Exploring generic scale-free networks*

Winfried Just† Hannah Callender‡ M. Drew LaMar§

December 23, 2015

This module is a companion module to the one on the preferential attachment model at this web site.[1] Here we study in more detail networks that are generic for a given network size and a given exponent of a power-law degree distribution. We explore predicted structural properties of such networks both mathematically and with IONTW.

1 Generic scale-free networks

This module is a continuation of our module [4] on the preferential attachment model.

Recall that the degree distribution of a given graph obeys a *power law* if

$$q_k = c_\gamma k^{-\gamma}, \tag{1}$$

where q_k is the probability that a randomly chosen node has degree k, and c_γ and γ are positive constants. As this formula makes sense only for $k > 0$, we will tacitly assume that the graph contains no isolated nodes, that is, $q_0 = 0$. Graphs with a power-law distribution of degrees are often called *scale-free networks.*

This phrase needs to be handled with care. Since $q_k > 0$ for all $k > 1$, Equation (1) could be literally true only if there were infinitely many nodes in the graph.[2] For graphs with finitely many nodes, (1) can be satisfied only *approximately.* If this is the case, we will write that the graph is *approximately a scale-free network.*

A number of constructions of approximately scale-free networks have been proposed in the literature. The *preferential attachment model* of Barabási and Albert [2] was studied in some detail in our module [4]. Here we focus on some properties that are common to all types of approximately scale-free networks and on properties of *generic* scale-free networks that are constructed by the procedure outlined in our module [3].

†Department of Mathematics, Ohio University, Athens, OH 45701 E-mail: mathjust@gmail.com

‡University of Portland E-mail: callende@up.edu

§The College of William and Mary E-mail: drew.lamar@gmail.com

[1] `http://www.ohio.edu/people/just/IONTW/`

[2] For graphs with infinitely many nodes it is not immediately clear how one would define the probability q_k, but this is an entirely different issue far beyond the scope of these modules. The recent research monograph [5] outlines a possible approach.

Recall how the latter construction works: For given $\gamma > 1$ and N, we construct a random graph $G_{SF}(N, \gamma)$ whose degree distribution approximates (1) by first randomly drawing degrees k_i according to the distribution (1), then attaching k_i "stubs" to node i, and finally connecting them randomly to form the edges. If N is sufficiently large, this will give a random network that is approximately scale-free and otherwise generic. We will refer to $G_{SF}(N, \gamma)$ as a *generic scale-free network* and drop the adverb "approximately" for easier readability.

2 Mathematical explorations of scale-free distributions and generic scale-free networks

2.1 Scale-free degree distributions

Let us assume that we are given a network with degree distribution (1). Since Equation (1) cannot be literally true in any network of finite size N, all estimates that we derive in this section need to be treated with some caution. They will at best be approximation for actual, finite-size networks with roughly scale-free degree distributions when the network size N is very, very large.

For such very, very large N, we can estimate c_γ from (1) by observing that the probabilities q_k must add up to 1.

$$1 = \sum_{k=1}^{\infty} c_\gamma k^{-\gamma} = c_\gamma \zeta(\gamma), \tag{2}$$

where $\zeta(\cdot)$ is the famous *Riemann zeta function.* For $\gamma \leq 1$ the series in (2) diverges and $\zeta(\gamma)$ is undefined. We will always assume that $\gamma > 1$. In this case, (2) implies that $c_\gamma \approx \frac{1}{\zeta(\gamma)}$.

For example, $\zeta(2) = \frac{\pi^2}{6} \approx 1.6449$, $\zeta(2.1852) \approx 1.5$, and $\zeta(3) \approx 1.2021$. It follows that the proportions of nodes with degree 1 in power-law distributions with parameters γ should satisfy

$$q_1 = \frac{1}{\zeta(\gamma)1^{-\gamma}} = \frac{1}{\zeta(\gamma)}. \tag{3}$$

The mean degree of a node i can be expressed as

$$\langle k \rangle = \sum_{k=1}^{\infty} c_\gamma k k^{-\gamma} = \sum_{k=1}^{\infty} c_\gamma k^{-\gamma+1}. \tag{4}$$

If $\gamma \leq 2$, the sum diverges. For finite networks with approximately such degree distributions this implies that the mean degree will increase without bound as the network size $N \to \infty$. For $\gamma > 2$, the mean degree will satisfy

$$\langle k \rangle = \frac{\zeta(\gamma - 1)}{\zeta(\gamma)}. \tag{5}$$

Recall from our module on the preferential attachment model that for many real-world networks approximately scale-free degree distributions with parameter $2 < \gamma < 3$ have been observed. However, for $\gamma < 3$ the variance of the degree distribution increases without bound as $N \to \infty$.

Exercise 1 *Estimate the variance of degree distribution for $\gamma > 3$ and show that if $\gamma \leq 2$, then the variance increases without bound as $N \to \infty$.*

Let us repeat our words of caution: The above calculations of means and variances strictly speaking apply only to infinite scale-free networks where (1) holds exactly. For finite N the degree distribution can only be approximately scale-free. Empirically studied networks may show such approximate distributions, but usually the fit is good only for k in an intermediate range, not for very large k or for k that are very close to 1. In this case we cannot estimate the mean or variance based on γ alone. For example, a study of the World Wide Web [1] found approximately scale-free distributions for the number of links pointing to a given web page j, with $\gamma_{in} \approx 2.1$, and the number of links embedded in a given web page i, with $\gamma_{out} \approx 2.45$. If we were to treat (5) literally, we would get different means for the number of incoming and the number of outgoing links. But a moment's though shows that these two means must be exactly equal. We just cannot compute them based on γ alone for the approximately scale-free distributions.

The variances are theoretically infinite for both γ_{in} and γ_{out}, but our calculations for the solution of Exercise 1 suggest that the variance of incoming links should be a lot larger than for outgoing links. This makes perfect sense as there is a physical limit on how many links one can embed in a single web page i, but the numbers of links that point to a given page j may wildly vary according to the page's popularity. Even for approximate power law distributions with $2 < \gamma \leq 3$, a smaller value of γ is a fairly reliable indicator of greater variance.

2.2 Hubs and maximum degree

Nodes with very large degrees in approximately scale-free networks are often called *hubs.* This terminology has its origin in the study of networks of airline connections, where the nodes represent airports, and an edge represents the existence of a direct flight between two airports. This informal definition of "hubs" does not specify how large the degree of a node needs to be so that it would qualify. We can give ourselves some flexibility by defining the *set of K-hubs* as $H(K) = \{i : k_i \geq K\}$. The relative size of $H(K)$ will be approximately equal to $P(k_i \geq K)$ as computed from (1).

Exercise 2 *Let K be fixed and let i be a randomly chosen node. Use an integral to estimate $P(k_i \geq K)$.*

It follows from Exercise 2 that the relative sizes of $H(K)$ will decrease gradually with K, so that there may be no obvious choice for the threshold for K above which we should consider $H(K)$ to represent the set of hubs.

Exercise 3 *Use your solution of Exercise 2 to estimate the median value of the maximum degree in $G_{SF}(N, \gamma)$.*

In Erdős-Rényi random graphs $G_{ER}(N, \lambda)$ the largest degree is expected to grow slower than $\ln(N)$. In contrast, your solution for Exercise 3 will show that the maximum degree in a generic scale-free network $G_{SF}(N, \gamma)$ scales like a power of N. This property is often colloquially referred to as "scale-free degree distributions have fat tails."

2.3 Connected components of $G_{SF}(N, \gamma)$

In contrast to approximately scale-free networks that are obtained from the preferential attachment model, generic scale-free networks are disconnected and have many small components.

Exercise 4 *Show that for large N the graph $G_{SF}(N, \gamma)$ will be disconnected with probability very close to 1 and estimate the mean number of connected components of size 2.*

3 Exploring generic scale-free networks with IONTW

Open IONTW, press **Defaults**, move the speed control slider to the extreme right, and change the following parameter settings:

network-type $\rightarrow$ **Generic Scale-free**
num-nodes: 300
lambda: 2.5

For this type of networks, the input parameter **lambda** controls the network parameter γ. After pressing **New** you would expect a network $G_{SF}(300, 2.5)$ to appear in the **World** window. This may or may not happen; sometimes you will see a blank **World** window and in the **Command Center** an error message

```
Degree sequence is not realizable as an undirected graph!
```

You will see this message quite often during your work in this module. It is due to the fact that the underlying algorithm first randomly draws a supposed degree sequence from the specified scale-free distribution and then checks whether there actually is a graph with this degree sequence. If you see the error message, simply press **New**, repeatedly if need be, until a network appears in the **World** window.

The network that eventually appears in the **World** window may at first not look that much different from, say, an Erdős-Rényi random graph. But look at the **Network Metrics** plot. It will show you the histogram of an approximately scale-free degree distribution: There will be a large number of nodes with degree 1, and then the height of the bars will rapidly decrease. Perhaps there will be an occasional uptick in the histogram; after all, the distribution will only be approximately scale-free.

The number on the horizontal axis that shows the maximum degree + 1 should be fairly substantial; record it for future comparison. Now press **Metrics** and inspect the data in the **Command Center**. You may notice some interesting things, but for now just record the mean degree.

Let us compare the scale-free network with an Erdős-Rényi network with the same mean degree and number of nodes. Choose

network-type $\rightarrow$ **Erdos-Renyi**
lambda: [the mean degree that you looked up for the scale-free network]

Press **New** and look at the histogram for the degree distribution in the **Network Metrics** plot. It will be dramatically different. Degree 2 should occur with the highest frequency and the maximum degree should be much smaller than for the instance of $G_{SF}(300, 2.5)$.

For approximately scale-free networks, the maximum degree would typically be on the order of $N^{1/(\gamma-1)}$, which works out to ≈ 45 for $N = 300$ and $\gamma = 2.5$. Due to random effects, the value you found for your network may be twice or only half as large, but it should be in the double or low triple digits. This contrasts sharply with the predictions for Erdős-Rényi random graphs $G_{ER}(N, \lambda)$ whose the largest degree is expected to be less than $\ln(N)$. Since for $\gamma > 1$ the function $N^{\frac{1}{\gamma-1}}$ grows much faster than $\ln(N)$, in power-law distributions the probability of extremely high values is much, much larger than in Poisson distributions with the same mean degrees. This is what the phrase "power-law distributions have fat tails" means.

Press **Metrics**, enlarge the **Command Center** by clicking on the double-arrow icon, and compare the other metrics for the two networks. Both graphs will have many connected components. But you may notice that the clustering coefficients are larger and the mean and maximum distances in the largest connected component are smaller in the scale-free network than in the Erdős-Rényi network.

The node with the highest degree in a scale-free network will definitely be a "hub," but there will be more than just one hub. Let us look at a smaller example. Choose

network-type $\rightarrow$ **Generic Scale-free**
num-nodes: 100
lambda: 2.5

Create a **New** network, and look at the picture in the **World** window. Can you make out the hubs? How many of the nodes would you classify as hubs? How are your observations related to Exercise 2?

Now let us study in more detail the properties of generic scale-free networks $G_{SF}(N, \gamma)$ with $N = 200$. Let us first list some numerical predictions that follow from our work in Section 2 for parameters $\gamma = 2.1, 2.5, 3, 5$.

- $\gamma = 2.1$
  - Proportion of nodes with degree 1: $q_1 = 0.6409$
  - Mean degree: $\langle k \rangle = 6.7840$
  - Maximum degree: ≈ 142

  - Number of connected components of size 2: ≈ 6

- $\gamma = 2.5$
  - Proportion of nodes with degree 1: $q_1 = 0.7454$
  - Mean degree: $\langle k \rangle = 1.9474$
  - Maximum degree: ≈ 34
  - Number of connected components of size 2: ≈ 29

- $\gamma = 3$
  - Proportion of nodes with degree 1: $q_1 = 0.8319$
  - Mean degree: $\langle k \rangle = 1.3684$
  - Maximum degree: ≈ 13
  - Number of connected components of size 2: ≈ 51

- $\gamma = 5$
  - Proportion of nodes with degree 1: $q_1 = 0.9644$
  - Mean degree: $\langle k \rangle = 1.0438$
  - Maximum degree: ≈ 3
  - Number of connected components of size 2: ≈ 89

We encourage students to figure out a way to use batch processing for these explorations but will describe here a simpler *ad hoc* procedure that gives some preliminary insights.

Set the speed control slider to the extreme right and press the `Clear` icon on the **Command Center** bar to delete all previously recorded data.

Now create 3 instances of $G_{SF}(200, \gamma)$ for each choice of $\gamma = 2.1, 2.5, 3, 5$. After creating each instance, press **Metrics** and separately record the maximum degree and the number of nodes with degree 1 that are shown in the **Network Metrics** plot. To facilitate data analysis, be sure to create exactly 3 instances for each parameter choice as the input parameters will not be shown in the **Command Center**.

Exercise 5 *Enlarge the* ***Command Center*** *with the double-arrow icon and analyze your data. Are the results roughly consistent with the theoretical predictions after taking into account the amount of variability between networks? For which properties did you find good agreement with the theoretical prediction, for which properties did you find large discrepancies?*

In your solution of Exercise 5 you will most likely find a good fit with the theoretical predictions in some cases and discrepancies in other cases. This is to be expected as $N = 200$ is not particularly big and the theoretical predictions were derived for very large N. You will most likely also observe large differences between networks with the same parameters.

Since your sample size is very small, these random fluctuations may have a large impact on the averages you compute. You could address the latter problem by analyzing large batches of networks, but our primary goal here is illustration of some patterns, not systematic research. You should be able to observe though that the general pattern is consistent with the theoretical predictions for networks $G_{SF}(N, \gamma)$: As γ increases, the expected mean and maximum degrees decrease, and the proportion of nodes with degree 1 as well as the number of connected components increase.

References

[1] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Internet: Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999.

[2] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[3] Winfried Just, Hannah Callender, and M. Drew LaMar. Exploring contact patterns between two subpopulations. `http://www.ohio.edu/people/just/IONTW/`.

[4] Winfried Just, Hannah Callender, M. Drew LaMar, and Ying Xin. The preferential attachment model. `http://www.ohio.edu/people/just/IONTW/`.

[5] László Lovász. *Large networks and graph limits.* American Mathematical Society, 2012.