Massachusetts Institute of Technology
Departments of Electrical Engineering & Computer Science and Economics
6.207/14.15 Networks
Spring 2017

Problem Set 3

Issued: Tuesday, March 14, 2017 **Due:** Thursday, April 06, 2017

Problem 3.1

[P 11.3 Newman] Consider a "line graph" consisting of n vertices in a line like this:

(a) Show that if we divide the network into two parts by cutting any single edge, such that one part has r vertices and the other has $n - r$, the modularity, Eq.(7.76), takes the value:

$$Q = \frac{3 - 4n + 4rn - 4r^2}{2(n-1)^2} \tag{1}$$

(b) Hence show that when n is even the optimal such division, in terms of modularity, is the division that splits the network exactly down the middle.

Solution:

(a) From the equation in Newman, we can write our modularity, Q, as:

$$Q = \frac{1}{2m} \sum_{i,j} B_{ij} \delta(c_i, c_j) \tag{2}$$

where m is the total number of edges, $B_{ij} = A_{ij} - \frac{k_i k_j}{2m}$, where k_i is the degree of node i and let $\delta(c_i, c_j) = \frac{1}{2}(s_i s_j + 1)$. Using the fact that $\sum_j B_{ij} = 0$ (Eq 11.41 in Newman), we can write our modularity as:

$$\begin{aligned} Q &= \frac{1}{4m} \sum_{ij} B_{ij} s_i s_j && (3) \\ &= \frac{1}{4m} \mathbf{s^T B s} && (4) \end{aligned}$$

Since we have two classes (partitions), we can assign all $s_i = 1$ if node i is in class 1, and $s_i = -1$ if node i is in class 2. Notice that $s_i s_j = 1$ if nodes i and j are in the same class and $s_i s_j = -1$ if they are in different classes. Using this, we break up our summation as follows:

$$
\begin{aligned}
Q = &\quad \sum_{j=1}^{r} \quad \sum_{i=1}^{r} B_{ij} s_i s_j + \sum_{i=r+1}^{n} B_{ij} s_i s_j \Bigg) + \sum_{j=r+1}^{n} \quad \sum_{i=1}^{r} B_{ij} s_i s_j + \sum_{i=r+1}^{n} B_{ij} s_i s_j \Bigg)\Bigg) && (5) \\
&= \frac{1}{4m} \quad \sum_{j=1}^{r} \sum_{i=1}^{r} B_{ij} - 2 \sum_{j=r+1}^{n} \sum_{i=1}^{r} B_{ij} + \sum_{j=r+1}^{n} \sum_{i=r+1}^{n} B_{ij} \Bigg) && (6) \\
&= \frac{1}{4m} \quad \sum_{j=1}^{r} \sum_{i=1}^{r} \left(A_{ij} - \frac{k_i k_j}{2m} \right) - 2 \sum_{j=r+1}^{n} \sum_{i=1}^{r} \left(A_{ij} - \frac{k_i k_j}{2m} \right) + \sum_{j=r+1}^{n} \sum_{i=r+1}^{n} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \Bigg) && (7) \\
& && (8)
\end{aligned}
$$

Notice that the sum over each A_{ij} is just the sum of the degrees of the corresponding nodes, and our degree products $k_i k_j$ take the following form:

$$
k_i k_j = \begin{cases} 1 & i = 1 \text{ or } i = n \text{ and } j = 1 \text{ or } j = n \\ 2 & i = 1 \text{ or } i = n \\ 2 & j = 1 \text{ or } j = n \\ 4 & \text{otherwise} \end{cases} \tag{9}
$$

Using the above gives us:

$$\sum_{j=1}^{r}\sum_{i=1}^{r} A_{ij} = 2(r-1) \tag{10}$$

$$\sum_{j=1}^{r}\sum_{i=r+1}^{n} A_{ij} = 1 \tag{11}$$

$$\sum_{j=r+1}^{n}\sum_{i=r+1}^{n} A_{ij} = 2(n-r-1) \tag{12}$$

$$\sum_{j=1}^{r}\sum_{i=1}^{r} k_i k_j = 4r^2 - 4r + 1 \tag{13}$$

$$\sum_{j=1}^{r}\sum_{i=r+1}^{n} k_i k_j = -4r^2 + 4rn - 2n + 1 \tag{14}$$

$$\sum_{j=r+1}^{n}\sum_{i=r+1}^{n} k_i k_j = 4(n-r)^2 - 4n + 4r + 1 \tag{15}$$

Plugging back into our equation for Q, setting $m = (n-1)$ and simplifying, we get our desired answer of:

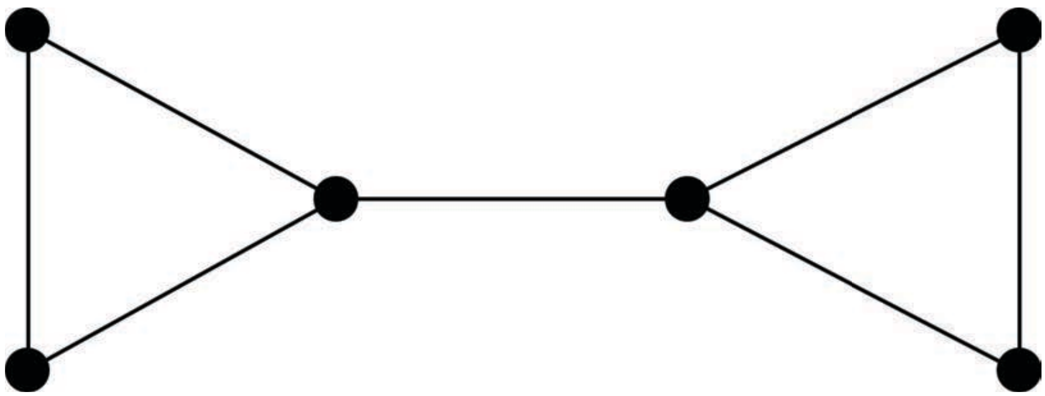
$$Q = \frac{3 - 4n + 4rn - 4r^2}{2(n-1)^2} \tag{16}$$

(b) We can find the value of r which maximizes our modularity, Q, by setting its derivative to 0. We have:

$$\frac{dQ}{dr} = \frac{2n - 4r}{(n-1)^2} \tag{17}$$

Setting $\frac{dQ}{dr} = 0$, and solving for r, we get $r = \frac{n}{2}$.

Problem 3.2
[P 11.4 Newman] Using your favorite numerical software for finding eigenvectors of matrices, construct the Laplacian and modularity matrix for this small network:



(a) Find the eigenvector of the Laplacian corresponding to the second smallest eigenvalue and hence perform a spectral bisection of the network into two equally sized parts.

(b) Find the eigenvector of the modularity matrix corresponding to the largest eigenvalue and hence divide the network into two communities.

You should find that the division of the network generated by the two methods is in this case, the same.

Solution:

(a) For parts (a) and (b) we use the following adjacency matrix:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{18}$$

We should get the following laplacian matrix:

$$L = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & -1 \\ -1 & 3 & -1 & 0 & 0 & -1 \\ 0 & -1 & 3 & -1 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & -1 & 2 & 0 \\ -1 & -1 & 0 & 0 & 0 & 2 \end{pmatrix} \tag{19}$$

The second smallest eigenvalue is: 0.438
The eigenvector corresponding to the second smallest eigenvalue is:

$$\begin{bmatrix} 0.46470513 \\ 0.26095647 \\ -0.26095647 \\ -0.46470513 \\ -0.46470513 \\ 0.46470513 \end{bmatrix} \tag{20}$$

This gives us the following partition vector:

$$\mathbf{s} = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \\ 1 \\ -1 \end{bmatrix} \tag{21}$$

(b) We should get the following modularity matrix:

$$\mathbf{B} = \begin{pmatrix} -0.28571429 & 0.57142857 & -0.42857143 & -0.28571429 & -0.28571429 & 0.71428571 \\ 0.57142857 & -0.64285714 & 0.35714286 & -0.42857143 & -0.42857143 & 0.57142857 \\ -0.42857143 & 0.35714286 & -0.64285714 & 0.57142857 & 0.57142857 & -0.42857143 \\ -0.28571429 & -0.42857143 & 0.57142857 & -0.28571429 & 0.71428571 & -0.2857142 \\ -0.28571429 & -0.42857143 & 0.57142857 & 0.71428571 & -0.28571429 & -0.28571429 \\ 0.71428571 & 0.57142857 & -0.42857143 & -0.28571429 & -0.28571429 & -0.28571429 \end{pmatrix} \tag{22}$$

The largest eigenvalue is: 1.732

The eigenvector corresponding to the largest eigenvalue is:

$$\begin{bmatrix} 0.44403692 \\ 0.32505758 \\ -0.32505758 \\ -0.44403692 \\ -0.44403692 \\ 0.44403692 \end{bmatrix} \tag{23}$$

This gives us the following partition vector:

$$\mathbf{s} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \tag{24}$$

For both methods, you should see that our graph is partitioned straight down the middle.

Problem 3.3

Consider an Erdös-Renyi random graph $G(n, p)$

(a) Let A_1 denote the event that node 1 has at least $l \in \mathbb{Z}^+$ neighbhors. Do we observe a phase transition for this event? If so, find the threshold function and explain your reasoning.

(b) Let B denote the event that a cycle with k edges (for a fixed k) emerges in the graph. Do we observe a phase transition of this event? If so, find the threshold function and explain your reasoning.

Solution:

(a) For any finite $l \in \mathbb{Z}^+$, phase transition is observed. To see this, consider the candidate threshold function $t(n) = \frac{r}{n}$ for any $r \in \mathbb{R}^+$. We need to prove for any l:

(i) $P(A_l|p(n)) \to 0$ if $\frac{p(n)}{t(n)} \to 0$

(ii) $P(A_l|p(n)) \to 1$ if $\frac{p(n)}{t(n)} \to \infty$

First assume that $\frac{p(n)}{t(n)} \to 0$. Denote the degree of node 1 by d_1. Since $\frac{p(n)}{t(n)} \to 0$, the expected degree satisfies:

$$\begin{aligned} \mathbb{E}[d_1] &= (n-1)p(n) && (25)\\ &= \frac{p(n)}{t(n)} t(n)(n-1) && (26)\\ &\cong \frac{p(n)}{t(n)} \frac{r(n-1)}{n} && (27)\\ \mathbb{E}[d_1] &\to 0 && (28)\\ & && (29) \end{aligned}$$

Note that this implies $P(A_l|p(n) \to 0$, since otherwise, the expected degree would be strictly positive.
Next assume that $\frac{p(n)}{t(n)} \to \infty$. It follows that $p(n) > \frac{r}{n}$ for any $r \in \mathbb{R}^+$ and sufficiently large n. The probability that A_l does not occur can be bounded as follows:

$$\begin{aligned} P(A_l^c|p(n)) &= \sum_{k=0}^{l-1} P(d_1 = k|p(n)) && (30)\\ &= \sum_{k=0}^{l-1} p(n)^k (1-p(n))^{n-1-k} \binom{n-1}{k} && (31)\\ &\leq \sum_{k=0}^{l-1} t(n)^k (1-t(n))^{n-1-k} \binom{n-1}{k} && (32)\\ &\leq \sum_{k=0}^{l-1} t(n)^k (1-t(n))^{n-1-k} \frac{n^k}{n!} && (33)\\ &= \sum_{k=0}^{l-1} \frac{r}{n}^k (1-\frac{r}{n})^{n-1-k} \frac{n^k}{n!} && (34)\\ &\cong \sum_{k=0}^{l-1} \exp(-r) \frac{r^k}{k!} && (35) \end{aligned}$$

Here the third line follows because if the graph was generated using $t(n)$ instead of $p(n)$, each link would be present with smaller probability and hence the probability that node 1 has less than l neighbors (the event A_l^c) would be larger. Since the above is true for any $r \in \mathbb{R}^+$, considering arbitrarily large r, it follows that:

$$P(A_l^c|p(n)) \to 0 \tag{36}$$

or equivalently:

$$P(A_l|p(n)) \to 1 \tag{37}$$

provided that $\frac{p(n)}{t(n)} \to \infty$

(b) We observe phase transition for this part as well. Similar to part (a), consider the candidate threshold function $t(n) = \frac{r}{n}$ for any $r \in \mathbb{R}^+$. We will prove that for a fixed k, the event B satisfies:

(i) $P(B|p(n)) \to 0$ if $\frac{p(n)}{t(n)} \to 0$

(ii) $P(B|p(n)) \to 1$ if $\frac{p(n)}{t(n)} \to \infty$

In order to prove (i), assume that $\frac{p(n)}{t(n)} \to 0$. Denote the number of distinct cycles on k nodes by C_k. Note that over n nodes, $\binom{n}{k}\frac{(k-1)!}{2}$ different cycles (of k nodes) can be observed and each cycle is realized with $p(n)^k$ probability. Therefore, the expectation of C_k can be found as:

$$\mathbb{E}[C_k] = \binom{n}{k}\frac{(k-1)!}{2}p(n)^k \tag{38}$$

Note that:

$$\begin{aligned}
\mathbb{E}[C_k] &\leq \frac{n^k}{2k}p(n)^k && (39)\\
&= \frac{n^k t(n)^k}{2k}\left(\frac{p(n)}{t(n)}\right)^k && (40)\\
&= \frac{r^k}{2k}\left(\frac{p(n)}{t(n)}\right)^k && (41)\\
& && (42)
\end{aligned}$$

hence $\mathbb{E}[C_k] \to 0$ as $\frac{p(n)}{t(n)} \to 0$. Note that this implies $P(B|p(n)) \to 0$, since otherwise, the expectation of C_k would be strictly positive.

Next assume that $\frac{p(n)}{t(n)} \to \infty$. In a graph with n nodes, there can be at most $N = \binom{n}{k}(k-1)! \cong c_0 n^k$ distinct cycles with k nodes (for a constant c_0). We enumerate these cycles and for $i \in 1...N$ define a random variable I_i such that $I_i = 1$ if the ith cycle is realized and 0 otherwise. Note that the probability that no cycle is realized satifies:

$$P \quad \sum_{i=1}^{N} I_i = 0 \Bigg) \leq \frac{\operatorname{var}\left(\sum_{i=1}^{N} I_i\right)}{\mathbb{E}\left[\sum_{i=1}^{N} I_i\right]^2} \tag{43}$$

$$\tag{44}$$

Calculating the expectations of the indicator variables we conclude that:

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^{N} I_i\right] &= \sum_{i=1}^{N} \mathbb{E}[I_i] && (45) \\ &= Np(n)^k && (46) \\ &\cong c_0 r^k \left(\frac{p(n)}{t(n)}\right)^k && (47) \end{aligned}$$

Next we upper bound var $\left(\sum_{i=1}^{N}\right)$. In order to do so, we use the following identity:

$$\operatorname{var} \quad \sum_{i=1}^{N} I_i \Bigg) = \sum_{i=1}^{N} \operatorname{var}(I_i) + \sum_{i \neq j} \operatorname{cov}(I_i, I_j) \tag{48}$$

$$\tag{49}$$

Using the properties of Bernoulli random variables, it follows that

$$\operatorname{var}(I_i) = p(n)^k (1 - p(n)^k) \tag{50}$$

On the other hand, $\operatorname{cov}(I_i, I_j) = \mathbb{E}[I_i I_j] - \mathbb{E}[I_i][I_j] = 0$ if the cycles i, j do not have an edge in common, since in this case the cycles are independent. Assume the cycles i and j have l common edges (hence $l+1$) common nodes. In this case:

$$\begin{aligned} \operatorname{cov}(I_i, I_j) &= \mathbb{E}[I_i I_j] - \mathbb{E}[I_i][I_j] && (51) \\ &= p(n)^l p(n)^{2k-2l} - p(n)^{2k} && (52) \\ &\leq p(n)^{2k-l} && (53) \end{aligned}$$

Also note that there are at most $\binom{n}{l+1}\binom{n}{2(k-l-1)}c_l$ such i, j pairs. This can be obtained first by identifying the common nodes, and then choosing the remaining nodes of both graphs and then considering ordergins of nodes in cycle (which is captured by the constant c_l). Combining the above, and calculating the sum of the covariances by condition on the number of common edges between i, j it follows that:

$$\text{var}\left(\sum_{i=1}^{N} I_i\right) \leq \sum_{i=1}^{N} p(n)^k(1-p(n)^k) + \sum_{l=1}^{k-1}\binom{n}{l+1}\binom{n}{2(k-l-1)}c_l p(n)^{2k-l} \tag{54}$$

$$\leq c_0 n^k p(n)^k + \sum_{l=1}^{k-1} n^{2k-l-1} c_l p(n)^{2k-l} \tag{55}$$

$$\leq c_0\left(\frac{p(n)}{t(n)}\right)^k + \sum_{l=1}^{k-1} c_l \frac{r^{2k-l}}{n}\left(\frac{p(n)}{t(n)}\right)^{2k-l} \tag{56}$$

Using the bound on variance we conclude:

$$P\left(\sum_{i=1}^{N} I_i = 0\right) \leq \frac{c_0 r^k\left(\frac{p(n)}{t(n)}\right)^k + \sum_{l=1}^{k-1} c_l \frac{r^{2k-l}}{n}\left(\frac{p(n)}{t(n)}\right)^{2k-l}}{\left(c_0 r^k\left(\frac{p(n)}{t(n)}\right)^k\right)^2} \tag{57}$$

Since $\frac{p(n)}{t(n)} \to 0$ this equation implies that

$$P\left(B^c | p(n)\right) = P\left(\sum_{i=1}^{N} I_i = 0\right) \to 0 \tag{58}$$

as claimed.

Problem 3.4

[P 12.6 Newman] We can make a simple random graph model of a network with clustering or transitivity as follows. We take n vertices and go through each distinct trio of three vertices, of which there are $\binom{n}{3}$, and with independent probability p we connect the members of the trio together using three edges to form a triangle, where $p = \frac{c}{\binom{n-1}{2}}$ with c constant.

(a) Show that the mean degree of a vertex in this network is $2c$.

(b) Show that the degree distribution is

$$p_k = \begin{cases} e^{-c}c^{k/2}/(k/2)! & \text{if } k \text{ is even,} \\ 0 & \text{if } k \text{ is odd.} \end{cases}$$

(c) Show that the clustering coefficient is $C = \frac{1}{2c+1}$.

(d) Show that when there is a giant component in the network, its expected size S, as a fraction of the network size, satisfies $S = 1 - e^{-cS(2-S)}$.

(e) What is the value of the clustering coefficient when the giant component fills half of the network?

Solution:

(a) For each vertex, there are $\binom{n-1}{2}$ pairs of others with which it cold form a triangle, and each triangle is present with probability $\frac{c}{\binom{n-1}{2}}$, for an average number of triangles c per vertex. Each triangle contributes two edges to the deree, so the average degree is $2c$.

(b) The probability p_t of having t triangles follows the binomial distribution:

$$\begin{aligned} p_t &= \binom{\binom{n-1}{2}}{t} p^t (1-p)^{\binom{n-1}{2}-t} && (59) \\ &\simeq e^{-c}\frac{c^t}{t!} && (60) \end{aligned}$$

where the final equality is exact in the limit of large n. The degree is twice the number of triangles and hence $t = k/2$ and:

$$p_k = e^{-c}\frac{c^{k/2}}{(k/2)!} \tag{61}$$

so long as k is even. Odd values of k cannot occur, so $p_k = 0$ for k odd.

(c) As shown above, there are on average c triangles around each vertex and hence nc is the total number of triangles in the network times three (since each triangle appears around three different vertices and gets counted three times).
The number of connected triples around a vertex of degree $k = 2t$ is $\binom{2t}{2} = t(2t-1)$ and there are np_t vertices with t triangles, with p_t as above. So the total number of connected triples is:

$$ne^{-c}\sum_{t=0}^{\infty} t(2t-1)\frac{c^t}{t!} = ne^{-c}(2c^2+c)e^c \tag{62}$$

$$= nc(2c+1) \tag{63}$$

(The sum is a standard one that can be found in tables, but it's also reasonably straightforward to do by hand if you know the right tricks)
Now the clustering coefficient is

$$C = \frac{(\text{number of triangles}) * 3}{(\text{number of connected triples})} \tag{64}$$

$$= \frac{nc}{nc(2c+1)} \tag{65}$$

$$= \frac{1}{2c+1} \tag{66}$$

(d) Let u be the probability that a vertex is not in the giant component. If a vertex is not in the giant component, then it must be that for each of the $\binom{n-1}{2}$ distinct pairs of other vertices in the network. Either (a) that pair does not form a triangle with our vertex (probability $1-p$) or (b) the pair does form a triangle (probability p) but neither member of the pair is itself in the giant component (probability u^2). Thus the analog of Eq. (12.12) for this model is

$$u = (1 - p + pu^2)^{\binom{n-1}{2}} \tag{67}$$

Putting $p = c/\binom{n-1}{2}$ and taking the limit of large n this becomes $u = e^{-c(1-u^2)}$. Putting $S = 1-u$ we then find that $S = 1 - e^{-cS(2-S)}$.

(e) Rearranging for c in terms of S we have

$$c = -\frac{\ln(1-s)}{S(2-S)} \tag{68}$$

and for $S = 1/2$ this gives $c = \frac{4}{3}\ln(2)$. Substituting into the expression for the clustering coefficient above then gives

$$C = \frac{1}{1 + \frac{8}{3}\ln(2)} \tag{69}$$

$$\cong 0.351 \tag{70}$$

Problem 3.5

Consider the variation on the small-world model proposed by Newman and Watts: consider a ring lattice with n nodes in which each node is connected to its neighbors k hops or less away. For each edge, with one probability p, add a new edge to the ring lattice between two nodes chosen uniformly at random.

(a) Find the degree distribution of this model.

(b) Show that when $p = 0$, the overall clustering coefficient of this graph is given by

$$Cl(g) = \frac{3k-3}{4k-2}.$$

(c) *(Optional for Bonus)*: Show that when $p > 0$, the overall clustering coefficient is given by

$$Cl(g) = \frac{3k-3}{4k-2}(1-p)^3.$$

Solution:

(a) Degree of vertex $= 2k +$ (number of shortcut edges attached to it). The number of shortcut edges is nk. For each such edge: add shortcut with probability p. There are nkp shorts on average. There are $2nkp$ ends of shortcuts on average. There are $2kp$ ends of shortcuts on average per vertex. Similar to Erdos-Renyi, we will have a Poisson distribution in the limit of large n:

$$p_s = e^{-2kp}\frac{(2kp)^s}{s!} \tag{71}$$

Therefore, the degree distribution is

$$p_d = \begin{cases} e^{-2kp}\frac{(2kp)^{d-2k}}{(d-2k)!} & d = 2k, 2k+1, \ldots \\ 0 & d < 2k \end{cases} \tag{72}$$

(b) First, we give the labels $1, \ldots, n$ to the nodes in a counter clockwise fashion starting from an arbitrary node. when $p = 0$, two nodes with labels u and v have an edge if they are at most k hops away, i.e., if $|u - v| \leq k$. We will compute the clustering coefficient by using the following definition:

$$Cl(g) = \frac{3 * (\text{triangles})}{\text{number of connected triples}} \tag{73}$$

The following expression gives the total number of triangles that agenet 1 forms with agents 2 to $k + 1$ (note that this is not the total number of triangles that agent 1 can form, since we are not cunting (yet) the triangles with agents $n - k + 1$ to n):

$$\sum_{i=3}^{k+1} \sum_{j=2}^{i-1} 1 = \frac{k(k-1)}{2} \tag{74}$$

From symmetry we conclude that the total number of triangles is simply $n\frac{k(k-1)}{2}$ (note that by considering the triangles that 1 forms just with agents 2 to $k + 1$ and continuing in the same fashion with agents 2 to n we avoided double-counting any triangles). Similarly, the number of triples that agent 1 forms with agents 2 to $2k + 1$ is given by:

$$\sum_{i=2}^{k+1} \sum_{j=i+1}^{k+i} 1 + 2\frac{k(k-1)}{2} \tag{75}$$

where the second term in the summation comes from the fact that each triangle is counted just once in the first term. Thus, the overall clustering coefficient is simply given by the ratio:

$$\frac{3k(k-1)}{2k^2 + 2k(k-1)} = \frac{3k - 3}{4k - 2} \tag{76}$$

(c) The overall clustering coefficient is defined as the average of the clustering coefficients of individual nodes in the graph. The clustering coefficient for agient i is defined as the ratio of all links between the neighbors of i over the number of all potential links between the neighbors. If agent i as n_i neighbors, then the number of potential links between i's neighbors is $\frac{n_i(n_i-1)}{2}$. As $n \to \infty$ the above definition is equivalent with the following:

$$\tilde{C}(g) = \frac{\text{expected number of links between the neighbors of a node}}{\text{expected number of potential links between the neighbors of a node}} \tag{77}$$

Next note that two neighbors of agent i that were connected at $p = 0$ will remain connected and linked with i with probability $(1-p)^3$ when $p > 0$. Thus, the expected number of links between the neighbors of a node is equal to $\frac{2k(k-1)}{2}(1-p)^3 + O(1/n)$, where the second term is negligible as $n \to \infty$ and corresponds to a new triangle being formed by two edges that were rewired and one that was not rewired, etc.
On the other hand, the expected number of potential links between the neighbors of a node remains the same as in the case of $p = 0$ and $\tilde{C}(g) = \frac{3k-3}{4k-2}(1-p)^3$.

Problem 3.6

Consider a society of n individuals. A randomly chosen node is infected with a contagious infection. Assume that the network of interactions in the society is represented by a configuration model with degree distribution p_k. Assume that any individual is immune independently with probability π. We would like to investigate whether the infection can spread to a nontrivial fraction of the society.

(a) Find a threshold for the immunity probability (in terms of the moments of the degree distribution) below which the infection spreads to a large portion of the population.

(b) What is this threshold for a k-regular random graph, i.e., a configuration model network in which all nodes have the same degree.

(c) What is this threshold for a power law graph with exponents less than 3, i.e., $p_k \sim k^{-\alpha}$ with $\alpha < 3$? The Internet graph (representing connections between routers) has a power law distribution with exponent $\sim 2.1 - 2.7$. What does this result imply for the Internet graph?

(d) Find the size of the infected population (you can assume that the infection spreads to a large portion of the population).

Solution:

(a) Degree distribution of a neighboring node:

$$\tilde{p}_n = \frac{kp_k}{< k >} \tag{78}$$

Thus,

$$\tilde{\mathbb{E}}[\text{number of children}] = \tilde{\mathbb{E}}[k-1] \tag{79}$$

$$= \left(\sum_k k\tilde{p}_k \right) - 1 \tag{80}$$

$$= \sum_k \frac{k^2 p_k}{< k >}) \Bigg) - 1 \tag{81}$$

$$= \frac{< k^2 >}{< k >} - 1 \tag{82}$$

Therefore, the expected number of infected children, λ, is:

$$\lambda \equiv (1-\pi) \left[\frac{< k^2 >}{< k >} - 1 \right] \tag{83}$$

Now emply branching process appoximation:

(i) if $\lambda < 1$, then the disease dies out after a finite number of stages

(ii) if $\lambda > 1$, then with positive probability, the disease persists by infecting a large portion of the population

Therefore

$$(1-\pi) \left[\frac{< k^2 >}{< k >} - 1 \right] > 1 \tag{84}$$

$$\pi < 1 - \frac{1}{\frac{<k^2>}{<k>} - 1} \tag{85}$$

$$= 1 - \frac{1}{\frac{<k^2>-<k>}{<k>}} \tag{86}$$

$$= 1 - \frac{< k >}{< k^2 > - < k >} \tag{87}$$

$$= \frac{< k^2 > -2 < k >}{< k^2 > - < k >} \tag{88}$$

(b) For $\bar{k}$-regular random graph:

$$\frac{< k^2 > -2 < k >}{< k^2 > - < k >} = \frac{\bar{k}^2 - 2\bar{k}}{\bar{k}^2 - \bar{k}} \tag{89}$$

$$= \frac{\bar{k} - 2}{\bar{k} - 1} \tag{90}$$

$$\bar{k} \geq 2 \tag{91}$$

(c) Power law: $p_k \sim k^{-\alpha}$, $\alpha < 3$. $< k^2 >= \infty$, so threshold is 1. Therefore, the internet ($\alpha \sim 2.1 - 2.7$) is robust. If you remove 98% of the nodes, it's still connected.

(d) We want to compute the size of the giant component. Consider a node and the event that this node is in the giant component, or equivalently, that the branching process does not die out. Let $\tilde{q}$ = probability that the branching process does not die out. Starting from neighboring node:

$$1 - \tilde{q} = \pi + (1 - \pi) \sum_{k=1}^{\infty} \tilde{p}_k (1 - \tilde{q})^{k-1} \tag{92}$$

where the first term is the probability that the neighbor is immune, and the second term (summation) is the probability that the neighbor is not immune (however, none of its other neighbors sustains the process).

$$1 - q = \sum_{k=0}^{\infty} p_k (1 - \tilde{q})^d \tag{93}$$

Where the right hand side is the probability that none of the neighbors manages to sustain the branching process.