23.1 Random Graphs

Random graphs are important theoretical tools. They are generally described in a probabilistic manner or as the result of some generative process. The use and study of random graphs serves two primary purposes.

The primary motivation for studying random graphs is that their analysis can provide insight into real-world empirically-observed phenomena and properties of real graphs. An analytic study can provide theoretic justification for observations like how real graphs tend to have small diameters or a single massive component.

Secondly, they can be used as *null models* for hypothesis testing. E.g., if we are observing or measuring some property of a real graph and we would like to better understand just how notable our empirical observations actually are, one thing we can do is compare our measurements to those obtained on a random graph with similar basic properties.

For example, *motif finding* is the study of a graph in order to identify subgraph structures that appear more often than would otherwise be expected. How can we actually define "more often than expected"? One way would be by comparing a random graph with some similar properties, such as degree distribution, to our network of interest. This is essentially a means to evaluate the *null hypothesis*, which is determining whether the observation we are making is due to some underlying reason or is simply random chance.

23.1.1 Erdös-Rényi Graphs

Erdö-Rényi graphs come in two flavors - the $G(n, m)$ model and the $G(n, p)$ model. The former model assumes a graph of n vertices and m edges, where these edges have (uniformly) randomly selected endpoints. The latter model assumes n vertices and the existence of an edge between any u, v vertex is determined via probability p, aka the *attachment probability*. Note that the $G(n, m)$ model can have multi-edges and self-loops while the $G(n, p)$ model will be simple. The generation of a $G(n, p)$ model also can be considered as a Bernoulli Process across all u, v vertex pairs.

We can use these models to understand why real graphs often have a single massive component or why many real graphs exhibit *small-world* properties. This model assumes *homogeneous mixing*, in that all vertices and their connections are statistically the same. This assumption helps simplify theoretical analyses, but this assumption is not particularly representative of almost any real naturally-arising graph.

23.1.2 Configuration Model

The configuration model addresses the above limitation of Erdö-Rényi graphs by generating a random graph that exhibits an exact degree sequence. Generation of configuration models is straightforward and somewhat similar to Havel-Hakimi graph generation. For each degree k in a given degree sequence, we construct a vertex with k *stubs*, which are "half-edges" that we wish to connect to each other. Randomly, two stubs among all vertices are selected and connected to construct an edge. This is repeated until no stubs remain. This process will result in loopy multi-graphs, as with the $G(n, m)$ model.

We can also consider attachment probabilities within the context of the configuration model. The probability of selecting a stub of some given vertex is proportional to the number of stubs (or degree) of that vertex. Specifically, given $2m = \sum_{v \in |V|} d(v)$, the probability of selecting any stub on vertex u with degree $d(u)$ is $\frac{d(u)}{2m}$. Thus, given two vertices u, v with degrees $d(u), d(v)$, a total of m edge selections, and the fact that we can select u, v or v, u in 2 different ways, the attachment probability between u, v is

$$p_{u,v} = \frac{d(u)}{2m}\frac{d(v)}{2m} \times 2 \times m = \frac{d(u)d(v)}{2m}$$

23.1.3 Chung-Lu Model

The Chung-Lu model is a variation of the configuration model. Generally, we consider it in terms of the attachment probabilities, as above. Between vertices u, v we have:

$$p_{u,v} = \frac{w_u w_v}{2m}$$

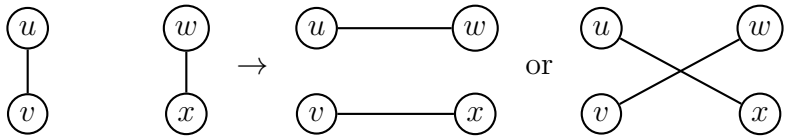
where w_u and w_v are weights associated with each vertex. Most commonly, these weights are simply considered as degrees, giving the same attachment probabilities as the configuration model. As with the configuration model, there is an inherent assumption that this model is only applicable to the class of loopy multi-graphs. While some theorists might claim that as $|V(G)| \to \infty$ the probability of self-loops and multi-edges approaches zero and therefore this model is applicable to simple graphs as well, these theorists are fools.

Let's discuss issues that can arise when considering these attachment probabilities in the context of simple graphs. Consider what happens when $w_u w_v > 2m$, which is not an uncommon occurrence in graphs with heavily skewed degree distributions or graphs that are suitably dense. In multi-graphs, this would imply the expectation of a non-zero number of multi-edges between u, v. In simple graphs, this probability is meaningless. In fact, the divergence between actual *uniformly random* simple graph attachment probabilities and the Chung-Lu or configuration model is quite substantial, even when $\forall u, v \in V : w_u w_v << 2m$.

23.1.4 Null Models

We'll define *null models* to be graphs that have some set of equivalent properties to a graph being studied, but are otherwise selected uniformly randomly from *all possible configurations* within the graph class defined by these properties. Uniformly random implies, in this context, that an explicit instantiation of such a null model will result in an unbiased selection of a single graph topology among all possible graph topologies (a combinatorially massive number) that fit the given properties. The specific properties are usually n, m, or an explicit degree sequence, among others.

For all graphs in the class of loopy multi-graphs, the configuration model or Chung-Lu model provide a means to realize explicitly or measure implicitly (i.e., by using attachment probabilities) null model statistics for a given degree sequence. For simple graphs, there is no direct generative method or closed-form equation for attachment probabilities. However, we can still uniformly select from the class of simple graphs with a given degree sequence by performing a Markov process of *double edge swaps* on any arbitrary graph with that degree sequence.



Double edge swaps take two edges in some graph $(u, v), (w, x)$ and swap their endpoints to construct edges $(u, w), (v, x)$ or $(u, x), (w, v)$. Note that these swaps would not alter the degrees of any of these vertices. By randomly selecting these edges and how they're swapped, not performing any swap that creates multi-edges or self loops, and repeating this a "suitably large" number of times, we can effectively produce a uniformly random sample from all possible simple graph topologies with some given degree distribution. How large our "suitably large" number actually is, is called the *mixing time* and is theoretically unknown in the general case.

Regardless, we can also use this approach to determine our "actual" simple attachment probabilities. To do this, we would generate some number of random samples, measure how often vertices within the various degree classes actually connect, and take the averages to be our $p_{i,j}$ attachment probabilities between some vertex of degree i and some vertex of degree j. However, his is a computationally expensive procedure.