COMPLEX NETWORKS, CSYS303

HW06 WRITE-UP

David W. Landay
University of Vermont
Graduate Student, Comp. Systems and Data Sci.

Problem 1

Given N labelled nodes and allowing for all possible number of edges m, what's the total number of undirected, unweighted networks we can construct? How does this number scale with N?

For a network of N nodes, the total number of ways to form an edge between any pair of nodes is simply

$$|E| = \binom{N}{2} \tag{1}$$

If we are concerned with constructing the unweighted, undirected networks, then the allowable number of edges m is a range of possible values defined as the following:

$$m = \{ 0, 1, 2, \ldots, |E| = \binom{N}{2} \}$$

$$0 \leq m \leq \binom{N}{2} \tag{2}$$

Therefore, the number of undirected, unweighted networks we can construct is determined by evaluating

$$\sum_{m=0}^{\binom{N}{2}} \binom{\binom{N}{2}}{m} \tag{3}$$

This number of networks scales as

$$2^{\binom{N}{2}} \tag{4}$$

Since $\binom{N}{2}$ can be re-written as

$$\binom{N}{2} = \frac{N!}{2!\,(N-2)!}$$

$$= \frac{N(N-1)(N-2)\dots 1}{2\,(N-2)!}$$

$$= \frac{N(N-1)\cancel{(N-2)\dots 1}}{2\,\cancel{(N-2)!}}$$

$$= \frac{1}{2}\,N(N-1)$$

then we can re-write **(4)** as

$$2^{\frac{N(N-1)}{2}} \tag{5}$$

Problem 2

Given N labelled nodes and a variable number of edges m, for what value of m do we obtain the largest diversity of networks? And for this m, how does the number of networks scale with N?

Given that $\binom{n}{k}$ is the binomial coefficient, the value of m for which we see the largest diversity of networks is

$$m = \frac{1}{2}\binom{N}{2} \tag{6}$$

This means that for a variable number of edges m, the largest diversity of networks is calculated to be

$$
\begin{aligned}
&\binom{\binom{N}{2}}{\frac{1}{2}\binom{N}{2}} \\
&= \binom{\frac{N(N-1)}{2}}{\frac{N(N-1)}{4}} \\
&= \frac{\frac{N(N-1)}{2}!}{\frac{N(N-1)}{4}!\left(\frac{N(N-1)}{2} - \frac{N(N-1)}{4}\right)!} \\
&= \frac{\frac{N(N-1)}{2}!}{\left(\frac{N(N-1)}{4}!\right)^2} && (7)
\end{aligned}
$$

We may use Stirling's approximation to simplify this expression and determine how the largest diversity of networks scales with the number of nodes. Recall that Stirling's approximation says the following:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \tag{8}$$

For simplicity, we will temporarily assign $A = N(N-1)$. Substituting factorial terms in **(7)** with **(8)**, we get

$$\binom{\binom{N}{2}}{\frac{1}{2}\binom{N}{2}} \sim \frac{\sqrt{2\pi\left(\frac{A}{2}\right)}\left(\frac{\left(\frac{A}{2}\right)}{e}\right)^{\left(\frac{A}{2}\right)}}{\sqrt{2\pi\left(\frac{A}{4}\right)}^{2}\left(\frac{\left(\frac{A}{4}\right)}{e}\right)^{2\left(\frac{A}{4}\right)}}$$

$$= \frac{\sqrt{2\pi\left(\frac{A}{2}\right)}\left(\frac{\left(\frac{A}{2}\right)}{e}\right)^{\left(\frac{A}{2}\right)}}{2\pi\left(\frac{A}{4}\right)\left(\frac{\left(\frac{A}{4}\right)}{e}\right)^{\left(\frac{A}{2}\right)}}$$

$$= \frac{\sqrt{2\pi}\left(\frac{A}{2}\right)^{\left(\frac{A}{2}\right)+\frac{1}{2}}\cancel{e^{-\left(\frac{A}{2}\right)}}}{2\pi\left(\frac{A}{4}\right)^{\left(\frac{A}{2}\right)+1}\cancel{e^{-\left(\frac{A}{2}\right)}}}$$

$$= \frac{\sqrt{2\pi}\left(2\,\frac{A}{4}\right)^{\left(\frac{A}{2}\right)+\frac{1}{2}}}{2\pi\left(\frac{A}{4}\right)^{\left(\frac{A}{2}\right)+1}}$$

$$= \frac{2^{\left(\frac{A}{2}\right)+\frac{3}{2}}}{\sqrt{2\pi\, A}}$$

Substituting A for $N(N-1)$, we get that

$$\binom{\binom{N}{2}}{\frac{1}{2}\binom{N}{2}} \sim \frac{2^{\frac{N(N-1)}{2}+\frac{3}{2}}}{\sqrt{2\pi\, N(N-1)}} \tag{9}$$

Given that $2^{\frac{N(N-1)}{2}+\frac{3}{2}} \gg \sqrt{2\pi\, N(N-1)}$, and $2^{\frac{N(N-1)}{2}} \gg \frac{3}{2}$, we get that the number of networks scales with

$$2^{\frac{N(N-1)}{2}} \tag{10}$$

Problem 3

We've seen that large random networks have essentially no clustering, meaning that locally, random networks are pure branching networks. Nevertheless, a finite,non-zero number of triangles will be present. For pure random networks, with connection probability $p = \frac{\langle k \rangle}{(N-1)}$, what is the expected total number of triangles as $N \to \infty$?

If p is the probability that an edge exists, and a triangle is formed by the existence of three edges between any three nodes, then we can first determine the probability of the existence of any three edges. That looks like this:

$$p^3 = \text{Probability of 3 nodes being connected}$$

$$= \left(\frac{\langle k \rangle}{(N-1)}\right)^3 \tag{11}$$

Next, we want to determine the number of ways to choose any three nodes. That is simply

$$\text{the number of ways to choose 3 nodes to connect} = \binom{\text{Cycle-length}}{2}\binom{N}{3} \tag{12}$$

We multiply by the number of ways to create a cycle of length n because depending on cycle-length, there might be multiple ways to form a cycle based on the number of ways to connect that number of nodes.

Therefore, the expected total number of triangles (3-cycles) is given by:

$$\binom{3}{2}\binom{N}{3}p^3$$

$$1 \times \binom{N}{3}\left(\frac{\langle k \rangle}{(N-1)}\right)^3 \therefore \tag{13}$$

This expression converges for large N (namely $N \to \infty$). First,

$$\binom{N}{3} = \frac{N!}{3!(N-3)!}$$

$$= \frac{N(N-1)(N-2)\cancel{(N-3)!}}{6\cancel{(N-3)!}}$$

$$= \frac{N(N-1)(N-2)}{6} \tag{14}$$

Combining **(13)** and **(14)**, we get

$$\left(\frac{\langle k\rangle^3}{(N-1)^3}\right)\left(\frac{N(N-1)(N-2)}{6}\right)$$

$$= \left(\frac{\langle k\rangle^3 N(N-2)}{6(N-1)^2}\right)$$

$$= \left(\frac{\langle k\rangle^3 (N^2-2N)}{6(N^2-2N+1)}\right) \tag{15}$$

As $N \to \infty$, we take the limit of **(15)** and find that it converges:

$$\lim_{N\to\infty}\left(\frac{\langle k\rangle^3 (N^2-2N)}{6(N^2-2N+1)}\right)$$

$$\to \frac{\langle k\rangle^3}{6} \therefore \tag{16}$$

Problem 4

Repeat the preceding calculation for cycles of length 4 and 5 (triangles are cycles of length 3).

For finding 4 and 5-cycles, the logic is the same. We want to determine the number of ways to form a 4 or 5-cycle (i.e: find the number of ways to choose 4 or 5 nodes to form edges between), and then determine the probability of the existence of any 4 or 5 edges in a network. Then, the number of 4-cycles or 5-cycles will be the product of these two values.

$$\binom{4}{2}\binom{N}{4}\left(\frac{\langle k\rangle}{(N-1)}\right)^4 \tag{17}$$

and...

$$\binom{5}{2}\binom{N}{5}\left(\frac{\langle k\rangle}{(N-1)}\right)^5 \tag{18}$$

Which converge to

$$\frac{6\langle k\rangle^4}{4!} \tag{19}$$

and ...

$$\frac{10\langle k\rangle^5}{5!} \tag{20}$$

$$\therefore$$

respectively

Problem 5

Show that the second moment of the Poisson distribution is

$$\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle$$

and hence that the variance $\sigma^2 = \langle k \rangle$

The Poisson distribution measures the probability of a given number of events occurring in a fixed interval of time/space, where events occur at a constant rate and are independent of previous outcomes. This probability is often written as

$$\text{Pr.}(X = k) = e^{-\lambda} \left(\frac{\lambda^k}{k!} \right)$$

where X is a random variable that denotes the number of successful outcomes (k) in a given interval.

Using a moment generating function, we can derive any n_{th} moment about the probability distribution function that defines a Poisson distribution. The n_{th} moment about a discrete probability distribution can be generated by the following:

$$\langle k^n \rangle = \mathbb{E}(k^n) = \sum_{k=0}^{\infty} P(X = k) k^n \tag{21}$$

Hence, the first moment about the Poisson distribution can be derived as follows:

$$\langle k \rangle \;=\; \sum_{k=0}^{\infty} k\, e^{-\lambda} \left(\frac{\lambda^{k}}{k!}\right)$$

factor out constants, and re-index the sum …

$$e^{-\lambda}\,\lambda \sum_{k=1}^{\infty} k \left(\frac{\lambda^{k\,-\,1}}{k!}\right)$$

$$e^{-\lambda}\,\lambda \sum_{k=1}^{\infty} \cancel{k} \left(\frac{\lambda^{k\,-\,1}}{\cancel{k}(k\,-\,1)!}\right)$$

$$e^{-\lambda}\,\lambda \Big[1 \,+\, \lambda \,+\, \frac{\lambda}{2!} \,+\, \ldots\Big]$$

Here, we recognize that the infinite sum $\Big[1 \,+\, \lambda \,+\, \frac{\lambda}{2!} \,+\, \ldots\Big]$ is the expansion of e^{λ}. Hence, we can substitute one term for the other and arrive at the following:

$$e^{-\lambda}\,\lambda \Big[1 \,+\, \lambda \,+\, \frac{\lambda}{2!} \,+\, \ldots\Big] \;=\; \cancel{e^{-\lambda}}\,\lambda\,\cancel{e^{\lambda}}$$

$$\Longrightarrow\; \langle k \rangle \;=\; \lambda \;\therefore \tag{22}$$

Finding the second moment of the Poisson distribution is a similar procedure:

$$\langle k^2 \rangle = \sum_{k=0}^{\infty} k^2 \, e^{-\lambda} \left(\frac{\lambda^k}{k!} \right)$$

$$= \sum_{k=0}^{\infty} \Big[k(k-1) + k \Big] \, e^{-\lambda} \left(\frac{\lambda^k}{k!} \right)$$

$$= \sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \left(\frac{\lambda^k}{k!} \right) + \sum_{k=0}^{\infty} k \, e^{-\lambda} \left(\frac{\lambda^k}{k!} \right)$$

$$= \sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \left(\frac{\lambda^k}{k!} \right) + \langle k \rangle$$

$$= \sum_{k=0}^{\infty} \cancel{k(k-1)} e^{-\lambda} \left(\frac{\lambda^k}{\cancel{k(k-1)}(k-2)!} \right) + \langle k \rangle$$

factor out constants, and re-index the sum …

$$= e^{-\lambda} \, \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \langle k \rangle$$

$$= e^{-\lambda} \, \lambda^2 \left[1 + \lambda + \frac{\lambda}{2!} + \ldots \right] + \langle k \rangle$$

$$= \cancel{e^{-\lambda}} \, \lambda^2 \, \cancel{e^{\lambda}} + \langle k \rangle$$

$$= \lambda^2 + \langle k \rangle$$

$$\implies \langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle \; \therefore \tag{23}$$

Problem 6

We've figured out in class that for large enough N (and $\langle k \rangle$ fixed) that a random network always has a Poisson degree distribution:

$$P(k;\lambda) = e^{-\lambda}\left(\frac{\lambda^k}{k!}\right) \tag{24}$$

where $\langle k \rangle = \lambda$. And as we've discussed, we don't find these networks in the real world (they don't arise due to simple mechanisms). Let's investigate this oddness a little further: Compute the expected size of the largest degree in an infinite random network given $\langle k \rangle$ and as a function of increasing sample size N. n other words, in selecting (with replacement) N degrees from a pure Poisson distribution with mean $\langle k \rangle$, what's the expected minimum value of the largest degree $\mathbf{min} k_{\text{max}}$?

A good way to compute $k_{\mathbf{max}}$ is to equate it to the value for which we expect $\frac{1}{N}$ of our random selections to exceed.

For simplicity, we will let $\mathbf{min} k_{\text{max}} = \tilde{k}$. We wish to find the expected value of $\tilde{k}$ for which the following is true:

$$\sum_{k=\tilde{k}}^{\infty} P(k;\lambda) = \frac{1}{N} \tag{25}$$

Hence, we are looking to find the following:

$$P(k \geq \tilde{k}) = \sum_{k=\tilde{k}}^{\infty} e^{-\lambda}\left(\frac{\lambda^k}{k!}\right) = \frac{1}{N} \tag{26}$$

First, we recognize that $e^{-\lambda}$ is a constant term, so we can place it outside of the sum. We can expand the sum to get:

$$P(k \geq \tilde{k}) = e^{-\lambda}\left[\frac{\lambda^{\tilde{k}}}{\tilde{k}!} + \frac{\lambda^{(\tilde{k}+1)}}{(\tilde{k}+1)!} + \frac{\lambda^{(\tilde{k}+2)}}{(\tilde{k}+2)!} + \ldots\right]$$

$$= e^{-\lambda}\,\frac{1}{\tilde{k}!}\left[\lambda^{\tilde{k}} + \frac{\lambda^{(\tilde{k}+1)}}{(\tilde{k}+1)} + \frac{\lambda^{(\tilde{k}+1)}}{(\tilde{k}+2)} + \ldots\right]$$

$$= e^{-\lambda}\,\frac{\lambda^{\tilde{k}}}{\tilde{k}!}\left[1 + \frac{\lambda}{(\tilde{k}+1)} + \frac{\lambda^2}{(\tilde{k}+2)} + \ldots\right]$$

Since we assume that k is large, then the expanded sum is negligible. Thus, we get:

$$P(k \geq \tilde{k}) \sim e^{-\lambda}\,\frac{\lambda^{\tilde{k}}}{\tilde{k}!}$$

$$\implies e^{-\lambda}\,\frac{\lambda^{\tilde{k}}}{\tilde{k}!} \sim \frac{1}{N}$$

As done in previous problems, we can approximate $\tilde{k}!$ using Stirling's approximation. We find the following:

$$e^{-\lambda}\,\frac{\lambda^{\tilde{k}}}{\tilde{k}!} \approx \frac{e^{-\lambda}\lambda^{\tilde{k}}}{\sqrt{2\pi\tilde{k}}\left(\frac{\tilde{k}}{e}\right)^{\tilde{k}}}$$

$$= e^{(-\lambda - \tilde{k})}\,\lambda^{\tilde{k}}\,\frac{1}{\sqrt{2\pi}}\,\tilde{k}^{(-\tilde{k} - \frac{1}{2})}$$

$$= e^{(-\lambda - \tilde{k})}\,\lambda^{\tilde{k}}\,\tilde{k}^{(-\tilde{k} - \frac{1}{2})},\ \text{since } \tilde{k} \text{ is large } \ldots$$

$$\implies e^{(-\lambda - \tilde{k})}\,\lambda^{\tilde{k}}\,\tilde{k}^{(-\tilde{k} - \frac{1}{2})} \sim \frac{1}{N}$$

Now we wish to write $\tilde{k}$ in terms of N. Continuing, we get:

$$
\begin{aligned}
&\ln\left[e^{(-\lambda\ -\ \tilde{k})}\,\lambda^{\tilde{k}}\,\tilde{k}^{(-\tilde{k}\ -\ \frac{1}{2})}\right]\ \sim\ \ln\left(\frac{1}{N}\right)\\
&\implies\ -\lambda\ -\ \tilde{k}\ +\ \tilde{k}\ln(\lambda)\ -\ (\tilde{k}\ +\ \frac{1}{2})\ln(\tilde{k})\ \sim\ \ln\left(\frac{1}{N}\right)\\
&\implies\ -\ \tilde{k}\ +\ \tilde{k}\ln(\lambda)\ -\ (\tilde{k}\ +\ \frac{1}{2})\ln(\tilde{k})\ \sim\ \ln\left(\frac{1}{N}\right)\\
&\implies\ \tilde{k}\ +\ (\tilde{k}\ +\ \frac{1}{2})\ln(\tilde{k})\ -\ \tilde{k}\ln(\lambda)\ \sim\ \ln(N)\\
&\implies \tilde{k} \lesssim \ln(N)\ \therefore
\end{aligned} \tag{27}
$$

So, the expected value of $\mathbf{min}k_{\text{max}}$ is strictly less then or similar to $\ln(N)$.