Universitat Pompeu Fabra Barcelona

Lecture 20 · Course Summary and Outlook

Networks, Crowds and Markets

Today's Lecture

1. Big picture: what problems we studied.
2. Static structure of networks.
3. Dynamics on networks (walks, PageRank, spreading).
4. Random graph models and communities.
5. Social networks, markets, and matching.
6. What to remember and what to read next.

Networks: the central theme

A **network** is a graph plus interpretation.

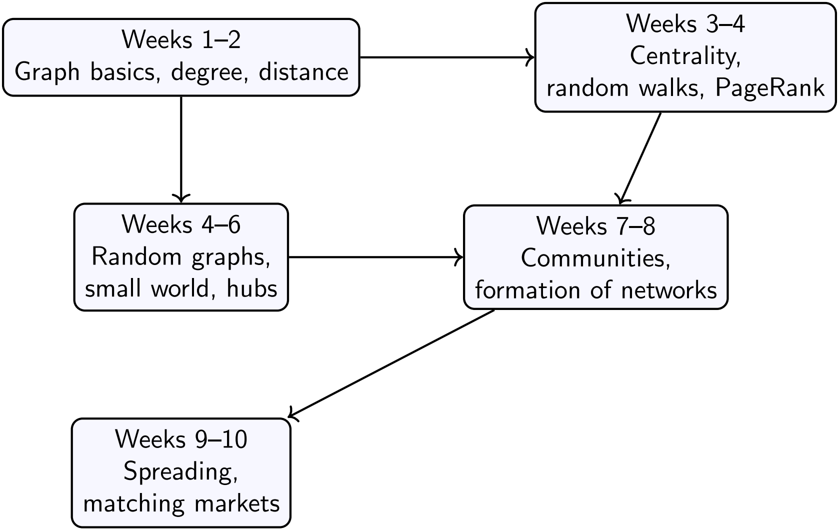
- Nodes: people, web pages, firms, cities, molecules.
- Edges: friendships, links, contracts, roads, interactions.

Two recurring questions

- **Structure**: what does the network look like?
- **Dynamics**: what processes live on top of it?

Throughout the course we moved back and forth between these two viewpoints.

A rough road map of the course



Basic objects (Weeks 1–2)

A (simple) graph

- Set of nodes V and edges $E \subseteq \{\{u, v\} : u, v \in V\}$.
- Degree of v: $d(v) = \#\{u : \{u, v\} \in E\}$.

Key examples

- Complete graph, path, cycle, star.
- Bipartite graphs (e.g. buyers–sellers, students–courses).

Real networks often mix these patterns (stars, clusters, bipartite parts).

Degree distribution became a critical concept for this course.

Distance, diameter, adjacency matrix

- **Shortest-path distance** $d(u, v)$: fewest edges in a path from u to v.
- **Diameter**: $\max_{u,v} d(u, v)$.
- **Average path length**: typical distance, more stable than diameter.

Adjacency matrix

$$A_{uv} = \begin{cases} 1 & \text{if there is an edge } u \sim v, \\ 0 & \text{otherwise.} \end{cases}$$

- Powers A^k count walks of length k.
- Matrix view is essential later (spectral methods, PageRank).

Centrality: who is "important"? (Weeks 3–4)

Different notions capture different roles:

- **Degree centrality**: many neighbours.
- **Closeness**: small average distance to others.
- **Betweenness**: lies on many shortest paths (broker).
- **Eigenvector centrality**: connected to important nodes.

No single centrality is "correct" in all contexts:

- choosing one is part of the modelling decision;
- we saw examples where different measures disagree.

Random walks and PageRank

Random walk on a graph

- Start from some node.
- At each step, move to a random neighbour.

Stationary distribution

- Long-run fraction of time spent at each node.
- On an undirected, connected graph: $\pi(v) \propto \deg(v)$.

PageRank idea

- Walk with occasional “teleportation” back to a random node.
- Removes issues with sinks; gives a global importance score.
- HITS: separates hubs and authorities.

Erdős–Rényi $G(n, p)$ and Thresholds (Weeks 4–6)

Erdős–Rényi $G(n, p)$

- n nodes; each edge present independently with prob. p.
- Degrees $\sim \mathrm{Bin}(n-1, p)$; avg. degree $\approx p(n-1)$.
- Serves as a baseline "null model" for comparison.

Threshold phenomena

- Giant component appears when avg. degree crosses ≈ 1.
- Many properties undergo sharp transitions: connectivity, isolated nodes, appearance of triangles, etc.

Beyond ER: Degree Heterogeneity

Real networks rarely look like $G(n, p)$:

- heavy-tailed degree distributions,
- presence of hubs,
- clustering far above p.

Configuration model

- Random graph with a **prescribed degree sequence**.
- Keeps heterogeneity, removes all other structure.

Preferential Attachment and Power Laws

Preferential attachment (BA model)

- New nodes attach to high-degree nodes with higher probability.
- Generates a heavy-tailed (power-law) degree distribution.
- Explains hubs and vulnerability to targeted attacks.

We contrasted ER (homogeneous) with BA (heterogeneous, hub-dominated).

Stochastic Block Models (SBM)

In Stochastic Block Model:

- Nodes belong to latent groups ("blocks").
- Edge probabilities depend on block memberships.
- Captures community structure in a probabilistic way.

Examples:

- assortative communities (dense inside, sparse between),
- core–periphery,
- bipartite-like structures.

Latent Space Models and ERGMs

Latent space models

- Nodes live in a hidden geometric space.
- Edge probability decreases with distance.
- Captures homophily and transitivity naturally.

ERGMs (briefly)

- Specify a probability distribution over networks using statistics like edges, triangles, degree sequence.
- Very flexible but difficult to fit if too flexible.

Clustering, small world, power laws

Clustering coefficient

- Probability that two neighbours of a node are connected.
- Real social networks: high clustering.
- $G(n, p)$: clustering $\approx p$, usually much smaller.

Small world

- Short path lengths *and* high clustering.

Hubs and power laws

- Degree distribution often heavy-tailed.
- Preferential attachment (BA): new edges attach to high-degree nodes.

Communities and modularity (Weeks 7–8)

Community (informal)

- Subset of nodes with many internal edges and fewer external edges.

Modularity

- Measures how much more edges fall inside groups than expected under a random graph with the same degree sequence.
- Many algorithms (e.g. greedy, spectral) aim to maximise modularity.

We saw that:

- Community detection is often heuristic.
- Hierarchical clustering is a computationally advantageous approach.

Mechanisms for (social) network formation

Why do networks look the way they do?

- **Homophily**: similar nodes tend to connect.
- **Triadic closure**: friend of my friend becomes my friend.
- **Preferential attachment**: rich get richer.
- **Constraints**: geography, capacity limits, institutional rules.

Networks form over time and they constantly change.

Static models (ER, BA, stochastic block models) can be viewed as snapshots of such dynamic processes.

Matching markets (Week 8)

Perfect matching

- A set of edges that pairs every buyer with exactly one good.
- Central concept in bipartite graphs.

Optimal assignment

- Each buyer i has a valuation v_{ij} for each good j.
- We choose a perfect matching that maximises total value.

Market–clearing prices

- Prices (p_j) such that every buyer prefers their assigned good.

This was our first example where *values on a bipartite graph* produce a global market outcome.

Compartmental models: SI, SIS, SIR (Weeks 9–10)

Compartments

- S: susceptible, I: infected, R: recovered (immune).

Three basic models

- SI: S $\rightarrow$ I, no recovery.
- SIS: S $\rightarrow$ I $\rightarrow$ S (no lasting immunity).
- SIR: S $\rightarrow$ I $\rightarrow$ R (permanent immunity).

Under homogeneous mixing we obtained:

- Logistic growth in SI.
- SIS with endemic vs disease-free regimes.
- SIR with a single epidemic peak and a final size.

Basic reproductive number and networks

Basic reproductive number

$$R_0 = \frac{\beta c}{\mu}$$

- β: transmission parameter, c: contact rate, μ: recovery rate.
- Roughly: expected number of secondary infections from one infected.

On networks:

- *Who* you meet depends on network structure.
- Hubs become super-spreaders.
- Same β, μ but different graphs $\Rightarrow$ very different curves $i(t)$.

Spreading on networks and rumors

Discrete-time SI / SIR on a network

- Infection moves along edges with probability p.
- For SIR, infected nodes also recover and do not return to S.

Rumor spreading (Maki–Thompson)

- States: Ignorant (I), Spreader (S), Stifler (R).
- An S who calls I turns I into S (rumor spreads).
- An S who calls S or R becomes R (loses interest).

Epidemics are stopped by *immunity*; rumors are stopped by *loss of interest*.

What you practised in tutorials and Colabs

- Constructing and visualising graphs in Python.
- Computing degree distributions and centrality measures.
- Simulating random graphs (ER, scale-free) and comparing to real networks.
- Implementing basic random walks and PageRank.
- Simulating SI/SIR on different network topologies.

The goal was not just to see formulas, but to *experiment*: change parameters, change graphs, and observe how behaviour changes.

Key conceptual takeaways

1. Local rules $+$ network structure $\Rightarrow$ complex global behaviour.
2. Average degree is not enough; degree *distribution* and hubs matter.
3. Random graph models provide useful baselines, but real networks often show more clustering and richer community structure.
4. Centrality is multi-faceted; "importance" depends on the question.
5. Epidemics, information, and markets can all be studied on the same graph-theoretic foundation.

Where to go from here

Possible next steps:

- More probability on graphs: percolation, branching processes.
- Statistical network models: ERGM, stochastic block models, latent-space models.
- Algorithmic side: scalable community detection, graph embeddings.
- Applications: recommender systems, causal inference on networks, spreading of opinions and behaviour.

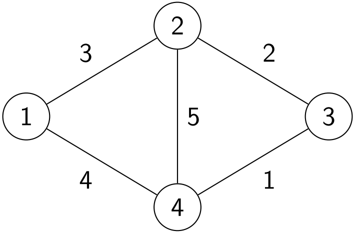
The methods you saw (linear algebra, probability, simulation) will reappear in many advanced topics.

Exam's Instructions & Structure

- Thursday, **11th December**. Starts 3pm to 5pm, 40.154.
- Formula sheets are permitted; two-sided, handwritten. **Sheets will be collected at the end of the exam**.
- Use of personal **non-programmable calculators** is permitted.
- Mobiles stored in a bag, strict no-cheating policy.
- 30 questions with short or very short answers.
- The questions will cover the material of the whole term.

Sample Question 1 (Weighted Closeness Centrality)

Consider the weighted graph:



(a) Write the weighted shortest-path distance between every pair of nodes (you may give them in a matrix).

(b) Which node has the highest weighted closeness centrality? Justify briefly using the distances.

Sample Question 2 (SIS Model and R_0)

Consider the SIS epidemic model

$$\frac{di}{dt} = \beta c \, (1 - i) \, i \; - \; \mu \, i,$$

where $i(t)$ is the fraction of infected individuals at time t, βc is the effective contact rate, and μ is the recovery rate.

Suppose $\beta c = 1.2$ and $\mu = 0.8$.

(a) Compute the basic reproductive number R_0.

(b) Based on this value of R_0, do you expect the infection to die out or persist in the long run? Briefly justify in one or two sentences.

Sample Question 3 (Configuration Model)

Consider a configuration model on 5 nodes with degree sequence

$$(d_1, d_2, d_3, d_4, d_5) = (4,\ 3,\ 2,\ 1,\ 0).$$

The total number of half-edges is 10.

(a) Give an approximate expression for the expected number of edges between nodes 1 and 3.

(b) Give an approximate expression for the expected number of edges between nodes 2 and 4.

Sample Question 4 (Random Graphs and Power Laws)

(a) ER vs. preferential attachment
Explain in two or three sentences how the degree distributions in $G(n, p)$ and in the Barabási–Albert preferential attachment model differ.

(b) Power-law moments
Consider a degree distribution with

$$P(K = k) \propto k^{-\alpha}, \qquad k \geq 1,$$

and exponent $\alpha = 2.4$.

(i) Is the *mean* degree finite or infinite? Briefly justify.

(ii) Is the *variance* of the degree finite or infinite? Briefly justify.

Sample Question 4 (Utility Maximisation)

Three buyers (1,2,3) and three goods (A,B,C). Valuations:

	A	B	C
1	7	5	2
2	3	9	4
3	6	4	8

Prices: $p_A = 4,\ p_B = 2,\ p_C = 6.$

(a) For each buyer, compute utilities $v_{ij} - p_j$.

(b) Which good does each buyer choose?

(c) Do these prices clear the market? Explain briefly.

Thank you for your work this term!