Exploring contact patterns between two subpopulations*

Winfried Just† Hannah Callender‡ M. Drew LaMar§

October 12, 2015

In this module[1] we introduce a construction of generic random graphs for a given degree sequence or degree distribution and explore whether mixing between hosts who belong to different subpopulations is assortative or disassortative.

1 Introduction

Sometimes a population of hosts can naturally be partitioned into two or more disjoint *subpopulations* that differ in certain characteristics. For example, the human population can be naturally partitioned into the subpopulation of residents of different countries. People in a given country tend to make more contacts with residents of the same country than with residents of a different country. We say that the mixing pattern in this case is *assortative.* On the other hand, when studying sexual contact networks, we would consider a partition of the population into males and females. Since most sexual contacts are made between members of different subpopulations for this partition, the mixing pattern in this case would be *disassortative.*

Of particular interest in disease modeling are partitions into subpopulations that differ by the the propensity of making contacts. In terms of the contact network we might consider a partition into subpopulation A of all hosts i who have degree $k_i > \Theta$ and subpopulation B of all hosts i who have degree $k_i \leq \Theta$, where Θ is a suitably chosen threshold. For such partitions it is less obvious than in the examples of the previous paragraph whether the mixing pattern should be considered assortative or disassortative. In this module we will explore one such example with IONTW and derive a mathematically rigorous condition for making this distinction.

---

†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`

2 Generic graphs for a given degree distribution

Suppose G is a given graph with *vertex set* $V(G) = \{1, \ldots, N\}$. For each nonnegative integer k, let Q_k count the number of nodes i with degree k, and let $q_k = \frac{Q_k}{N}$. Thus Q_k represents the *number* of nodes with degree k, while q_k represents the *proportion* of nodes with degree k. While the symbols Q_1, Q_2, Q_3 in other contexts may denote quartiles, here they have a different meaning.

Since the maximum possible degree in any graph with N nodes is $N-1$, the mean degree $\langle k \rangle$ can be computed in this terminology as:

$$\langle k \rangle = \frac{1}{N} \sum_{k=0}^{N-1} k Q_k = \sum_{k=0}^{N-1} k q_k = \sum_{k=0}^{\infty} k q_k. \tag{1}$$

As every edge contains two nodes, we have

$$2\#(E(G)) = \sum_{k=0}^{N-1} k Q_k, \tag{2}$$

where $\#(E(G))$ denotes the number of edges in the graph G. Thus (1) implies the following useful expression:

$$\langle k \rangle = \frac{2|E(G)|}{N}, \qquad \text{or, equivalently,} \qquad \langle k \rangle N = 2|E(G)|. \tag{3}$$

In instances of Erdős-Rényi random graphs $G_{ER}(N, \lambda)$, the degree distribution $\bar{q}$ will be close to binomial and can be approximated for large N by a Poisson distribution or a normal distribution [1]. For many empirically studied contact networks, very different degree distributions have been found. One would want to study disease transmission on such networks by using different types of random networks with degree distributions that more closely match the empirically observed ones. We present here a general construction that gives random graphs whose degree distributions closely match a given one, but which are generic in all other aspects.

Let us first assume we have N nodes and the degree k_i of each node i is given. Let $\bar{k} = (k_1, \ldots, k_N)$ be the sequence of these degrees. A graph $G_{SQ}(N, \bar{k})$ with the given degree **S**e**Q**uence can be constructed as follows: First attach k_i half-edges or *stubs* on each node i. Then randomly draw pairs of stubs, without replacement, and connect the pairs to form actual edges.

Now assume that we are given only a probability distribution $\bar{q} = (q_1, q_2, \ldots)$ of the degrees of a random graph so that q_k represents the probability that the degree of a randomly chosen node i is k, that is, $P(k_i = k) = q_k$. Then we can construct a graph $G_D(N, \bar{q})$ with the given degree **D**istribution by first drawing k_i randomly from this distribution, independently for each i, and then constructing $G_D(N, \bar{q}) = G_{SQ}(N, \bar{k})$ for the resulting sequence $\bar{k}$ of degrees.

This does not always work (think about a situation where we accidentally draw $k_i > N-1$ or if $\sum \bar{k}$ turns out to be odd, so that (2) will be violated). Even for legitimate degree sequences the construction is somewhat difficult to implement on a computer, as we may sometimes draw two stubs of the same vertex or stubs for two vertices that have already been connected earlier by an edge. These difficulties do not need to overly concern us here as our goal is mathematical analysis of the resulting graph, not implementation.

3 Exploring an example with IONTW

Open IONTW, press **Defaults** and choose

network-type → Custom Distribution

Consider a small class of 10 students, 4 of whom are majoring in biology and 6 of whom are majoring in math. Each of the math majors has exactly 2 friends in this class, and each of the biology students has exactly 3 friends in this class. To create such a network with IONTW, write two lines

```
degrees
[3 3 3 3 2 2 2 2 2 2]
```

into a plain text file and save the file under the name `degrees.txt` in the same directory in which you keep IONTW. In the **World** window a network with 10 nodes should appear. Press **Labels** and figure out who are the math majors and how this info is coded in the file `degrees.txt`. Repeat a few times by pressing **Load** and opening `degrees.txt`.

Do the networks that you see in your **World** window exhibit assortative or disassortative mixing?

Actually, the networks that you create in this way should, *on average,* neither be assortative nor disassortative. The reason is that it is an instance of $G_{SQ}(10, \bar{k})$, where $\bar{k}$ is the degree sequence that you specified in the file `degrees.txt`. The code creates 3 stubs for each bio major and 2 stubs for each math major and then combines them randomly into edges, regardless of the major of the students that are represented by the nodes of each stub. However, due to random fluctuations, each particular instance you actually see may exhibit slightly assortative or slightly disassortative mixing. Let us figure out which is the case for our particular instance that is displayed in Figure 1.

We had changed the labels by entering in the the following commands in the `observer>` line:

```
ask turtles with [count link-neighbors = 3] [ set label "bio" ]
ask turtles with [count link-neighbors = 2] [ set label "math" ]
```

Let us count the three types of edges in the instance that we see in our **World** window:

$$\begin{aligned} &3 \text{ edges of type } \{bio, bio\} \\ &6 \text{ edges of type } \{bio, math\} \\ &3 \text{ edges of type } \{math, math\} \end{aligned} \tag{4}$$

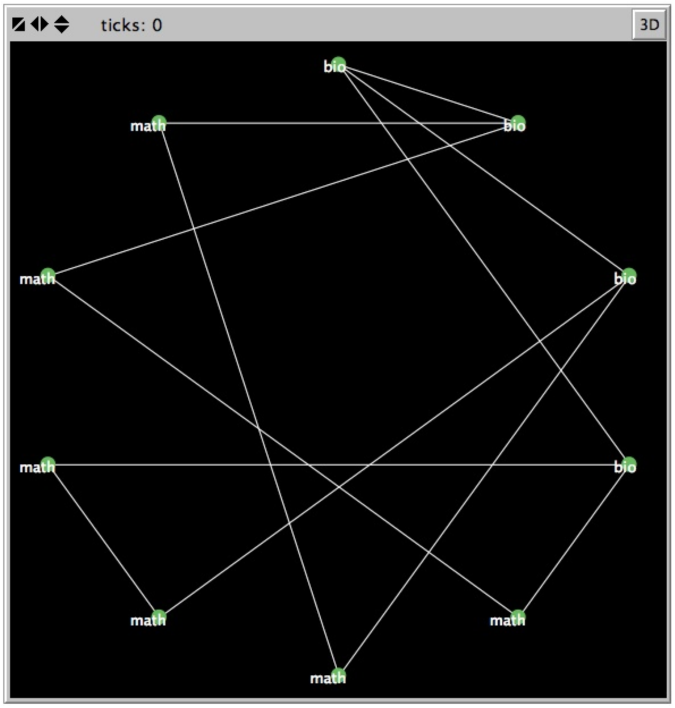

Figure 1: The graph of friendships between bio and math majors.

Now it certainly looks like this network might be disassortative as both the bio and the math majors formed more friendships with a student who majors in the other field. But we should not jump to conclusions too quickly. After all, there are more math majors in this class that the bio majors could select their friends from, and the bio majors are more likely to form friendships in the first place.

How would the proportions need to look like if the mixing were neither assortative nor disassortitive? We could perhaps estimate these proportions by calculating means for many instances that our software creates. But this would be very tedious. Let us instead try to derive a formula by analyzing the construction of the random graphs $G_{SQ}(10, \bar{k})$.

It is worth considering this problem in a more general context. Suppose we have a large population of size N that is partitioned into two disjoint subpopulations of bio and math majors, of sizes B and M respectively. Moreover, assume that the degree of each node that represents a bio major is k and the degree of each node that represents a math major is ℓ. Then the bio subpopulation would contribute Bk stubs and the math subpopulation would contribute $M\ell$ stubs. Let $p = \frac{Bk}{Bk+M\ell}$ denote the proportion of stubs attached to nodes that represent bio majors, and let $q = \frac{M\ell}{Bk+M\ell}$ denote the proportion of stubs attached to nodes that represent math majors.

Exercise 1 *Find the values of* B, M, k, ℓ, p *and* q *for the degree distribution in our example of* $G_{SQ}(10, \bar{k})$.

Now consider how an edge is formed in the construction of $G_{SQ}(N, \bar{k})$: By randomly drawing two stubs and joining them. The sampling needs to be done without replacement here, which adds some complications. But except near the end of the construction we could approximate it very well by sampling without replacement, which should get the following probabilities of creating edges of the three different types:

$$P(\{bio, bio\}) = p^2, \quad P(\{bio, math\}) = 2pq, \quad P(\{math, math\}) = q^2. \tag{5}$$

Exercise 2 *Have you seen such proportions before in your biology classes? What were these proportions called and what is the connection with the context in which you saw them?*

Now we might consider a given network assortative if the observed proportion of edges $\{bio, math\}$ is less than $2pq$, and disassortative if this proportion is larger than than $2pq$.

Exercise 3 *(a) Which counts would correspond to* (5) *for $G_{SQ}(10, \bar{k})$?*

(b) Were we correct in classifying the network with the counts given in (4) *as disassortative?*

(c) Now test whether the network that you see in the ***World*** *window of* your *IONTW interface is assortative or disassortative.*

References

[1] 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.

Sample solutions for the exercises
Exploring contact patterns between two subpopulations

Sample solution for Exercise 1: Since there are 4 bio majors and 6 math majors, we have $B = 4$ and $M = 6$. Each node that represents a bio major has degree $k = 3$ and each node that represents a math major has degree $\ell = 2$. Thus $p = \frac{Bk}{Bk+M\ell} = \frac{12}{12+12} = \frac{1}{2} = q$. ☐

Sample solution for Exercise 2: Equation (5) is the same as the Hardy-Weinberg Law. Read "stub at bio node" as "allele A," "stub at math node" as "allele a," "edge $\{bio, bio\}$" as AA homozygote," "edge $\{bio, math\}$" as "heterozygote" and you will see the connection.

If the phenotypic effects of the gene under consideration cause carries of allele A to preferentially mate with other carriers of allele A and carriers of allele a to preferentially mate with other carriers of allele a, we will see *assortative mating* and fewer heterozygotes than what would be predicted by the Hardy-Weinberg law. ☐

Sample solution for Exercise 3: (a) By the result of Exercise 1 we have $p^2 = q^2 = \frac{1}{4}$ and $2pq = \frac{1}{2}$. Since there are 12 edges total, the counts of the edges should be exactly as in (4).

(b) No. The network should be considered neither assortative nor disassortative. ☐