Exploring random regular graphs with IONTW*

Winfried Just† Hannah Callender‡ Drew LaMar§

March 10, 2015

In this module[1] we introduce and explore the structure of random regular graphs. Moreover, we compare the predictions of SIR-models on random regular contact networks with the predictions of corresponding models on Erdős-Rényi networks.

1 Random regular graphs

A graph G is *k-regular* if every node has the same degree k. Graphs that are k-regular for some k are called *regular.*

Let us look at some examples. Open IONTW, click **Defaults**, and set

num-nodes: 5

After clicking **New** you will see a picture of the complete graph K_5 in the **World** window. In this graph each node i has degree $k_i = 4$; it is a 4-regular graph. More generally, for every N the complete graph K_N with N nodes is $N-1$-regular.

Now choose

network-type → **Empty Graph**

and click **New** again. You will see the empty graph $\bar{K}_5$. Each node in an empty graph has degree 0. Empty graphs are 0-regular.

For a third example, choose

network-type → **Nearest-neighbor 1**
num-nodes: 10
d: 2

After clicking **New** you will see a *one-dimensional nearest neighbor graph* in the **World** window. These graphs are denoted here by $G^1_{NN}(N,d)$; the one that is displayed in your **World** window is $G^1_{NN}(10,2)$. It is a 4-regular graph. You may want to look at the 2-regular graph $G^1_{NN}(20,1)$ and the 6-regular graph $G^1_{NN}(20,3)$ by changing

network-type → **Nearest-neighbor 1**

†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]This version was originally posted at `https:qubeshub.org/iontw`

num-nodes: 20
d: 1, 3

Notice that all regular graphs that we have looked at so far have a very rigid structure. Now suppose empirical data on a large contact network of N hosts indicate that every node has degree k, but that otherwise there does not seem to be a discernible structure to the contacts. How should we model disease transmission on such a network? If, for example, $k = 4$, then using $G^1_{NN}(N, 2)$ as a model of the (only partially known) contact network would be dangerous. While $G^1_{NN}(N, 2)$ is a 4-regular graph with the same number of nodes and degree distribution as the contact network, it has a very special structure that seems to be absent in the contact network. As we will see in the forthcoming module on clustering coefficients, this structure may strongly influence the dynamics of disease transmission and give misleading predictions about transmission on the actual contact network. To be on the safe side, we want to base our model of disease transmission on networks that assume nothing beyond the empirically verified property that the degree of each node is k.

In the module *Exploring contact patterns between two subpopulations* we have introduced a general construction of a random graph $G_{SQ}(N, \bar{k})$ that gives instances with a specified degree ***S**e**Q**uence* $\bar{k} = (k_1, \ldots, k_N)$. In all other respects, these graphs are completely generic. Here we will use this construction for the special case when $\bar{k} = (k, \ldots, k)$, that is, when $k_i = k$ for all i. The resulting graphs $G_{SQ}(N, \bar{k})$ are k-regular but otherwise completely random. They will henceforth be denoted by $G_{Reg}(N, k)$ and called *random (k-)regular graphs.*

To illustrate this construction, consider a population of 40 residents of the U.S., each of whom is married to another person in this population. Assume that we list the population in increasing order of their Social Security numbers. Let G_{spouse} be the graph that represents marital relationships in this population by edges. The information that we have given you allows you to deduce that G_{spouse} is a 1-regular graph, but not to whom host 10 is married.

Now choose the following parameter settings:

network-type $\rightarrow$ **Random Regular**
num-nodes: 40
lambda: 1

After clicking **New** you will see an instance of $G_{Reg}(40, 1)$. Note that the input field **lambda** controls the parameter k in the construction of $G_{Reg}(N, k)$. Think about the graph in the **World** window as representing the information about the graph G_{spouse} that we gave you above. Click **Labels** and see to whom host 10 is supposed to be married. Now click **New** a few more times and see how the graph, and the spouse of host 10, will change from instance to instance. Would it be fair to say that the construction of $G_{Reg}(40, 1)$ embodies *only* the information that we gave you about the graph G_{spouse}?

2 The connected components of random regular graphs

This part of the module is more theoretical and will be of interest primarily to students of mathematics. It can be skipped entirely or in part by students who are primarily interested in disease transmission.

It is interesting to compare the expected properties of $G_{Reg}(N,k)$ and $G_{ER}(N,k)$. The mean degree is k in both $G_{Reg}(N,k)$ and $G_{ER}(N,k)$. But in an Erdős-Rényi random graph $G_{ER}(N,k)$ only the mean degree is equal to k; for individual nodes the degree may be larger or smaller. One might think though that in other respects $G_{Reg}(N,k)$ and $G_{ER}(N,k)$ should be quite similar, but it turns out that there are important differences.

One important difference is that while Erdős-Rényi random graphs $G_{ER}(N,k)$ can always be constructed when $k \leq N-1$, random regular graphs $G_{Reg}(N,k)$ do not exist for some choices of $k < N$.

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

network-type $\rightarrow$ **Random Regular**
num-nodes: 10
lambda: 5

Click **New** to create an instance of $G_{Reg}(10,5)$. Now change

num-nodes: 11

and click **New** again. The *World* window will stay blank and the **Command Center** will display an error message. Experiment with various combinations of even and odd values of **num-nodes** and **lambda** to see when you get a network and when you get an error message.

Exercise 1 *Why is it impossible to have a k-regular graph with N nodes when both k and N are odd?*

Erdős-Rényi random graphs $G_{ER}(N,\lambda)$ with $\lambda > 1$ are predicted to contain a giant component that comprises about $\varrho(\lambda)N$ nodes. Recall from our module *Exploring Erdős-Rényi random graphs with IONTW* that $\varrho(1.5) = 0.5828$, $\varrho(2) = 0.7968$ and $\varrho(3) = 0.9405$. As a warm-up, let us remind ourselves how we can test these predictions with IONTW. Choose the following parameter settings:

model-time: $\rightarrow$ **Discrete**
infection-prob: 1
end-infection-prob: 1
network-type $\rightarrow$ **Erdos-Renyi**
num-nodes: 200
lambda: 2
auto-set: On

Use **New** to create a network and make sure that it gets initialized to one infectious node in an otherwise susceptible population. After you click **Go**, the nodes in the connected

component of the initially infectious node j^* will turn grey. Run about 10 simulations. Use **New** to create a new network for each of them. See how the theoretical predictions match what you see in the **World** window and **Disease Prevalence** plot at the end of each. Then set

lambda: 3

and run 10 simulations as before. Do you see a clear distinction between a giant connected component and small connected components? Did you observe any moderate-sized components?

Now we are ready to explore random regular graphs. Since $G_{ER}(200, k)$ and $G_{Reg}(200, k)$ will have roughly the same mean degree k and the connectivity is supposed to be random in respects other than the restriction on the degree distribution, one might expect that the structure of these two types of random graphs should be similar. In particular, one might expect similar distributions of the sizes of the connected components.

Let's explore whether this prediction holds up. Choose

network-type $\rightarrow$ **Random Regular**

and keep all other parameters fixed. This will set up a graph $G_{Reg}(200, k)$, where k corresponds to the input parameter **lambda**. Repeat the previous experiments in this subsection, first with $k = 3$ and then with $k = 2$.

Are your observations similar or are they different from what you observed for Erdős-Rényi random networks?

It will be better to collect more data to make sure that what you saw is not a fluke. Using the template that is provided in the instructions for the modules at this web site, conduct two batch processing experiments of 100 runs each for $G_{Reg}(200, k)$ with $k = 2, 3$ that record the numbers of removed nodes at the end of each run. For these experiments, keep **auto-set** switched **On** and use

Measure runs using these reporters:

```
count turtles with [removed?]
length(item 0 conncomp)
```

Setup commands:

```
new-network
compute-shortest-paths
```

Be sure to retain your data for further analysis.

Exercise 2 *(a) Sort the data from the lowest to the highest values in the output column* `count turtles with [removed?]`*.*

(b) Describe the observation in terms of the distribution of the sizes of the connected component of the randomly chosen initially infectious node j^*.

(c) The data in the last output column, `length(item 0 conncomp)`*, give you the sizes of the largest component. How are they related to the data in the column that you analyzed in point (b)? Describe the distribution of the data in this column.*

(d) Compare your observations with the theoretical predictions for Erdős-Rényi random graphs: Does it appear to be the case that in the graphs $G_{Reg}(200, k)$ for $k = 2, 3$ there is a dichotomy between many very small components in addition to one giant component that comprises roughly a fixed fraction of all nodes?

In Exercise 2 you almost certainly observed significant discrepancies between the sizes of the connected components of a randomly chosen node j^* in graphs $G_{Reg}(N, k)$ and the predictions of these sizes for $G_{ER}(N, k)$ that were discussed in our module *Exploring Erdős-Rényi random graphs with IONTW* at this web site. For $k \geq 3$ it can be proved that the graphs $G_{Reg}(N, k)$ will a.a.s. be connected, so that the largest component encompasses the entire graph(see, for example, Section 7.6 of [1]). Is this consistent with your observations for $G_{Reg}(200, 3)$ in Exercise 2?

In Section 1 you already looked at instances of $G_{Reg}(N, 1)$ and saw that all connected components have size 2. Again, this is different from the picture for $G_{ER}(N, 1)$, where many isolated nodes are predicted and the size of the largest connected component should scale proportionally to $N^{2/3}$ [1, 8].

The most interesting case occurs for $k = 2$. Let us explore a few small instances. Use the following parameter settings:

network-type $\rightarrow$ **Random Regular**
num-nodes: 7
lambda: 2
auto-set: Off

Click **New** repeatedly and look at about 20 instances of $G_{Reg}(7, 2)$.

Exercise 3 *(a) Are all the graphs that you see connected?*
(b) Explain why each of these graphs must be a disjoint union of cycles.
(c) Find the possible sizes of the connected components of $G_{Reg}(7, 2)$ and $G_{Reg}(200, 2)$.

It is not immediately clear though what the exact distribution of the sizes of these components of the largest component of $G_{Reg}(N, 2)$ should be. In fact, there are several different methods for generating instances of random regular graphs, and the answer may depend on which particular method is used. Different methods for generating random graphs with specified degree sequences, and in particular, random regular graphs, are reviewed in [5, 6].

IONTW performs a Markov Chain Monte Carlo edge switching procedure [7] to randomize and get an approximate uniform realization after first constructing the graph using Havel-Hakimi algorithm [2, 3]. Now look again at the data that you collected for Exercise 2. These data give you an idea about the distribution of the sizes of the connected component and the largest connected component for this method of generating $G_{Reg}(200, 2)$.

3 Disease transmission on random regular networks

Now let us explore the spread of infectious disease on random regular graphs. As a warm-up, let us first revisit our old friends, the Erdős-Rényi graphs $G_{ER}(200, 2)$.

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: 0.75
end-infection-prob: 1
network-type $\rightarrow$ **Erdos-Renyi**
num-nodes: 200
lambda: = 2
auto-set: On

This sets up a next-generation SIR-model. In Module 6 of [4] we found that for next-generation SIR-models the predictions for the spread of infections on Erdős-Rényi contact networks were practically identical to the predictions under the uniform mixing assumption for the given value of R_0.

Use **New** to create a network and then click **Metrics.** Scroll up the **Command Center** to look up `R0` and verify that R_0 for our model is reasonably close to 1.5.

If everything seems fine, we are ready to **Go**. We would expect to see minor outbreaks about 42% of the time and major outbreaks with final size near $\varrho(1.5) = 0.5828$ about 58% of the time. Do a few runs to see whether this works out as expected.

Now set

lambda: 3
infection-prob: 0.5

This should give the same value $R_0 \approx 1.5$. Check by clicking **New** and then **Metrics**, do a few test runs, and see whether the predictions still seem to play out as expected.

Now choose

network-type $\rightarrow$ **Random Regular**

and retain all other parameter settings. Check with **New** and **Metrics** that $R_0 \approx 1.5$.

Before we run any simulations, let us take a minute to think what we should expect: $R_0 = 1.5$ is the same as previously; so is the mean degree $\langle k \rangle = 3$. Thus we might perhaps get similar results as for Erdős-Rényi networks. But in view of the results that were discussed in Section 2, we would also expect the connected component of the index case to be the entire graph. Perhaps this might cause more frequent and larger major outbreaks on networks $G_{Reg}(200, 3)$ than on networks $G_{ER}(200, 3)$? Let's see.

Run about 10 exploratory simulations and observe what is going on in the **World** window and the **Disease Prevalence** plot. Are your observations consistent with your expectations? If not, check whether all parameter settings are as specified. If not, correct and run another 10 exploratory simulations. Do you still not see what one might expect?

Let's get serious and run a batch of 100 simulations with these settings. Follow the template that is provided in the instructions for using our modules at this web site to define and run a **New** batch processing experiment for the current parameter settings. Use the following specifications:

Measure runs using these reporters:

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

Setup commands:

```
new-network
```

Exercise 4 *(a) Sort your output columns from lowest to highest.*

(b) Do you see a distinct gap between major and minor outbreaks? Do you see many outbreaks with final size near 0.58*?*

(c) Does the structure of $G_{Reg}(N, 3)$ *appear to make outbreaks larger or smaller relative to what we might predict for* $G_{ER}(N, 3)$*?*

(d) Does the structure of $G_{Reg}(N, 3)$ *appear to make outbreaks larger or smaller relative to what we might predict under the uniform mixing assumption?*

Puzzled? Let us see what happens in $G_{Reg}(200, 2)$. Change the following settings:

lambda: 2
infection-prob = 0.75

Click **New** and the **Metrics** to verify that $R_0 = 1.5$. Next repeat the exploratory simulations in explorations that you had done prior to the batch processing experiment. Do you observe any outbreaks that one might classify as "major?"

Now run a similar batch processing experiment for the current settings as you had done for Exercise 4. Be sure to either create a **New** experiment or to edit the value for both `"lambda"` in the dialogue window.

Exercise 5 *(a) Sort your output columns from lowest to highest.*

(b) Do you see a distinct gap between major and minor outbreaks? Do you see many outbreaks with final size near 0.58*?*

(c) Does the structure of $G_{Reg}(N, 2)$ *appear to make outbreaks larger or smaller relative to what we might predict for* $G_{ER}(N, 2)$*?*

(d) Does the structure of $G_{Reg}(N, 2)$ *appear to make outbreaks larger or smaller relative to what we might predict under the uniform mixing assumption?*

What is going on here? Try to find an explanation for the observed discrepancies between the predicted spread of the infection on $G_{ER}(200, 3), G_{ER}(200, 3)$ and the observed pattern for $G_{Reg}(200, 3), G_{Reg}(200, 2)$. We will give the explanation in the forthcoming module on the so-called *replacement number.*

References

[1] Bela Bollobás. *Random graphs*. Cambridge University Press, second edition, 2001.

[2] S. Hakimi. On realizability of a set of integers as degrees of the vertices of a linear graph. *Journal of the Society for Industrial and Applied Mathematics*, 10:496–17506, 1962.

[3] V. Havel. A remark on the existence of finite graphs. *Casopis Pest. Mat*, 80:477–17480, 1955.

[4] 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*. Academic Press, 2015.

[5] R. Milo, N. Kasztan, S. Itzkovitz, M. E. J. Newman, and U. Alon. On the uniform generation of random graphs with prescribed degree sequences. *Preprint*, 2003. Available at ArXiv `http://arxiv.org/abs/cond-mat/0312028`.

[6] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:026118, 2001.

[7] A. Rao, R. Jana, and S. Bandyopadhyay. A markov chain monte carlo method for generating random (0,1)-matrices with given marginals. *Sankhya Ser. A*, 58:225–17242, 1996.

[8] Joel Spencer. The giant component: The golden anniversary. *Notices of the AMS*, 57(6):720–724, 2010.