The preferential attachment model*

Winfried Just† Hannah Callender‡ M. Drew LaMar§ Ying Xin¶

December 24, 2015

Many empirically studied networks have approximately so-called *power-law* or *scale-free* degree distributions. In Section 1 we formally define such distributions and explore some of their properties. We also introduce and briefly compare two methods for constructing random networks with approximately power-law degree distributions: generic scale-free networks and the preferential attachment model. In Sections 2 and 3 we explore disease transmission on networks that are obtained from the preferential attachment model and implications for designing effective vaccination strategies.

1 Scale-free networks

In regular graphs all nodes have the same degree and in Erdős-Rényi random graphs all nodes have degrees that are very close to the mean degree $\langle k \rangle$. In contrast, many empirically studied real-world networks contain a few nodes whose degree is very large relative to $\langle k \rangle$ (so-called *hubs*), while the vast majority of nodes have degrees that are substantially smaller than $\langle k \rangle$. The particular degree distributions in these examples often resemble a *power law* aka *scale-free* degree distribution. Since such distributions have been found in many large real-world networks, such as the World Wide Web, neural and social networks, in nets of citations of scientific papers, etc. [5], the study of scale-free networks is a vibrant topic of current research.

While many excellent sources such as [3] give popular accounts of this topic, misconceptions persist. We begin this section with a mathematically rigorous definition of power-law distributions and scale-free networks and describe some of their properties. Next we compare two different methods for constructing random scale-free networks: *generic networks* with a given degree distribution and the famous *preferential attachment model.* In the remaining sections of the model we will focus on models of disease transmission on contact networks that have been constructed according to the preferential attachment model. In

†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

¶Ohio University E-mail: yx123812@ohio.edu

a companion module we will explore in more detail properties of generic scale-free models and of disease transmission on them.

1.1 Scale-free degree distributions and scale-free networks

Consider a degree distribution $\bar{q} = (\dots, q_k, \dots)$, where q_k denotes the probability that a randomly chosen node i has degree $k_i = k$. This degree distribution obeys a *power law* if

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

where 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, $q_0 = 0$.

Note that (1) implies that we always have

$$\frac{q_k}{q_{2k}} = \frac{c_\gamma k^{-\gamma}}{c_\gamma (2k)^{-\gamma}} = 2^\gamma, \tag{2}$$

regardless of whether k is a single-digit number or on the order of hundreds or thousands. The ratio is the same at all scales of measuring k. This is one reason why 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[1]. 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.* For many large real-world networks, such approximately scale-free degree distributions have been reported, usually with parameters γ such that $2 < \gamma < 3$ (see Chapter 3 of [8] for reprints of some classical case studies and [3] for a nontechnical description of many examples).

It is worth taking a few minutes to consider how one might detect that a given graph has approximately a power law degree distribution. If (1) were to hold exactly, then a log-log plot of q_k as a function of k would show all data points on a straight line with slope $-\gamma$. A log-log plot of the actual degree distribution of a given graph will not usually have all points on a straight line, but for many empirically observed real-world networks the log-log plot shows surprisingly accurate straight-line approximations for k in some intermediate range. The approximation cannot be accurate for very large k, and it will often not be valid for k that are close to 1. Still, if the approximation is very good for k in a broad intermediate range, the graph can be considered to be an (approximately) scale-free network. The parameter γ can then be estimated from the slope of the regression line.

Figure 1 shows two examples of approximately scale-free networks. In the left panel, the regression line was found for degrees $k \leq 7$; in the right panel, the regression line was found for degrees $k \leq 20$. Within this range, we get a very good fit. The data for higher values of k do not match this pattern.

[1]For graphs with infinitely many nodes it is not immediately clear how one would define the probability q_k. While this issue is far beyond the scope of this module, interested students may want to consult the advanced monograph [7] for a promising approach to studying what happens in the limit when $N \to \infty$.

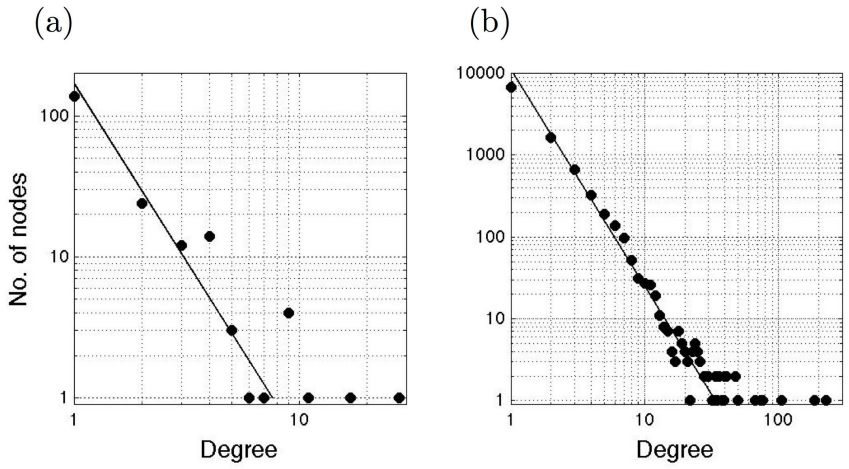

Figure 1: Log-log plots for degree distributions in approximately scale-free networks. (a) A network with $N = 100$ nodes. The slope of the regression line is -2.51975. (b) A network with $N = 10,000$ nodes. The slope of the regression line is -2.65123.

For a given γ, one can calculate the value of the constant c_γ in (1) and the mean degree $\langle k \rangle$. In the forthcoming module *Exploring generic scale-free networks* we will guide you through these calculations. The formulas that one can obtain in this way are interesting from the mathematical point of view. But since these calculations assume that (1) is *exactly* true, they are not directly applicable to the study of real-world networks that are approximately scale-free. For a given real-world network G it will only be true that for some positive constant c_G we have $q_k \approx c_G k^{-\gamma}$ for k that are neither too large nor too close to 1. The value of c_G will depend on the particular network and may be very different from the theoretical value c_γ that we will derive in the module *Exploring generic scale-free networks.*

In approximately scale-free networks there are many more nodes with very large degrees than in Erdős-Rényi random graphs that were studied in Module 6 of the online appendix of [6] and our module *Exploring Erdős-Rényi random graphs with IONTW* at this web site[2] (see Exercise 2 below). This property is colloquially described by saying that power-law distributions have *fat tails.* 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

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

airline connections, where the nodes represent airports, and an edge represents the existence of a direct flight between two airports. As you will see, hubs play an important role in the spread of pathogens on a network.

However, in a given network there is usually no obvious cut-off for the minimum degree that a node needs to have so as to qualify as a hub. We might 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 1 *Let K be fixed and let i be a randomly chosen node. Use an integral to estimate $P(k_i \geq K)$.*

It follows 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. Choosing a reasonable cut-off value for the intended application of the model is part of the art of modeling.

Exercise 2 *Use your solution of Exercise 1 to estimate the median value of the maximum degree in $G_{SF}(N, \gamma)$.*

1.2 Generic scale-free networks

For given $\gamma > 1$ and N, we can use the procedure outlined in Section 2 of the module *Exploring contact patterns between two subpopulations* at this web site[3] to construct a random graph $G_{SF}(N, \gamma)$ whose degree distribution approximates (1): First randomly draw degrees k_i according to the distribution (1), then attach k_i stubs to i, and finally connect 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.

The properties of $G_{SF}(N, \gamma)$ will be explored in detail in the forthcoming module *Exploring generic scale-free networks.* Let us just mention here that for large N the graph $G_{SF}(N, \gamma)$ will be disconnected with probability very close to 1 and we expect a significant proportion of the nodes to reside in connected components of size 2.

1.3 The preferential attachment model

Generic random scale-free graphs $G_{SF}(N, \gamma)$ are convenient mathematical abstractions that can be relatively easily studied. It is not clear to what extent their properties mirror those of approximately scale-free networks that are found in the real world. Take, for example, connectedness. The airline transportation network that represents direct flights as edges between airports (nodes) is presumably connected. At least that is its purpose. In this respect, the airline network is very different from from $G_{SF}(N, \gamma)$, for which we predicted a large number of connected components of size 2. On the other hand, a sexual contact

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

network presumably would contain a large number of such components. At least this is the idea of marriage.

Moreover, the construction of $G_{SF}(N, \gamma)$ may or may not reflect the actual processes by which connections (edges) are formed in nature. As an alternative to starting from an assumed degree distribution as in the construction of $G_{SF}(N, \gamma)$, we might try to model the *process* by which actual connections are formed, and then mathematically derive the expected properties of the resulting network, including the degree distribution. The *preferential attachment model* of Barabási and Albert [4] does exactly that. It is motivated by the growth of the World Wide Web. One can model this network by considering individual web pages as nodes i whose degree k_i is the number of other web pages that contain a link to page i. Since links are directed, the resulting structure is a directed graph rather than a graph, but we will ignore this difference in our discussion of the model.

The size of the WWW keeps growing all the time. Web pages that already have many links pointing to them, such as Google or YouTube, accumulate new links at a higher rate than less popular ones, such as the home page `http://www.ohio.edu/people/just/IONTW/` of this module. In other words, the rich keep getting richer. Barabási and Albert [4] modeled this pattern by assuming that we start with a small network of m_0 nodes and then successively add new nodes, one at a time. Each new node will be connected to m of the previous nodes in such a way that the probability of the new node connecting to an already existing one i is proportional to k_i in the current network. They then showed that after a long time of growth the network should exhibit approximately a power-law degree distribution as in (1) with parameter $\gamma = 3$.

For specificity, here we will always assume that the m_0 initial nodes form a complete graph. The resulting networks with N total nodes will be denoted by $G_{PA}(N, m_0, m)$.

1.4 Comparing generic scale-free networks with networks obtained from the preferential attachment model

The fact that the preferential attachment model can only generate networks with $\gamma \approx 3$ imposes limits on its applicability. Empirical studies of actual networks often indicate values of γ that are strictly between 2 and 3. For example, for the distribution of the number of links that point to a given web page, the paper [2] reported a value $\gamma \approx 2.1$. In [1] the authors point out that processes other than attachment of a new node play a role in network formation. For example, links can be subsequently broken and/or be replaced by new links between existing nodes. The preferential attachment model can be modified so that by carefully choosing the parameters that govern these processes of rewiring, the value of γ can be tuned to a specified values that are different from 3. Several other modifications of the preferential attachment model have been studied. For readers who wish to learn more about this topic we recommend [8] as an excellent introduction to the literature. Here we will restrict ourselves to working with the original version of the preferential attachment model.

It is interesting to compare the philosophies behind the constructions of $G_{PA}(N, m_0, m)$ and $G_{SF}(N, 3)$. Consider a randomly chosen node i. In $G_{SF}(N, \lambda)$, we randomly draw k_i

from a power-law distribution and attach k_i stubs to node i. For a given stub at node j, the probability that this stub will be combined with a stub for node i to form an edge $\{i, j\}$ is proportional to k_i. Thus the degree of node i can be considered a measure of *attractiveness* of node i in $G_{SF}(N, \lambda)$. This is actually very similar to what happens in the construction of $G_{PA}(N, m_0, m)$, except that the attractiveness of node i for node j is its degree in the subgraph that is induced by nodes $1, \ldots, j-1$, not its degree in $G_{PA}(N, m_0, m)$ itself. However, since in the construction of $G_{PA}(N, m_0, m)$ the rich keep getting richer, a node that has already a relatively high degree at the time when node j gets added will very likely also have a relatively high degree at the end of the construction. Thus the difference between the two models is rather subtle.

Let us look at the two constructions from the point of view of the growth of the WWW. If you design a new web page j, your decision of whether or not to embed a link to an existing page i will be driven by two factors: How *interesting* page i is, and whether you have actually visited it. Your probability of having visited it may be assumed roughly proportional to the number of links pointing towards it (the *popularity* of the web page). Whether or not you find the page interesting is a subjective judgement, but it may be reasonable to assume that more links will point to web pages that more people find interesting. Let us swallow our pride: Most surfers of the web find YouTube more interesting than our pages on network models of disease transmission.

If we were to model the WWW by $G_{SF}(N, \lambda)$, we would in effect assume that all web pages have an intrinsic attractiveness and then generate random links accordingly. If we choose $\gamma = 2.1$, this might give us a picture of the WWW that conforms even more closely than $G_{PA}(N, m_0, m)$ to the degree distribution reported in [2]. But it would remain entirely unclear why the distribution of "intrinsic attractiveness" should follow a power law, or why the parameter γ should be 2.1. The generic model may give us a realistic network $G_{SF}(N, 2.1)$, but it does not *explain* the mechanisms by which the distribution arises. In contrast, in the construction of $G_{PA}(N, m_0, m)$, each of the initial nodes starts out equally attractive, and there is no assumption that some nodes may be intrinsically more interesting than others. Subsequent changes in attractiveness are entirely driven by the current popularity. The approximately power-law distribution of the nodes *emerges* as the network grows, and the postulated mechanism is sufficient to explain the distribution and to *predict* the value $\gamma = 3$.

Since the philosophical differences between the two constructions are very subtle, we should expect that $G_{PA}(N, m_0, m)$ and $G_{SF}(N, 3)$ will have rather similar properties, and in fact they do in many respects. In at least one aspect though there is a striking difference: The construction of $G_{PA}(N, m_0, m)$ implies that the resulting graph will always be connected, while $G_{SF}(N, 3)$ usually has many connected components of size 2.

This seems to make $G_{SF}(N, 3)$ a better candidate for modeling sexual contact networks, but it is not immediately clear why this should be the case. Let us consider for example the following network G_{EB}. Its nodes represent all people who ever lived and were sexually active on the Island of Eternal Bliss. This island was first settled in 1632 by a group of m_0 refugees from the Thirty Year's War. An edge in $G_{EB}(N)$ connects two islanders if,

and only if, they ever had sexual contact. In this model no rewiring of edges takes place. The network only changes when new sexual partnerships are formed. One would naturally assume that the islanders choose their sexual partners based on attractiveness. This process of network growth certainly looks similar to the preferential attachment model.

Exercise 3 *Explain which aspects of the growth of G_{EB} may differ substantially from the assumptions of the preferential attachment model. In particular, explain which of these aspects are likely to make G_{EB} disconnected. Are there some important aspects in which G_{EB} would be expected to differ from $G_{SF}(N, \lambda)$ as well as from $G_{PA}(N, m_0, m)$?*

2 Exploring the preferential attachment model with IONTW

Let us see how the preferential attachment model works. Open IONTW, click **Defaults**, and change the following parameter settings:

network-type $\rightarrow$ **Preferential Attachment**
num-nodes: 20
lambda: 4
d: 2

Move the speed control slider to a very slow setting; adjust for comfortable viewing as needed. Press **New**, sit back, relax, and enjoy the movie. When the movie ends, press **Scale** to get a better view of the network.

Exercise 4 *How are the input parameters **lambda** and **d** related to the parameters of the network $G_{PA}(N, m_0, m)$?*

Now use what you have learned in the previous exercise to set up creation of a network $G_{PA}(20, 1, 1)$.

The graph $G_{PA}(N, 1, 1)$ is always a tree and thus contains exactly $N - 1$ edges (see, for example, Exercise 9.42 of the online appendix to [6]). Thus the mean degree in a tree with N nodes is equal to $\langle k \rangle = \frac{2(N-1)}{N}$; for large N it is close to but a little smaller than 2. For small N we get the following values:

N	1	2	3	4	5	6	7	8	9
$\langle k \rangle$	0	1	1.33	1.5	1.6	1.67	1.71	1.75	1.78

(3)

Exercise 5 *Upon pressing **New** you will see how the graph $G_{PA}(N, 1, 1)$ is built by adding one node at a time. Record for each step the current degree of the node to which the new node gets attached. Here the phrase "current degree" refers to the degree before the new node gets attached.*

Does the new node always get attached to the node with the current highest degree or at least to a node with above-average degree? Explain in your own words in what sense the attachment procedure is "preferential."

Now let us consider the degree distributions in networks that are obtained from the preferential attachment model. In the default setting **plot-metric** → **Degree Distribution** you should see in the **Network Metrics** plot a distribution that is similar to the one predicted by (1), with nodes of degree 1 having the highest frequency, and q_k getting smaller as k increases. Since the preferential attachment model predicts $\gamma \approx 3$, in view of (2) we would expect that $q_2 \approx \frac{q_1}{8}$ and $q_4 \approx \frac{q_1}{64}$. Moreover, the solution of Exercise 2 indicates that the maximum degree should scale like $\sqrt{N}$, that is, should roughly increase by a factor of 1.4 whenever we double N. The value of the maximum degree in a given network can be found in the **Network Metrics** plot that shows $1 + \max k_i$ on the right of the horizontal axis for the default setting **plot-metric** → **Degree Distribution**.

For $N = 20$ this pattern will not show up clearly though, due to finite size effects. Also, to get somewhat reliable data, we should average over several instances of networks. Move the speed control slider to the extreme right, and choose

num-nodes: 40, 80, 160

For each choice of **num-nodes**, create three networks with **New**. For each network, record the maximum degree and the approximate percentages of nodes with degrees $k = 1, 2, 4$. You can estimate the latter by moving the cursor to the top of the relevant bars in the **Network Metrics** plot and recording the value shown on the vertical axis. Do the results appear consistent with the predictions of the preferential attachment model?

When trying to answer the above question, you may have found that samples of size 3 are not large enough to draw reliable conclusions. Let us investigate larger batches of networks.

For each of the settings

infection-prob: 0
end-infection-prob: 1
num-nodes: 60, 120, 240, 480

define and run a batch processing experiment with 100 repetitions using the following specifications:

Measure runs using these reporters:

```
max [count link-neighbors] of turtles
count turtles with [count link-neighbors = 1]
count turtles with [count link-neighbors = 2]
count turtles with [count link-neighbors = 4]
```

Setup commands:

```
new-network
```

Exercise 6 *Analyze the output files by computing the means of the proportions of nodes with degrees $k = 1, 2, 4$ and the mean of the maximum degree for each case. Do these means reasonably well conform to the predictions that we derived in Subsection 1.1?*

As we already explained in Subsection 1.1, even for relatively large N you cannot get perfect agreement with the degree distribution (1) that gives positive probabilities for all $k \geq 1$.

The distribution will only be approximately scale-free. Most likely, there will be occasional upticks in the histogram that you see in the **Network Metrics** plot. Figure 1 is actually based on the degree sequence of two network $G_{PA}(N, 1, 1)$ that were created with our software and shows this pattern. Note also that the slope of the regression line is -2.51975 for $N = 100$ and -2.65123 for $N = 10,000$ nodes. The discrepancy from the theoretically predicted value $-3 = -\gamma$ of [4] appears to be due to finite size effects. The theoretical predictions were made under the assumption of very large network size N, and a value of $N = 10,000$ may be too small to get a good approximation to $\gamma = 3$. Note that the approximation is somewhat better for the larger of the two values of N that we explored.

3 Disease transmission on networks obtained from the preferential attachment model

In scale-free networks, there are a few highly connected nodes (the hubs), while most nodes have very small degrees. This naturally suggests the following questions:

(a) Will this structure speed up or slow down the spread of infections compared to what the corresponding compartment level models would predict?

(b) Will it make major outbreaks more or less likely?

(c) If major outbreaks do occur, will they tend to have larger or smaller sizes than in corresponding compartment-level models?

(d) Which vaccination strategies will work best if the contact network is approximately scale-free?

Here we will explore questions (a)–(c) for contact networks that are constructed under the preferential attachment model, and in the next section we will do the same for question (d).

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

model-time $\rightarrow$ **Discrete**
infection-prob: 0.1519
end-infection-prob: 1
network-type $\rightarrow$ **Preferential Attachment**
num-nodes: 200
lambda: 4
d: 4
auto-set: On

This sets up a next-generation SIR-model on networks $G_{PA}(200, 4, 4)$.

Use **New** to create a network and then click **Metrics**. Scroll up the **Command Center** to look up `R0` and verify that $R_0 = 1.2$ for our model.

If everything looks fine, we are ready to **Go**. Do a few runs to see whether we get major or minor outbreaks. Most likely, you will see some outbreaks that you would confidently classify as major and as well as some that you would confidently classify as minor.

Now, let us have a glance at the corresponding compartment-level models. Change the settings to:

infection-prob: 0.006031
network-type → **Complete Graph**

Click **New**, then click **Metrics**. Look up `R0` and convince yourself that we still have $R_0 = 1.2$. Click **Go** and do a few simulation runs. What do you observe? Do major outbreaks become more likely to occur? Are your observations consistent with your expectations?

Now, it's time to get serious and run batches of 100 simulations, one each for both of the parameter settings for which we did preliminary explorations. Use the following specifications:

Measure runs using these reporters:

```
count turtles with [removed?]
ticks
```

Setup commands:

```
new-network
```

Exercise 7 *In the output files of both of your experiments, sort the data by the column labeled* `count turtles with [removed?]` *and find a meaningful grouping of your data into those corresponding to minor and major outbreaks.*

(a) Record the proportion of simulations in which you observed a major outbreak.

(b) Calculate the mean duration of major outbreaks.

(c) Calculate the mean final size for the group of major outbreaks.

(d) How do the results differ between the two models? Do the observed differences appear to be statistically significant or could they be attributed to random fluctuations?

Let us increase the value of R_0 and see whether we can get similar observations. Explore the following parameter settings:

infection-prob: 0.25
network-type → **Preferential Attachment**

and

infection-prob: 0.0099246231
network-type → **Complete Graph**

For both settings, create a network with **New**, click **Metrics**, and verify that $R_0 = 1.975$.

Exercise 8 *With each set of settings, run a batch of 100 simulations using the same specifications as before. Repeat the data analysis of Exercise 7 for the outputs.*

Now we can try to answer the questions that were mentioned at the beginning of this section:

(a) Will this structure speed up or slow down the spread of infections compared to what the corresponding compartment-level models would predict?

(b) Will it make major outbreaks more or less likely?

(c) If major outbreaks do occur, will they tend to have larger or smaller final sizes than in corresponding compartment-level models?

Well, based on the observations in our simulations, we cannot answer the latter two questions for sure, because for different parameter settings we did not observed statistically significant differences or a consistent pattern of differences between the two types of model for the two choices of R_0. However, based on our results, it seems like the mean duration of major outbreaks on contact networks obtained from the preferential attachment model is always shorter than that of outbreaks on the corresponding compartment-level models. Thus, a reasonable guess is that this structure will speed up the spread of infectious compared to what the corresponding compartment-level models would predict.

Exercise 9 *Give an intuitive explanation for this answer to question (a).*

We already know that in networks constructed according to the preferential attachment model, some nodes will have much larger degrees than others, while in complete graphs that correspond to compartment-level models, every node has the same degree. Thus one may want to investigate for the preferential attachment model whether the degree of the index case will make a difference. What would be your guess?

Let's use simulations to check whether you guessed correctly. Change the settings as follows:

infection-prob: 0.25
network-type $\rightarrow$ **Preferential Attachment**
min-deg: 9

The last option ensures that the index case will always be chosen among the nodes with degree at least 9.

Exercise 10 *Run a batch of 100 simulations with the specifications we used before, and analyze your output by answer the same questions as in Exercise 7.*

Wow! With the degree of the index case lager than 8, an outbreak is very likely to be major, almost for sure!

Thus, we really don't want the index case to be a node with relatively higher degree. Also, the subsequent spread of the infectious disease may crucially depend on secondary infections caused by nodes with high degrees. Wouldn't it be great if these nodes don't contribute to propagating the infection?

This gives us a clue of how one might design a good vaccination. We will investigate in the next section how well this might work.

3.1 Vaccination strategies for scale-free contact networks

The networks $G_{PA}(N, m_0, m)$ are connected. As a warm-up for this section, let us explore what removal of a certain fraction of randomly chosen nodes does to the connectedness of the network.

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

network-type $\rightarrow$ **Preferential Attachment**
num-nodes: 50
lambda: 2
d: 2

Press **New** to create an instance of $G_{PA}(50, 2, 2)$, and then press **Scale** for better viewing in the **World** window.

The graph you see is connected. Let us try to make it disconnected by randomly removing some nodes. Choose the parameter settings

set-state-to $\rightarrow$ **Removed**
set-state-by $\rightarrow$ **Number of nodes**
num/frac: 10
min-deg: 0

Press **Set** to immunize 10 randomly chosen nodes and see what happens. Repeat a few times by pressing first **Reset** and then **Set**. Do the subgraphs induced by the green nodes tend to stay connected except for the occasional isolated node?

Now let us immunize 25 randomly chosen nodes. Set

num/frac: 25

Press **Reset** and then **Set** a few times. What do you observe?

You may see a few more isolated green nodes, but most likely the vast majority of the green nodes will still form a connected subgraph, despite the fact that we have removed half of all nodes! You have discovered a very important property:

In scale-free networks, random removal of nodes is unlikely to destroy connectedness of the largest component.

Next let us remove 10 of the hubs, where we will consider a node a hub if it has degree at least 4. Set

num/frac: 10
min-deg: 4

Press **Reset** and then **Set** a few times. What do you observe?

Now the subgraphs induced by the green nodes have a relatively large number of small connected components. You have discovered another very important property:

In scale-free networks, random removal of a certain fraction of the hubs will destroy the largest connected component.

These properties are very important if one wants to prevent transportation or communication networks with approximately scale-free degree distributions from becoming disconnected. At any given time some fraction of all internet servers is down, but this does not perceptibly affect the functionality of the internet. But a targeted attack on a significant proportion of the hubs could disrupt the internet.

Here we are not interested in *preserving* connectivity though. The purpose of targeting vaccination at a certain subset of nodes is precisely to *disrupt* the connectivity of the unvaccinated hosts as much as possible. Let us see whether we can put the properties that you just discovered to good use in designing optimal vaccination strategies for scale-free networks.

In the remainder of this section we will examine several possible vaccination strategies for $G_{PA}(200, 4, 4)$. Change the following parameter settings:

infection-prob: 0.25
end-infection-prob: 1
num-nodes: 200
lambda: 4
d: 4
auto-set: Off

This will set up a next-generation SIR-model with $\langle k \rangle = 7.9$ and $R_0 = (0.25)(7.9) \approx 2$.

First let us run a batch processing experiment to get baseline data for the situation where no control measures whatsoever are implemented. For this and all subsequent batch processing experiments in this module use

Repetitons: 100
Measure runs using these reporters:

```
count turtles with [removed?]
```

Make sure your output file names are short and distinctive so that in the subsequent data analysis you can easily see what is where. Here you may want to use something like "no-vacc."

For this first experiment, use

Setup commands:

```
new-network
ask n-of 1 turtles with [susceptible?]  [become-infectious]
```

Now run the experiment. It may take a while before it is completed. Get yourself a cup of coffee. Text your friends. Call your dad.[4]

Since $R_0 \approx 2$, the herd immunity thresholds for the compartment-level version of this model would be $HIT \approx 0.5$. Let us get a second set of baseline data for vaccinating $200HIT = 100$ randomly chosen nodes. **Edit** your previous experiment by inserting a second line in your **Setup commands** that you will then modify in subsequent experiments:

Setup commands:

[4]This message brought to you by Correebugs, KeepShort, and Revizon, proud sponsors of excellence in publishing.

```
new-network
ask n-of 100 turtles [ become-removed ]
ask n-of 1 turtles with [susceptible?]  [become-infectious]
```

Run the experiment and keep your data for subsequent analysis.

Next let us assume that we have only 40 or 20 doses of vaccine. We will test what happens if we use them for vaccinating randomly chosen nodes. **Edit** your previous experiment by changing

Setup commands:

```
new-network
ask n-of 40 turtles [ become-removed ]
ask n-of 1 turtles with [susceptible?]  [become-infectious]
```

Run the experiment and then another one for vaccinating only 20 instead of 40 randomly chosen nodes.

Now let us see what happens if we use our 20 doses of vaccine to vaccinate 20 randomly chosen *hubs.* As there is no obvious cutoff for the minimum degree above which a node would qualify as a hub, let us first consider all nodes i with degree $k_i \geq 7$ as hubs, and then all nodes i with degree $k_i \geq 10$.

Edit your previous experiment by changing

Setup commands:

```
new-network
ask n-of 20 turtles with [count link-neighbors > 6][ become-removed ]
ask n-of 1 turtles with [susceptible?]  [become-infectious]
```

Run the experiment, and then run another one with

Setup commands:

```
new-network
ask n-of 20 turtles with [count link-neighbors > 9][ become-removed ]
ask n-of 1 turtles with [susceptible?]  [become-infectious]
```

Finally, let us examine the effect of vaccinating *all* nodes i with degree $k_i \geq 10$. **Edit** your previous experiment by changing

Measure runs using these reporters:

```
count turtles with [removed?]
count turtles with [count link-neighbors > 9]
```

Setup commands:

```
new-network
ask turtles with [count link-neighbors > 9][ become-removed ]
ask n-of 1 turtles with [susceptible?]  [become-infectious]
```

Now that you have run enough experiments to bring up your caffeine levels to high mental alertness, catch up on your texting, and chat with your entire family, you are ready to focus on analyzing the data.

Exercise 11 *(a) Analyze your data. For each data set, sort the output column from lowest to highest. Try to discern a distinct gap between minor and major outbreaks. If there is no gap, record the minimum, maximum, and overall mean. If there is a gap, record the number of outbreaks that you classified as minor and the maximum number of hosts who were removed in the minor outbreaks. For the major outbreaks, record the minimum and maximum numbers, as well as the means. Express these means in terms of the proportion of unvaccinated hosts who experienced infection.*

(b) How would you describe the effectiveness of the various vaccination strategies that you explored?

References

[1] Réka Albert and Albert-László Barabási. Topology of evolving networks: local events and universality. *Physical review letters*, 85(24):5234, 2000.

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

[3] Albert-László Barabási. *Linked: How Everything is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. Basic Books, second edition, 2014.

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

[5] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Structure of growing networks with preferential linking. *Physical Review Letters*, 85:4633–4636, 2000.

[6] Winfried Just, Hannah Callender, and M Drew LaMar. Disease transmission dynamics on networks: Network structure *vs.* disease dynamics. In Raina Robeva, editor, *Algebraic and Discrete Mathematical Methods for Modern Biology*, pages 217–235. Academic Press, 2015.

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

[8] Mark Newman, Albert-László Barabási, and Duncan J Watts. *The structure and dynamics of networks*. Princeton University Press, 2006.