MATH6142 Complex Networks
May Exam 2020

- *Solution Problem 1*

(a) A weighted undirected network is a network in which all the links are undirected and weighted. This means that each link is associated to a real or integer number (the weight of the link) indicating the intensity of the interaction. Moreover each link describes a symmetric interaction therefore if i is linked to j also j is linked to i with the same weight. A weighted network can be characterized by and edge list of links (i, j, w_{ij}) where we can link exclusively the links with $i < j$ or equivalently we list all ordered pairs of nodes (i, j) connected by a link with weights $w_{ij} = w_{ji}$. **(2.5 marks)**[**Bookwork**]
An example of a weighted network is an airport network where each node is an airport, each link between two airports indicate a flight connection and the weight indicate the number of return flights among the two connected airports flying each week. **(2.5 marks)**[**Bookwork**]

(b) A *bipartite network* $G_B = (V, U, E)$ is a network formed by two non overlapping sets of nodes U and V and by a set of links E, such that every link joins a node in V with a node in U. **(2.5 marks)**[**Bookwork**]
An example of bipartite network is a collaboration network in which the two sets of non-overlapping nodes are the set of scientists and the set of published scientific articles, the links connect each scientist to the articles that he/she authored. **(2.5 marks)**[**Bookwork**]

(c) The degree centrality of a node is given by its degree, i.e. the number of links incident to it. An example in which the degree centrality can give a relevant proxy of the importance of a node is the degree of a node in online social media such as Facebook where the friendship is reciprocal. On Twitter, where the links are directed a good indication of the centrality of a node is the in-degree centrality indicating the number of followers of a Twitter account. **(2.5 marks)**[**Unseen**]
The betweenness centrality ranks nodes according to their betweenness, and nodes with higher betweenness are the nodes that are traversed by many shortest paths when we consider all shortest paths between each pair of nodes in the network. Therefore nodes with high betweenness are important in the prediction of congestions in a road network, because crossing (nodes) with high betweenness will

be more likely to be congested.
(2.5 marks)[**Unseen**]

- *Solution Problem 2*

a) The network is directed because the adjacency matrix is not symmetric. **(2 marks)**[**Unseen**]
The network is shown in Figure 1. **(3 marks)**[**Unseen**]

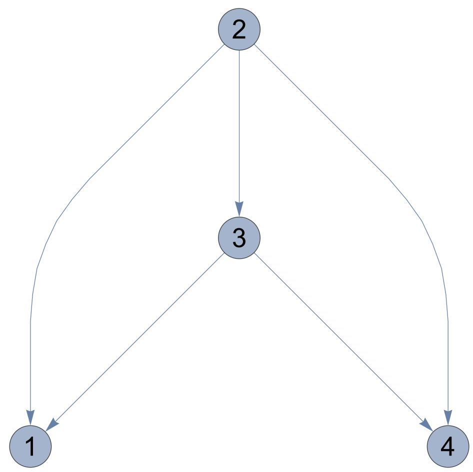

Figure 1: The directed network with adjacency matrix $\mathbf{A}$.

b) The in-degree sequence is $\{2, 0, 1, 2\}$. **(2.5 marks)**[**Unseen**]
The out-degree sequence is $\{0, 3, 2, 0\}$. **(2.5 marks)**[**Unseen**]

c) The in-degree distribution is given by $P^{in}(0) = 1/4, P^{in}(1) = 1/4, P^{in}(2) = 1/2, P^{in}(3) = 0$. **(2 marks)** [**Unseen**]
The out-degree distribution is given by $P^{out}(0) = 1/2, P^{out}(1) = 0, P^{out}(2) = 1/4, P^{out}(3) = 1/4$ for $k = 1, 2$. **(2 marks)** [**Unseen**]

d) The eigenvector centrality $\mathbf{x}$ can be found as following.
Given the initial guess $\mathbf{x}^{(0)} = \frac{1}{N}\mathbf{1}$ given by

$$\mathbf{x}^{(0)} = \frac{1}{4}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}. \tag{1}$$

Let us calculate the result of the iteration

$$\mathbf{x}^n = \mathbf{A}\mathbf{x}^{n-1}. \tag{2}$$

(1 mark)[**Bookwork**]

We have

$$\mathbf{x}^{(1)} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \frac{1}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 2 \\ 0 \\ 1 \\ 2 \end{pmatrix},$$

(1 mark)[**Unseen**]

$$\mathbf{x}^{(2)} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \frac{1}{4} \begin{pmatrix} 2 \\ 0 \\ 1 \\ 2 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

(1 mark)[**Unseen**]

$$\mathbf{x}^{(3)} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \frac{1}{4} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

(1 mark)[**Unseen**]

Therefore

$$\mathbf{x}^{(n)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{3}$$

for every $n \geq 3$. **(2 marks)**[**Unseen**]

It follows that

$$\mathbf{x} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{4}$$

(2 marks)[**Unseen**]

e) The Katz centrality $\mathbf{x}$ can be expressed in matrix form as

$$\mathbf{x} = \beta \sum_{n=0}^{\infty} \alpha^n \mathbf{A}^n \mathbf{1} \tag{5}$$

where $\mathbf{1}$ is the N-dimensional column vector of elements $\mathbf{1}_i = 1$, forall $i \in \{1, 2, \ldots, N\}$. For our matrix we have

$$\begin{aligned} \mathbf{A}^0 &= \mathbf{I}, \\ \mathbf{A}^1 &= \mathbf{A}. \end{aligned} \tag{6}$$

(2 marks)[**Unseen**]

Moreover we have

$$\mathbf{A^2} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

(2 marks)[**Unseen**]

and

$$\mathbf{A}^3 = \mathbf{A^2 A} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

It follows that

$$\mathbf{A}^n = \mathbf{A}^3 = \mathbf{0} \tag{7}$$

for every $n \geq 3$. **(2 marks)**[**Unseen**]

The Katz centrality can then be evaluated as

$$\mathbf{x} = \beta \left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \alpha \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} + \alpha^2 \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \right] \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

which gives

$$\mathbf{x} = \begin{pmatrix} 1 + 2\alpha + \alpha^2 \\ 1 \\ 1 + \alpha \\ 1 + 2\alpha + \alpha^2 \end{pmatrix} \tag{8}$$

(2 marks)[**Unseen**]

- *Solution Problem 2*

a) The probability that a node i is connected to a node j in the $\mathbb{G}(N,p)$ ensemble is given by p.
The probability that a node i is not connected to a node j in the $\mathbb{G}(N,p)$ random graph ensemble is given by

$$1 - p.$$

(1 mark)- [Bookwork].

Assuming independence of the two events (presence of the link, and node connected to it not belonging to the giant component), the probability that a node i is connected to a node that is not in the giant component is given by

$$p(1-S).$$

(1 mark)-[Bookwork].

Given the fact that each link of a random graph is drawn independently, the probability that a node is not in the giant component $1-S$ is given by

$$1-S = \prod_{j=1|j\neq i}^{N} [1-p+p(1-S)] = [1-pS]^{N-1} \tag{9}$$

(2 marks)-[Bookwork].

By inserting in Eq. (9)

$$p = c/(N-1)$$

where c indicates the average degree, assumed to be independent of the network size N, we get

$$\begin{aligned} S &= 1-(1-pS)^{N-1} \\ &= 1-\left(1-\frac{c}{N-1}S\right)^{N-1} \\ &\simeq 1-e^{-cS} \end{aligned} \tag{10}$$

where the last expression is valid in the limit $N \gg 1$ and where we have used the limit

$$\lim_{N\to\infty}\left(1+\frac{a}{N}\right)^{N} = e^{a}. \tag{11}$$

(3 marks)-[Bookwork].

b) The functions $y(S) = S$ and $\tilde{y}(S) = 1-e^{-cS}$ are both monotonically increasing.
These functions cross at $S = 0$, therefore $S = 0$ is always a solution of the Eq. (9). **(2 marks)-[Bookwork]**
We observe that $\tilde{y}(S)$ is non-decreasing, has maximum slope at $S = 0$ and satisfies $\tilde{y}(S) < 1$. In fact

$$\frac{d\tilde{y}(S)}{dS} = ce^{-cS} \tag{12}$$

and

$$0 < \frac{d\tilde{y}(S)}{dS} = cs^{-cS} \leq \left.\frac{d\tilde{y}(S)}{dS}\right|_{S=0} = c. \tag{13}$$

Moreover

$$\tilde{y}(S) = 1 - e^{-cS} < 1. \tag{14}$$

(3 marks)-[Bookwork]

Therefore a non-trivial solution $S > 0$ of the equation $S = 1 - e^{-cS}$ emerges when the functions $y(S)$ and $\tilde{y}(S)$ are tangent to each other at $S = 0$. **(2 marks)-[Bookwork]**
This condition implies

$$\begin{aligned} \left.\frac{d\tilde{y}(S)}{dS}\right|_{S=0} &= 1 \\ c &= 1. \end{aligned} \tag{15}$$

Therefore the critical average degree for having a giant component is $c = 1$. **(3 marks)-[Bookwork]**

c) Let us impose that all nodes, except one, belong to the giant component, i.e.

$$S = 1 - \frac{1}{N}. \tag{16}$$

(2 marks)-[Bookwork].
In this case the average degree should be $c = \langle k \rangle \simeq \ln N$ for $N \gg 1$. In fact by inserting Eq. (16) into Eq. (9) we have

$$1 - \frac{1}{N} = 1 - e^{-cS} = 1 - e^{-c(1-1/N)} \simeq 1 - e^{-c} \tag{17}$$

where in the last expression we have put

$$1 - 1/N \simeq 1,$$

for $N \gg 1$. **(4 marks)-[Bookwork]**
Therefore we get

$$\frac{1}{N} = e^{-c} \tag{18}$$

or equivalently

$$c = \ln(N). \tag{19}$$

(2 marks)-[Bookwork].

- *Solution Problem 4*
  We consider a scale-free network with degree distribution

$$P(k) = Ck^{-\gamma} \tag{20}$$

valid for $m \leq k \leq K$ and $\gamma > 1$.

(a) In the continuous approximation the normalization condition reads

$$\int_m^K Ck^{-\gamma} = 1. \tag{21}$$

(2 marks)-[Bookwork].

By performing the integral we get

$$C\frac{1}{\gamma-1}\left(m^{1-\gamma}-K^{1-\gamma}\right)=1 \tag{22}$$

Therefore we obtain the expression for the normalization constant C given by

$$C=(\gamma-1)\left[m^{1-\gamma}-K^{1-\gamma}\right]^{-1}. \tag{23}$$

(3 marks)-[Bookwork].

(b) The first moment $\langle k\rangle$ is given by

$$\langle k\rangle=\int_m^K Ck^{1-\gamma}. \tag{24}$$

(1 mark)-[Bookwork].

By performing the integral we get

$$\langle k\rangle=\begin{cases}\frac{C}{2-\gamma}\left(K^{2-\gamma}-m^{2-\gamma}\right) & \text{for} \quad \gamma\neq 2\\ C\ln\left(\frac{K}{m}\right) & \text{for} \quad \gamma=2\end{cases} \tag{25}$$

(1.5 marks)-[Bookwork].

The second moment $\langle k^2\rangle$ is given by

$$\langle k^2\rangle=\int_m^K Ck^{3-\gamma}. \tag{26}$$

(1 mark)-[Bookwork].

By performing the integral we get

$$\langle k^2\rangle=\begin{cases}\frac{C}{3-\gamma}\left(K^{3-\gamma}-m^{3-\gamma}\right) & \text{for} \quad \gamma\neq 3\\ C\ln\left(\frac{K}{m}\right) & \text{for} \quad \gamma=3\end{cases} \tag{27}$$

(1.5 marks)-[Bookwork].

(c) Power-law networks have power-law exponents $\gamma\in(2,3]$ and they are characterising by having a finite average degree $\langle k\rangle$ and a diverging second moment $\langle k^2\rangle$ of the degree distribution as $N\to\infty$ and as the cutoff $K\to\infty$. Given this property, scale-free networks are a very good approximation of the vast variety of real networks characterized by large fluctuations in the degree of nodes, i.e. networks in which it is not rare to find nodes with a number of links orders of magnitude bigger than the average. On the contrary Poisson networks cannot account for these large fluctuations because they are characterized by having a standard deviation of the degree distribution $\sigma = \sqrt{\langle k \rangle}$, so they cannot allow for large fluctuations in nodes degrees. **(5 marks)-[Unseen]**.

- *Solution Problem 5*

  (a) *Complex networks are neither completely ordered (like lattices) nor completely disordered (like random graphs). Using your own words: (a) Discuss the main properties of real-world complex networks such as social networks or the Internet,* SOL: most of real-world complex networks do not have homogeneous or Poisson degree distributions (like lattices or random graphs) but instead have a power law degree distribution (i.e. they are scale-free networks) with hubs. Also, real-world networks are usually small-world: at the same time they have high clustering and fulfil the small-world distance property (SWDP). **(5 marks)-[Unseen]**.

  (b) *Discuss how these resemble or differ from classic synthetic models such as lattices or random graphs, and why the differences are so important in, for instance, epidemic spreading.* SOL: lattices tend to have high clustering but dont display the SWDP, and have a homogeneous degree distribution. Random graphs have a Poisson degree distribution, display the SWDP but have vanishing clustering. So any of them have power law degree distribution, and thus are unable to explain the presence of hubs (nodes with very large degree) that play an important role in epidemic spreading (when a disease reaches a hub, it spreads massively, and if the network does not have such super spreaders the spreading is usually slower). **(5 marks)-[Unseen]**.

  (c) *Discuss also what other synthetic models beyond lattices or random graph models have been proposed recently to be able to recover the properties we observe in real-world networks.* SOL: The Barabasi-Albert model was introduced to recover power law degree distributions, a property we observe in real-world networks [HERE THE STUDENT MIGHT EXPLAIN IN GENERAL TERMS THE MECHANISMS OF THE MODEL, BUT WITHOUT GOING INTO MUCH DETAIL]. The Watts-Strogatz model was introduced to recover high clustering and SWDP, two properties which show up together in real-world networks [HERE THE STUDENT MIGHT EXPLAIN IN GENERAL TERMS THE MECHANISMS OF THE MODEL, BUT WITHOUT GOING INTO DETAILS] **(5 marks)-[Unseen]**.