Chapter 6
Characterization and Coarsening of Autonomous System Networks: Measuring and Simplifying the Internet

Alberto Garcia-Robledo
Cinvestav-Tamaulipas, Mexico

Arturo Diaz-Perez
Cinvestav-Tamaulipas, Mexico

Guillermo Morales-Luna
Cinvestav-IPN, Mexico

ABSTRACT

This Chapter studies the correlations among well-known complex network metrics and presents techniques to coarse the topology of the Internet at the Autonomous System (AS) level. We present an experimental study on the linear relationships between a rich set of complex network metrics, to methodologically select a subset of non-redundant and potentially independent metrics that explain different aspects of the topology of the Internet. Then, the selected metrics are used to evaluate graph coarsening algorithms to reduce the topology of AS networks. The presented coarsening algorithms exploit the k-core decomposition of graphs to preserve relevant complex network properties.

INTRODUCTION

Recent years have witnessed the rise of Network Science, defined as the study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena (Committee on Network Science for Future Army Applications, 2005), thanks to: (1) the modernization and automation of data acquisition techniques using computer systems, (2) the growing availability of ever-increasing computing resources that enable the analysis of large volumes of data, (3) the availability

DOI: 10.4018/978-1-4666-9964-9.ch006
and openness of rich databases in different disciplines, and (4) the growing interest in holistic approaches to explain the behavior of complex systems as a whole (Albert & Barabási, 2002). Complex networks have been used to model the macroscopic structure of massive technological phenomena, including the topology of the Internet at the Autonomous System (AS) level.

An AS is a group of host IP addresses, known as routing prefixes, that share common routing policies. AS’s interact with each other through a massive network of links to form an AS Network (ASN) that communicates dozens to millions of hosts around the world. From a technological perspective, the measurement of the Internet can be considered a Big Data problem that typically involves the analysis of the Internet topology for the discovery (and inference) of AS connections. ASN links, viewed as a whole, reveal important details of the functional organization and evolution of the Internet (Cho, 2012). As the Internet has evolved, it has achieved the characteristics of a complex network, which can be used to model the macroscopic structure of massive technological phenomena.

A variety of existing Internet measurement initiatives propose complex network metrics to characterize different topological aspects of ASN’s for a wide variety of purposes (Angeles et. al., 2007). Unfortunately, the selection of a definitive set of metrics is hampered by the existence of a wide variety of redundant metrics that unintentionally explain the same aspects of complex networks. There are efforts that study the pair-wise correlation between metrics to select a subset of non-redundant metrics (Costa et. al., 2010; Li et. al., 2011; Jamakovic & Uhlig, 2008; Bounova & de Weck, 2012; Filkov et. al., 2009). However, it is not clear if existing results based on complex network datasets from different domains are valid for the characterization of ASN’s. The characterization of an ASN should be performed in terms of non-redundant metrics, i.e. metrics that hold little correlation, inferred from measurements on ASN datasets characterized through size-independent metrics, as the authors will later observe in this Chapter.

On the other hand, the massive size of complex networks like the Internet topology introduces the need for mechanisms that reduce the size of real-world graphs for faster processing. Given a complex network, can the graph be quickly zoomed out? Is there a smaller equivalent representation of a large graph that preserves its topological properties, such as its degree distribution and its average degree? Graph coarsening methods combine nodes to shrink large graphs while preserving topological properties of interest.

It has been reported that the topology of the Internet shows a core (Carmi et. al., 2007), integrated of well-connected hubs that represent the backbone of the Internet. The $k$-core decomposition defines a hierarchy of nodes that enables the differentiation between core and non-core components of a network. The authors will take advantage of this differentiation to design a new graph coarsening framework that preserve relevant complex network properties.

The objective of this Chapter is two-fold: (1) to present a data-driven methodology to obtain non-redundant complex networks metrics, and (2) to apply the methodology on a practical scenario: the simplification of the Internet graph at the AS level.

The rest of this Chapter is organized as follows. First, the graph coarsening and metrics selection problems are defined, and their relationship is discussed. Later, a variety of degree, clustering, distance, centrality, and scaling complex network metrics are reviewed. Then, the correlations among the presented metrics on evolving Internet networks are presented. Non-redundant complex network metrics should provide roughly the same amount of information about ASN’s than the full set of metrics. To test this claim, the researchers assess clusters of ASN’s through Principal Component Analysis (PCA).

Later, the authors exploit non-redundant complex network metrics, the concept of node centrality and the coreness of graphs, to develop and evaluate coarsening algorithms based on the $k$-core decomposition