Towards a Topology Generator Modeling AS Relationships

Xenofontas Dimitropoulos
Georgia Tech
fontas@ece.gatech.edu

George Riley
Georgia Tech
riley@ece.gatech.edu

Dmitri Krioukov
CAIDA
dima@caida.org

Ravi Sundaram
Northeastern
koods@ccs.neu.edu

1 Introduction

The development of realistic topology generators is a problem that has attracted significant attention the last few years. There exist a number of competing approaches to construction of random graphs reproducing important properties of real networks. Unfortunately, all the existing approaches do not capture an inherent aspect of many real networks: links in real networks represent relationships, which can be of several types reflecting different properties. Such link type inhomogeneity is present in the AS topology of the Internet, which is a intermix of customer-to-provider (c2p), peer-to-peer (p2p) and sibling-to-sibling (s2s) relationships. It also exists in social networks (different types of social relationships), in biological networks (different types of protein interactions) and in many other real networks. Node relationships not only provide additional information on the network topology structure, but more significantly, they represent an indispensable aspect of real networks, and this aspect needs to be properly modeled for many practical applications. Consider for example the case of using synthetic AS topologies to simulate a new routing protocol. In order to realistically simulate routing in a synthetic AS topology, we need to take into account how policies affect routing decisions. However, this is not possible with existing topology generators since they solely model the network connectivity properties without incorporating any insight about the relationships between nodes. In this work we address this problem by introducing a framework for modeling different relationships found in real networks. We use our framework to model AS relationships in the Internet and to develop an AS topology generator that generates AS graphs with realistic connectivity properties as well as realistic AS relationships.

2 Modeling AS relationships

To represent different types of relationships we use a graph G with edges labeled with one of T possible colors c_j, $1 \leqslant j \leqslant T$. Colors reflect edges with different relationships or, more generally, properties. For example, the c2p, p2p and s2s relationships of an AS graph can be represented with three different colors. Each edge in the graph G is either *undirected* or *directed*. An edge is undirected when its color represents a symmetric relationship like p2p, whereas it is directed when its color represents an asymmetric relationship like c2p. Thus, the graph G can effectively represent networks with different types of symmetric or asymmetric node relationships. We call such graphs *annotated graphs* and introduce the following definitions: 1) the $c_{j_1}^{un}$-degree of a vertex is the number of adjacent undirected edges with color c_{j_1}. 2) the $c_{j_2}^{in}$-degree ($c_{j_3}^{out}$-degree) of a vertex is the number of adjacent in-edges (out-edges) with color c_{j_2} (c_{j_3}). We collectively denote these degrees as c_j^*-degrees.

Most of the current state-of-the-art topology generators attempt to reproduce the degree distribution of the modeled network. The degree distribution captures only the diversity of degrees in a graph. For our topology generator, we choose to reproduce the following properties: 1) The degree distribution. 2) The c_j^*-degree distributions, which are a natural generalization of the degree distribution. 3) The correlations between different c_j^*-degrees. To illustrate the significance of the c_j^*-degree correlations consider the following example: it is well-known that large tier-1 ASes have a large number of customers, i.e., large customer-degree, and no providers, i.e., zero provider-degree. We capture such patterns by reproducing the c_j^*-degree correlations. These three properties are collectively captured by the joint distribution (JD) of the c_j^*-degrees.

We model the JD of an AS topology using *copulas* [3]. A copula is a statistical tool that models the dependence structure of a joint distribution and separates this structure from the marginal distributions. In our case, the dependence structure of JD is the correlations between different c_j^*-degrees, whereas the marginal distributions of JD are the c_j^*-degree distributions. To find the appropriate copula and marginal distributions we collect eight AS topologies from RouteViews in six month intervals between 01/18/2002 and 07/18/2005. We infer c2p and p2p relationships using the heuristics in [2, 1]. We represent c2p and p2p relationships using two colors. Moreover, c2p edges are directed and p2p edges undirected, which yields three c_j^*-degrees: the customer-, provider-, and peer-degrees of an AS. Using the collected topologies we first construct an empirical copula that effectively captures the exact correlations we observed in our data. Then, we fit the customer-degree distribution using the probability weighted moments method with a generalized Pareto distribution (GPD). Similarly, we fit the peer-degree distribution with a pair of GPDs, one for the body and one for the tail of the distribution. We find that the provider-degree distribution has almost invariant quantiles at the first six points of the distribution, i.e., zero to five providers, but we were not able to fit the tail of the distribution.

Finally, we construct a random graph with a given number of vertices N using the following process: 1) We first use the models of the previous paragraph to generate the numbers of expected customer, peer and provider adjacencies for each of the N vertices. 2) Then, for each vertex and for each expected adjacency we introduce an colored stub, which is undirected for peer adjacencies, in-directed for customer adjacencies and out-directed for provider adjacencies. 3) Finally, we conduct a random matching between stubs of the same color and of compatible directions to get the final graph. If we cannot find a match for a stub, which may happen if the sum of in-degrees is not equal to the sum of out-degrees, we ignore it.

3 Evaluation

We use our method to generate a set of synthetic AS topologies with c2p and p2p relationships. Then, we compare the customer-, provider- and peer-degree distributions of the synthetic topologies with the corresponding distributions of real AS topologies annotated with inferred AS relationships. We find that our topology generator accurately reproduces the observed customer-, provider- and peer-degree distributions. We also compare the JDs of the synthetic topologies with the corresponding JDs of the real topologies and verify that our synthetic topologies preserve the observed correlations between customer-, provider- and peer-degrees. To illustrate a part of our experiments, in Figure 1 we plot the complimentary cumulative distribution function (CCDF) of the customer-, provider- and peer-degrees. The empty points show the distributions observed in a real AS topology collected on 07/18/2005, whereas the solid points depict the same distributions in a synthetic topology of the same size. The close match between the true and the generated distributions demonstrates the effectiveness of our approach.

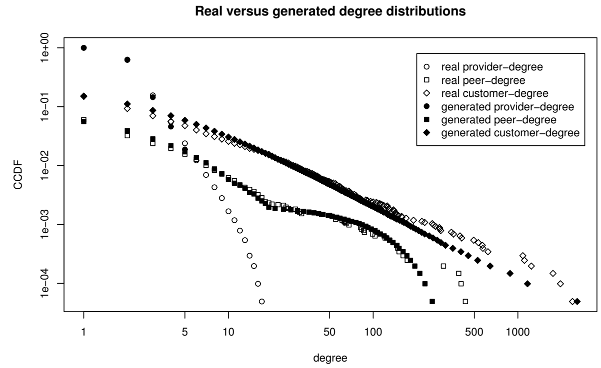

Figure 1: Customer-, peer- and provider-degree distributions of a real graph and of a synthetic graph of the same size.

4 Conclusions

In this work we address the problem of generating synthetic graphs with realistic node relationships. We introduce a framework for modeling node relationships observed in real networks and for synthetic reproduction of important properties associated with these relationships. We focus on modeling the AS relationships in the Internet and demonstrate first results of our topology generator that accurately reproduces c2p and p2p relationships.

References

[1] X. Dimitropoulos, D. Krioukov, M. Fomenkova, B. Huffaker, kc claffy, and G. Riley. Inferring AS relationships: Peering links resurected. In *under submission*, 2005.

[2] X. Dimitropoulos, D. Krioukov, B. Huffaker, kc claffy, and G. Riley. Inferring AS relationships: Dead end or lively beginning? In *Proceedings of 4th Workshop on Efficient and Experimental Algorithms (WEA' 05)*, May 2005.

[3] R. B. Nelson. An introduction to copulas. *Springer-Verlag Lecture Notes in Statistics*, 139:216, 1999.