# ENGG2430D Tutorial 10

BY ZHIBO YANG

*April 8, 2015*

## 1 Introduction

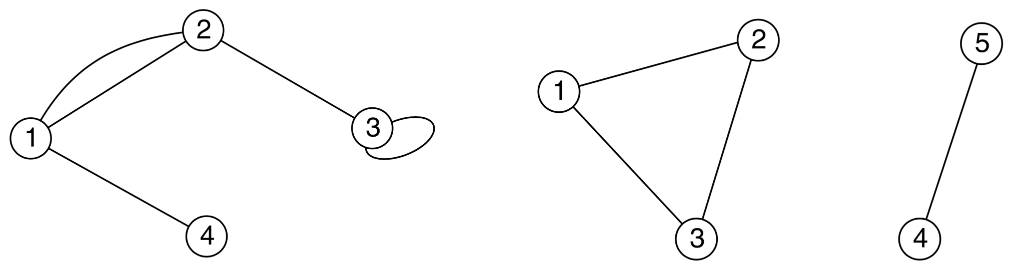
In this tutorial, we will discuss over a classical application of probability thoery, *(probabilistic) random graph*. Basically, this will involve conditinal probability, expectation of random variables, properties of binomial random variable and Poisson random variable, etc.

The material is organized as follows: first I will introduce to you some basic notation of graphs in section 2; then the setting of random graph is specified in section 3, also we will derive some interesting properties of random graph using what you have learned in the class.

## 2 Basic Notation of Graph

A graph basically consists of two major components, *vertices (or nodes)* and *edges*, which are denoted by $V$ and $E$ in the following. A graph can be catogorized as *directed* and *undirected* depending on whether the edges, represented by pairs of vertices, have direction or not. But note that in our settings, the edges have no direction. For instance, if $V = \{1,\ 2,\ 3,\ 4\}$ and $E = \{(1,2), (1,4), (2,3), (3,3)\}$, we represent this graph as follows in Figure 1(a). As is shown, there can be multiple edges between two vertices (1 and 2), and loops (from node 3 to itself) in the graph.

We say a graph is *connected* if there is a path between any pair of vertices in the graph. Mathematically speaking, the graph $G = <V, E>$ is connected iff for any two nodes $i, j \in V$, there exists a sequence of nodes, $i, i_1, i_2, ..., i_k, j$ such that $(i, i_1), (i_1, i_2), ..., (i_k, j) \in E$. See Figure 1 for illustration.




(a) (b)

**Figure 1.** Random Graphs: (a) connected graph; (b) disconnected graph

## 3 Random Graph

Now consider the following graph $G = <V, E>$ where $V = \{1, 2, ..., n\}$ and $E = \{(i, X(i)), i = 1, 2, ..., n\}$. The $X(i)$ are *independent* random variables such that

$$P\{X(i) = j\} = \frac{1}{n}, \qquad j = 1, 2, ..., n$$

That is, each vertex in the graph $G$ is randomly connected to one of the vertices (including itself) with equal probabilities. We refer graph $G$ as *random graph*.

In this tutorial, we focus on deriving the *probability that the random graph $G$ described above is connected*, denoted as $P\{G \text{ is connected}\}$.

For better derivation, let us assume without loss of generality that the graph starts at vertex 1, thus follows the sequence of vertices, $1, X(1), X^2(1), ..., X^n(1)$, where $X^n(1) = X(X^{n-1}(1))$; and define a random variable $N$ which represents the first $k$ such that $X^k(1)$ is not a new node, i.e.,

$$N = 1\text{st } k \text{ such that } X^k(1) \in \{1, X(1), X^2(1), ..., X^{k-1}(1)\}$$

This means starting from vertex 1, the vertex $k$ goes back to one of the vertices previously visited, see Figure 2 for illustration.

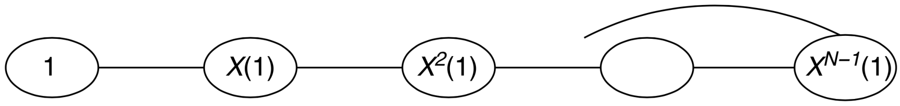



**Figure 2.** Supernode

Therefore, using conditional probability, we can rewrite $P\{G \text{ is connected}\}$ as follows,

$$P\{G \text{ is connected}\} = \sum_{k=1}^{n} P\{G \text{ is connected}|N = k\}P\{N = k\} \tag{1}$$

Looks like we are making our life worse because we now have two items to calculate? No, not if we can proof the following lemma.

**Lemma 1.** *Given a random graph* $G = <V, E>$ *where* $V = \{0, 1, 2, ..., m\}$ $(|V| = m + 1)$ *and* $E = \{(i, Y(i)), i = 1, 2, ..., m\}$, *where*

$$Y(i) = \begin{cases} j & \text{with probability } \frac{1}{m+k}, \quad j = 1, ..., m \\ 0 & \text{with probability } \frac{k}{m+k} \end{cases}$$

*where $k$ is a positive constant, then we have*

$$P\{G \textit{ is connected}\} = \frac{k}{m+k}$$

**Proof.** (**Induction**) As it can be shown that the above lemma holds true when $m = 1$ for all $k$. Now we assume lemma 1 also holds true for all values less than $r$ for all $k$, i.e., $m < r$, in the following we prove lemma 1 also holds true for $m = r$ for all $k$.

Let us first condition on the number of edges $(i, Y(i))$ where $Y(i) = 0$ which is denoted by $Z$. then we have

$$P\{G \text{ is connected}\} = \sum_{z=0}^{r} P\{G \text{ is connected}|Z = z\}\binom{r}{z}\left(\frac{k}{r+k}\right)^z\left(\frac{r}{r+k}\right)^{r-z} \tag{2}$$

since $Z$ is actually a binomial random variable with parameter $\left(r, \frac{k}{r+k}\right)$.

In the above setting, given $Z = z$, we consider the set $\{(i, Y(i))|Y(i) = 0\}$ as one *supernode* 0. Thus we have a new graph $G' = <V', E'>$ with $V' = \{0, 1, 2, ..., m'\}$ and $E' = \{(i, Y'(i)), i = 1, 2, ..., m'\}$, where

$$Y'(i) = \begin{cases} j & \text{with probability } \frac{1}{m'+z}, \quad j = 1, ..., m' \\ 0 & \text{with probability } \frac{z}{m'+z} \end{cases}$$

and $m' = r - z < r$. Since $m' < r$, so by our hypothesis the probability that $G'$ is connected is $\frac{z}{r}$. Therefore, we have

$$P\{G \text{ is connected}|Z = z\} = \frac{z}{r}$$

and by (2) we have

$$P\{G \text{ is connected}\} = \sum_{z=0}^{r} \frac{z}{r}\binom{r}{z}\left(\frac{k}{r+k}\right)^z\left(\frac{r}{r+k}\right)^{r-z} = \frac{1}{r}E(Z) = \frac{k}{r+k}$$

This completes the proof. □

Go back to our problem (1), by applying lemma 1, we have

$$P\{G \text{ is connected}|N=k\}=\frac{k}{n} \tag{3}$$

because we can regard there $k$ nodes $1, X(1), X^2(1), ..., X^{k-1}(1)$ as one *supernode*, within which the nodes are connected to each other; and no edges is emanated from these nodes. The situation is the same as if we have $n-k+1$ nodes, one supernode and $n-k$ ordinary nodes. And each ordinary node is connected to the supernode with probability equals to $\frac{k}{n}$ and all ordinary nodes with probability equals to $\frac{1}{n}$. This is exactly what described in lemma 1.

With (1) and (3), we can derive that

$$P\{G \text{ is connected}\}=\sum_{\boldsymbol{k=1}}^{\boldsymbol{n}}\frac{k}{n}P\{N=k\}=\frac{E(N)}{n} \tag{4}$$

To compute $E(N)$, we define the indicator variable $I_i$ $(i\geqslant 1)$ such that

$$I_i=\begin{cases}1, & \text{if } i\leqslant N\\ 0, & \text{if } i>N\end{cases}$$

Hence,

$$N=\sum_{i=1}^{\infty}I_i$$

Therefore, we have

$$E(N)=E\left(\sum_{i=1}^{\infty}I_i\right)=\sum_{i=1}^{\infty}E(I_i)=\sum_{i=1}^{\infty}P\{N\geqslant i\} \tag{5}$$

The event that $N\geqslant i$ means the nodes $1, X(1), X^2(1), ..., X^{i-1}(1)$ are all different (no repetition in these nodes), which follows that

$$P\{N\geqslant i\}=\frac{(n-1)}{n}\frac{(n-2)}{n}\cdots\frac{(n-i+1)}{n}=\frac{(n-1)!}{(n-i)!n^{i-1}}$$

and so, from (4) and (5), we finally get

$$P\{G \text{ is connected}\}=(n-1)!\sum_{i=1}^{n}\frac{1}{(n-i)!n^i}=\frac{(n-1)!}{n^n}\sum_{j=0}^{n-1}\frac{n^j}{j!}$$

the second equality follows from letting $j=n-i$ and $0\leqslant j\leqslant n-1$ since $1\leqslant i\leqslant N\leqslant n$.

**PS:** There are some other variations of random graph, and a lot more properties can be further explored, please refer to [1] and [2] for details.

# 4 Reference


1. ERDdS, P., and A. R&WI. "On random graphs I." *Publ. Math. Debrecen* 6 (1959): 290-297.
2. Gilbert, E. N. Random Graphs. Ann. Math. Statist. 30 (1959), no. 4, 1141–1144.