Main Examination period 2023 – May/June – Semester B

# MTH6142: Complex Networks SOLUTIONS

**Duration: 2 hours**

The exam is intended to be completed within **2 hours**. However, you will have a period of **4 hours** to complete the exam and submit your solutions.

**You should attempt ALL questions. Marks available are shown next to the questions.**

All work should be **handwritten** and should **include your student number**. Only one attempt is allowed – **once you have submitted your work, it is final**.

In completing this assessment:

- You may use books and notes.
- You may use calculators and computers, but you must show your working for any calculations you do.
- You may use the Internet as a resource, but not to ask for the solution to an exam question or to copy any solution you find.
- You must not seek or obtain help from anyone else.

When you have finished:

- scan your work, convert it to a **single PDF file**, and submit this file using the tool below the link to the exam;
- e-mail a copy to **maths@qmul.ac.uk** with your student number and the module code in the subject line;

**Examiners: V. Latora, G. Bianconi**

**Question 1 [50 marks].**
Consider the undirected network $G = (V, E)$ with $N = 15$ nodes described by the following set of links:

$$E = \{(1,2),(1,3),(2,4),(2,5),(3,6),(3,7),(4,8),(4,9),$$
$$(5,10),(5,11),(6,12),(6,13),(7,14),(7,15)\}.$$

(a) Calculate the number of links $L$ in the network, and draw the network. Is the network connected? Which type of network is this? (*Explain your answer*). **[7]**

(b) Write down the degree sequence, the degree distribution and the average degree of the network G. **[7]**

(c) Compute the diameter $D$ of the network $G$, and list all the pairs of nodes which are at distance $D$ (*Hint: it is not necessary to compute all the node distances to work out the diameter of the graph*). **[6]**

(d) Evaluate the betweenness centrality for all the nodes in the network $G$. Use the following definition of betweenness centrality of a node $i$:

$$b_i = \sum_{r=1}^{N}\sum_{s=1}^{N}\frac{n^i_{rs}}{n_{rs}}$$

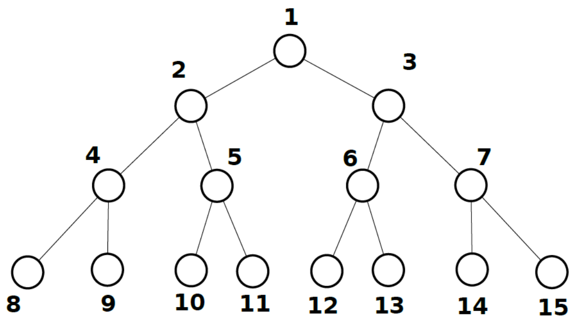
where $n_{rs}$ is the number of shortest paths between node $r$ and node $s$, while $n^i_{rs}$ is the number of shortest paths between node $r$ and node $s$ passing through node $i$. *Hint: (Notice that, according to this definition, paths starting from node i or ending in node j are also counted).* **[11]**

(e) Which are the nodes with the largest possible values of betweenness centrality in the network, and which are those with the largest values of closeness centrality and of degree centrality? **[9]**

(f) Suppose you can add a link to connect two nodes of the network $G$ to obtain another network $G'$. List one link choice that will create a triangle in the network $G'$, one link choice that will create a cycle of length 4, and link choice that will create a cycle of length 5. How many are all the possible link choices that will guarantee that the network $G'$ has a cycle of length $\ell = 7$. **[10]**

**Solutions:**

(a) The network drawing is: [**4**]




The network has $L = 14$ links [**1**]

Being a connected network with $N = 15$ nodes and $L = N - 1$ links it is therefore a tree [**2**]

(b) Degree sequence: $\{2, 3, 3, 3, 3, 3, 3, 1, 1, 1, 1, 1, 1, 1, 1\}$ [**2**]

The graph $G$ is a tree with $N = 15$ nodes. There are 8 nodes that are leaves, i.e. have degree equal to 1 (namely $8; 9; 10; 11; 12; 13; 14; 15$), hence $p_1 = \frac{8}{15}$. There is only one node with degree equal to 2, i.e., node 1, so $p_2 = \frac{1}{15}$. Finally, there are 6 nodes with degree equal to 3 (namely, $2; 3; 4; 5; 6; 7$), hence $p_3 = \frac{6}{15}$. In conclusion we have:

$$p_k = \begin{cases} \frac{8}{15} & \text{for} \quad k = 1, \\ \frac{1}{15} & \text{for} \quad k = 2, \\ \frac{6}{15} & \text{for} \quad k = 3, \\ 0 & \quad\quad\;\; \text{otherwise.} \end{cases}$$

[**3**]

The average degree is

$$\langle k \rangle = \frac{2K}{N} = \frac{28}{15}$$ [**2**]

Alternatively, the average degree can be computed using the node degree distribution $p_k$ as:

$$\langle k \rangle = \sum_{k=0}^{\infty} k \cdot p_k = \frac{8}{15} + 2 \cdot \frac{1}{15} + 3 \cdot \frac{6}{15} = \frac{28}{15}.$$

(c) The diameter of the network is $D = 6$. In fact, the largest distance between two nodes in the graph is equal to $D = 6$, since, for instance, node 8 and node 15 are both at distance 3 from node 1, and given that there is only one path connecting 8 to 15 (remember that in a tree there exists exactly one path among each pair of nodes), that path has to pass by node 1, and its length will be equal to $D = 6$. [**3**]

The set of nodes which are at distance equal to $D = 6$ is the Cartesian product of the two sets $V_1 = (8; 9; 10; 11)$ and $V_2 = (12; 13; 14; 15)$. [**3**]

(d) In the case of a tree, since there is only one path among each pair of nodes, the definition of betweenness $b_i$ of node $i$ reads:

$$b_i = \sum_{r=1}^{N} \sum_{s=1}^{N} n_{rs}^i$$ **[3]**

where $n_{rs}^i = 1$ if the shortest path between node $r$ and node $s$ pass through node $i$, otherwise $n_{rs}^i = 1$.

Now, it might be tempting to think that node 1 is the one with maximal betweenness, but this is not the case, as we will show. In fact, all the $7 \cdot 7 = 49$ undirected paths connecting each of the 7 nodes in the sub-tree rooted at node 2 to each of the 7 nodes in the sub-tree rooted at node 3 pass through node 1. At the same time any shortest path connecting two nodes in the same sub-tree is not passing through node 1. Moreover with this definition of betweenness of node $b_i$ we have also to consider that all nodes starting or ending in $i$ pass by node $i$, which makes a total of other $(2N - 1) = 29$ contributions

Hence, the (unnormalised) betweenness centrality of node 1 is equal to:

$$c_1^B = 2 \times 49 + 29 = 127$$ **[2]**

Node 2 and node 3 will obviously have equal betweenness centralities, due to the symmetry of the graph. In particular, we have:

$$c_2^B = c_3^B = 2 \times [(6 \times 8) + 3 \times 3] + 29 = 2 \times 57 + 29 = 143$$ **[2]**

since e.g. node 2 mediates all the paths between the 6 nodes in the sub-tree rooted at 2 and the 8 remaining nodes, but it also mediates the paths between nodes $(4; 8; 9)$ and nodes $(5; 10; 11)$.

With a similar argument we have that nodes $(4; 5; 6; 7)$ have betweenness centrality equal to:

$$c_4^B = c_5^B = c_6^B = c_7^B = 2 \times [2 \times 12 + 1 \times 1] + 29 = 2 \times 25 + 29 = 79$$ **[2]**

since they mediate the $12 \times 2 = 24$ paths between the nodes in the sub-tree below them and the rest of the graph, plus the path between the two nodes in the sub-tree below them.

Nodes $(8; 9; 10; 11; 12; 13; 14; 15)$ have betweenness centrality equal to $2N - 1 = 29$, as any leaf of a tree. **[2]**

(e) The nodes with the largest betweenness centrality are then nodes 2 and 4. **[2]**

The nodes with the largest closeness centrality is node 1. E.g. $d_1 < d_2$ where $d_1 = 2 \cdot 1 + 4 \cdot 2 + 8 \cdot 3 = 34$ and $d_2 = 3 \cdot 1 + 5 \cdot 2 + 2 \cdot 3 + 4 \cdot 4 = 35$. **[5]**

The nodes with the largest degree centrality are the six nodes $2; 3; 4; 5; 6; 7$. **[2]**

(f) E.g. adding link (2,3) or link (4,5) will create a triangle in the network $G'$. [**2**]

E.g. adding link (2,6) will create a cycle of length 4 [**2**]

E.g. adding link (5,6) will create a cycle of length 5. [**2**]

To create a cycle of length $\ell = 7$ in the network $G'$ we need to add a link between a node in $V_1 = (8; 9; 10; 11)$ and a node in $V_2 = (12; 13; 14; 15)$. We thus have 16 different possibilities. [**4**]

**Question 2 [50 marks].**
Consider the Erdös-Rènyi random graph ensemble $\mathbb{G}_{N,L}$ and assume that the number of links grows as a function of $N$, as $L(N) = \alpha N$, where $\alpha$ is a positive real constant (independent on $N$).

(a) Write down the general expression, as a function of $N$ and $\alpha$, for the probability of finding a given graph $G = (V, E)$ in the ensemble $\mathbb{G}_{N,L(N)}$. [**8**]

(b) Fix now $N = 5$ and consider the three cases $\alpha_1 = 0.8$, $\alpha_2 = 1$, and $\alpha_3 = 1.2$. Evaluate, in each of the three cases, the probability of finding a given graph $G = (V, E)$ with 5 nodes in the ensemble $\mathbb{G}_{N,L(N)}$. [**6**]

(c) Find a general expression for the average node degree $\langle k \rangle$ in a graph of the ensemble $\mathbb{G}_{N,L(N)}$, and evaluate the average node degree $\langle k \rangle_1$, $\langle k \rangle_2$ and $\langle k \rangle_3$ for the graphs in each of the three ensembles considered in point (b) (i.e. by fixing $N = 5$ and $\alpha_1 = 0.8$, $\alpha_2 = 1$, $\alpha_3 = 1.2$ respectively). [**7**]

(d) Find the function $p(N)$ of the Erdös-Rènyi random graph ensemble $\mathbb{G}_{N,p}$ which best approximate the Erdös-Rènyi random graph ensemble $\mathbb{G}_{N,L(N)}$. [**5**]

(e) For the Erdös-Rènyi random graph ensemble $\mathbb{G}_{N,p(N)}$ found in point (d), write down the distribution $P(L)$ for the total number of links $L$ in a graph. Calculate the variance of such distribution and find an expression for the ratio $r$ between the standard deviation and the expectation value of $L$. [**8**]

(f) Consider the Erdös-Rènyi random graph ensemble $\mathbb{G}_{N,p(N)}$ found in point (d), assume $N \to \infty$, and find the condition on $\alpha$ for the graphs to have a giant component. Find the same result using the Molloy-Reed criterion. [**8**]

(g) Consider the Erdös-Rènyi random graph ensemble $\mathbb{G}_{N,p(N)}$ found in point (d), assume $N \to \infty$, and find the expected numbers of triangles and 4-cliques as a function of $\alpha$. [**8**]

**Solutions:**

(a) By definition of Erdös-Rènyi random graph ensemble $\mathbb{G}_{N,L}$, all the graphs with $N$ nodes and $L$ links have the same probability to appear. Hence, the probability that a given graph $G = (V, E)$ is sampled from $\mathbb{G}_{N,L}$ is:

$$P(G) = \begin{cases} \frac{1}{Z} & \text{iff } |V| = N \text{ and } |E| = L \\ 0 & \text{otherwise} \end{cases}$$

where $Z$ is equal to the total number of graphs with $N$ nodes and $L$ links:

$$Z = \binom{M}{L} = \frac{M!}{L!(M-L)!} \qquad \text{where} \quad M = \binom{N}{2} = \frac{N(N-1)}{2}$$

which is the number of different ways in which we can choose $L$ pairs to connect among the possible $M$ pairs.

Hance, the probability of finding a given graph $G = (V, E)$ in the ensemble $\mathbb{G}_{N,L(N)}$ is

$$P(G) = \begin{cases} \frac{1}{Z} & \text{iff } |V| = N \text{ and } |E| = \alpha N \\ 0 & \text{otherwise} \end{cases} \qquad [\mathbf{4}]$$

where

$$Z = \binom{M}{L(N)} = \frac{M!}{(\alpha N)!(M-\alpha N)!} \qquad \text{where} \quad M = \binom{N}{2} = \frac{N(N-1)}{2} \qquad [\mathbf{4}]$$

(Notice: we therefore need to assume that $\alpha$ is such that $\alpha N$ is a natural number).

(b) We have: $M = \binom{5}{2} = \frac{5 \cdot 4}{2} = 10$ pairs of nodes that we can connect.

The number of links in the three cases is:

$L_1 = \alpha_1 N = 0.8 \cdot 5 = 4$,

$L_2 = \alpha_2 N = 5$

$L_3 = \alpha_3 N = 6$

Hence, the probability of finding a graph $G = (V, E)$ of 5 nodes in $\mathbb{G}_{5,4}$ is:

$$P(G) = \binom{10}{4}^{-1} = \frac{4!\ 6!}{10!} = \frac{1}{210} = 0.0048$$

if the graph $G$ has $|E| = 4$ links, while $P(G) = 0$ otherwise **[2]**

The probability of finding a graph $G = (V, E)$ of 5 nodes in $\mathbb{G}_{5,5}$ is:

$$P(G) = \binom{10}{5}^{-1} = \frac{5!\ 5!}{10!} = \frac{1}{252} = 0.0040$$

if the graph $G$ has $|E| = 5$ links, while $P(G) = 0$ otherwise **[2]**

The probability of finding a graph $G = (V, E)$ of 5 nodes in $\mathbb{G}_{5,6}$ is:

$$P(G) = \binom{10}{6}^{-1} = \frac{6!\ 4!}{10!} = \frac{1}{210} = 0.0048$$

if the graph $G$ has $|E| = 4$ links, while $P(G) = 0$ otherwise **[2]**

(c) The average node degree is:

$$\langle k \rangle = \frac{2K(N)}{N} = \frac{2\alpha N}{N} = 2\alpha \qquad [\mathbf{4}]$$

We therefore have: $\langle k \rangle_1 = 1.6$, $\langle k \rangle_2 = 2$, $\langle k \rangle_3 = 2.4$. [**3**]

(d) There are two ways to do this. Either we acknowledge that the probability of connecting two nodes in $\mathbb{G}_{N,L}$ is $L/M$. Hence:

$$p(N) = \frac{\alpha N}{N(N-1)/2} = \frac{2\alpha}{N-1} \qquad [\mathbf{5}]$$

Or we acknowledge that, on average, the graphs in $\mathbb{G}_{N,p}$ will have $pM$ links and we set $pM = L$, which leads to same result as above.

(e) In $\mathbb{G}_{N,p(N)}$, the total number of links $L$ can vary and is binomially distributed. Since $p(N) = \frac{2\alpha}{N-1}$, we have:

$$P(L) = P(\#\text{ links} = L) = \binom{M}{L} p^L (1-p)^{M-L} = \binom{N(N-1)/2}{L} \left(\frac{2\alpha}{N-1}\right)^L \left(1 - \frac{2\alpha}{N-1}\right)^{M-L}$$

[**4**]

$$\bar{L} = Mp = \frac{N(N-1)}{2}\frac{2\alpha}{N-1} = \alpha N$$

$$\sigma^2(L) = Mp(1-p) = \frac{N(N-1)}{2}\frac{2\alpha}{N-1}\left(1 - \frac{2\alpha}{N-1}\right) = \alpha N \left(1 - \frac{2\alpha}{N-1}\right) \qquad [\mathbf{2}]$$

$$r = \frac{\sqrt{\sigma^2(L)}}{\bar{L}} = \frac{\sqrt{\sigma^2(L)}}{\alpha N} = \frac{\sqrt{1 - \frac{2\alpha}{N-1}}}{\sqrt{\alpha N}} = \sqrt{\frac{1}{\alpha N} - \frac{2}{N(N-1)}} \qquad [\mathbf{2}]$$

(f) The graphs of $\mathbb{G}_{N,p(N)}$ have a giant component if $\langle k \rangle = p(N-1)$ is such that:

$$\lim_{N\to\infty} \langle k \rangle > 1$$

We therefore have:

$$\lim_{N\to\infty} \langle k \rangle = \lim_{N\to\infty} \frac{2\alpha}{N-1}(N-1) = 2\alpha > 1$$

if $\alpha > 0.5$. [**4**]

The Molloy and Reed criterion says that almost every graph in the ensemble will have a giant component iff the degree distribution satisfies

$$\langle k^2 \rangle - 2\langle k \rangle > 0$$

Using the Poisson approximation for the degree distribution we have:

$$\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle = 4\alpha^2 + 2\alpha$$

Hence:

$$\langle k^2 \rangle - 2\langle k \rangle = 4\alpha^2 - 2\alpha = 2\alpha(2\alpha - 1)$$

which is positive for $\alpha > 1/2$. [**4**]

(g) For large $N$, $p$ can be approximated as

$$p = \frac{2\alpha}{N^z} \text{ with the exponent } z = 1$$ [**2**]

Therefore, the expected number of triangles is:

$$\mathcal{N}^{triangles} = \frac{(2\alpha)^3}{3!} = \frac{4}{3}\alpha^3$$ [**3**]

while the expected number of 4-cliques $\mathcal{N}^{4-cliques}$ is 0, because $z = 1 > 2/3$. [**3**]

---

**End of Paper.**