Degrees & intro to NetworkX

Introduction to complex networks theory

ADA 2022/2023

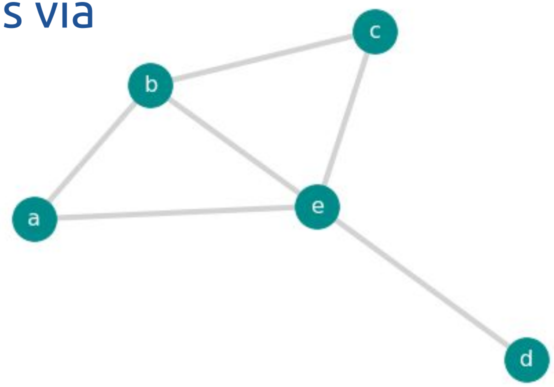
Marija Mitrović Dankulov mitrovic@ipb.ac.rs
Aleksandra Alorić aleksandra.aloric@gmail.com

Aims of this lecture

- Network representation
- Introduce basic node centrality - its degree
- Introduce degree distribution
- Math refresher: matrices, probabilities, distributions
- Introduce networkx and some numpy

Networks

- Reminder: nodes or vertices, connected via links or edges
- Mathematically: G = (V, E)
  - Where V is set of all vertices and E is set of edges between them
  - For a graph in this slide V = {a, b, c, d} and E = {(a,b), (a,e), (b,c), (b,e), (c,e), (e,d)}
- This means that one logical way to store a network is via lists of edges
- Alternatively, we can use adjacency matrix



Math refresher: Matrices

$$\begin{array}{rcl} & \text{3 columns} & \\ & \downarrow \quad \downarrow \quad \downarrow & \\ A = & \begin{bmatrix} -2 & 5 & 6 \\ 5 & 2 & 7 \end{bmatrix} & \begin{array}{l} \leftarrow \\ \leftarrow \end{array} \text{2 rows} \end{array}$$

- Matrix dimensions: rows×columns
- Notation: row i, column j of matrix A are denoted A_{ij}
- Useful for representing data (think about image pixel values), solving systems of equations, we'll see application in network science

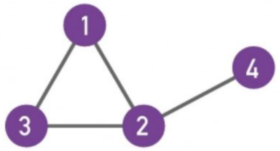
If needed: [Intro to matrices]()

Adjacency matrix

- Network with N nodes is represented with an NxN adjacency matrix, keeping information about all connections

All networks with 4 nodes:

$$A_{ij} = \begin{matrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{matrix}$$



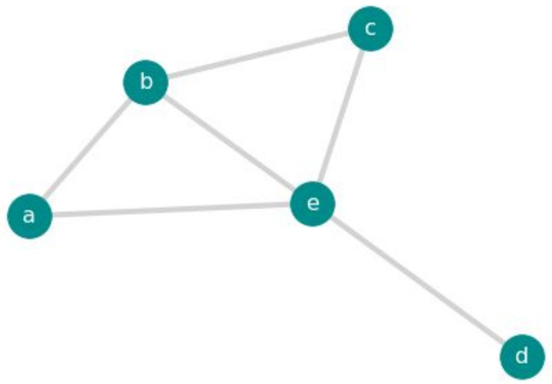
$$A_{ij} = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

Simple, undirected networks:

$$\boxed{A_{ij} = A_{ji} \qquad A_{ii} = 0}$$

Adjacency matrix

- Network with N nodes is represented with an NxN adjacency matrix, keeping information about all connections



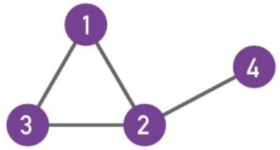
	a	b	c	d	e
a	0	1	0	0	1
b		0	1	0	1
c			0	0	1
d				0	1
e			1	1	0

Adjacency matrix

- For undirected networks, total number of links: $L = \frac{1}{2} \sum_{i,j=1}^{4} A_{ij}$

- Number of neighbours: $k_i = \sum_{j=1}^{4} A_{ij}$

- Test it yourself:



$$A_{ij} = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

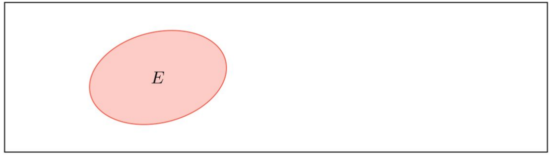
numpy matrices

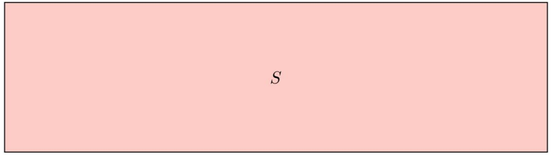
Node degree

- Node's total number of neighbours
- Degree centrality - (or just degree) is one simple measure of node importance in the network
- Nodes with high degree can play important role in spreading processes (they have large audience)
- Often in network visualisations nodes with high degree are marked with larger marker

Math refresher: Probabilities

- Probability properties:
  - probability is between 0 and 1 included
  - probability that at least one of the elementary events in the entire sample space will occur is 1
  - probability of a union of mutually exclusive elements is a sum of elements' probabilities
  - see figures
- Drawing two heart cards
  - What's the total number of 2 card combinations: **Sample space**
  - What's the total number of 2 card combinations when both are hearts: **Event**





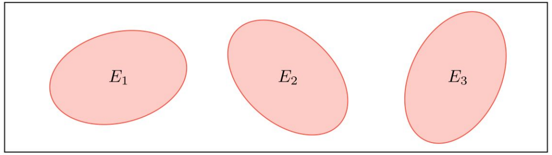
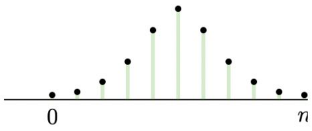
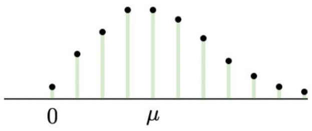
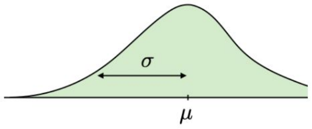
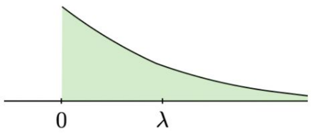


Image source: https://stanford.edu/~shervine/teaching/cs-229/refresher-probabilities-statistics

Math refresher: Random variable

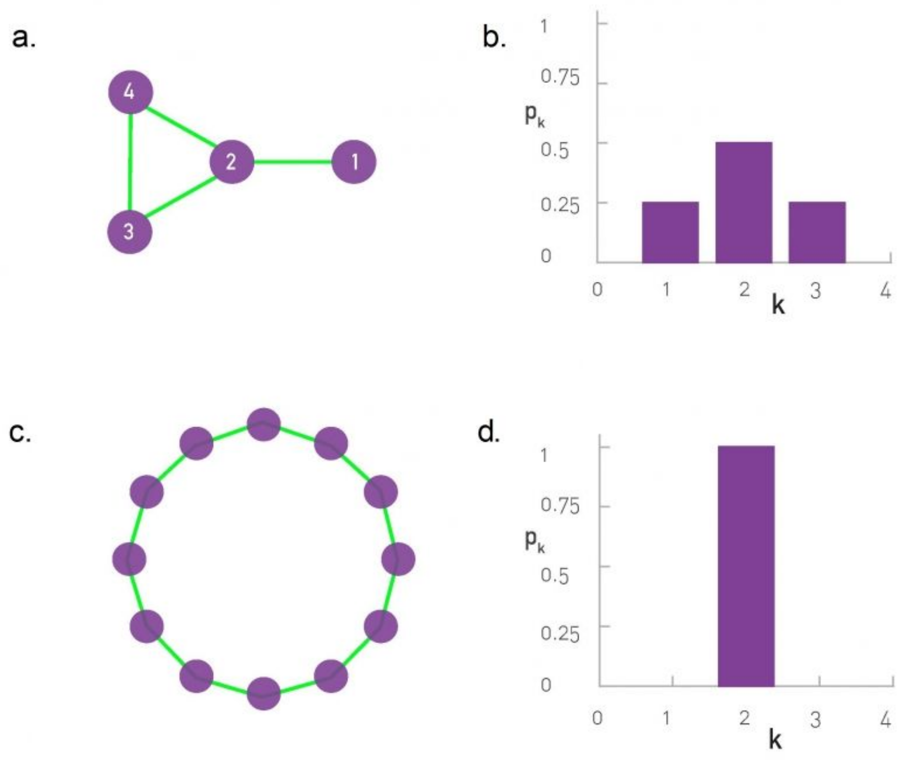
- Formal way to think about experiment (measurement) outcomes
  - Think about outcomes of tossing a coin, drawing a card from a deck, a student's results on the test…
  - Discreet (pass/fail, or number of points on the test) vs continuous (time it takes for student to do the test)
- Probability mass and density functions: maps between possible realisations of random variable and their probability
  - Useful when thinking about a model behind your data
- Data distribution - distinct values of the variable you measured and their occurrence count, or frequency of occurrence
  - Mean
  - Median
  - Mode
  - Variance, standard deviation

Some distributions that can be handy

Type	Distribution	PDF	$\psi(\omega)$	$E[X]$	$\mathrm{Var}(X)$	Illustration
(D)	$X \sim \mathcal{B}(n,p)$	$\binom{n}{x} p^x q^{n-x}$	$(pe^{i\omega} + q)^n$	np	npq	
(D)	$X \sim \mathrm{Po}(\mu)$	$\frac{\mu^x}{x!} e^{-\mu}$	$e^{\mu(e^{i\omega}-1)}$	μ	μ	
(C)	$X \sim \mathcal{U}(a,b)$	$\frac{1}{b-a}$	$\frac{e^{i\omega b} - e^{i\omega a}}{(b-a)i\omega}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
(C)	$X \sim \mathcal{N}(\mu,\sigma)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$e^{i\omega\mu - \frac{1}{2}\omega^2\sigma^2}$	μ	σ^2	
(C)	$X \sim \mathrm{Exp}(\lambda)$	$\lambda e^{-\lambda x}$	$\frac{1}{1-\frac{i\omega}{\lambda}}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	

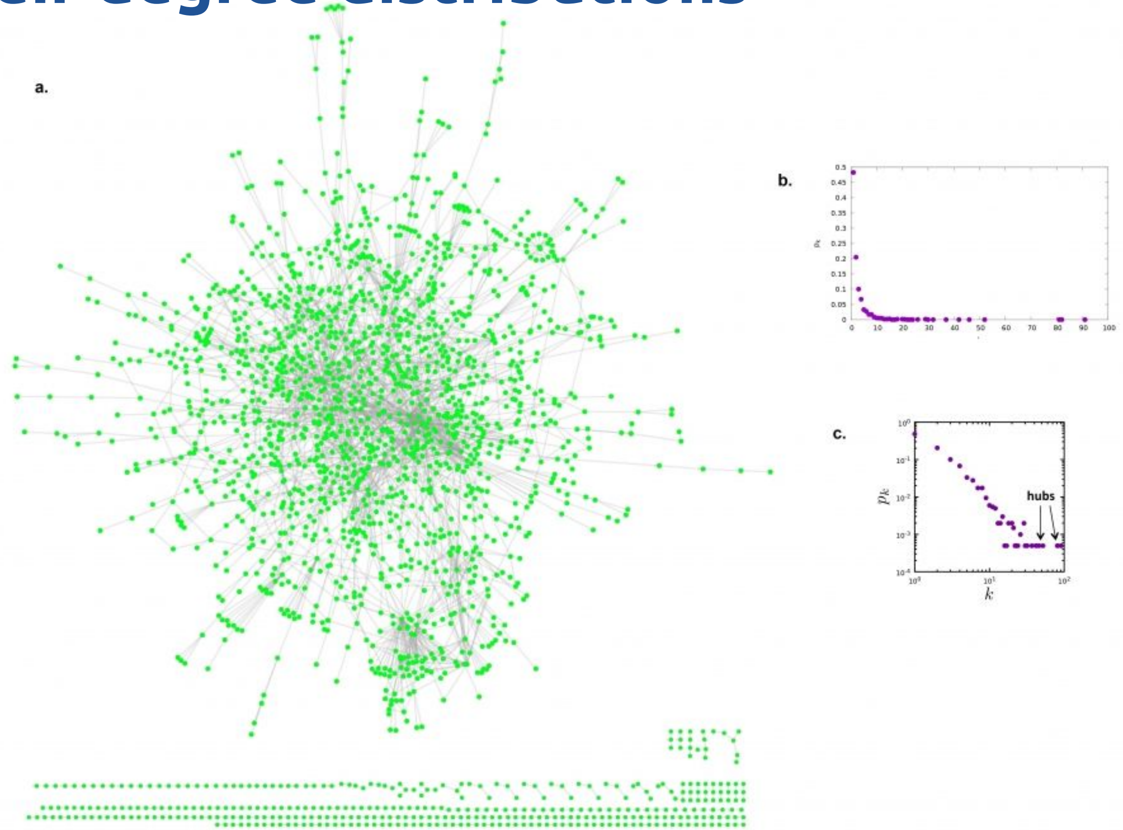
Degree distributions

- If we collect data about the degree of every node in a network, we can study the degree distribution
- On the right, simple degree distributions are shown
- It turns out that in reality degree distributions are more interesting, and often share some similarities, e.g.



Real networks and their degree distributions

- Although we study normal (Gaussian) distribution most of the time, distributions in reality rarely look Gaussian
- Here, a more typical distribution in an example degree distribution of yeast protein network
- Why is that important?
    - We keep thinking that mean/median/mode are all the same (which is true for Gaussian)
    - But in distributions like this one, all these three values are different, and mean doesn't tell us about some "typical" or frequent degree



Protein interaction network of yeast
Image source: http://networksciencebook.com/chapter/2#degree

Hubs

- Nodes with very large degree (neighbourhood)

- Right: retweet network of a highly viral fabricated news report, nodes (twitter accounts with most retweets) highlighted by size and layout

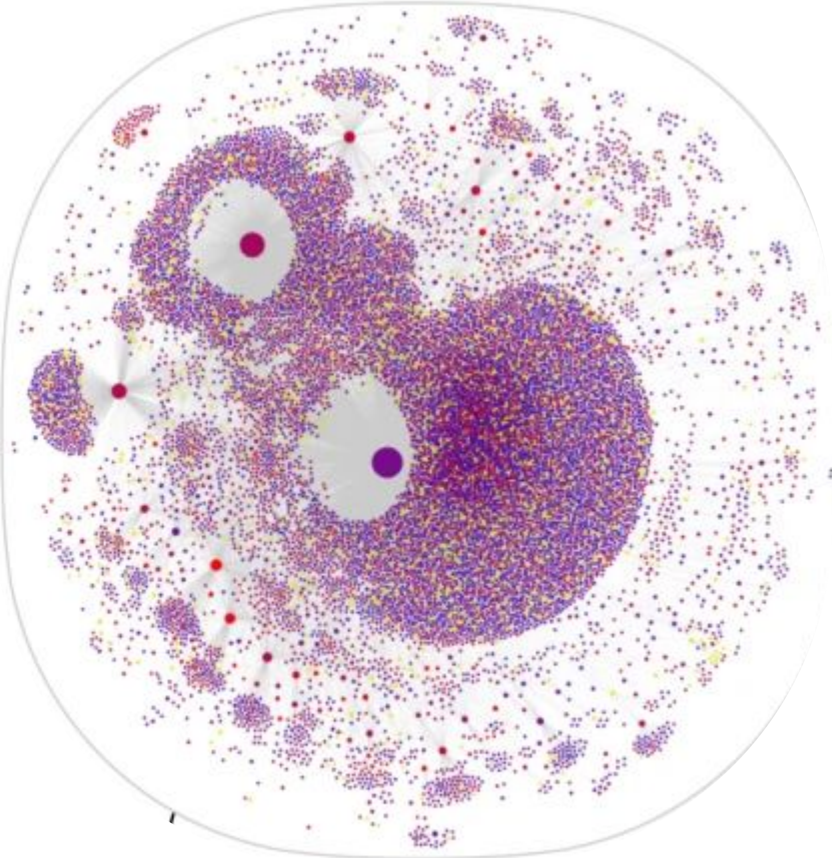


Image source:
https://www.nature.com/articles/s41467-018-06930-7

Further reading

- Network science book: sections 2.1 to 2.4
- Probability & statistics refresher
- 3blue1brown probability videos (not strictly related to what we talked about but very useful for thinking about data and probability in general)
- Advanced, but very interesting scientific discussion about degree distributions in real networks:
  - Scale-free networks are rare https://arxiv.org/abs/1801.03400
  - Response: Love is all you need

Maths questions

- What's maximal degree in a network with N nodes?
- If the number of links in a network is L, what is the average degree?
- If the network with N nodes has L links, what is the probability that two randomly selected nodes are connected?

Homework

Submission link