fault tolerant architecture using Deep Reinforcement Learning. Poster presentation at: LANCOMM Student Workshop , May 2019, Gramado-RS, Brasil. Simpósio Brasileiro de Redes de Computadores (SBRC), 2019.

3. Characterizing complex networks using Information Theory quantifiers. Poster presentation at: LANET, Aug 2019, Cartagena, Colômbia. II Latin American Conference on Complex Networks, 2019.

4. A detailed characterization of complex networks using Information Theory. Published at Scientific Reports, v. 9, p. 16689, in Nov, 2019.

5. Towards self-healing SDNs for dynamic failures. Poster presentation at Khipu, Dec 2019, Montevideo, Uruguay. Latin American Meeting In Artificial Intelligence, 2019.

6. Mapping Network Traffic Dynamics in the Complexity-Entropy Plane. IEEE Symposium on Computers and Communications (ISCC) 2020.

Here it is a list of papers produced related to the general area studied and in collaboration with other researchers (chronological order):

1. Algoritmo para redução de dados em redes de sensores baseado em Teoria da Informação. Paper presented at: Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP) 2019.

2. Uma arquitetura de aprendizagem profunda para desagregação de energia (**best paper**). Paper presented at: Escola Regional de Computação Bahia, Alagoas e Sergipe (ERBASE) 2019.

3. A Deep Learning approach for Energy Disaggregation considering Embedded Devices. Paper presented at Brazilian Symposium on Computing Systems Engineering (SBESC) 2019.

Chapter 1

Introduction

Nature is full of interdependent components coupled together, acting as one big and complex system. Molecules and organisms are the most ancient example of the interconnected system observed in nature, inspiring scientists throughout the centuries. As humans evolved, society grew more complex and interconnected than ever. Independent people living together as one bigger system (i.e., cities) changed the design of living; before, each small settlement produced its own food, and it had its own habits, it was simpler to understand and control. A few travelers walked and rode in between cities; they were unusual people, distributing information about other cultures, and starting an irreversible interconnection process.

With the advent of better transportation techniques, humans started to travel faster, and the system started to become more complicated than any royalty or governor was able to understand. Such challenges guided the curiosity of a few, which started to design tools and frameworks in a search to understand how these complex systems behaved. One of the earliest well-formulated problems of this nature is the *traveling salesman problem* (TSP) from around 1832 with no mathematical formulation. The TSP raised the following question: "Given a list of cities and the distances between them, what is the shortest possible route that visits each city *once*, and returns to the origin city?"

Another critical problem in networking history is the "small-world problem". Travers and Milgram [44] discuss the occasional experience of meeting someone we share a mutual acquaintance, raising a question: "Starting with any two people in the world, what is the probability that they will know each other?" In a small-world network, two random nodes can reach out to each other in a relatively small number of hops, i.e., a small-world network has a relatively small average path length (see Sec. 3.4.2). Nevertheless, communication amongst different societies amazes social psychologists, communication specialists, and political scientists. There is a clear impact when two isolated communities connect, *it changes the structure and dynamics of society* [51]

The search to improve the communication in society – lowering the average path length information takes to travel – led the development of the most extensive networked system

ever engineered: the Internet [33, Ch.1]. The Internet connects billions of devices through communication links and switches. Similarly, as humans abide by laws maintaining a well-functioning society, currently, Internet devices must respect communication protocols and design principles to "survive".

On the Internet, hosts or end-systems do not communicate directly with each other; they send their data through communication links and packet switches. There are many types of communication links, each different link transfer data at different transmission rate measured in bits/second. When an end-system sends data to another system, it segments its data into smaller pieces called *packets*. The route that these packets will take, it will depend upon the packet switches; these packet switches can be link-layer switches (Layer-2 switches) or routers (Layer-3 switches). Usually, routers are in the core of the network, while link-layer switches are "closer" to the end-systems; the end-systems access the Internet through these collections of packet switches and communications links that establish the Internet Service Providers (ISPs). Each ISP runs independently conforming to Internet protocols such as IP (Internet Protocol), BGP (Border-Gateway Protocol), and MPLS (Multi-Protocol Label Switching).

Into the Internet, the World Wide Web (WWW) was born. In a similar manner to the Internet, the WWW relied on protocols such as the Hypertext Transfer Protocol (HTTP) to function accordingly. In the 1990s, the WWW was a collection of documents (webpages) available on the Internet and connected through virtual links pointing to other documents. Before Google, there was no easy way to map the WWW content. Like the Internet, researchers assumed the WWW had random structure until Barabási and Albert [2] saw it differently. Using web crawlers, they mapped the WWW structure and observed that, instead of a random network, they obtained something differently. According to the studies of Random Graph Theory inspired in the Erdős and Rényi model [22], a random graph presents a *degree distribution* (histogram of the node degrees) resembling a Poisson distribution. At the same time, Barabási and Albert observed a degree distribution that presented power-law characteristics on its tail. Similarly, Faloutsos et al. [24] also observed a power-law behavior in the degree distribution of the Internet. Observing that this was no random discovery, Barabási and Albert went further and proposed a model [10] that better described the "scale-free networks" (i.e., networks that degree distributions resemble a power-law), considering growth and preferential attachment (see Sec. 3.4.3).

The discovery that the Internet and WWW were scale-free networks changed the perspective when evaluating these network characteristics, and started a brand-new research field of *network science*. As the Internet grew and became significantly crucial in our society, there were severe concerns about the robustness and resilience of these networks. Barabási and Albert observed that scale-free networks were robust to random attacks – removing random nodes – while they were susceptible to target attacks in the hubs, quickly dissolving the connectivity of its end-systems [3].

Nonetheless, with the development of network science, it became clear that it was necessary an understanding of the network structure to design reliable networks as if some node presents a failure, its load shifts to other nodes, stressing the network behavior; this applies to computer networks, transportation, power-grids, and others. By understanding the network characteristics and modeling its dynamical processes, we can uncover how network structure and dynamics affect the robustness of the system. Therefore, in this work, one of our several contributions is to *carry an extensive analysis of synthetic network topologies, proposing a simple method to characterize network topologies (e.g., scale-free, small-world, or random) using Information Theory quantifiers.*

In our scenario, when discussing the network dynamics, we mainly refer to traffic, as the only dynamical process evaluated. Usually, network dynamics can refer to mobility, depending upon the nature of the system; this is not our case, since the focus is on the backbone of ISPs, and generally, they have static topologies.

Concerning network traffic dynamics, it was a similar situation to what we saw in network topologies. Initially, the researchers considered network traffic random, following a Poisson distribution. However, Leland et al. [37] collected several days of network traffic and observed a self-similar nature in Ethernet traffic. Self-similarity describes a phenomenon where some stochastic processes present high-variability and long-term correlations, usually captured by a heavy-tailed distribution or power-laws (see Sec. 4.1). This characteristic uncovered the idea that network traffic presents patterns, and these patterns can be classified and predicted. Lakhina et al. [35] classified general network traffic in three categories: (i) long periodic volumes; (ii) bursty events; and (iii) noise. Considering these findings, Nucci et al. [47] present traffic generation models considering heavy-tailed probability distributions. Thus, in this work, another of our contributions *is to evaluate if the existent traffic models reproduce the dynamics of real network traffic.*

At this point, it is clear that although computer networks became more complex than ever, they are far from random. Assessing network data, we can observe distinct patterns, from structure to dynamics. These patterns must be taken into consideration when planning, designing and deploying a network. However, what if we only identify the problem after the network is already deployed? Can we optimize an operational network running thousand of critical services? This scenario represents the actual one for Internet Service Providers.

With the Internet dynamics changing quickly, it becomes costly to maintain a critical network that generally consists of proprietary hardware struggling to attend new demands. Traditional IP networks consist of several protocols coupled together into a circuit or "black-box" software. This scenario led to the development of Software-Defined Networks (SDNs) and the OpenFlow protocol; SDNs enable quick automation of network management and a centralized view of the network state. Therefore, with the SDNs, ISPs would be able to gather their data efficiently and optimize its functioning through

software APIs. The possibility of having centralized management (i.e., a controller) of the network, has set a new trend for ISPs: Network automation through Machine Learning (ML) and Artificial Intelligence (AI). Using data from past events and continually collecting information about the network state, we can characterize the network features, create models and even use trial-and-errors methods as we can design flexible software solutions to optimize the network.

Overall, it is essential that ISPs store their data, providing a comprehensive evaluation of their structure and dynamics, and automate as many operations as possible. Otherwise, they will fall behind, losing vital resources such as time and money in a very competitive industry. Failures may happen in many levels of the system's architecture, and the development of a comprehensive fault management solution is challenging [67]. Our contributions rely on developing an information-theoretic characterization of network topologies and traffic. It is our understanding that such characterization can guide future developments of network automation solutions. Computer networks are complex systems that generates a lot of data, and information-theoretic approaches give us the possibility of lowering this burden, and focusing on the network underlying characteristics.

We tackle the network problems in two distinct aspects: *structure* and *dynamics*. For this reason, we structure this thesis proposal as follows. Chapter 2 enlighten our discussion by presenting the state-of-the-art when dealing with computer networks and software-defined networks. Chapter 3 evaluates distinct network structures and topologies, proposing an information-theoretic approach to simplify the discussion on network characteristics. Our results enhance the discussion on the characterization of complex networks in general and may inspire the development of future optimization solutions for network control. Chapter 4 evaluates the characteristics of the traffic models in the Complexity-Entropy plane, which may further help us identify traffic patterns. Chapter 5 wraps our discussion and explores the future directions for our work.

Chapter 2

State of the art

Before diving into our development, it is critical that we discuss the state-of-the-art in each of the fields we will discuss as follows. Considering the Complex Networks domain, we discuss the characterization of network topologies in Section 2.1, and its applications to the domain of Computer Networks. In Section 2.2, we discuss the most classical discoveries of traffic characterization in Computer Networks, and the current trends of applications.

2.1 Characterization of network topologies

As discussed in Chapter 1, the problems of network characterization is the key to other aspects such as information spreading [50] and fault tolerance [3]. When identifying the network's structural characteristics, we can predict how the network will react to distinct processes. Therefore, for computer networks and the Internet, it is of huge importance to know how the structure will react when unexpected events happen.

The process of characterizing a network can be long and tedious, there are many metrics to quantify distinct properties of the network, and for each class (e.g., random), there are many considerations. Zanin et al. [68] discuss a long process of data mining and how to extract meaningful information from network datasets. In this way, many researchers attempt to simplify the characterization process by using "universal" metrics. Humphries et al. [30] proposed the so-called *small-world-ness* measure to distinguish if a network is small-world or not. Schieber et al. [54] proposed a distance measure that compares two distinct networks, giving the likelihood that two distinct networks present similar characteristics. Wiedermann et al. [65] proposed the usage of Complexity-Entropy plane for complex networks, by adapting the idea behind network entropy [16] to develop the Statistical Complexity quantifier. Many other studies characterize a collection of networks, in the attempt to develop a broader characterization [31, 19, 69].

Based on existing methods for network characterization, there are several studies to characterize Internet-like topologies. Siganos et al. [56] evaluated the presence of power-

law degree distribution in distinct router topologies, extending the studies developed in [24]. Canbaz et al. [18] enlighten one of the biggest problems in the characterization of the Internet, which is the measurement process. Bakhshaliyev et al. [6] went further and studies the usual paths that many end-to-end services take in intra-domain networks, observing that around 78% of the paths does not use the shortest path only, indicating load balancing and rerouting strategies. More recent, Canbaz et al. [17] did a thorough evaluation of Internet-like topologies and proposed a tool for mining Internet data, both topology, and traffic.

2.2 Characterization of network traffic

Different from network topology, the characterization of network traffic is much more common, and intersects with the studies of network tomography. As we introduced in Chapter 1, Leland et al. [37] and Willinger et al. [66] changed the perspective when evaluating network traffic. There was much more effort to develop a successful model of network traffic. Garret and Willinger [27] discussed the usage of fractional Brownian motion to model the self-similarity aspects of network traffic, creating a traffic generation approach. Grossglauser et al. [29] proposed an attractive fluid traffic model whose parameters are controllable, and its resulting process is self-similar. Most of these models require a sophisticated deep understanding of stochastic processes and their analytical properties.

Duffield et al. [21] discussed the existent dependence between the traffic generation model and the quality of service observed in the queueing systems; the outcomes may vary according to the model used, resulting in a biased analysis. Furthermore, Bolotin et al. [14] proposed strategies to control the network functioning, considering the long-range dependence and self-similar nature of network traffic. They use ON/OFF sources and heavy-tailed distributions to generate synthetic traffic, identifying critical points, and preparing the network according to the model prediction.

Even though [27] and [29] proposed better models to reproduce network traffic. Researchers continue to use ON/OFF sources based on heavy-tailed distributions due to its simpler formulation [47, 61, 11, 60]. Therefore, considering these studies, we conduct a characterization of these models in Chapter 4, and we propose a straightforward analysis using the Complexity-Entropy plane to provide a better overview of synthetic and real traffic dynamics.

Chapter 3

Characterizing network topologies

We propose the Shannon-Fisher plane to a simple and efficient characterization of complex networks compared with the standard models of the literature [25]. Using both the Network Entropy and the Fisher Information Measure, we can identify several clusters of distinct network properties such as *small-world*, *scale-free*, and *randomness*.

Nonetheless, the extent of our contributions in characterizing networks is the following:

- The proposal of the Shannon-Fisher plane for network characterization.
- The impact of matrix reordering in the Fisher Information Measure
- A comprehensive evaluation of many complex network models

3.1 Notation and definitions

A graph can take many forms depending upon the system being represented. The most common representation is the *simple* graph $G(V, E)$, consisting of nodes $v_i \in V$, and edges $e_{uv} : (u, v) \in E$ with no self-loops $(u, u) \notin E$, no direction $(u, v) = (v, u)$, and neither weights on its edges. These properties facilitate when evaluating a network; but, it also limits the power of generalization when analyzing the system's underlying dynamics. Generally, simple graphs are more useful when we want to understand the system's structure, i.e., how the components arrange themselves in space.

In this thesis, we will generally represent a graph as adjacency matrix $\boldsymbol{A}$ with $N \times N$ dimensions, where $|V| = N$ is the number of nodes, and $|E| = M$ is the number of edges. For a simple graph, each (i, j) element takes $A_{i,j} = 1$ if a connection exists between nodes i and j, otherwise, $A_{i,j} = 0$. The absence of self-loops is represented as $A_{i,i} = 0$, $\forall i = 0, \ldots, N-1$; and, for undirected graphs $\boldsymbol{A} = \boldsymbol{A}^T$.

The node degree k_i indicates the amount of connections node i contains. Considering the adjacency matrix $\boldsymbol{A}$, we can calculate the node degree $k_i = \sum_j A_{i,j}$, therefore, $0 \leq k_i \leq N-1$. If a node i has degree $k_i = 0$, we say this is an isolated node; if $k_i = N-1$, we

have a fully connected node. The number of connections of a node may be an indicator of its relevance. In order to measure how the number of connections is distributed for a network, we define the degree distribution $P(k)$. The degree distribution measures the fraction of nodes with degree k in the network, i.e., $P(k) = N_k/N$, where N_k indicates the number of nodes with degree k.

Sometimes, it is useful to refer to sparse or dense graphs using a measure of link density. For simple graphs, the link density is the fraction of existing links (i.e., $\sum_i^N k_i = 2M$) over the maximum possible links (i.e., $N(N-1)/2$):

$$\xi = \frac{1}{N(N-1)} \sum_i^N k_i = \frac{2M}{N(N-1)}. \tag{3.1}$$

The clustering coefficient measures the presence of clusters in a networks. As discussed earlier, this is an important property of *small-world* networks, where the nodes tend to cluster easily. The clustering coefficient can be measured locally or globally. The local clustering coefficient quantifies how the neighborhood of a node compares to a *clique*. A clique is a subset of vertices $\mathcal{C} \subseteq V$ where every two distinct vertices are adjacent. Nevertheless, the local clustering coefficient for an undirected graph is defined as the fraction of nodes in the neighborhood $N_i = \{v_j : e_{ij} \in \mathcal{E} \vee e_{ji} \in E\}$ over the maximum numer of links in the neighborhood for an undirected graph (i.e., $k_i(k_i-1)/2$):

$$C_i = \frac{2|e_{jk} : (v_j, v_k) \in N_i, e_{jk} \in \mathcal{E}|}{k_i(k_i-1)} \tag{3.2}$$

Another understanding of the clustering coefficient is counting the number of triangles in the neighborhood, compared to the number of pairs in a three-node neighborhood subset, i.e., a triplet. In this sense, the local clustering coefficient can be defined as

$$C_i = \frac{\lambda_G(v)}{\tau_G(v)}, \tag{3.3}$$

where $\lambda_G(v)$ is the number of triangles on $v \in \mathcal{V}$, and $\tau_G(v)$ is the number of connected pairs in a triplet.

Finally, both definitions are the same, but they can be used differently for calculating the global clustering coefficient. The global clustering coefficient is also known as transitivity, defined as

$$C^{\Delta} = \frac{3 \times \text{number of closed triplets}}{\text{number of all triplets}}. \tag{3.4}$$

Another alternative of the global clustering coefficient is based on averaging the clus-

tering for single nodes. Then, the clustering for the whole network is the average

$$\bar{C} = \frac{1}{N} \sum_{i=1}^{N} C_i. \tag{3.5}$$

3.2 Network Entropy

Network Entropy is based on the classical Shannon Entropy for discrete distributions. Burda et al. [16] proposed a measure of Network Entropy based on the probability that a random walker goes from node i to any other node j. This probability distribution $P^{(i)}$ is defined for each node i and has entries

$$p_{i \to j} = \begin{cases} 0, & \text{for } a_{ij} = 0, \\ 1/k_i, & \text{for } a_{ij} = 1. \end{cases} \tag{3.6}$$

It is easy to observe that $\sum_j p_{i \to j} = 1$ for each node i.

Based on the probability distribution $P^{(i)}$, the entropy for each node can be defined as

$$\mathcal{S}^{(i)} \equiv \mathcal{S}[P^{(i)}] = -\sum_{j=1}^{N-1} p_{i \to j} \ln p_{i \to j} = \ln k_i. \tag{3.7}$$

with $\mathcal{S}^{(i)} = 0$ if node i is disconnected.

After calculating the entropy for each node, we then calculate the normalized node entropy by

$$\mathcal{H}^{(i)} = \frac{\mathcal{S}[P^{(i)}]}{\ln(N-1)} = \frac{\ln k_i}{\ln(N-1)}. \tag{3.8}$$

Finally, the *normalized Network Entropy* is calculated averaging the normalized node entropy over the whole network as

$$\mathcal{H} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{H}^{(i)} = \frac{1}{N \ln(N-1)} \sum_{i=1}^{N} \ln k_i. \tag{3.9}$$

The normalized Network Entropy is maximal $\mathcal{H} = 1$ for fully connected networks, since $p_{i \to j} = (N-1)^{-1}$ for every $i \neq j$ and the walk becomes fully random, i.e., jumps from node i any other node j are equiprobable. The walk becomes predictable in a sparse network because it limits the possibility of jumps. The sparser the network, the lower becomes its Network Entropy.

The normalized Network Entropy $\mathcal{H}$, hence, quantifies the heterogeneity of the network's degree distribution, with lower values for nodes with lower degrees and higher values for nodes with higher degrees. For example, peripheral nodes present lower $\mathcal{H}^{(i)}$ than hubs. Entropy, thus, ranges from $\mathcal{H} \to 0$ (sparse networks) to $\mathcal{H} \to 1$ (fully con-

nected networks).

3.3 Network Fisher Information Measure

The *normalized Fisher Information Measure (FIM)* [53] for a node i is given by

$$\mathcal{F}^{(i)}[P^{(i)}] = \frac{1}{2}\sum_{j=1}^{N-1}\left[\sqrt{p_{i\to j+1}} - \sqrt{p_{i\to j}}\right]^2 \ . \tag{3.10}$$

The normalized *network Fisher Information Measure* is given by

$$\mathcal{F} = \frac{1}{N}\sum_{i}\mathcal{F}^{(i)}[P^{(i)}] \ . \tag{3.11}$$

If the system under study is in a very ordered state, i.e., a sparse network, almost all $p_{i\to j}$ values are zeros, we have Shannon Entropy $\mathcal{H} \to 0$ and normalized Fisher's Information Measure $\mathcal{F} \to 1$. On the other hand, when a very disordered state represents the system under study, that is when all $p_{i\to j}$ values are similar, we obtain $\mathcal{H} \to 1$ and $\mathcal{F} \to 0$. We can then define a Shannon-Fisher plane, which can also be used to characterize Complex Networks.

However, a critical problem emerged, and it drifted our focus in this approach. The Fisher Information Measure considers the node indices to build the random-walk-based distribution, if we permute these indices without altering the network structure, our results for Network Fisher may change (see Fig. 3.1). This dependency is problematic when using our proposal. Nonetheless, we have found an adequate solution that enhances our results for both synthetic and real networks; and does not compromise the initial discussion, only improves it, but it comes with an additional computational cost.

3.3.1 Matrix reordering problem

As the Fisher Information Measure is sensitive to the ordering of node labels, we want to find an order φ that maximizes the amount of information that we can extract from the system using the adjacency matrix $\mathbf{A}$, i.e., an optimal representation $\mathbf{A}^*$ that it can reveal the most patterns if they do exist.

An ordering, or order, is a bijection $\varphi(v) \to i$ from $v \in V$ to $i \in N = \{1, \ldots, n\}$ that associates a unique index to each vertex. We denote one specific ordering from the set of all possible orderings as φ^*. Usually, a network comes with an arbitrary ordering that we call *initial order*, denoted $\varphi_0(v)$ to distinguish from a computed order. A *transformation* from one ordering to another is called a permutation $\boldsymbol{\pi}$. Formally, a permutation is a

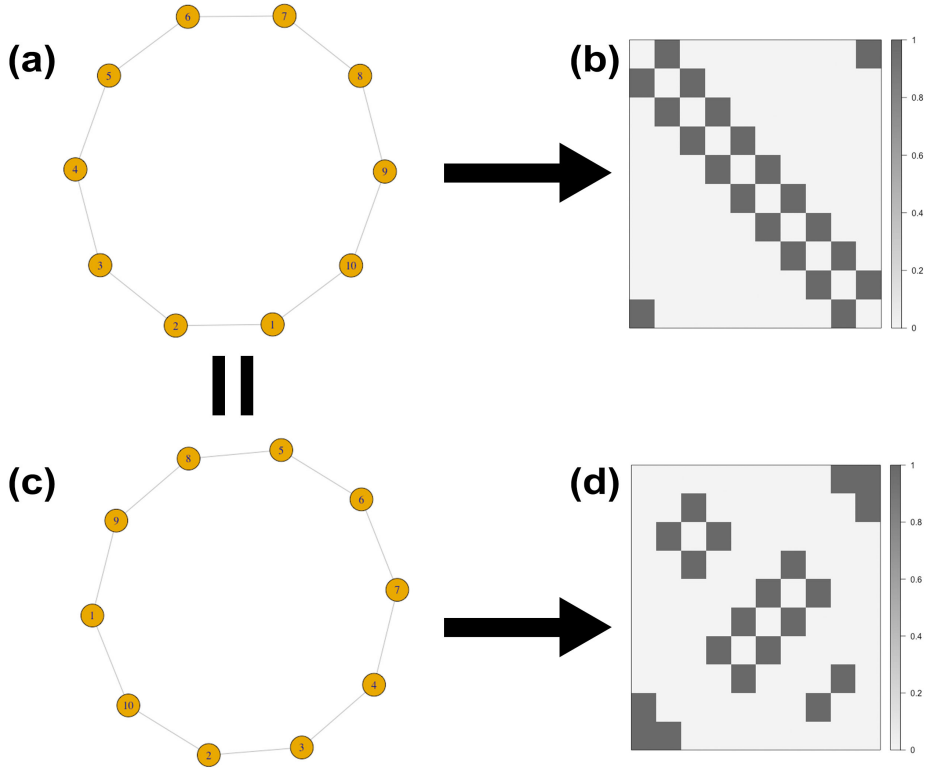


Figure 3.1: (a) shows the 1-ring topology with $N = 10$ and original order of the lattice. (a) shows a very ordered adjacency matrix for 1-ring with ordered labels. (b) shows the 1-ring topology with $N = 10$ after randomly permuting the node labels. (c) shows the resulting matrix after shuffling the node labels for the 1-ring topology.

bijection $\boldsymbol{\pi}(x) \to y$ such that:

$$\boldsymbol{\pi}(x_i) = y_i, \ (x, y) \in N^2 \text{ where } y_i = y_j \implies i = j. \tag{3.12}$$

Each permutation is implemented as a vector containing n distinct indices in N. We denote S the set of all possible permutations $n!$ for n. A reordering of an undirected network G consists in computing one permutation $\boldsymbol{\pi} \in S$ that maximizes or minimizes an objective function $q(\boldsymbol{\pi}, G)$, such that:

$$\arg\min_{\boldsymbol{\pi} \in S} q(\boldsymbol{\pi}, G). \tag{3.13}$$

For each permutation $\boldsymbol{\pi}$, we may have a different value of Network Fisher Information Measure $\mathcal{I}(\boldsymbol{\pi}, G) \in \mathcal{F}$, where $\mathcal{F}$ is the set of all possible FIM for a given network G and permutation $\boldsymbol{\pi}$.

From our previous results [25], we observe that there exists a pattern of transitions between k-ring and random networks, where k-ring is the most ordered matrices with

block-diagonal patterns and the lowest FIM values, except for $k = 1$. Therefore, when the number of connections increases, the adjacency matrix $\mathbf{A}$ starts to saturate with ones, and the Network Fisher decreases. Thus, we choose a permutation $\boldsymbol{\pi}$ that minimizes the FIM $\mathcal{I}$ for a given network G, such that:

$$\boldsymbol{\pi}^* = \arg\min_{\boldsymbol{\pi} \in S} \mathcal{I}(\boldsymbol{\pi}, G). \tag{3.14}$$

We denote our ideal FIM as $\mathcal{I}^* \equiv \mathcal{I}(\pi^*, G)$, as $\boldsymbol{\pi}^*$ is the optimal permutation for our problem. Now that we know what to find, we need to define how to do it. Finding the best possible solution for our problem is immediate if we run all the possible permutations, and we choose the one with the lowest possible value of FIM. However, this is not feasible, as we have $n!$ permutations for each undirected network G. For this, there are several algorithms for matrix reordering or *seriation*.

The *block-diagonal pattern* is one of the most sought-after matrix patterns. It consists of coherent rectangular areas that appear in ordered matrix whenever strongly connected components or cliques are present in the underlying topology. Initially, we are focused on finding the best possible solution, and for this task, the *Optimal-Leaf-Ordering* is the best algorithm as it finds an exact solution [15, 8]. However, it is the most expensive technique with a time of complexity of $\mathcal{O}(n^2 \log(n))$ and memory complexity of $\mathcal{O}(n)$.

Thus, we use this exact solution for $N < 10000$, but for $N > 10000$, we chose a sub-optimal algorithm that focuses on the angular order of eigenvectors [26]. Although the exact solution (obviously) performs better than the sub-optimal algorithm, the sub-optimal algorithm produces a very consistent result. It is our understanding that this algorithm gives us more information than the natural ordering of the system, then, enhancing our results.

3.4 Results: Synthetic Networks

We analyze the behavior of Information Theory quantifiers when applied to Random (RN), Small World (SWN), and Scale-Free networks (SFN). We simulated independent instances of these networks for several parameters and then analyzed how their Network Entropy and Fisher Information Measure vary. These synthetic networks may present some degree of *stochasticity* related to its parameters setting, which results in variations of the quantifiers; for this, when we observe variations in any measure $\mathcal{X}$, we represent it by its average value $\overline{\mathcal{X}}$ along with its sampling standard deviation $s_{\mathcal{X}}$. These variations, when too small, can be hard to distinguish in figures, but their numerical results should make this clearer.

3.4.1 Random Networks: Erdős-Rényi

Boccaletti et al. [13] state that: "*the term random graph refers to the disordered nature of the arrangement of links between different nodes.*" According to Ref. [46], Solomonoff and Rapoport [57] initiated the study upon the nature of random graphs, but Erdős and Rényi [22] are most known by observing the properties of networks as they increase the number of random connections, thus, defining an ensemble of graphs $G(N, M)$, with N nodes and M edges. Later, Gilbert [28] described an alternative method for generating random graphs by defining an ensemble of graphs $G_{N,p}$ with N nodes connecting randomly according to a *linking probability* p that is analogous to the *link density*.

Although Erdős-Rényi (ER) random graphs are well studied in network science, they often fail at describing the essential properties of real networks. An important aspect of ER graphs, is the critical point identifying the emergence of a giant component in a random network $p_c = \ln N/N$. This property is derived analytically from the $G_{N,p}$ ensemble. For a graph in $G_{N,p}$, the distribution of the degree k of any particular node v is binomial:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}, \tag{3.15}$$

where N is the number of nodes in the graph, and p is the linking probability. For a large N, and $Np =$ constant, this resembles a Poisson distribution. Nonetheless,

$$P(k) \to \frac{(np)^k e^{-np}}{k!} \text{as } n \to \infty \text{ and } np = \text{constant}. \tag{3.16}$$

Erdős and Rényi [23] described the behavior of $G_{N,p}$ for distinct values of p. They observed the following:

- If $Np < 1$, then $G_{N,p}$ is very *unlikely* to have any connected components of size larger than $O(\ln(N))$.
- If $Np = 1$, then $G_{N,p}$ is very *likely* to have a largest component whose size is of order $N^{2/3}$.
- If $Np \to c > 1$, where c is a constant, then $G_{N,p}$ is likely to have a unique *giant component*, and no other component will have more than $O(\ln N)$ vertices.
- If $p < \frac{(1-\epsilon)\ln N}{N}$, then $G_{N,p}$ is very likely to contain isolated nodes.
- If $p > \frac{(1+\epsilon)\ln N}{N}$, then $G_{N,p}$ is very likely to be connected.

Thus, $\ln N/N$ is considered a critical threshold for the connectedness of $G_{N,p}$. Other properties of the Erdős-Rényi model are described more precisely in the asymptotic limits of $N \to \infty$.

We analyzed fifty independent ER graphs $G_{N,p}$ for every combination of $N = \{50, 1000, 10000\}$ and $p \in \{0, 0.001, 0.002, \ldots, 0.99, 1\}$ making, thus, a total of $50 \times 3 \times 1001$ graphs. Figure 3.2 shows the variation of the Shannon Entropy (Fig. 3.2a) and Fisher Information Measure (Fig. 3.2b) with respect to the link density, while Figure 3.2c depicts the relationship in between the Shannon Entropy and Fisher Information Measure.

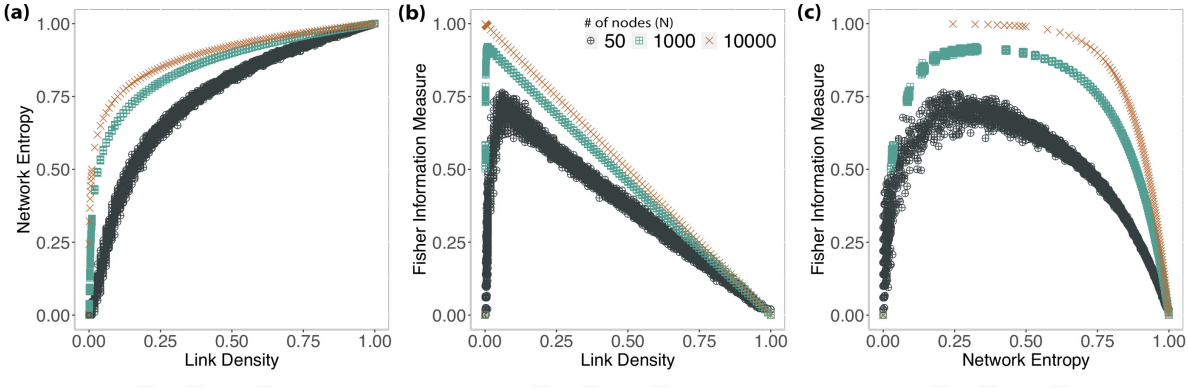


Figure 3.2: Results showing the relationship of Shannon Entropy and Fisher Information Measure with link density (**(a)** and **(b)**), and between Fisher Information Measure and Shannon Entropy **(c)** for 50 independent Erdős-Rényi networks except when $N = 10000$. The dark-green circles correspond to $N = 50$; the light-green squares to $N = 1000$; and the orange crosses to $N = 10000$.

Figure 3.2a shows how the Shannon Entropy $\mathcal{H}$ varies with respect to the link density ξ. The variation starts steep, then saturates. This may enhance the sensibility of Shannon-Fisher plane for sparse networks, but it may not be sensitive to denser graphs. The relationship between $\mathcal{H}$ and ξ also depends on the number of nodes N. The Shannon Entropy increases, for the same link density, with N. However, the rate of this growth decreases with N.

Figure 3.2b suggests that the Fisher Information Measure presents two distinct regimes for ER graphs as a function of their link density. Initially, this measure grows steadily: for $p = 0$ the network starts is totally disconnected; as p increases, it reaches the critical threshold p_c that is relative to the network number of nodes N, after which the measure decreases in a quasi-linear fashion $\mathcal{I}^* \approx b - ap$, where a and b are constants that changes according to the network size N. For $N = 50$, the critical point is $\overline{p_c} = 0.06$ with standard deviation $s_{p_c} = 0.02$ and $\overline{\mathcal{I}^*} = 0.70$, $s_{\mathcal{I}^*} = 0.04$; for $N = 1000$, $\overline{p_c} = 0.01$, $s_{p_c} = 0.002$ with $\overline{\mathcal{I}^*} = 0.912$, $s_{\mathcal{I}^*} = 0.003$; for $N = 10000$, $\overline{p_c} = 0.001$ with $\overline{\mathcal{I}^*} = 0.999$ and no variation observed. As the linking probability for ER graphs is analogous to the link density, this also stands for the link density, so $\overline{\mathcal{I}^*} \approx b - a\xi$ for every $\xi > \xi_c$. This result relates to the expected phase transitions in random graphs at $p_c > \ln N/N$ [23], as the network will almost surely be connected.

Figuree 3.2c shows the relationship between the Shannon Entropy and the Fisher Information Measure. As expected, the larger the network is, the less the variability

observed. For this reason, we will only present results for $N = 1000$ hereinafter.

3.4.2 Small-world Networks: Watts-Strogatz

While the ER graphs have clearly established properties, it often does not account for real networks behavior. They do not generate local clustering, lacking the presence of triplets, inducing a low clustering coefficient. Even though, in general, a ER network has relatively low average path length in order of $L \sim O(\ln N)$.

In order to address this problem, Watts and Strogatz [64] (WS) proposed a model to build graphs $G_{N,k}$ that can reproduce this *small-world* property with a high clustering coefficient. Small-world networks present an intrinsic characteristic of having a relatively small average path length between nodes [59]. Starting with a k-ring network with N nodes and a probability β. The rewiring consists of removing existing edges and connecting to another random node. When $\beta = 0$, we have a ring lattice, and for $\beta = 1$, it produces a random graph. The lattice structure of the model when $\beta = 0$ produces a locally clustered network, while the rewiring mechanism reduces the average path length by creating shortcuts. The algorithm introduces about $\beta Nk/2$ of non-lattice edges. Even though produces a random graph it does not actually approach the ER model as every node will be connected to at least $k/2$ vertices when $\beta = 1$.

For a ring lattice, the average path length $L = N/2k$. In the limiting case of $\beta \to 1$, the graph approaches a random graph with $L = \ln N / \ln k$. In the region of $0 < \beta < 1$, the average path length diminishes quickly when increasing β.

Similarly, the clustering coefficient to the ring lattice $C = 3(k-2)/4(k-1)$ as k grows. When $\beta \to 1$, the clustering coefficient is of the same order as the classical random graphs, $C = k/(N-1)$. For intermediate values, the model produces networks with the *small-world* property and a nontrivial clustering coefficient [13].

In the attempt to perform quantitative analysis for clustering coefficient in small-world networks, and considering a relationship with the average path length, Humphries et al. [30] proposed the small-world-ness S^{Δ}, defined as

$$S^{\Delta} = \frac{C^{\Delta}/C^{\Delta}_{\text{rand}}}{L/L_{\text{rand}}}, \tag{3.17}$$

where C^{Δ} and L are, respectively, the clustering coefficient (transitivity) and average path length, and C^{Δ}_{rand} and L_{rand} are the results computed for an ensemble of 100 ER networks, simulated with the same link density ξ as the real network. With this approach, Humphries et al. [30] state that for $S^{\Delta} > 1$, the network can be considered small-world.

Here, we evaluate the behavior of the Watts-Strogatz model with the Shannon-Fisher plane, in order to understand how this model varies according to its parameters. Figure 3.3 shows the same analysis as presented previously for random networks. Figure 3.3a shows

that the relationship between Network Entropy and link density is consistent with the observed in Random Networks and that there is a little variation concerning β. Therefore, the Network Entropy $\mathcal{H}$ by itself does not provide information to identify different Small-World models.

Figure 3.3b shows the relationship between the Fisher measure and link density, indexed by the rewiring probability β (shades of blue). As expected, the behavior in the limit $\beta = 1$ is the same (linear decay) as the one observed for RN; cf. Figure 3.3b. There is a lower bound, which corresponds to k-rings (red dots). We see in Figure 3.3 that for $k = 1$, we start with a ring lattice where $\mathcal{H} = 0.1$, $\mathcal{I}^* = 0.5$, and increasing β, we see $\mathcal{I}^*$ increasing until it reaches a random graph.

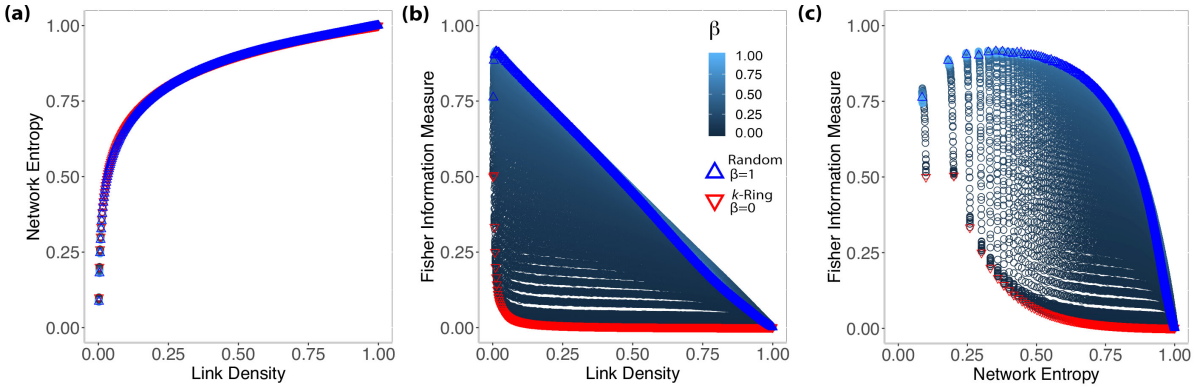


Figure 3.3: Relationship between Shannon Entropy and Fisher Information Measure with link density (**(a)** and **(b)**), and between Fisher Information Measure and Shannon Entropy for Watts-Strogatz networks **(c)**. We restricted the analysis to $N = 1000$, $k \in \{1, 2, 3, \ldots, 499, 500\}$ and $\beta \in \{0, 0.001, 0.002, \ldots, 0.99, 1\}$; the downward red triangles correspond to k-rings ($G_{N,k}$ with $\beta = 0$); the upwards blue triangles are random graphs ($G_{N,k}$ with $\beta = 1$). The blue gradient from dark to light corresponds to the rewiring probability β: the intensity of the blue color is inversely proportional to the value of β.

3.4.3 Scale-free Networks

The literature often uses *scale-free* networks as models for real networks. They have a degree distribution that can be fitted by a power-law, i.e., $P(k) \sim k^{-\gamma}$, where γ is the degree exponent usually in $2 \leq \gamma \leq 3$, as for $\gamma > 3$ the scale-free property can easily be confused with random networks [9]. We will evaluate the Barabási-Albert [10] model (BA) for evolving *scale-free* networks, as it has two essential features: network growth and the preferential attachment mechanism.

For network growth, at each time step t, we insert new nodes with m links connecting with N_0 existing nodes in the network. We create these links according to a probability given by the preferential attachment: the probability that a node i connects with j is

proportional to the actual degree of node i:

$$\Pi^{(i)} = \frac{k_i}{\sum_j k_j}. \tag{3.18}$$

In this way, the preferential attachment (PA) induces *hubs* (highly connected nodes), and peripheral communities, where nodes have a similar degree. We know that the Barabási-Albert model is unable to reproduce all the diversity existing for scale-free networks, as it captures only the power-law with degree exponent $\gamma = 3$. Therefore, throughout the years, we found many variations of this model. In this work, we extend our analysis for non-linear preferential attachment, the fitness property, the aging property, and, finally, the configuration model.

Non-linear preferential attachment

Krapivsky et al. [32] introduced a non-linear preferential attachment that creates different regimes for the network according to an exponent α controlling the network topology. The non-linear PA is given by

$$\Pi^{(i)} = \frac{k_i^\alpha}{\sum_j k_j^\alpha}. \tag{3.19}$$

For $\alpha \neq 1$, the growth model stops resulting in a power-law degree distribution. There are, thus, three different growth regimes:

- The sublinear regime ($\alpha < 1$) has a power-law with an exponential cutoff, where the preferential attachment is not strong enough to produce a pure power-law degree distribution.
- The linear regime ($\alpha = 1$) has a pure power-law behavior corresponding to the *Barabási-Albert* [10] model, with a resulting $\gamma = 3$.
- The superlinear regime ($\alpha > 1$) presents a particular behavior where the network condensates, i.e., very few nodes win all connections; it also does not result in a power-law degree distribution.

We mapped outcomes of the BA model with a non-linear preferential attachment using the Krapivsky's model onto the Shannon-Fisher plane, as shown in Figure 3.4a. For $\alpha = 0$, we have a random network, since $\Pi^{(i)} = 1$ for every i, the network no longer obeys the preferential attachment mechanism, just the evolving growth property; the result is $\overline{\mathcal{H}} = 0.072$, $s_{\mathcal{H}} = 0.001$ and $\overline{\mathcal{I}^*} = 0.87$, $s_{\mathcal{I}^*} = 0.001$. Increasing α in steps of 0.01 changes the network's regime slowly, and we see this transition in the Shannon-Fisher plane until it reaches $\alpha = 1$. In the linear regime $\overline{\mathcal{H}} = 0.063$, $s_{\mathcal{H}} = 0.001$ and $\overline{\mathcal{I}^*} = 0.874$, $s_{\mathcal{I}^*} = 0.001$. In the superlinear regime $\mathcal{H} \to 0$ and $\mathcal{I}^*$ starts decreasing above $\alpha > 1.4$, as seen in

Figure 3.4b. Figure 3.4a shows no intersections between Barabási-Albert networks (BA) and any other network class (e.g. Random or Small-world).

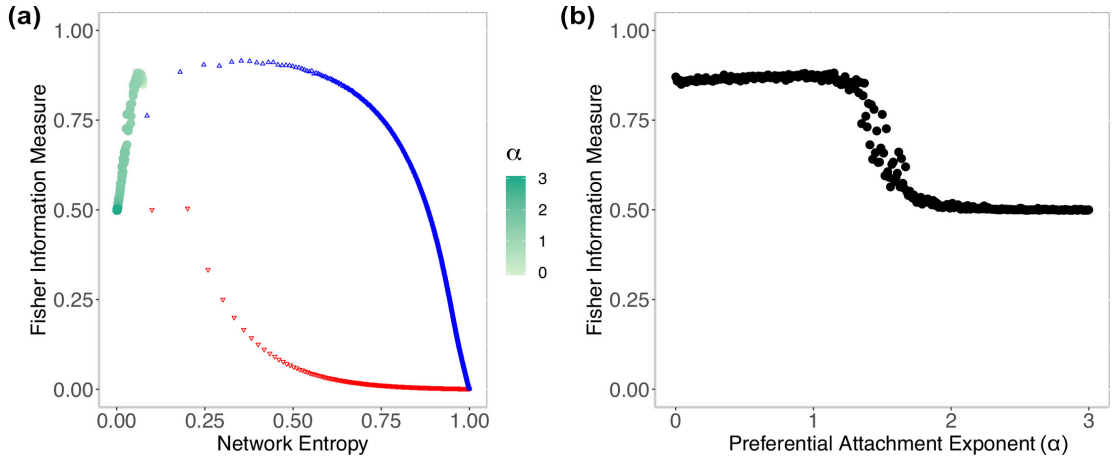


Figure 3.4: **(a)** Barabási-Albert networks, non-linear preferential attachment with $N = 1000$, and $\alpha \in [0, 3]$. For the sake of visualization, we plot the red downward triangles representing $G_{N,k}$ with $\beta = 0$, i.e., k-ring graphs; blue upward triangles are $G_{N,k}$ with $\beta = 1$, i.e., random graphs. **(b)** shows how changing α causes disturbances in the Fisher Information Measure, when evaluating the Barabási-Albert model with non-linear PA.

Similar to earlier sections, we evaluated the link density ξ along with Network Entropy $\mathcal{H}$ and Fisher Information Measure $\mathcal{I}^*$. This time, the results with link density in comparison with Network Entropy, shown in Figure 3.5b, have more interesting behavior. Although the link density does not change ($\xi = 0.002$), $\mathcal{H}$ absorbs the changes and when increasing α, $\mathcal{H} \to 0$. In Figure 3.5a, we observe how Fisher Information Measure $\mathcal{I}^*$ against link density ξ produces varying results, as in between $0 < \alpha < 1.4$, $\mathcal{I}^*$ increases, then, for $\alpha > 1.4$, the values of $\mathcal{I}^*$ decrease.

Fitness property

Some networks have nodes that create connections with more ability, e.g., a popular web page. Usually, these nodes gain relationships faster than common nodes. The Bianconi-Barabási model [12, 1] describes this property named *fitness*. We can model it using the preferential attachment considering a fitness coefficient η_i alongside the node degree k_i:

$$\Pi^{(i)} = \frac{\eta_i k_i}{\sum_j \eta_j k_j}. \tag{3.20}$$

In Eq. 3.20, the dependence of $\Pi^{(i)}$ on η_i models the fact that even younger nodes can acquire links faster if they have sufficiently higher fitness than older nodes. Therefore, we draw 30 networks with $N = 1000$, considering a uniform distribution for the fitness η_i of each node i. For this, we do not expect a perfect power law, but we expect $\gamma = 2.255$, asymptotically.

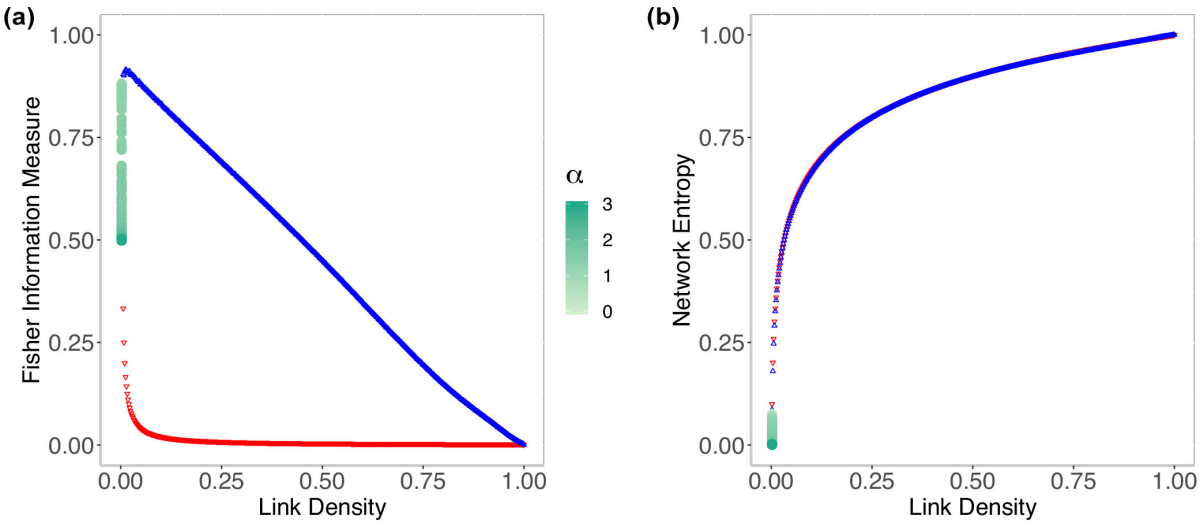


Figure 3.5: (a) Relationship between link density and Fisher Information Measure for Barabási-Albert networks using a non-linear preferential attachment; the gradient indicates how the preferential attachment exponent α changes. (b) Relationship between the Network Entropy and link density, where $\xi = 0.002$ for any α. To help the visualization of the region where Barabási-Albert networks stand in relation to the other synthetic networks, red downward triangles represent $G_{N,k}$ with $\beta = 0$, i.e., k-ring graphs; blue upward triangles are $G_{N,k}$ with $\beta = 1$, i.e., random graphs.

Figure 3.6 shows the Fitness model falls into a region similar to BA, but with some networks closer to Random rather than Scale-Free.

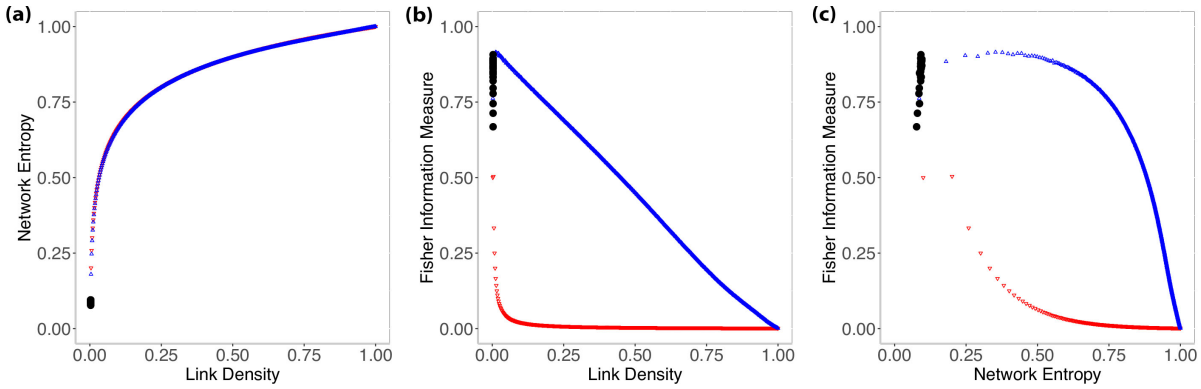


Figure 3.6: Relationship between Shannon Entropy and Fisher Information Measure with link density **(a)** and **(b)**, and between Fisher Information Measure and Shannon Entropy (c) for Biaconi-Barabási (Fitness model). Black points indicate the thirty networks with $N = 1000$ generated using a uniform distribution for the fitness scores of each node.

Aging property

Another aspect we can also consider for a scale-free network is the *aging* property [20]. Regularly, for the Barabási-Albert model, we account only for the node degree or, as seen before, the fitness coefficient. However, what happens when a node starts to reduce

the rate of acquiring new links with time? This aging process causes the nodes to lose relevance; thus, it changes the network structure and dynamics. We can model this property considering:

$$\Pi^{(i)}(k_i, t - t_i) \sim k(t - t_i)^{-\nu}, \tag{3.21}$$

where ν is a parameter controlling the dependence of the attachment probability on the node's age. According to ν, we can define three scaling regimes:

- If $\nu < 0$, new nodes will connect to older nodes. If $\nu \to -\infty$, each new node connects to the oldest node, resulting in a condensed network or hub-and-spoke topology. Hence, we have a more heterogeneous network with a few hubs and many peripheral nodes.

- If $\nu > 0$, nodes connect to younger nodes. By aging, nodes lose the ability of preferential attachment. In this case, the network tends to be more homogeneous.

- For $\nu > 1$, the aging effect dominates the preferential attachment effect, the network loses its scale-free property, and it eventually approaches a random network. When $\nu \to \infty$, each node connects to its immediate predecessor.

For evaluating the aging property, we generate distinct networks with $N = 1000$ and $\nu \in \{-3.0, -2.9, -2.8, ..., 2.9, 3.0\}$ with 30 replications of each setting; thus, we have a total of 18030 networks. Figure 3.7a shows the results for networks with a growing Network Entropy $\mathcal{H}$ and a steady link density $\xi = 0.002$, the same result as for BA model. Figure 3.7b shows the results for the aging property, and once more, we can observe the "oscillation" that happens to all the other scale-free models previously discussed. Figure 3.7c presents the results considering the Network Entropy $\mathcal{H}$ and Fisher Information Measure $\mathcal{I}^*$, where we can see the Aging model transition in the plane according to its scaling regimes.

The numerical results for the Aging model are the following: for $\nu = -3$, $\overline{\mathcal{H}} = 0.024$, $s_{\mathcal{H}} = 0.0014$ and $\overline{\mathcal{I}^*} = 0.74$, $s_{\mathcal{I}^*} = 0.032$ i.e., we have a condensed network; when $\nu > 0$, $\mathcal{H}$ and $\mathcal{I}^*$ continue to grow until $\nu = 1$, wherefore $\overline{\mathcal{H}} = 0.072$, $s_{\mathcal{H}} = 0.0009$ and $\overline{\mathcal{I}^*} = 0.857$, $s^*_{\mathcal{I}} = 0.004$; for $\nu > 1$, $\mathcal{H}$ grows steadily and $\mathcal{I}^*$ decays, reaching a random regime. Finally, we noticed that the scale-free regime expected for $\nu \in [0, 1]$ is observed in the Shannon-Fisher plane, where the values for $\overline{\mathcal{H}} = 0.063$, $s_{\mathcal{H}} = 0.005$ and $\overline{\mathcal{I}^*} = 0.866$, $s_{\mathcal{I}^*} = 0.007$.

The configuration model

A recurrent problem is "how do we generate networks with an arbitrary $P(k)$?". For this, we use the configuration model, also known as a random network with a pre-defined degree sequence [45]. According to Ref. [9], the algorithm consists of the following steps:

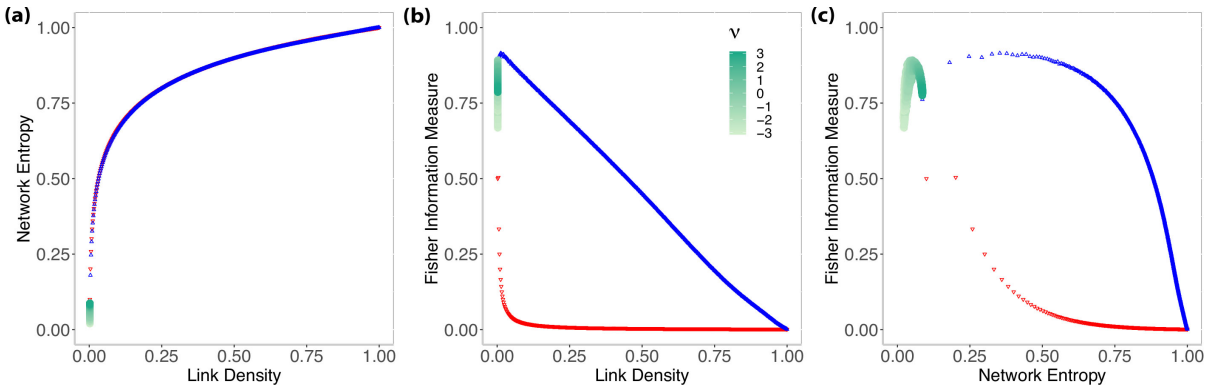


Figure 3.7: Relationship between Shannon Entropy and Fisher Information Measure with link density **(a)** and **(b)**, and between Fisher Information Measure and Shannon Entropy **(c)** for the Aging model. The gradient indicates the aging exponent $\nu \in [-3, 3]$ and how its growth controls the network scaling regimes.

1. Assign a degree to each node as stubs or half-links. We have to start from an even number of stubs; otherwise, we will have unpaired stubs.

2. Randomly selects a pair of half-links and connects them; then randomly choose another pair from the remaining $2L - 2$ half-links.

3. Repeat this process until paired up all stubs. Depending on how we pair them up, we may obtain distinct networks. Some networks include cycles, self-loops, or multi-links. In this work, we consider only simple graphs, thus, after generating the network for a degree sequence, we "simplify" the graph, removing self-loops and multi-links.

As we are trying to reproduce scale-free properties using the configuration model, we assign a pure power-law distribution $P(k) = k^{-\gamma}$ with $\gamma \in [2, 5]$. For this model, we expect that for $2 \leq \gamma \leq 3$, the network is in the scale-free regime; when $\gamma > 3$, the network starts to condensate, as the distribution has a steep curve. It means that few nodes have most of the links, and most nodes have few links. Such networks present structure and dynamics more similar to a hub-and-spoke topology. It is interesting to assess the configuration model; now, the networks built using a power-law with $\gamma \in [2, 3]$ represent a small area of the Shannon-Fisher plane (Fig. 3.8a), where the allegedly Scale-Free networks present themselves. This behavior provides an interesting perspective when evaluating real networks.

Finally, we evaluate these networks with $N = 1000$ using a pure power-law distribution with $\gamma \in \{2.0, 2.1, 2.2, ..., 4.9, 5.0\}$; as we cannot guarantee that networks with the same degree exponent have the same topology, we replicate this experiment 30 times, then, we have 1312 networks. In Figure 3.8a we have that $\overline{\xi} = 0.001$, with $s_\xi = 0.0006$ while the Network Entropy $\mathcal{H}$ decreases as we increase the degree exponent. This behavior is similar for all the other scale-free models when we are in the condensed regime. In Figure 3.8c,

we observe how FIM $\mathcal{I}^*$ is capturing the changes, wherefore the degree exponent $\gamma \in [2, 3]$ we have $\overline{\mathcal{H}} = 0.045$, $s_{\mathcal{H}} = 0.017$ and $\overline{\mathcal{I}^*} = 0.886$, $s_{\mathcal{I}^*} = 0.026$; and for $\gamma \in (3, 5]$, we have $\overline{\mathcal{H}} = 0.010$, $s_{\mathcal{H}} = 0.005$ and $\overline{\mathcal{I}^*} = 0.952$, $s_{\mathcal{I}^*} = 0.018$.

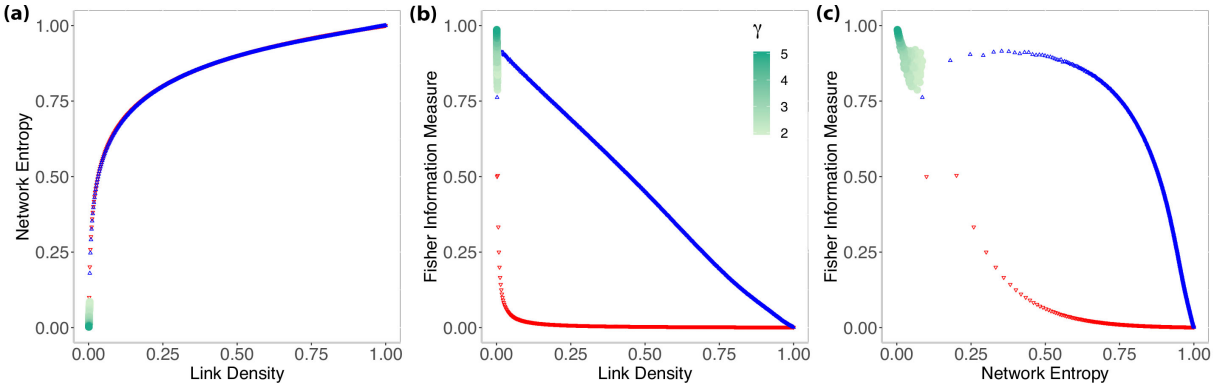


Figure 3.8: Relationship of Shannon Entropy and Fisher Information Measure with link density **(a,b)**, and between Fisher Information Measure and Shannon Entropy **(c)** for the configuration model with a degree distribution following a pure power-law $P(k) \sim k^{-\gamma}$. The gradient indicates the degree exponent $\gamma \in [2, 5]$ and how it controls the network scaling regimes.

3.5 Discussion

Complex networks have many faces, thus attempting to label them considering a single network property may be misleading. Real networks have many components and distinct interactions among them, for example, a scale-free network may have peripheral communities that lead to the small-world structure. Our proposal quantifies network structure and dynamics, considering a simplified plot. We show consistent results with other network features when we apply this methodology to synthetic networks.

The Shannon-Fisher plane enhances our ability to evaluate complex networks:

- The transition that the Watts-Strogatz model exhibits in between k-ring and random graphs, leading us to define the small-world region;
- The two distinct regimes for the Erdős-Rényi model when reaching the critical linking probability;
- The three regimes for the non-linear preferential attachment on scale-free networks and distinct growth models, which transits between random, scale-free and condensed networks;
- The fitness model's behavior when we consider a uniform distribution for each node's fitness, and how it has similar features to the Barabási-Albert model;

- The effect of aging for scale-free networks and how the aging exponent can control the system's behavior in the same manner to what happens with the non-linear preferential attachment;
- And finally, how we can generate networks with a pure power-law considering distinct degree exponent.

Chapter 4

Characterizing network traffic

Network traffic plays a critical role in network planning and control. The researchers assume that traffic from Ethernet and other IP-related networks have a self-similar nature: high-variability and long-term correlations. Many studies try to model these characteristics for simulation and further optimization. One of the most straightforward approaches to model these characteristics is to consider ON/OFF sources (packet-train), where ON- and OFF-periods are i.i.d., generated with random heavy-tailed distributions. Using information theory quantifiers, in particular the Causality Complexity-Entropy Plane, we show that heavy-tailed distributions do not capture most of the network traffic dynamics. They only reproduce the stochastic dynamics of traffic, which accounts for one of the smallest parts. We conduct this study by observing the Abilene dataset, fitting the LogNormal and LogLogistic distributions, and evaluating them onto Causality Complexity-Entropy Plane in comparison with $1/f$-noise, which is one of the most observed long-term correlated noises in nature stochastic processes. Also, to enhance our illustrated results, we use the k-nearest-neighbors (kNN) to classify the real and generated traffic according to the results obtained.

4.1 Self-similarity and network traffic

The evaluation of network traffic is the subject of extensive analysis for decades. Initially, it was common knowledge that network traffic was similar to a Poisson process with finite variance and aggregated dynamics close to white noise. Leland et al. [37] identified two distinct characteristics of Ethernet traffic collected from local area networks (LAN): (i) they identified the *bursty* traffic nature: high-volume demands within short-time periods with a substantial positive variance; and (ii) they observed periodic volume within daily periods, named "ripples". Hence, they showed that Poisson-related processes were not ideal for modeling network traffic, as traffic time-series were *self-similar*. This achievement shifted the previously randomness assumed, towards pattern identification into network traffic time-series. These discoveries helped the network researchers to address traffic

modeling and performance evaluation with a new perspective.

Self-similarity [40] implies a stochastic process with high-variability and long-term correlation that can be measured by the Hurst exponent [41]. High-variability can be represented by heavy-tailed distribution with infinite variance, while long-term correlation implies a self-correlation that presents a power-law decay. In simple words, a stochastic process is self-similar with the proper self-similarity degree given by the Hurst exponent, if it presents the same correlation structure for distinct time-scale.

Willinger et al. [66, 58] discussed the self-similarity and long-range dependent nature regarding strictly alternating ON/OFF sources (or packet-train). The ON-period (or packet-train length) is the time of sending the packets; the OFF-period (or inter-train distances) is when the source does not send any packet. Both ON- and OFF-periods are i.i.d. with infinite variance and linearly independent between themselves. Then, ON/OFF sources, in the aggregated traffic, present characteristics of self-similar stochastic processes. As an example of reference self-similar processes [49], we have: $1/f$-noise, fractional Brownian motion (fBm), and fractional Gaussian noise (fGn) processes.

Many studies continued exploring the notion of self-similarity within network traffic. Lakhina et al. [35] extended the fact that the Poisson models can not replicate the dynamics of LAN traffic. Using principal component analysis (PCA), they evaluate the Sprint and Abilene datasets. From a long-range perspective, they show that most of the network traffic is periodic with daily cycles (the *ripples*); and the remaining traffic consists of bursty events and noise. This work corroborated the existence of the three traffic components (or *categories*), with precise pictorial results showing the traffic behavior.

Several researchers use this approach of using PCA to characterize traffic, leading to further studies of anomaly detection using machine learning techniques. Lakhina et al. [34] applied the subspace method to the three different types of flow traffic identified, being able to tell when some particular anomaly behavior appeared. So far, all these studies evaluated traffic obtained from aggregating the origin-destination flow statistics.

Nucci et al. [47] assess the same datasets from Lakhina et al. Using this data, they fit a few probability distribution functions (e.g., Log-normal and Log-logistic), showing that heavy-tailed curves can describe most of the aspect identified at [37, 35]. The community extensively acknowledges this work, and regularly, it guides the simulation parameters of computer networks in many environments: Vishwanath et al. [62] developed a traffic generator that captures traffic of distinct services in the network, automatically extracts the distribution parameters and it reproduces the traffic structure with statistical significance. Similarly, Benson et al. [11] introduced a similar analysis for the traffic of data center networks. More recent, Varet et al. [60] developed a traffic generation tool using the concept of ON/OFF sources [66] and the fitted models of [47].

In this chapter, we question the dynamical properties captured by fitted distributions of traffic flows with information-theory-based quantifiers. Usually, these distributions fit

the empirical cumulative distribution functions with a low error rate considering K-S (Kolmogorov-Smirnov) and C-S (Chi-Squared) tests. Although low error with relevant p-values indicates a good fit, it does not guarantee that the dynamics of the systems are similar, and as [47] pointed out, this fit may not account for noisy events that can usually happen.

Rosso et al. [52] evaluates chaotic systems and compared them with correlated and non-correlated stochastic processes in the causal complexity-entropy plane (CCEP). This bidimensional representation plane can distinguish the chaotic or stochastic nature of these systems. Aquino et al. [5] applied this methodology to characterize vehicle velocities that, so far, were represented by Poisson-related models, showing that these velocities present a long-term correlated nature, similar to $1/f$-noise. A contribution similar to what was done by Willinger et al. [37] concerning network traffic flow time-series. Later, the same methods were successfully applied to characterize the electric load consumption with the same purpose [4].

We suggest the use of information-theory quantifiers for distinguishing traditional and self-similar traffic. Our main question is:

"*Does existent traffic models reproduce the dynamics of real network traffic?*"

Nevertheless, we compared the real data from origin-destination traffic flows with the proposed models in the causal complexity-entropy plane. According to our results, these models do not appear to reproduce the dynamics of real data.

We observe the distinct dynamics presented by real traffic data, in comparison with the models used in network planning and simulation. Further, we show that real data have similar dynamics to ideal self-similar processes (e.g., $1/f$-noise, fBm, and fGn) with proper parameter setting, confirming the existing discussion about network traffic in the CCEP. Also, we observe that existing fitted models, according to the results of CCEP, have dynamics closer to white noise, rather than long-term correlated processes. Thus, these results can be beneficial for the development of decision-making solutions concerning traffic engineering. Although the existing models satisfy planning requirements, it is not sufficient to guide the comparison of existing solutions.

4.2 Time-series and information theory quantifiers

The characterization of the traffic flow time series takes two steps: first, we apply the Bandt-Pompe technique [7], transforming the time series into a symbolic sequence histogram that retains time causal information; second, we map this sequence onto the Causality Complexity-Entropy Plane (CCEP) [52], and its results indicate its underlying physical nature.

Bandt and Pompe [7] presented a simple method that compares neighboring values given: a time series $\mathcal{X}(t) = \{x_t : t = 1, \ldots, T\}$; an embedding dimension $D \geq 2$ ($D \in \mathbb{N}$), for practical purposes $D = 2, \ldots, 7$; and an embedding delay time $\tau \in \mathbb{N}$ that determines the time separation between the data. Then, we can find the $D!$ ordinal patterns of length D. Let us take this simple example presented by Bandt and Pompe [7]. Consider a time series with seven values:

$$\mathcal{X} = \{4, 7, 9, 10, 6, 11, 3\}, \tag{4.1}$$

we can find four pairs which $x_t < x_{t+1}$ and two pairs for which $x_t > x_{t+1}$. Hence, four pairs will be transformed to the symbolic representation $\pi = 01$ ($x_t < x_{t+1}$), while the remaining two will be represented by $\pi = 10$ ($x_t > x_{t+1}$). In order to have a unique result, if we had $x_t = x_{t+1}$, it would be treated the same as if $x_t < x_{t+1}$. This example considered $D = 2$ (and $\tau = 1$), let us take an $D = 3$, $\tau = 1$ example, also given by Ref. [7]. First, we have this sequence $(4, 7, 9)$ and $(7, 9, 10)$ that represent the permutation symbol $\pi = 012$ since they are in increasing order. Then, we may have $(9, 10, 6)$ and $(6, 11, 3)$ having the permutation symbol $\pi = 201$ since $x_{t+2} < x_t < x_{t+1}$, and $(10, 6, 11)$ has $\pi = 102$, and so forth.

The definition stands that for all the $D!$ possible permutations $\{\pi\}$ of order D, their associated relative frequencies can be computed by the number of times that we found this particular order sequence (permutation symbol) in the time series divided by the total number of sequences. Hence, the relative frequency of each symbol π is given by

$$p(\pi) = \frac{\#\{t \,|\, t \leq T - D, (x_{t+1}, \ldots, x_{t+D}) \text{ has type } \pi\}}{T - (D-1)\tau}, \tag{4.2}$$

where $\#$ means the cardinality (number) of the set. This should estimate the frequency of π as good as possible for a finite time-series. The ordinal patterns associated with $P \equiv \{p(\pi)\}$ are invariant with respect to nonlinear monotonic transformations. Accordingly, nonlinear drifts or scaling artificially will not modify the estimation of quantifiers.

We evaluate the normalized Shannon Entropy [55] and the Statistical Complexity [36] of the frequency distribution P obtained with the Bandt-Pompe technique over time series of network traffic flows. Since this method consists of counting order sequences of permutation symbols, the former quantifier is named *Permutation Normalized Shannon Entropy* and the latter *Permutation Statistical Complexity*. Although, for the sake of simplicity, from this point further, we refer to both quantifiers only as: *Shannon Entropy* and *Statistical Complexity*.

Let $\theta : \Omega \to \mathbb{R}$ be a discrete random variable with $T < \infty$ possible values and $\Omega = \{\xi_i : i = 1, \ldots, T\}$ and probability distribution $P = \{p_j \geq 0 : j = 1, \ldots, N\}$, whereas $\sum_{j=0}^{N} p_j = 1$. The Shannon Entropy $\mathcal{S}[P] = -\sum_{j=1}^{N} p_j \ln p_j$, where, by convention, $0 \ln 0 = 0$. This quantifier is related to the amount of uncertainty described by P

and its underlying physical process.

Thus, we can define permutation Shannon Entropy as $\mathcal{S}[P] = -\sum p(\pi) \ln p(\pi)$. It is easy to follow that $0 < \mathcal{S}[P] < \ln D!$, where the lower bound is observed when we have a strictly increasing or decreasing sequence, and the upper bound is obtained for an i.i.d. sequence (a completely random system) where all the $D!$ permutations appear with the same probability, i.e., a uniform distribution. The uniform distribution is taken as reference distribution P_e, which is given by $p_j = 1/N$ for every $1 \leq j \leq N$. It is immediate that $\mathcal{S}[P_e] = \ln N$. Given the frequency distribution, $N = D!$, hence, $\mathcal{S}[P_e] = \ln D!$, describing the situation of less information about the observed time-series.

Further, Lamberti et al. [36] introduced the Statistical Complexity $\mathcal{C}_{JS}$ and the Jensen-Shannon $\mathcal{Q}_{JS}$ disequilibrium measure; the disequilibrium is the "distance" between a given probability distribution P and the equilibrium reference distribution P_e:

$$\begin{aligned} \mathcal{Q}_{JS}[P, P_e] &= Q_0 \cdot \mathcal{J}_S[P, P_e] \\ &= Q_0 \cdot \left\{ \mathcal{S}\left[\frac{P + P_e}{2}\right] - \frac{\mathcal{S}[P] + \mathcal{S}[P_e]}{2} \right\}; \end{aligned} \tag{4.3}$$

with $\mathcal{J}_S[P, P_e]$ being the Jensen-Shannon divergence [38]. This information-theoretical-based divergence measure is intimately related to the Kullback-Leibler relative entropy. Also, Q_0 is a normalization constant equal to the inverse of the maximum possible value of $\mathcal{J}_S$ so that $\mathcal{Q}_{JS} \in [0, 1]$. Q_0 is obtained when one of the probabilities of P is equal to one, and the remaining are equal to zero.

Statistical Complexity measure $\mathcal{C}_{JS}$ can detect essential details of the underlying dynamics that gives rise to the observations. We define this measure according to the functional product proposed by López-Ruiz et al. [39]:

$$\mathcal{C}_{JS} = \mathcal{H}_{\mathcal{S}}[P] \cdot \mathcal{Q}_{JS}[P, P_e], \tag{4.4}$$

where

$$\mathcal{H}_S[P] = \mathcal{S}[P]/\mathcal{S}^{\max} \tag{4.5}$$

is the normalized Shannon entropy $\mathcal{H}_S \in [0, 1]$ with $\mathcal{S}^{\max} = \mathcal{S}[P_e]$, and the disequilibrium $\mathcal{Q}_{JS}$ is the Jensen-Shannon disequilibrium measure.

Finally, we analyze the time-series using a bidimensional diagram, named Complexity-Entropy Plane (CCEP). The CCEP is obtained by plotting $\mathcal{C}_{JS}$ (vertical axis) versus $\mathcal{H}_S$ (horizontal axis) [52]. For a given value of $\mathcal{H}_S$, the range of possible $\mathcal{C}_{JS}$ values, in this plane, varies between $\mathcal{C}_{JS}^{\min}$ and $\mathcal{C}_{JS}^{\max}$ [42]. The term "causality" reminds the fact that temporal correlations between successive samples are taken into account through $P = \{p(\pi)\}$; the PDF computed with the Bandt-Pompe methodology is used to evaluate both Information Theory quantifiers $\mathcal{H}_S$ and $\mathcal{C}_{JS}$.

Rosso et al. [52] showed that this tool is particularly efficient at distinguishing between

the chaotic and stochastic nature of time series since the permutation quantifiers have distinct behaviors for different types of dynamics. For regular (periodic) processes, both quantifiers have small values, close to zero. Totally uncorrelated stochastic processes are close to $\mathcal{H}_S \approx 1$ and $\mathcal{C}_{JS} \approx 0$. Correlated stochastic processes with $f^{-\kappa}$ power spectrum ($0 < \kappa \leq 3$) are characterized by intermediate $\mathcal{H}$ with $\mathcal{C}$ values between $\mathcal{C}_{JS}^{\min}$ and $\mathcal{C}_{JS}^{\max}$. We found similar results for the self-similar stochastic processes: fractional Brownian motion (fBm) and fractional Gaussian noise (fGn) [52].

4.3 Network traffic flow characterization

We consider the Abilene dataset [35] as the real data obtained from origin-destination traffic flow measures. This dataset contains information from 121 origin-destination pairs, measured at each 5 minutes (12 times per hour), for 24 hours and 7 days, giving us the total of $12 \times 24 \times 7 = 2016$ samples for each origin-destination pair.

Initially, this dataset was used by Lakhina et al. [35] to identify distinct flow categories. Later, Nucci et al. [47] discussed the underlying characteristics of this data. They proposed a dynamic stationary model that can be described by:

$$X(i, j, t) = \bar{x}(i, j, T) + W(i, j, t), \tag{4.6}$$

where for (i, j)-th flow, the mean of the origin-destination flow is given by $\bar{x}(i, j, T)$. During the period T, $W(i, j, t)$ captures the random fluctuation in $(T - 1) \leq t \leq T$, as a zero mean random process whose variance needs to be determined. Later, this base model is simplified to:

$$X(i, j, t) = \bar{x}(i, j, T). \tag{4.7}$$

For this, we assume the random fluctuations to be so small, that the dynamic of the system is not disturbed; we show that this is not the case. Nonetheless, for $X(i, j, t)$, they estimate distinct distribution fittings that consider the mean values of origin-destination flows for each hour. In this work, we are interested in the 5-minutes granularity, as the Abilene dataset is already a representation of complex network traffic, reducing this data, even more, would reduce the amount of information dangerously.

Therefore, we choose the recommended distributions of Nucci et al. [47]: *LogNormal* and *LogLogistic*. We estimate the distribution parameters using the Maximum Likelihood Estimation method (MLE). A random variable X has a log-normal distribution if the random variable $\ln X$ has a normal distribution. In the same fashion, a random variable X has a log-logistic distribution, if the random variable $\ln X$ has a logistic distribution.

Using the proper log-likelihood functions and the estimator, we were able to achieve the fit displayed in Figure 4.1. For statistical significance, we tested our distributions with the Kolmogorov-Smirnov test that indicates the quality of the fitting [43]. We accept the

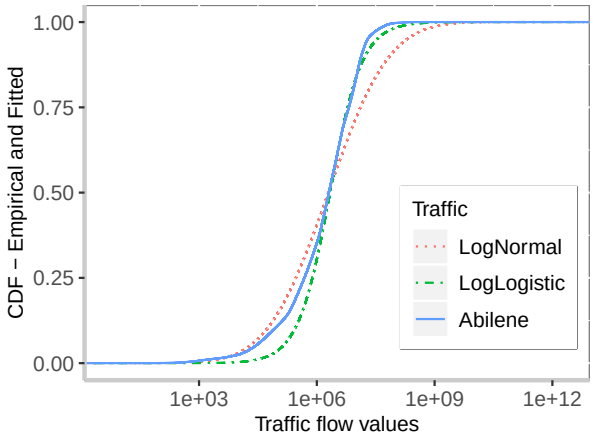


Figure 4.1: Traffic values for the Abilene dataset and the fit of the LogNormal and LogLogistic distributions.

fitting hypothesis if the p-value has significance level $\alpha < 0.05$. Observing the table 4.1, we can see that our fitting parameters have *failed* the hypothesis testing.

Table 4.1: K-S test results for traffic flow values.

Distributions	Parameters	K-S	
		Test value	p-value
LogNormal	$\mu = 14.48$, $\sigma = 2.81$	0.15594	<2.2e-16
LogLogistic	$\alpha = 0.94$, $\beta = 14.58$	0.086064	<2.2e-16

This estimation attempts to give us a proper way of generating mean traffic flow values for a given scenario, but it is not enough. Several research papers have discussed the presence of long-term correlation that is not captured by fitting a distribution. Even though, Nucci et al. [47] suggest adding a zero-mean Gaussian noise for the intervals that the real data, it is not enough. Using CCEP, we observed that (Section 4.4) the Abilene dataset, in comparison with the random distributions fitting, present a distinct dynamics; in fact, the dataset present dynamics closer to the $1/f^{\kappa}$-noise.

The $1/f^{\kappa}$-noise phenomena is widely found in nature, and it is often related with long-memory processes and long-term correlations for $f^{-\kappa}$ with $0.5 \leq \kappa \leq 1.5$. This noise is an intermediate between the well understood white noise ($\kappa = 0$) with no correlation in time and random walk (*Brownian motion*) noise with no correlation between increments. There are no simple, even linear stochastic differential equations generating signals with $1/f^{\kappa}$-noise. The widespread occurrence of this phenomenon suggests that a generic mathematical explanation might exist. The standard definition of $1/f^{\kappa}$-noise [63] refers to a phenomenon with spectral density given by

$$S(f) = constant/f^{\kappa}, \tag{4.8}$$

where f is the frequency. To generate a process with constant spectral density, we use a random uniform distribution.

4.4 Results

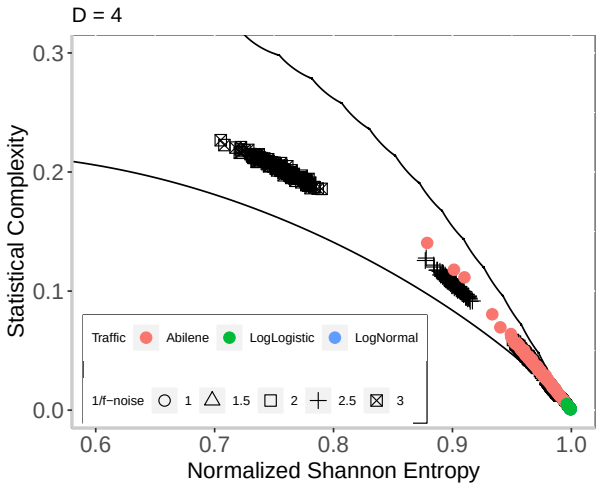
We characterize the traffic flow time series using the Bandt-Pompe symbolization with $D = \{4, 5, 6\}$ and $\tau = 1$. Considering each origin-destination flow and the fit obtained for the LogNormal and LogLogistic distribution, we generated 121 samples with 2016 length each, using the obtained parameters in Table 4.1. We compare the results from the Abilene dataset, LogNormal, and LogLogistic, alongside $1/f^{\kappa}$-noise with $\kappa = \{0, 0.5, 1, 1.5, 2, 2.5, 3\}$, although $\kappa = \{0, 0.5\}$ are left out of the planes to improve visualization, as they are really close to $\kappa = 1$ and amongst each other.

The $1/f^{\kappa}$-noise has important properties to notice. First, for $\kappa = 0$, we have white noise, with the constant power spectrum (entirely random state); for $0.5 < \kappa < 1.5$, we have a pink noise, as one of the most common phenomena observed in nature, easily observable in music and arts; for $\kappa > 1.5$ we have distinct "colors" of noise with a highly correlated structure. With this information, we proceed to evaluate the time-series in the CCEP, as shown in Figure 4.2.

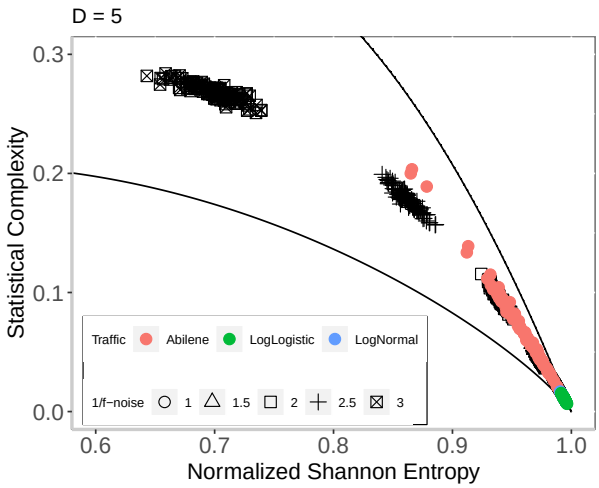
We change D to extend our analysis for distinct ordinal pattern lengths considering that $M \gg D!$, as $M = 2016$. Figure 4.2 shows that the variation of D impacts the spreading of the results, but the observed dynamics is maintained. For the $1/f^{\kappa}$-noise and $D = 5$ (Fig. 4.2b): with $\kappa = 0$, the average Shannon Entropy is $\mathcal{H} = 0.994$ and the average Statistical Complexity is $\mathcal{C}_{JS} = 0.011$; when $\kappa = 0.5$, $\mathcal{H} = 0.993$ and $\mathcal{C}_{JS} = 0.012$; $\kappa = 1.0$, $\mathcal{H} = 0.988$ and $\mathcal{C}_{JS} = 0.210$; $\kappa = 1.5$, $\mathcal{H} = 0.974$ and $\mathcal{C}_{JS} = 0.044$; $\kappa = 2.0$, $\mathcal{H} = 0.940$ and $\mathcal{C}_{JS} = 0.094$; $\kappa = 2.5$, $\mathcal{H} = 0.862$ and $\mathcal{C}_{JS} = 0.178$; and finally for $\kappa = 3$, $\mathcal{H} = 0.699$ and $\mathcal{C}_{JS} = 0.268$.

When evaluating the results in the CCEP for the Abilene dataset with $D = 5$, $\tau = 1$, we obtained the respective average values: $\mathcal{H} = 0.958$ and $\mathcal{C}_{JS} = 0.068$. However, simply observing the plots in Figure 4.2, we can see that Abilene has a considerable variance, much higher than LogNormal and LogLogistic. The average results for LogNormal in the CCEP are: $\mathcal{H} = 0.994$ and $\mathcal{C}_{JS} = 0.010$; and for LogLogistic, we have $\mathcal{H} = 0.993$ and $\mathcal{C}_{JS} = 0.001$. Nevertheless, to a more robust comparison in between the traffics and $1/f^{\kappa}$-noise, we use the k-Nearest-Neighbors (kNN) algorithm to "classify" the given traffic in respect to the distinct spectrum of correlated noise.

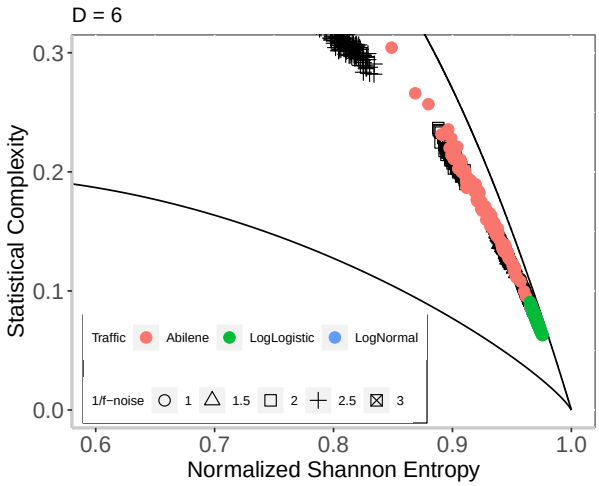
We trained a kNN classifier with $k = 3$ and Euclidean distance for all the distinct settings of $1/f$-noise with $D = \{4, 5, 6\}$ and $\kappa = \{0, 0.5, 1, 1.5, 2, 2.5, 3\}$. Then, we "predict" the κ exponent according to the training dataset that considers the Shannon Entropy $\mathcal{H}$, Statistical Complexity $\mathcal{C}_{JS}$ and the ordinal pattern length D, $\tau = 1$. Table 4.2 presents the results with the relative frequencies of each traffic data in the corresponding noise



(a) $D = 4$, $\tau = 1$



(b) $D = 5$, $\tau = 1$



(c) $D = 6$, $\tau = 1$

Figure 4.2: Causality Complexity-Entropy Plane for the traffic flow values with distinct D values.

category. We can see that the Abilene dataset has less than ten percent of its data with dynamics similar to $\kappa \leq 1$, while LogNormal and LogLogistic has 100% of its data on this region. In fact, Abilene has 92% of its data in between $1.5 \leq \kappa \leq 2.5$.

Table 4.2: Classification of the traffic time-series according to $1/f$-noise exponent κ with relative quantities.

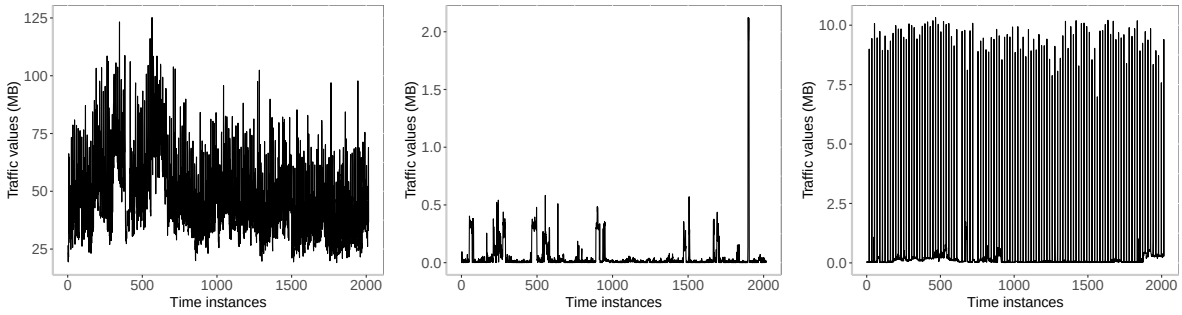
Traffic data	$1/f$ exponent κ						
	0	0.5	1	1.5	2	2.5	3
Abilene	0.01	0.01	0.06	0.48	0.42	0.03	0.00
LogLogistic	0.57	0.41	0.02	0.00	0.00	0.00	0.00
LogNormal	0.59	0.39	0.02	0.00	0.00	0.00	0.00

This result is understandable, as LogNormal and LogLogistic is a random-based distribution, considering Normal and Logistic processes. This observation implies that even though these distributions fit the data correctly, they do not reproduce the system's *self-similar* dynamics. Therefore, we understand that using these distributions models for simulation and reproducing real network scenarios may not be accurate. However, a random stochastic nature is even harder to control; therefore, planning to random traffic is better when addressing network capacity, as rare scenarios may appear without a pattern.

Finally, for a pictorial analysis, we can see Figure 4.3 with various origin-destination traffic values. At first, we pick the traffic with the highest Statistical Complexity value (Fig. 4.3a) this traffic contains the most amount of information content, i.e., *patterns*; we observe that this time-series has a close to periodic nature. Figure 4.3b shows the traffic time-series with the lowest Shannon Entropy, which implies the most ordered traffic flow. Figure 4.3c exhibits the traffic time-series with the highest Shannon Entropy; therefore, the most chaotic traffic with the least amount of existing patterns. These figures ensure our expectations, and we can see that network traffic is quite complicated, and it contains many distinct dynamics that a simple distribution fitting does not capture.

4.5 Discussion

Many studies discuss the existing self-similar and complex dynamics of network traffic. To our best knowledge, there is no discussion that these traffics are more complex than any mathematical model can describe. Currently, flow-based models using packet-trains are the most used approaches with ON- and OFF-periods. Heavy-tailed distributions typically describe these periods [48]. In this paper, we show that this may be an inaccurate approach that considers "random" traffic patterns. Our work uses the Causality Complexity Entropy Plane, enhancing network traffic characterization with a simple and efficient method. This characterization exhibits the differences between the traffic gener-



(a) Highest Statistical Complexity (b) Lowest Shannon Entropy (c) Highest Shannon Entropy and lowest Complexity

Figure 4.3: Traffic time series for the origin-destination flow. These Shannon Entropy and Statistical Complexity results were calculated with $D = 5$ and $\tau = 1$.

ated using the distributions and the real dataset. It is clear that even though a correlated nature is present in real traffic, these distribution fittings do not reproduce this. The distributions we evaluate are usually applied to generate both ON- and OFF- periods in packet-train models. Although the discussion of OFF-periods models in comparison with real data is left out, we expect to explore the characteristics of more complex packet-train models traffic in our next steps.

Understanding the dynamics of these time-series is critical when trying to predict the best approach to optimize a network that knows its current state, but it can not ensure the proper future scenario. Besides that, this work will guide our future simulated scenarios to avoid considering unrealistic traffic. The proposal of a model that captures the structure and dynamics of real traffic time-series remains an open-issue. Nevertheless, the calculations of Statistical Complexity and Shannon Entropy are quite efficient, giving us the possibility of evaluating traffic in real-time with a sliding window approach. Alongside recent breakthroughs of programmables networks can be an exciting scenario for our discovery, as we can calculate these information theory quantifiers when monitoring networks through APIs and reprogramming the network configuration based on the results.

Chapter 5

Conclusions

Computer networks are complex systems composed of many devices that regulate and determine how the network functions. Observing the network behavior can be tricky, as we have a considerable amount of data that is continually changing and increasing. The current state-of-the-art machine-learning models still can not fully model a computer network adequately. This often happens because these systems' features, such as topology or traffic, have enough complexity to become a problem of itself. In this work, we believe information-theoretic solutions can low the burden of complexity in many of these problems. Instead of developing a solution for monitoring topology or traffic time-series, we can instead monitor how the information-theory quantifiers change within a certain period.

Even though monitoring is quite feasible, changing the network behavior based on data can be tricky, as many devices have closed software that does not inter-operate amongst them. Software-defined networking attempts to solve this issue by bringing the network configuration from hardware to software and allowing even more flexible monitoring solutions.

Nonetheless, in this, we conducted an extensive analysis of network traffic and topologies. In summary, our contributions are:

- We propose the Network Fisher Information Measure for characterizing the network topology.
- We carried an extensive analysis of many network models to show the feasibility of monitoring network topologies through information-theory quantifiers.
- The identification of a network as *random*, *small-world* or *scale-free* can ensure the development of fault-management solution, as we know the failures to expect in each of these models.
- We evaluate network traffic time-series using the Complexity-Entropy plane.

- We show that the usage of heavy-tailed distributions in traffic generation can be misleading, and we confirm the literature discussion upon it.

Finally, we believe the information-theoretic framework applied for the monitoring and evaluation of Computer Networks can bring many advantages to the current state-of-the-art. These quantifiers are sensitive to many phase transitions within these systems, being very prone to develop solutions of failure detection and even prediction.

Unfortunately, we were unable to develop information-theoretic solutions for current SDNs infrastructure. Many of our attempts were frustrated by many instabilities of the open-source platforms. Due to the novelty of SDN and the many challenges expected when trying to communicate with network devices using only software, we understand these limitations. Although we hope that as the community gains strength and matures its knowledge soon, our analysis's applications can be many.

References

[1] Adamic, L. A., and Huberman, B. A. Power-law distribution of the world wide web. *Science 287*, 5461 (2000), 2115–2115.

[2] Albert, R., Jeong, H., and Barabási, A.-L. Diameter of the world-wide web. *nature 401*, 6749 (1999), 130–131.

[3] Albert, R., Jeong, H., and Barabási, A.-L. Error and attack tolerance of complex networks. *nature 406*, 6794 (2000), 378–382.

[4] Aquino, A. L., Ramos, H. S., Frery, A. C., Viana, L. P., Cavalcante, T. S., and Rosso, O. A. Characterization of electric load with Information Theory quantifiers. *Physica A: Statistical Mechanics and its Applications 465* (2017), 277–284.

[5] Aquino, A. L. L., Cavalcante, T. S. G., Almeida, E. S., Frery, A. C., and Rosso, O. A. Characterization of vehicle behavior with information theory. *European Physical Journal B 88*, 10 (2015).

[6] Bakhshaliyev, K., Canbaz, M. A., and Gunes, M. H. Investigating characteristics of internet paths. *ACM Transactions on Modeling and Performance Evaluation of Computing Systems (TOMPECS) 4*, 3 (2019), 1–24.

[7] Bandt, C., and Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. *Physical Review Letters 88*, 17 (apr 2002), 174102.

[8] Bar-Joseph, Z., Gifford, D. K., and Jaakkola, T. S. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics 17*, suppl_1 (2001), S22–S29.

[9] Barabási, A.-L. *Network science*. Cambridge University Press, 2016.

[10] Barabási, A.-L., and Albert, R. Emergence of scaling in random networks. *Science 286*, 5439 (1999), 509–512.

[11] Benson, T., Akella, A., and Maltz, D. A. Network traffic characteristics of data centers in the wild. *Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC* (2010), 267–280.

[12] Bianconi, G., and Barabási, A.-L. Competition and multiscaling in evolving networks. *EPL (Europhysics Letters) 54*, 4 (2001), 436.

[13] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. Complex networks: Structure and dynamics. *Physics Reports 424*, 4-5 (2006), 175–308.

[14] Bolotin, V., Coombs-Reyes, J., Heyman, D., Levy, Y., and Liu, D. Ip traffic characterization for planning and control. In *Proc. ITC* (1999), vol. 16, pp. 425–436.

[15] Brandes, U. Optimal leaf ordering of complete binary trees. *Journal of Discrete Algorithms 5*, 3 (2007), 546–552.

[16] Burda, Z., Duda, J., Luck, J.-M., and Waclaw, B. Localization of the maximal entropy random walk. *Physical review letters 102*, 16 (2009), 160602.

[17] Canbaz, A. M. *Internet Topology Mining: from Big Data to Network Science*. PhD thesis, University of Nevada, Reno, 2018.

[18] Canbaz, M. A., Thom, J., and Gunes, M. H. Comparative analysis of internet topology data sets. *2017 IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS 2017*, May (2017), 635–640.

[19] Colizza, V., Flammini, A., Maritan, A., and Vespignani, A. Characterization and modeling of protein-protein interaction networks. *Physica A: Statistical Mechanics and its Applications 352*, 1 (2005), 1–27.

[20] Dorogovtsev, S. N., and Mendes, J. F. F. Evolution of networks with aging of sites. *Physical Review E 62*, 2 (2000), 1842.

[21] Duffield, N., Lewis, J., O'Connell, N., Russell, R., and Toomey, F. Predicting quality of service for traffic with long-range fluctuations. In *Proceedings IEEE International Conference on Communications ICC'95* (1995), vol. 1, IEEE, pp. 473–477.

[22] Erdős, P., and Rényi, A. On random graphs. *Publicationes Mathematicae 6* (1959), 290–297.

[23] Erdös, P., and Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci 5*, 17-61 (1960), 43.

[24] Faloutsos, M., Faloutsos, P., and Faloutsos, C. On power-law relationships of the internet topology. *ACM SIGCOMM computer communication review 29*, 4 (1999), 251–262.

[25] Freitas, C. G., Aquino, A. L., Ramos, H. S., Frery, A. C., and Rosso, O. A. A detailed characterization of complex networks using information theory. *Scientific reports 9*, 1 (2019), 1–12.

[26] Friendly, M. Corrgrams: Exploratory displays for correlation matrices. *The American Statistician 56*, 4 (2002), 316–324.

[27] Garrett, M. W., and Willinger, W. Analysis, modeling and generation of self-similar vbr video traffic. *ACM SIGCOMM computer communication review 24*, 4 (1994), 269–280.

[28] Gilbert, E. N. Random graphs. *The Annals of Mathematical Statistics 30*, 4 (1959), 1141–1144.

[29] Grossglauser, M., and Bolot, J.-C. On the relevance of long-range dependence in network traffic. *IEEE/ACM transactions on networking 7*, 5 (1999), 629–640.

[30] Humphries, M. D., and Gurney, K. Network 'small-world-ness': a quantitative method for determining canonical network equivalence. *PloS One 3*, 4 (2008), e0002051.

[31] Ikehara, K., and Clauset, A. Characterizing the structural diversity of complex networks across domains. *arXiv preprint arXiv:1710.11304* (2017).

[32] Krapivsky, P. L., Redner, S., and Leyvraz, F. Connectivity of growing random networks. *Physical Review Letters 85*, 21 (2000), 4629.

[33] Kurose, J. F. *Computer networking: A top-down approach featuring the internet, 3/E*. Pearson Education India, 2005.

[34] Lakhina, A., Crovella, M., and Diot, C. Characterization of network-wide anomalies in traffic flows. In *Proceedings of the 2004 ACM SIGCOMM Internet Measurement Conference, IMC 2004* (2004), pp. 201–206.

[35] Lakhina, A., Papagiannaki, K., Crovella, M., Diot, C., Kolaczyk, E. D., and Taft, N. Structural analysis of network traffic flows. *Performance Evaluation Review 32*, 1 (2004), 61–72.

[36] Lamberti, P., Martin, M., Plastino, A., and Rosso, O. Intensive entropic non-triviality measure. *Physica A: Statistical Mechanics and its Applications 334*, 1-2 (mar 2004), 119–131.

[37] Leland, W. E., Taqqu, M. S., Willinger, W., and Wilson, D. V. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on networking 2*, 1 (1994), 1–15.

[38] Lin, J. Divergence Measures Based on the Shannon Entropy. *IEEE Transactions on Information Theory 37*, 1 (1991), 145–151.

[39] López-Ruiz, R., Mancini, H., and Calbet, X. A statistical measure of complexity. *Physics Letters A 209*, 5-6 (dec 1995), 321–326.

[40] Mandelbrot, B. Self-similar error clusters in communication systems and the concept of conditional stationarity. *IEEE Transactions on Communication Technology 13*, 1 (1965), 71–90.

[41] Mandelbrot, B. B., and Taqqu, M. S. *Robust R/S analysis of long run serial correlation*. IBM Thomas J. Watson Research Division, 1979.

[42] Martin, M. T., Plastino, A., and Rosso, O. A. Generalized statistical complexity measures: Geometrical and analytical properties. *Physica A: Statistical Mechanics and its Applications 369*, 2 (sep 2006), 439–462.

[43] Massey Jr, F. J. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association 46*, 253 (1951), 68–78.

[44] Milgram, S. The small world problem. *Psychology today 2*, 1 (1967), 60–67.

[45] Molloy, M., and Reed, B. A critical point for random graphs with a given degree sequence. *Random structures & Algorithms 6*, 2-3 (1995), 161–180.

[46] Newman, M. *Networks*. Oxford university press, 2018.

[47] Nucci, A., Sridharan, A., and Taft, N. The problem of synthetically generating IP traffic matrices: Initial recommendations. *Computer Communication Review 35*, 3 (2005), 19–31.

[48] Olivier, P., and Benameur, N. Flow level ip traffic characterization. In *Teletraffic Science and Engineering*, vol. 4. Elsevier, 2001, pp. 25–36.

[49] Park, K., and Willinger, W. *Self-Similar network traffic and performance evaluation*. Wiley & Son, 2000.

[50] Pastor-Satorras, R., and Vespignani, A. Epidemic spreading in scale-free networks. *Physical review letters 86*, 14 (2001), 3200.

[51] Pirenne, H. *Medieval Cities: Their Origins and the Revival of Trade-Updated Edition*. Princeton University Press, 2014.

[52] Rosso, O. A., Larrondo, H. A., Martin, M. T., Plastino, A., and Fuentes, M. A. Distinguishing noise from chaos. *Physical Review Letters 99*, 15 (oct 2007), 154102.

[53] Sánchez-Moreno, P., Yánez, R., and Dehesa, J. Discrete densities and Fisher information. In *Proceedings of the 14th International Conference on Difference Equations and Applications. Difference Equations and Applications. Istanbul, Turkey: Bahçesehir University Press* (2009), pp. 291–298.

[54] Schieber, T. A., Carpi, L., Díaz-Guilera, A., Pardalos, P. M., Masoller, C., and Ravetti, M. G. Quantification of network structural dissimilarities. *Nature Communications 8* (2017), 13928.

[55] Shannon, C. E. A Mathematical Theory of Communication. *Bell System Technical Journal 27*, 3 (jul 1948), 379–423.

[56] Siganos, G., Faloutsos, M., Faloutsos, P., and Faloutsos, C. Power laws and the as-level internet topology. *IEEE/ACM Transactions on networking 11*, 4 (2003), 514–524.

[57] Solomonoff, R., and Rapoport, A. Connectivity of random nets. *The Bulletin of Mathematical Biophysics 13*, 2 (1951), 107–117.

[58] Taqqu, M. S., Willinger, W., and Sherman, R. Proof of a fundamental result in self-similar traffic modeling. *ACM SIGCOMM Computer Communication Review 27*, 2 (1997), 5–23.

[59] Travers, J., and Milgram, S. The small world problem. *Psychology Today 1*, 1 (1967), 61–67.

[60] Varet, A., and Larrieu, N. Realistic network traffic profile generation: Theory and practice. *Computer and Information Science 7* (2014), 1–16.

[61] Vishwanath, K. V., and Vahdat, A. Swing: Realistic and responsive network traffic generation. *IEEE/ACM Transactions on Networking 17*, 3 (2009), 712–725.

[62] Vishwanath, K. V., and Vahdat, A. Swing: Realistic and responsive network traffic generation. *IEEE/ACM Transactions on Networking 17*, 3 (2009), 712–725.

[63] Ward, L. M., and Greenwood, P. E. 1/f noise. *Scholarpedia 2*, 12 (2007), 1537. revision #137265.

[64] Watts, D. J., and Strogatz, S. H. Collective dynamics of 'small-world'networks. *Nature 393*, 6684 (1998), 440.

[65] Wiedermann, M., Donges, J. F., Kurths, J., and Donner, R. V. Mapping and discrimination of networks in the complexity-entropy plane. *Physical Review E 96*, 4 (2017), 042304.

[66] Willinger, W., Taqqu, M. S., Sherman, R., and Wilson, D. V. Self-similarity through high-variability: Statistical analysis of Ethernet lan traffic at the source level. *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, SIGCOMM 1995 5*, 1 (1995), 100–113.

[67] Yu, Y., Li, X., Leng, X., Song, L., Bu, K., Chen, Y., Yang, J., Zhang, L., Cheng, K., and Xiao, X. Fault management in software-defined networking: A survey. *IEEE Communications Surveys & Tutorials 21*, 1 (2018), 349–392.

[68] Zanin, M., Zunino, L., Rosso, O. A., and Papo, D. Permutation entropy and its main biomedical and econophysics applications: a review. *Entropy 14*, 8 (2012), 1553–1577.

[69] Zufiria, P. J., and Barriales-Valbuena, I. Entropy characterization of Random Network Models. *Entropy 19*, 7 (jun 2017), 321.