EXPONENTIAL RANDOM GRAPH MODELS FOR THE LAYPERSON

RYAN DEMUSE

1. Motivation

Before we discuss anything slightly resembling exponential random graph models, let's first talk about what they may be good for and why we should care about them. Firstly, an exponential random graph models can be used to generate a random instance of a graph. Now, when we say a graph, we do not mean the graph of a function that you are used to from an algebra or calculus class such as the one to the right. Instead, we mean a *combinatorial graph* or, simply, a *graph*. A graph $G = (V, E)$ is a collection of *vertices*, V, and *edges*, E, which are ordered pairs of vertices indicating a relationship between two vertices.

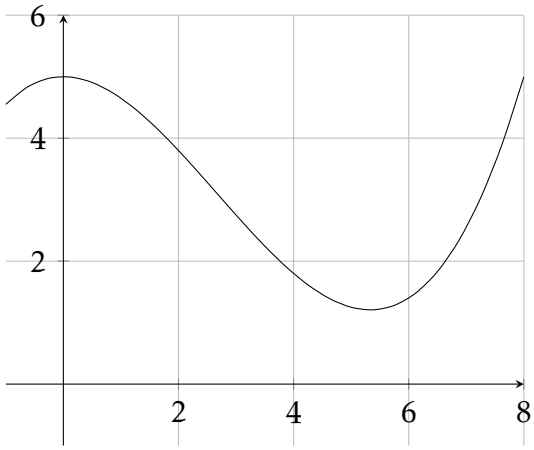
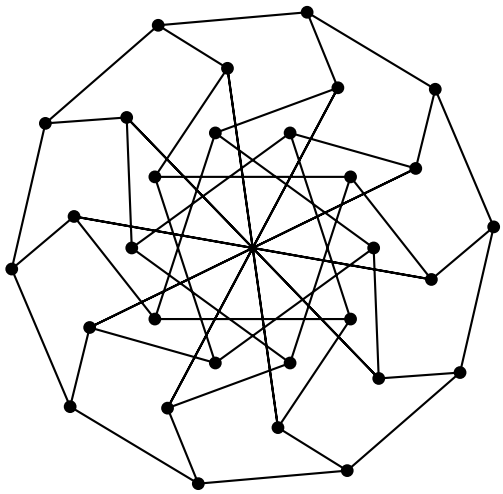


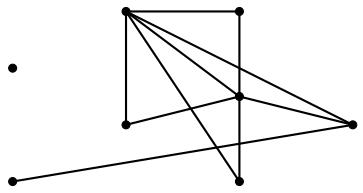
Figure 1. Graph of a function.



A graph $G = (V, E)$ looks more similar to the object to the left. The filled dots are the vertices and the lines connecting the filled dots are the edges. This graph, known as the *Tutte-Coxeter graph* is highly structured in many mathematical ways, which may be unsurprising because it looks very structured. Unfortunately for those who study graphs, *graph theorists* do not usually have the pleasure of always studying such nicely ordered strucutres. In reality, most graphs are highly disordered and unstructured, or, at the very least, they *appear* to be disordered.

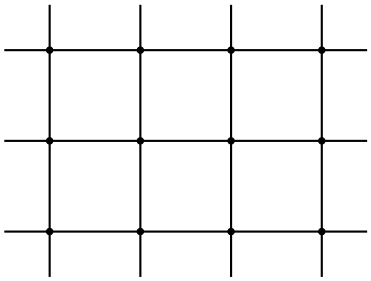
Date: November 18, 2021.

Consider the graph below. This graph is nowhere near as nice as the Tutte-Coxeter graph from above. It has an *isolated vertex*, a vertex that does not have any edges connecting it to other vertices. Some vertices are only connected to one or two other vertices, and some are connected to four different vertices.



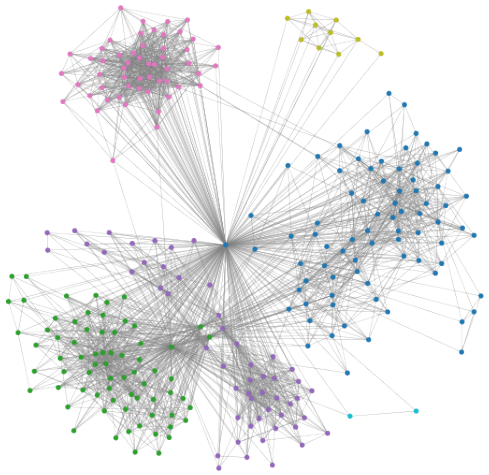
There are many more graphs that look like this one than the Tutte-Coxeter graph. Graphs can be used for many different things. They can be used as a model for road or transportation networks, they can be used to study the relationships between individuals on a social network, or even the connections between neurons in the brain.

But this leads us to a question, what does a typical graph look like? Well, that is a really hard question to answer and leads us toward the theory of exponential random graphs. What a typical graph looks like really depends on the context. A "typical" graph in each of these preceeding scenarios all look different from one another. For example, in a city, a typical road network looks a lot like this:



By a "typical" network, we mean something that looks like a common example. Here, each dot represents an intersection and each edge represents a road between intersections. Something important that this example illustrates is that we do not have a sense of how long the distance is between each intersection along a road. A graph is an abstract mathematical structure where we could include this information or omit it. If we wanted to, we could label each edge with the length between intersections, but this is a more complex situation.

To contrast our road network example, let's look at an example of what a typical social network graph might look like.



Here we have what appears to be a single node in the center with connections to many vertices expanding out from that central node. Furthermore, there are clusters of vertices that are connected to many other vertices within those clusters. The single node in the center could be a person with the edges connecting out to their friends. The clusters on the outskirts

are their friend groups. Some groups are highly connected to one another such as the lower groups, and some are only connected within themselves, such as the groups in the upper part of the graph. This is an example of what somebody's personal facebook or twitter friend graph may look like.

2. Features of Social Networks

As we can see, social network graphs are big and complicated, as well as being, seemingly, unorganized. However, there are some characteristics that social network graphs all tend to share. Notice that vertices tend to be a part of a highly connected group. By highly connected, we mean a group of vertices where every vertex is connected to almost all of the other vertices in that group. This is a network feature that we will refer to as *clustering*. Furthermore, it is a reasonable assumption that a friend of my friend is more likely to be a friend of mine than a random stranger. We refer to this feature as *transitivity*. These are two essential network features that many social network graphs share [citation needed here].

Now, suppose that you want to study social network graphs. To be precise, let's say that you want to find the average number of friends that a person has on a social network. Not just a person on facebook, but the average number of friends a person has on any social network, including twitter, instagram, tiktok, etc. How would you go about doing this? One way this could be done is by taking every single person on all of the social networks and adding up how many friends they have, then dividing by the number of people on those social networks. In actuality, this is infeasible. There are over 1 *trillion* connections on facebook alone! We would have to store these huge networks somehow and analyze them by traversing through them to count the number of connections. This requires vast amounts of memory and computing power. Unfortunately, this is the only way for us to know that we have gotten the exact number we are looking for. What if we relax our standards a little bit? Maybe we are okay getting really close to the average number of friends a person has. Perhaps we are perfectly happy with a small range that is *most likely* the average number of friends someone may have.

If we are okay giving up a little, we don't actually need to store the networks themselves, nor do we have to traverse through them! This is where the power of expoenential random graph models materializes. The idea is to create a *model* that closely mimics the behavior of the social networks that we care about. A mathematical model of a social network is a function that takes in a social network graph and assigns it a number based on how closely it imitates the behavior of facebook or twitter. A number close to 1 means that the graph is excellent at emulating the desired behavior while a value near 0 means the graph is lousy at it.

3. Exponential Random Graph Models

So why use exponential random graph models, surely there are other ways to model social networks? Absolutely, there are many different ways from Erdős-Rényi or stochastic block model to random geometric graphs. It just so happens that exponential random graph models display

those key features that are present in social networks, clustering and transitivity.

Let's take a step back to determine where we are. We have replaced traversing through multiple social networks with trillions of connections to a single function that evaluates graphs based on how well they emulate social networks. So how do we take this function and compute the average number of friends a person has? The exponential random graph model has the general form

$$f(G) = Z^{-1} e^{T(G)}, \tag{1}$$

where G is a graph that input into the function f, T is a function on the graph that is determined so that it gives higher numbers to graphs that look like social networks, and Z makes sure those numbers are between 0 and 1. If you have never encountered it before e is a special mathematical constant known as *the natural number* or *Euler's number*. To compute the average number of friends an individual has, we find the number of edges in each graph on a certain number of vertices weighted by f, then divide by the total number of possible graphs on n vertices. Equationally, this is equivalent to

$$\text{Avg}(n) = 2^{-\binom{n}{2}} \sum_{G} a(G) f(G), \tag{2}$$

where n is the number of vertices in our social network and $a(G)$ counts the average number of friends for a vertex in G. This now reduces to a problem of sotring massive social networks and traversing through them while counting the number of connections that they have to calculating the value of a function for a particular number of people in the social network.

I make this sound like we have made this problem a lot easier to solve... In actuality, we haven't really. There are a lot of graphs on a given number of vertices. A lot. For example, the possible number of social networks involving 10 people is

$$35,184,372,088,832$$

That is over 35 *trillion*. It takes a computer nearly 6 days to simply count to 35 trillion, let alone compute a function that many times. This is even harder than the problem that we started with. Where do we go from here?

4. Modern Methods

Thankfully, there has been extensive research into exponential random graph models thanks to the likes of Persi Diaconis, Sourav Chatterjee, Charles Radin, Mei Yin, and many others. Among the large number of results that have expanded our understanding of exponential random graphs and their uses (as well as advances that have been made in mathematics as a whole from research into exponential random graphs), there are a few that make answering our question much easier and faster. Let's take a step back to Equation (1) and modify it a little bit. We really aren't changing anything, but it makes it easier to state some of the results that we will need along the way. Rewrite f as

$$f(G) = e^{n^2(T(G) - \psi_n)}. \tag{3}$$

Essentially, what we have done is taken the constant Z, renamed it ψ_n, placed it in the exponent, and scaled by n^2. Why n^2? There are about n^2 possible edges in a graph on n vertices. The function f is what is known as a probability measure. That means, when we sum up the values of f

over all G on n vertices, we get 1. Using this, we can write an equation to find the value of ψ_n,

$$\psi_n = \frac{1}{n^2} \log \sum_G e^{T(G)}. \tag{4}$$

The attentive reader will notice that this is the same sum from $\text{Avg}(n)$. That means that ψ_n is just as hard to calculate! This is where some awesome research advances will help us out. As n gets larger and larger, the values of ψ_n get closer and closer to a particular number. Let's call this number ψ_∞. Mathematically, we can write this as

$$\psi_\infty = \lim_{n\to\infty} \psi_n.$$

Chatterjee and Diaconis determined that even though ψ_n is difficult to calculate for any value of n and gets harder and harder as n gets larger, ψ_∞ is much easier to calculate. To be precise,

$$\psi_\infty = \max_G \{T(G) - I(G)\} \tag{5}$$

where I is some other function that is note important here. What is important though is the fact that we are no longer taking a sum of exponentially many things. Instead, we are finding the maximum value of a function. There are much, much faster methods to finding a maximum of a function than taking a sum. We have finally found a way to make this process faster; however, finding ψ_∞ is not quite the same as finding $\text{Avg}(n)$. That is okay though because there is another result that says that there is a small number of graphs that attain this maximum value for ψ_∞. Furthermore, the graphs that attain this value are much more likely to occur than any others, that is, they are graphs G such that $f(G)$ is very close to 1 and all other graphs have values very close to 0. Let's call this collection of graphs G^*. Finally, we can use this much smaller set of graphs to calculate $\text{Avg}(n)$ as

$$\text{Avg}(n) \approxeq \frac{1}{|G^*|} \sum_{G\in G^*} a(G),$$

since $f(G) \approxeq 1$ for G in the set G^*.

5. Problem solved?

Now, I have skipped over some of the mathematical rigor and abused a few concepts to make it easier to explain and understand. A rigorous treatment and to fully understand the intricacies involved would require an undergraduate degree in mathematics as well as several years of graduate mathematics in probability and measure theory, combinatorics, graph theory, and graph limit theory. But let me walk through what we have done and where we are now.

Originally, we wanted to determine the average number of friends an individual has on any social network. These social networks are massive, with billions of people and trillions of connections between them. Instead of storing these massive structures on a computer and walking through trillions of nodes, we create a model, or a function, that accurately emulates the behavior of social networks. We can then equationally write down how to find the average number of friends a person has. It turned out that this was just as hard, if not more difficult, to compute. Using modern results of exponential random graph models, we determined that there is a very small number of graphs that are likely to look like social networks. We then use this

small set of graphs to compute the average number of friends for an individual.

If there were only one thing that I would like a reader to take away from this short article, it is that there is always a give and take with mathematics. In the end, we gave away some of the accuracy of our solution, and, in return, we were able to get an answer much quicker than we could possibly hope of getting the exact answer.

So, have we solved the problem? Well, that is a complicated question. We started pretty early with a way to solve the problem, but it was slow. There are two aspects to solving the problem: getting an answer very close to the correct one, and doing this quickly. In an essence, we haven't solved the problem, we just came up with a fast way to estimate the correct solution. In our modern age of computer algorithms in the software dominated world, estimating things fast is often more important than taking forever to get the exact answer. Research is always striving to get more accurate solutions faster than before. This problem is no different.

6. Is that all they can do?

This is a good example of how exponential random graph models can solve practical problems. This is not the only thing they can be used for. It has been shown that exponential random graph models are excellent at not only modeling social networks, but also economic networks and even networks of neurons in the human brain.

Exponential random graph models are a mathematical tool that are used to analyze anything that can be represented as a graph. There are infinitely many possible models. Finding the correct one, or at the very least, a good one, is another difficult question. Nevertheless, they are an interesting tool that can be used to study many aspects of the world indirectly, saving us storage and time.