

# The friendship paradox\*

Winfried Just<sup>†</sup> Hannah Callender<sup>‡</sup> M. Drew LaMar<sup>§</sup>

December 23, 2015

In this module we introduce the so-called friendship paradox and illustrate how it affects disease transmission on networks that exhibit this phenomenon.

## 1 The friendship paradox

Why do your friends have more friends than you do? The question may sound offensive. We don't even know you. How can we assume than you have fewer friends than your friends have on average? Because most people do. This so-called *friendship paradox* has first been described and studied in [1]. It does seem counterintuitive: If we are talking about the *average* number of friends of *average* friends of an *average* person, shouldn't this average out to the average number of friends of an average person?

Enough loose talk about averages that makes the average person's head spin. Let's steady our thoughts with some solid mathematical definitions. Consider a graph  $G$  that represents friendships between persons numbered  $1, \dots, N$ . The degree  $k_i$  of node  $i$  represents the number of  $i$ 's friends. The “average” number of friends of a randomly chosen person can be most naturally interpreted as the mean degree  $\langle k \rangle$  that is given by

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i. \quad (1)$$

For a fixed  $i$  who has at least one friend the mean number of friends of  $i$ 's friends, denoted by  $\langle k_f \rangle^i$ , can be calculated as

$$\langle k_f \rangle^i = \frac{1}{k_i} \sum_{\{j: \{i,j\} \in E(G)\}} k_j. \quad (2)$$

For  $i$  with  $k_i = 0$  the notion of “mean number of friends of  $i$ 's friends” is meaningless. We leave  $\langle k_f \rangle^i$  undefined in this case. Let  $N_{1+}$  denote the number of nodes  $i$  of degree  $k_i \geq 1$ .

---

\*©Winfried Just, Hannah Callender, and M. Drew LaMar 2014

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

<sup>‡</sup>University of Portland E-mail: callende@up.edu

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

If  $N_{1+} \geq 1$  we can define the mean of the mean number of friends of friends of a randomly chosen node  $i$  as

$$\langle k_f \rangle = \frac{1}{N_{1+}} \sum_{\{i: k_i \geq 1\}} \langle k_f \rangle^i. \quad (3)$$

In this terminology, we can express the friendship paradox as the strict inequality

$$\langle k_f \rangle > \langle k \rangle. \quad (4)$$

Inequality (4) is a mathematically rigorous statement, but is it true? Actually, not in all graphs  $G$ . If it is, then we will write that  $G$  exhibits the friendship paradox with excess  $\langle k_f \rangle - \langle k \rangle$ .

Let us look at two illustrative examples. Open IONTW, click **Defaults**, and choose

**network-type** → **Nearest-neighbor 1**

**num-nodes:** 9

**d:** 2

Create a network by pressing **New**. The graph that you see in the **World** window is an example of a *one-dimensional nearest neighbor network* and will be denoted by  $G_{NN}^1(9, 2)$ .

**Exercise 1** Calculate  $\langle k \rangle$  and  $\langle k_f \rangle$  for the graph  $G_{NN}^1(9, 2)$ . Does this graph exhibit the friendship paradox?

Now change

**network-type** → **Nearest-neighbor 2**

**num-nodes:** 6

**d:** 1

Create a network by pressing **New**. The graph that you see in the **World** window is an example of a *two-dimensional nearest neighbor network* and will be denoted by  $G_{NN}^2(6, 1)$ .

**Exercise 2** Calculate  $\langle k \rangle$  and  $\langle k_f \rangle$  for the graph  $G_{NN}^2(6, 1)$ . Does this graph exhibit the friendship paradox?

We will rigorously prove in Section 4 that the inequality  $\langle k_f \rangle \geq \langle k \rangle$  holds in all graphs and that the strict inequality (4) holds in *most* graphs. For now, you may want to do the following exercise:

**Exercise 3** (a) Find a common-sense explanation for the fact that some graphs do exhibit the friendship paradox.

(b) Form a conjecture about sufficient and necessary conditions for the structure of graphs that do not exhibit the friendship paradox.

## 2 The friendship paradox and models of disease transmission on contact networks

The friendship paradox certainly does look surprising, but does it have anything to do with disease transmission on networks? A lot, actually. Consider a next generation *SIR*-model on a network  $G$ . In such a model,  $R_0 \approx b\langle k \rangle$ . Assume for simplicity that the exact equality  $R_0 = b\langle k \rangle$  holds and that the network size  $N$  is very large. Consider an outbreak that is started by an index case  $j^*$  in an otherwise susceptible population. Then  $R_0 = b\langle k \rangle$ .

Recall the definition of the replacement number  $R_1^{st}$  from the previous module<sup>1</sup>. It is the mean number of secondary infections that will be caused by an average host who is infectious at time 1 in state  $st$ . Whenever we use this notation, we implicitly assume that there is at least one infectious host in state  $st$ . As we have seen, even for very large population sizes  $N$ , the number  $R_1^{st}$  may be significantly smaller than  $R_0$ . The reason is that if nodes tend to have small degrees, each host who is infectious at time  $t = 1$  has at least one adjacent host who is no longer susceptible.

Let us carefully consider what happens at time  $t = 1$ : All nodes that are infectious at time  $t = 1$  are adjacent to  $j^*$ . If  $j^*$  was randomly chosen, then on average the infectious nodes at time  $t = 1$  will have  $\langle k_f \rangle$  adjacent nodes, one of whom is  $j^*$ . Thus we get the following upper bound for next-generation *SIR* models:

$$R_1^{st} \leq b(\langle k_f \rangle - 1) = R_0 \frac{\langle k_f \rangle - 1}{\langle k \rangle}. \quad (5)$$

In  $k$ -regular graphs we always have  $\langle k_f \rangle = \langle k \rangle = k$ , and the right-hand side of (5) is identical with the estimate  $R_{ub} = R_0 \frac{\langle k \rangle - 1}{\langle k \rangle}$  that you will have discovered in the module on the replacement number. For random regular graphs  $G_{Reg}(N, k)$  with sufficiently large  $N$  we then found that  $R_t^{st} \approx R_0 \frac{k-1}{k}$  for sufficiently small  $t$ . This gave a prediction of slower initial growth of an outbreak than what the uniform mixing assumption would predict for the given value of  $R_0$ .

For networks other than random regular graphs, the situation may be much more complicated. First of all  $R_0 \frac{\langle k_f \rangle - 1}{\langle k \rangle}$  may be smaller, equal to, or larger than  $R_0$ , depending on the excess in the friendship paradox for the given contact network. In the latter case, the first line of Equation (1) of our module on the replacement number will be violated!

Second, the estimate  $R_1^{st} \leq b(\langle k_f \rangle - 1)$  does not always imply that  $R_1^{st} \approx b(\langle k_f \rangle - 1)$  for sufficiently large  $N$ , not even when the graph is regular. We will examine this phenomenon in our module on clustering coefficients.

Third, we cannot automatically generalize the inequality  $R_1^{st} \leq b(\langle k_f \rangle - 1)$  to  $R_t^{st} \leq b(\langle k_f \rangle - 1)$  for all  $t > 1$ . For example,  $R_2^{st}$  depends on the mean number of friends of friends of friends of a randomly chosen index case. As we will illustrate in a later module, for some network types the latter number may significantly exceed  $\langle k_f \rangle$ .

In many types of networks though, the estimate (5) can be generalized to

---

<sup>1</sup>The replacement number, posted at <http://www.ohio.edu/people/just/IONTW/>

$$R_t^{st} \approx b(\langle k_f \rangle - 1) \tag{6}$$

for sufficiently large  $N$  and sufficiently small  $t$ .

If the approximation (6) is valid in the initial stages of an outbreak, the arguments of our module on the replacement number apply, and  $R_0 \frac{\langle k_f \rangle - 1}{\langle k \rangle}$  instead of  $R_0$  becomes a reliable predictor for the expected initial growth of an outbreak and the probability  $z_\infty$  that introduction of one index case into an otherwise susceptible population will cause only a minor outbreak. In particular, if (6) is valid and  $\langle k_f \rangle - \langle k \rangle = 1$ , we might expect the spread of diseases on such networks to closely match the predictions derived under the uniform mixing assumption. This is exactly what we observed in our explorations of next-generation models based on Erdős-Rényi networks<sup>2</sup>!

**Exercise 4** Consider a next-generation SIR-model on an Erdős-Rényi network  $G_{ER}(N, \lambda)$ . Assume that  $N$  is very large relative to  $\lambda$ .

(a) Show that  $\langle k_f \rangle - \langle k \rangle \approx 1$ .

(b) Assume an initial state with one index case in an otherwise susceptible population and let  $\varepsilon > 0$ . Show that for any given  $t > 0$  the probability  $P(|R_t^{st} - b(\langle k_f \rangle - 1)| < \varepsilon)$  approaches 1 as  $N \rightarrow \infty$  so that (6) becomes a valid approximation. Hint: It can be shown that for any given  $t$  and probability  $q < 1$ , there exists a bound  $B(t, \lambda, q)$  such that with probability at least  $q$  the total number of nodes that are no longer susceptible at time  $t$  is at most  $B = B(t, \lambda, q)$ , regardless of population size  $N$ . You may want to use this result in your argument rather than deriving it yourself.

### 3 Exploring the effect of the friendship paradox on disease transmission with IONTW

In this section we explore disease transmission on some networks that exhibit the friendship paradox with large excess.

Open IONTW, click **Defaults**, move the speed control slider to the extreme right, and choose

**network-type** → **Regular Tree**

**lambda:** 1

**d:** 9

Press **New** to create a star tree with  $N = 10$  nodes and then make it look nice by pressing **Spring**, waiting until it has taken a nice shape, pressing **Spring** again and then **Scale** to make it better fit the **World** window. Press **Labels** and recall that the *root* is labeled 0 by NETLOGO and the other nodes are numbered from 1 to 9. The tree in your **World** window is an example of a *star tree*  $G_{ST}(N)$  with  $N = 10$  nodes,  $N - 1 = 9$  leaves, and one node, the root, with degree  $N - 1 = 9$ .

---

<sup>2</sup>See modules *Exploring random regular graphs with IONTW* and *The replacement number* at this web site <http://www.ohio.edu/people/just/IONTW/>

**Exercise 5** (a) Calculate  $\langle k \rangle$  and  $\langle k_f \rangle$  as well as the excess  $\langle k_f \rangle - \langle k \rangle$  for this network.  
(b) Generalize the result of part (a) to star trees  $G_{ST}(N)$  with arbitrary numbers  $N \geq 2$  of nodes.

Let us explore a tree with  $d = 9$  but more levels. Change

**lambda: 2**

Press **New** to create a tree with  $N = 91$  nodes and then make it look nice by pressing **Spring**, waiting until it has taken a nice shape, pressing **Spring** again and then **Scale** to make it better fit the **World** window.

**Exercise 6** Calculate  $\langle k \rangle$  and  $\langle k_f \rangle$  as well as the excess  $\langle k_f \rangle - \langle k \rangle$  for this network.

Wow! The excess in the friendship paradox for each of the networks that we have explored so far is much larger than the mean degree! Does our claim that “if you are like most people, your friends have more friends than you do” still sound outrageous?

Let us explore how such a large excess might influence the spread of infectious diseases on the regular tree with 91 nodes.

Use the following parameter settings to set up a next-generation *SIR*-model:

**model-time** → **Discrete**

**infection-prob:** 0.4

**end-infection-prob:** 1

**auto-set:** **On**

Press **New** to make one node infectious. Press **Metrics** and look up and record the value of  $R_0$  in the **Command Center**. It should be clearly less than 1.

As we explained in Section 2, the mean value  $\langle R_1^{st} \rangle$  will be larger than  $R_0$  in this model. Here the mean is taken over all states  $st$  with at least 1 infectious host that can occur at time  $t = 1$  when the initial state contains exactly one infectious host in an otherwise susceptible population,

**Exercise 7** Use the results of Section 2 and Exercise 6 to calculate  $\langle R_1^{st} \rangle$  for this model.

Now set up and run a batch processing experiment for the current parameter settings following the template that is given in the instructions<sup>3</sup> on how to use our modules. Work with the following specifications:

Define a **New** experiment.

**Repetitions:** 100

**Measure runs using these reporters:**

count turtles with [removed?]

**Setup commands:**

new-network

---

<sup>3</sup>Posted at <http://www.ohio.edu/people/just/IONTW/>

**Exercise 8** *Open your output file and order the column with the header `count turtles with [removed?]` from largest to smallest. Record the maximum and the mean final sizes of the observed outbreaks.*

Now let us compare the results with those for corresponding models on contact networks that are random regular graphs  $G_{Reg}(91, 2)$  with the same number of nodes. Note that for these graphs the mean degree 2 is even slightly larger than the mean degree  $\langle k \rangle$  that you found in Exercise 6 for the regular tree of the previous batch processing experiment. This should translate into an almost identical but even slightly larger value of  $R_0$  compared with the current model. Change

**network-type** → **Random Regular**

Press **New** to create a network, then **Metrics**, look up the value of  $R_0$  in the **Command Center** and compare it with the value that you found for the previous model.

Set up and run a batch processing experiment for the current parameter settings with the following specifications:

Define a **New** experiment.

**Repetitions:** 100

**Measure runs using these reporters:**

count turtles with [removed?]

**Setup commands:**

new-network

**Exercise 9** (a) *Open your output file and order the column with the header `count turtles with [removed?]` from largest to smallest. Record the maximum and the mean final sizes of the observed outbreaks.*

(b) *Compare your findings with the ones of Exercise 8. How does the structure of the regular tree appear to influence the dynamics of the model?*

In our next example we will have  $R_0 = 0.65$ . Before we introduce the example itself, let us get a baseline idea about the predictions for an *SIR*-model with the uniform mixing assumption for this value of  $R_0$ . Change the following parameter settings:

**infection-prob:** 0.0066

**network-type** → **Complete Graph**

**num-nodes:** 100

Create a **New** network. Then press **Metrics** and look up the value of  $R_0$  for this model in the **Command Center**. It should be very close to and actually slightly larger than 0.65. Recall that for an *SIR*-model with  $R_0 < 1$  under the uniform mixing assumption only minor outbreaks are predicted. Let us see how this prediction works out in a relatively small population of size  $N = 100$ .

Set up and run a batch processing experiment for the current parameter settings with the following specifications:

Define a **New** experiment.

**Repetitions:** 100

**Measure runs using these reporters:**

count turtles with [removed?]

**Setup commands:**

new-network

**Exercise 10** *Open your output file and order the column with the header count turtles with [removed?] from largest to smallest. Record the maximum and the mean final sizes of the observed outbreaks, as well as the numbers of runs where at least 10 hosts experienced infection and of those runs where no secondary infections whatsoever occurred.*

Now let us study our second example of disease transmission on a large network that exhibits the friendship paradox. Change

**infection-prob:** 0.1

If you have not already done so, download the sample input file `degreesFP.txt` from our web site<sup>4</sup> and save it in the *same directory* where you keep IONTW. Press **Load** and open this file. The network that you will see in the **World** window is a generic graph  $G_{SQ}(100, \bar{k})$  for the degree sequence  $\bar{k}$  that specifies degree  $k_i = 2$  for each node  $i = 0, \dots, 74$  and degree  $k_i = 20$  for each node  $i = 75, \dots, 99$ . Generic graphs  $G_{SQ}(N, \bar{k})$  were defined in our module *Exploring contact patterns between two subpopulations*. You may want to press **Labels** and **Update** the **Degree Distribution** to see how the specified degree sequence relates to the picture in the **World** window.

This graph has a mean degree of  $\langle k \rangle = 6.5$ . We will see in Exercise 16 of the next section that this graph exhibits the friendship paradox with rather large excess.

Press **Metrics** and look up the value of  $R_0$  for this model in the **Command Center**. It should be equal to 0.65.

Let us run some preliminary explorations of disease transmission in this network. Press **Set** to introduce one infectious node, then **Go**. Examine the **Disease Prevalence** plot to see what happened in this outbreak. Repeat about 10 times by first pressing **Reset**, then **Set**, and then **Go**. Pay attention to both the information in the **Disease Prevalence** plot and in the **World** window. The latter will show you which node becomes initially infectious, and which nodes experience infection during the simulated outbreak. Formulate a tentative conjecture about this connection.

Now change

**min-deg:** 3

This will have the effect that the initially infectious node is randomly chosen from among the nodes that have degree larger than 2. In our network, all of these nodes have degree 20. Repeat the previous explorations for the new settings.

---

<sup>4</sup><http://www.ohio.edu/people/just/IONTW/>

**Exercise 11** *Formulate a conjecture about the relationship between the final sizes of outbreaks and the choice of the initially infectious node based on these explorations. Also write down your observations about the set of nodes that experience infection during outbreaks.*

Now let us try to confirm your conjecture of Exercise 11 with three batch processing experiments for the current parameter settings. Set up and run the first experiment with the following specifications:

Define a **New** experiment.

**Repetitions:** 100

**Measure runs using these reporters:**

count turtles with [removed?]

**Setup commands:**

```
load-from-file "degreesFP.txt"
```

```
ask n-of 1 turtles [become-infectious]
```

In the dialogue box **Run options** set

**Simultaneous runs in parallel:** 1

The specifications given above assume that you did save the file `degreesFP.txt` in the same directory where you keep `IONTW`. If you prefer saving this file in a different directory, such as the one named `examples`, you will need to modify the first line of **Setup commands** as follows:

```
load-from-file "examples/degreesFP.txt"
```

In this experiment, the initially infectious node will be chosen randomly from among all 100 nodes.

For the second experiment, choose the following specifications:

**Duplicate** the previous experiment and then **Edit** it as follows:

Choose a new suggestive **Experiment name**.

Replace

```
ask n-of 1 turtles [become-infectious]
```

with

```
ask n-of 1 turtles with [count link-neighbors > 2] [become-infectious]
```

Then run the experiment with

**Simultaneous runs in parallel:** 1

In this experiment, the initially infectious node will be chosen randomly from among the 25 nodes with degree  $> 2$ , that is, with degree 20.

For the third experiment, choose the following specifications:

**Duplicate** the previous experiment and then **Edit** it as follows:

Choose a new suggestive **Experiment name**.

Replace

```
ask n-of 1 turtles with [count link-neighbors > 2] [become-infectious]
```

with

```
ask n-of 1 turtles [count link-neighbors <3] [become-infectious]
```

Then run the experiment with

### Simultaneous runs in parallel: 1

In this experiment, the initially infectious node will be chosen randomly from among the 75 nodes with degree 2.

**Exercise 12** (a) For each output file order the column with the header `count turtles with [removed?]` from largest to smallest. Record the maximum and the mean final sizes of the observed outbreaks, as well as the numbers of runs where at least 10 hosts experienced infection and for those runs where no secondary infections whatsoever occurred.

(b) Compare the results of these findings between the three experiments and with the results that you obtained in Exercise 10 for the same value of  $R_0$  under the uniform mixing assumption.

(c) How would you interpret the observations you made in point (b) in view of the fact that the models of the contact networks in the last three experiments exhibit the friendship paradox with large excess? How well does your conjecture of Exercise 11 hold up?

## 4 The proof of the friendship paradox

The material in this section will primarily be of interest to advanced undergraduate and graduate students of mathematics.

Let  $G$  be a graph. Think of the edges of  $G$  as representing friendships between people. Recall from Section 1 that the friendship paradox asserts that  $\langle k_f \rangle > \langle k \rangle$ , where  $\langle k \rangle$  can be interpreted as the mean number of friends of a randomly chosen  $i$  and  $\langle k_f \rangle$  can be interpreted as the mean number of friends of the friends of a randomly chosen  $i$ .

**Theorem 1** *The inequality  $\langle k_f \rangle \geq \langle k \rangle$  holds in every graph  $G$ . Moreover,  $\langle k_f \rangle = \langle k \rangle$  if, and only if,  $G$  contains no isolated nodes and  $k_i = k_j$  for every edge  $\{i, j\} \in E(G)$ .*

**Exercise 13** Show that  $\langle k_f \rangle$  can be expressed as

$$\langle k_f \rangle = \frac{1}{N_{1+}} \sum_{\{i,j\} \in E(G)} \frac{k_j}{k_i} + \frac{k_i}{k_j}. \quad (7)$$

**Exercise 14** Prove Theorem 1.

Thus the only graphs that do *not* exhibit the friendship paradox are graphs in which all connected components are regular. In these graphs every friend of  $i$  has exactly the same number of friends as  $i$  has.

In your solution for Exercise 3 you may have found the following explanation of the friendship paradox: Consider randomly chosen nodes  $i$  and  $j$ . Then  $j$  is more likely to

be  $i$ 's friend if  $j$  has a lot of friends. In other words, the friends of a randomly chosen  $i$  tend to have more than the average number of friends, exactly as the friendship paradox predicts. For a mathematically formal version of this argument, consider a generic random graph  $G_{SQ}(N, \bar{k})$  with a given degree sequence. These graphs were introduced in our module *Exploring contact patterns between two subpopulations* at this web site<sup>5</sup>. Let  $i$  be a randomly chosen node  $i$  in this graph. In the construction of  $G_{SQ}(N, \bar{k})$  we attach  $k_i$  stubs to it. Consider a given stub  $stb$  and another node  $j$ . Then  $stb$  could be linked with any one of the  $k_j$  stubs at  $j$ . Thus the probability that stub  $stb$  will eventually form part of an edge  $\{i, j\}$  is proportional to  $k_j$ . In particular,  $i$  is *more* likely to be adjacent to nodes that have above-average degrees.

By taking the argument of the previous paragraph one step further we can derive a nice estimate for the expected value of the excess  $\langle k_f \rangle - \langle k \rangle$  in  $G = G_{SQ}(N, \bar{k})$  (or  $G = G_D(N, \bar{q})$  for the corresponding degree distribution  $\bar{q}$ ). A pair  $\{i, j\}$  will become an edge in  $G_{SQ}(N, \bar{k})$  if, and only if, *some* stub at  $i$  will be linked with *some* stub at  $j$ . The probability that a given stub at  $i$  will be linked with a given stub at  $j$  is approximately equal to  $\frac{1}{\sum_{i=1}^N k_i}$ . For  $k_i, k_j \ll N$  this implies that the probability that  $\{i, j\}$  becomes an edge is roughly equal to the product  $\frac{k_i k_j}{\sum_{i=1}^N k_i}$ . Since  $\sum_{i=1}^N k_i = \langle k \rangle N$ , we get

$$P(\{i, j\} \in E(G)) \approx \frac{k_i k_j}{\sum_{i=1}^N k_i} \approx \frac{k_i k_j}{\langle k \rangle N}. \quad (8)$$

Substituting (8) in (7) gives the following estimate of the expected value of  $\langle k_f \rangle$ :

$$\begin{aligned} \langle k_f \rangle &= \frac{1}{N_{1+}} \sum_{\{i, j\} \in E(G)} \frac{k_j}{k_i} + \frac{k_i}{k_j} \\ &\approx \frac{1}{2N_{1+}} \sum_{i=1}^N \sum_{j \neq i}^N P(\{i, j\} \in E(G)) \left( \frac{k_j}{k_i} + \frac{k_i}{k_j} \right) \\ &\approx \frac{1}{2N_{1+}} \sum_{i=1}^N \sum_{j=1}^N P(\{i, j\} \in E(G)) \left( \frac{k_j}{k_i} + \frac{k_i}{k_j} \right) \\ &\approx \frac{1}{2N_{1+}} \sum_{i=1}^N \sum_{j=1}^N \frac{k_i^2 + k_j^2}{\langle k \rangle N}. \end{aligned} \quad (9)$$

Here we needed to divide by 2 since each edge joins two stubs and will be considered twice in the summation.

The approximation in the second line of (9) will be valid if the probability of creating a loop  $\{i, i\}$  in the process of linking the stubs will be very small relative to the probability of creating *bona fide* edges. This will usually (but not always!) be the case.

<sup>5</sup><http://www.ohio.edu/people/just/IONTW/>

**Exercise 15** Consider a degree sequence  $\bar{k}$  with  $Q_0 = 0$  or a degree distribution  $\bar{q}$  with  $q_0 = 0$ . Use the above observations to prove that for large  $N$  the graphs  $G = G_{SQ}(N, \bar{k})$  or  $G = G_D(N, \bar{q})$  will have the following properties:

$$\begin{aligned} \langle k_f \rangle &\approx \frac{\text{Var}(k)}{\langle k \rangle} + \langle k \rangle, \\ \langle k_f \rangle - \langle k \rangle &\approx \frac{\text{Var}(k)}{\langle k \rangle}, \end{aligned} \tag{10}$$

where  $\text{Var}(k)$  denotes the variance of the degree distribution.

For example,  $k$ -regular random graphs are graphs of the form  $G = G_D(N, \bar{q})$  with  $q_k = 1$ . In these graphs we have  $\text{Var}(k) = 0$  and (10) confirms that these graphs do not exhibit the friendship paradox.

**Exercise 16** Use the result of Exercise 15 to estimate  $\langle k_f \rangle$  and the excess  $\langle k_f \rangle - \langle k \rangle$  for a graph  $G_{SQ}(100, \bar{k})$  that has 75 nodes of degree 2 and 25 nodes of degree 20.

The result of Exercise 15 applies only to graphs that are very similar to  $G = G_{SQ}(N, \bar{k})$  or  $G = G_D(N, \bar{q})$  for the given degree distribution. Recall from our module *Exploring contact patterns between two subpopulations* that graphs  $G = G_{SQ}(N, \bar{k})$  are neither assortative nor disassortative by degree. In contrast, consider a graph  $G$  with several connected components that are  $k$ -regular, but not for the same  $k$ . Theorem 1 implies that  $G$  will not exhibit the friendship paradox although  $\frac{\text{Var}(k)}{\langle k \rangle}$  will be positive. Such graphs are *completely assortative by degree*. We can see that for graphs that exhibit strong assortativity by degree (10) might substantially overestimate the magnitude of the excess  $\langle k_f \rangle - \langle k \rangle$ .

How about strong disassortativity by degree? Will (10) tend to underestimate  $\langle k_f \rangle - \langle k \rangle$ ? Consider a star tree  $G_{ST}(N)$  with  $N - 1$  leaves and one node, the root, with degree  $N - 1$ . You already calculated  $\langle k_f \rangle$  for  $G_{ST}(N)$  in Exercise 5. Note that star trees  $G_{ST}(N)$  with  $N > 2$  are completely *disassortative* by degree, as for each edge  $\{i, j\}$  one of the nodes must be a leaf with degree 1 while the other node must be the root with degree  $N - 1 > 1$ .

**Exercise 17** Compute  $\frac{\text{Var}(k)}{\langle k \rangle} + \langle k \rangle$  for  $G_{ST}(N)$  and compare the result with the value of  $\langle k_f \rangle$  that you obtained in your solution of Exercise 5.

## References

- [1] Scott L Feld. Why your friends have more friends than you do. *American Journal of Sociology*, pages 1464–1477, 1991.