1)

a. For a given node i, Define an indicator variable Δ_j for whether there is an edge between i, j. The edge exists as long as at least one triangle is created with i, j and another node. Then $P(\Delta_j) = 1 - (1-p)^{n-2}$. $E\sum_{j\neq i}\Delta_j = \sum_{j\neq i} E\Delta_j = \sum_{j\neq i} 1 - (1-p)^{n-2} = (n-1)(1-(1-p)^{n-2})$

   Apply the binomial theorem and substitute $p = \frac{c}{\binom{n-1}{2}} = \frac{2c}{(n-1)(n-2)}$:

   $= (n-1)\left(1 - \sum_{i=0}^{n-2}\binom{n-2}{i}\left(\frac{-2c}{(n-1)(n-2)}\right)^i\right)$

   Take out the first two terms of the summation:

   $= (n-1)\left(\frac{(n-2)2c}{(n-1)(n-2)} - \sum_{i=2}^{n-2}\left(\frac{(n-2)!*(-2c)^i}{i!(n-2-i)!(n-1)^i(n-2)^i}\right)\right)$

   $\lim_{n\to\infty}\frac{n!}{n^n} = 0$ so, for large n, the summation goes to 0:

   $\approx 2c$

b. For large n, it is very unlikely for placed triangles to overlap. If no triangles overlap, then each node must have and even number of edges and $P(k) = 0$ if k is odd. If k is even, we can express it as $k = 2l, l \in Z$ where l is also the number of triangles involving a given node i. This becomes a binomial distribution where $n' = \binom{n-1}{2}$, which is the maximum value for l. For large n, we can approximate this using a poisson distribution with $\lambda = n'p = c$. Then $p_l = \frac{e^{-c}c^l}{l!}$. Finally, $p_k = \frac{e^{-c}c^{\frac{k}{2}}}{\left(\frac{1}{2}k\right)!}$

c. As before, we use the assumption that for large n, triangles have near-0 probability of overlapping. There are $\binom{n}{3} = \frac{n(n-1)(n-2)}{6}$ possible triangles, each occurring with probability $\frac{2c}{(n-1)(n-2)}$. On average, there will be $\binom{n}{3} * \frac{2c}{(n-1)(n-2)} = \frac{nc}{3}$ triangles. For clustering, these are counted 3 times, for each node in the triangle. Then the numerator is nc. For the number of possible triangles, we count nc triangles plus the number of incomplete triangles. For a given node i with two triangles i, j, k and i, l, m, we have four incomplete triangles by taking i along with combinations of $\{j,k\} \times \{l,m\}$. From part (a), we know that each node has on average c triangles. Then we have 4 open triangles for each of the c^2 pairs of triangles. If we add across all nodes, we have $4nc^2$. However this double counts open triangles, so in total there are $2nc^2$ open triangles. Then $C = \frac{nc}{nc+2nc^2} = \frac{1}{1+2c}$.

2) We can approximate the in-degree distribution using the power law with $\alpha = 2 + \frac{a}{c}$. Then $a = 30(3-2) = 30$.

a. 30. This is the average in-degree of all nodes. This is equal to the average out-degree of all nodes, which is 30.

b. Here we can use the closed form solution for the indegree distribution:

   $P(k=0) = \frac{\Gamma(0+a)\Gamma\left(2+\frac{a}{c}\right)}{\Gamma\left(0+a+2+\frac{a}{c}\right)} * \frac{\Gamma\left(a+1+\frac{a}{c}\right)}{\Gamma(a)\Gamma\left(1+\frac{a}{c}\right)} = \frac{\Gamma(3)\Gamma(32)}{\Gamma(33)\Gamma(2)} = \frac{2*8.22e33}{1*2.63e35} = \frac{1}{561}$.

c. Here we can use the CCDF of the power-law approximation:

   $P(k \geq 100) \approx 100^{-\left(1+\frac{a}{c}\right)} = 100^{-2} = \frac{1}{10{,}000}$.

3)

a. Suppose we add a new node, i. $P(i \to j) = \frac{1}{n}$ for all existing nodes j. The expected number of new edges to degree-k nodes is:

$E[i \to j | k_j = q] = np_q(n) * \frac{c}{n} = cp_q(n)$

After adding the new node, the expected number of degree-k nodes is:

$(n+1)p_q(n+1) = np_q(n) + cp_{q-1}(n) - cp_q(n)$

$(n+1)p_0(n+1) = np_0(n) + 1 - cp_0(n)$

$p_q = \lim_{n\to\infty} p_q(n) = \lim_{n\to\infty} p_q(n+1) = cp_{q-1} - cp_q = \frac{c^q}{(1+c)^{q-1}}$

Proof by induction:

$p_0 = 1 - cp_0 \Rightarrow p_0 = \frac{1}{1+c}$

Suppose $p_m = \frac{c^m}{(1+c)^{m-1}}$

$p_{m+1} = c * \frac{c^m}{(1+c)^{m-1}} - cp_{m+1} \Rightarrow (1-c)p_{m+1} = \frac{c^{m+1}}{(1+c)^{m-1}} \Rightarrow p_{m+1} = \frac{c^{m+1}}{(1+c)^m}$

QED.

b. $p_q = \frac{c^q}{(1+c)^{q-1}} = (1+c) * \left(\frac{c}{c+1}\right)^q = C\left(\frac{c}{c+1}\right)^q$

4) See `3_4.ipynb`

a. $m_{11} = m_{22} = 10, m_{12} = m_{21} = 1, \kappa_1 = \kappa_2 = 11.$

$L = \frac{1}{2}\left(2 * 10 \ln\left(\frac{10}{121}\right) + 2 * 1 \ln\left(\frac{1}{121}\right)\right) \approx -29.728$

$e^L = 1.228e - 13$

b. Symmetry is exhibited by 2,3 and by 6,7

i. 1 - $m_{11} = 6, m_{22} = 10, m_{12} = 3, \kappa_1 = 9, \kappa_2 = 13, L \approx -32.94$
ii. 2/3 - $m_{11} = 4, m_{22} = 10, m_{12} = 4, \kappa_1 = 8, \kappa_2 = 14, L \approx -33.75$
iii. 4 - $m_{11} = 6, m_{22} = 12, m_{12} = 2, \kappa_1 = 8, \kappa_2 = 14, L \approx -31.91$
iv. 5 - $m_{11} = 12, m_{22} = 4, m_{12} = 3, \kappa_1 = 15, \kappa_2 = 7, L \approx -33.26$
v. 6/7 - $m_{11} = 10, m_{22} = 6, m_{12} = 3, \kappa_1 = 13, \kappa_2 = 9, L \approx -32.94$
vi. 8 - $m_{11} = 10, m_{22} = 4, m_{12} = 4, \kappa_1 = 14, \kappa_2 = 8, L \approx -33.75$

c. $g_1 = \{1,2,3,4\}$

$g_2 = \{5,6,7,8\}$

$c_1 = 2$

$c_2 = 3$

$c_3 = 3$

$c_4 = 3$

$c_5 = 4$

$c_6 = 2$

$c_7 = 2$

$c_8 = 3$

$\omega = \begin{bmatrix} 0.625 & 0.0625 \\ 0.0625 & 0.625 \end{bmatrix}$

5)

a. $\hat{P} = \begin{bmatrix} 0.57142857 & 0.14285714 \\ 0.14285714 & 0.22222222 \end{bmatrix}$

b. $L = -25.80578231485554$

$\widehat{g_1} = \{3,5,6,7,9\}, \widehat{g_2} = \{1,2,4,8,10\}$

$$\hat{P} = \begin{bmatrix} 0.16 & 0.6 \\ 0.6 & 0.08 \end{bmatrix}$$

c. $\hat{P}_{rs}$ represents the probability of making a group between groups r, s. To promote assortativity, we want $P_{r=s}$ to be large and $P_{r\neq s}$ to be small. Then we can add a constraint that $\min_r P_{rr} > \max_{r\neq s} P_{rs}$. This ensures that groups will always be more likely to have edges with themselves than with other groups.
$L = -27.727369144977594$
$\widehat{g_1} = \{8,9,10\}, \widehat{g_2} = \{1,2,3,4,5,6,7\}$
$$\hat{P} = \begin{bmatrix} 0.22222222 & 0.14285714 \\ 0.14285714 & 0.57142857 \end{bmatrix}$$

6)

a. $P(A_{ik} = A_{jk} = 1) = p^2$, We can model the number of common neighbors as a binomial distribution with $p' = p^2, n' = n - 2$. Then the mean is $n'p' = p^2(n-2)$.

b. $P(A_{ij} = A_{ik} = A_{jk} = 1) = p^3$. For a specific triangle, we can model its occurrence with an indicator variable Δ_i. Thus $E\Delta_i = p^3$. There are $N = \binom{n-1}{2}$ triangles that include a given node i. WLOG, suppose that $\Delta_1 \dots \Delta_N$ are these triangles. Then we want $E\sum_{i=1}^{N}\Delta_i = \sum_{i=1}^{N} E\Delta_i = \sum_{i=1}^{N} p^3 = Np^3 = \binom{n-1}{2}p^3$

c. Define an indicator variable Δ_i for the i^{th} k-clique. $E\Delta_i = p^{\binom{k}{2}}$ as we must have an edge between every pair of nodes in the clique. There are $\binom{n}{k}$ possible cliques so $E\sum_{i=1}^{\binom{n}{k}}\Delta_i = \binom{n}{k}p^{\binom{k}{2}}$.

d. Define an indicator variable Δ_i for the i^{th} k-star. The k-star must have $k-1$ edges from the center to the spokes and $\binom{k-1}{2}$ missing edges between any two spokes. Then $E\Delta_i = p^{k-1}(1-p)^{\binom{k-1}{2}}$. There are $\binom{n}{k}$ possible stars so the expected number of stars is $\binom{n}{k}p^{k-1}(1-p)^{\binom{k-1}{2}}$.