Random Network Models Based on Density

Glencora Borradaile[1], Theresa Migler[2], and Gordon Wilfong[3]

[1] School of EECS, Oregon State University
[2] California Polytechnic State University, San Luis Obispo
[3] Bell Labs

1 Introduction

We present random graph models based on the *density distribution* of a graph. The *density decomposition* of a graph is a hierarchical partition of the vertices into regions of uniform density [1]. The density decomposition is is unique in the sense that a given network has exactly one density decomposition. The number of vertices in each region defines a distribution of the vertices according to the density of the region to which they belong, that is, a *density distribution*. Although density is closely related to degree, the density distribution of a particular network is not necessarily similar to the degree distribution of that network. For example, in many synthetic networks, such as those generated by popular network models (e.g. preferential attachment and small worlds), the density distribution is very different from the degree distribution. On the other hand, in real networks, the density and degree distributions are measurably similar [1].

Based on this observation, that real networks have similar density and degree decompositions and that many synthetic networks have dissimilar density and degree decompositions, we develop an abstract model, that, given a particular density distribution, produces a network having that density distribution. Applied naïvely, given a density distribution of a real network, this model generates networks with realistic average path lengths (average number of hops between pairs of vertices) and degree distributions; that is similar to the given real network. In addition to having short average path lengths, large-scale, real networks also tend to have high *clustering coefficients* [3]. The clustering coefficient of a vertex v is the ratio of the number of pairs of neighbors of v that are connected to the number of pairs of neighbors of v; the clustering coefficient of a network is the average clustering coefficient of its vertices. Our model, naïvely applied, unfortunately, but not surprisingly, results in networks with very low clustering coefficients. However, we show that applying the abstract model in a more sophisticated manner, using ideas from the small world model of Watts and Strogatz [5], results in much higher clustering coefficients suggesting that real networks may indeed be *hierarchies of small worlds*. Our hierarchies of small worlds specification is just one way to tune our abstract model; our model is quite flexible allowing for the easy incorporation of other network generation techniques.

A key observation that distinguishes our model from other network models is our *qualitatively different* treatment of vertices. That is, our model begins by assigning vertices to levels of the density decomposition. This sets vertices qualitatively apart from each other; for example, a vertex assigned to a dense level of the decomposition is treated very differently from a vertex assigned to a sparse level of the decomposition.

2 Results

The density decomposition partitions a graph into *rings*, $R_0, R_1, \ldots R_k$. These rings divide the graph into regions of uniform density in the following sense: The graph obtained by identifying the vertices in $\cup_{j \geq i} R_j$ and deleting the vertices in $\cup_{j<i} R_j$ will always have density between $i-1$ and k [1]. The density decomposition is summarized by the density distribution $\rho = (|R_0|, |R_1|, \ldots |R_k|)$, the number of vertices in each ring.

Given a density distribution ρ, we can generate a network with n vertices having this density distribution using the following *abstract model*:

Input: density distribution ρ and target size n
Output: an network G with n vertices and density distribution ρ
1: Initialize G to be a network with empty vertex set V
2: **for** $i = |\rho|, \ldots, 0$ **do**
3: $\quad R_i \leftarrow$ set of ρ_i vertices
4: $\quad$ add R_i to V
5: $\quad$ **for** each vertex $v \in R_i$ **do**
6: $\quad\quad$ connect i vertices of V to v

Using this generic model, we propose two specific models, the *random density distribution model* (RDD) and the *hierarchical small worlds model* (HSW), by specifying how the neighbors are selected in Step 6 of the abstract model.

For the RDD model, we choose i vertices from V uniformly at random in Step 6. We use this to model four varied networks: AS, DBLP, EMAIL, and TRUST (Figure 1). For each given network, we generate another random network having the given network's number of vertices and density distribution. Remarkably, although we are only specifying the distribution of the vertices over a density decomposition, the resulting degree distributions of the RDD networks are very similar to the original networks they are modeling. Further, the average path lengths of the RDD networks are realistic, within 2 of the average path lengths of the original networks However, the clustering coefficients of the RDD networks are unrealistically low (Figure 1).

We provide a more sophisticated model which addresses the unrealistically low clustering coefficients of the RDD model by generating a small world (SW) [5] network among the vertices of each ring of the density decomposition. In the hierarchical small worlds (HSW) model, for vertices in R_i, we create a small world network on $|R_i|$ vertices: Order R_i cyclically; for each $v \in R_i$, with probability p, connect each of the i vertices before v in this order to v; if $c \leq i$ neighbors for v are selected in this way, select $i-c$ vertices uniformly at random from $\cup_{j>i} R_j$ (or R_i if this is the densest ring) and connect these to v. Clearly, this is a specification of neighbor selection for Step 6 of the abstract model. As with the SW model, the HSW model provides a similar trade-off between clustering coefficient and average path length, although the relationship is less strong. In addition, we observe a similar trade-off between p and degree distribution: as p increases, the degree distribution approaches that of the original network. This is in sharp contrast to the SW model which have degree distributions far from the original (normal vs. close to power law).

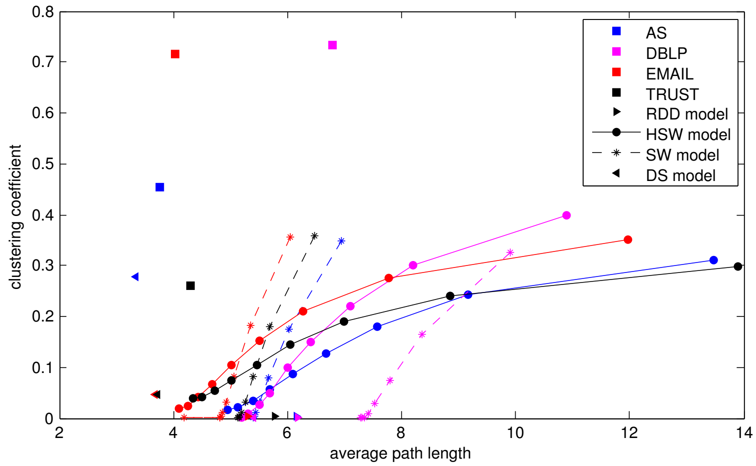

Fig. 1. We model four varied networks: In the AS network vertices represent autonomous systems and two autonomous systems are connected if there is a routing agreement between them [7]. In the DBLP network, vertices represent computer scientists and two computer scientists are connected if they have at least one co-authored paper [6]. In the EMAIL network vertices represent Enron email addresses and two addresses are connected if there has been at least one email exchanged between them [2]. In the TRUST network vertices represent `epinions.com` members and two members are connected if one trusts the other [4]. We give the clustering coefficient versus average path length for RDD, HSW, SW (Small Worlds) and DS (Degree Sequence) models. For HSW we let $p = 0.1, 0.2, \ldots, 0.9$. Colors indicate the network being modelled. Squares denote the data for the original networks.

References

1. Glencora Borradaile, Theresa Migler, and Gordon Wilfong. Density decompositions of networks. In *Complex Networks IX*, volume 9, pages 15–26, 2018.
2. Bryan Klimt and Yiming Yang. Introducing the Enron Corpus. In *First Conference on Email and Anti-Spam*, 2004.
3. Mark Newman. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA, 2010.
4. Maryam Tahajod, Azadeh Iranmehr, and Nasim Khozooyi. Trust management for semantic web. In *Computer and Electrical Engineering, 2009. ICCEE '09. Second International Conference*, volume 2, pages 3–6, 2009. http://snap.stanford.edu/data/.
5. Duncan Watts and Steven Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):409–10, 1998.
6. Jaewon Yang and Jure Leskovec. Defining and evaluating network communities based on ground-truth. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, MDS '12, pages 3:1–3:8, New York, NY, USA, 2012. ACM. http://snap.stanford.edu/data/.
7. Yu Zhang. Internet AS-level Topology Archive. http://irl.cs.ucla.edu/topology/.