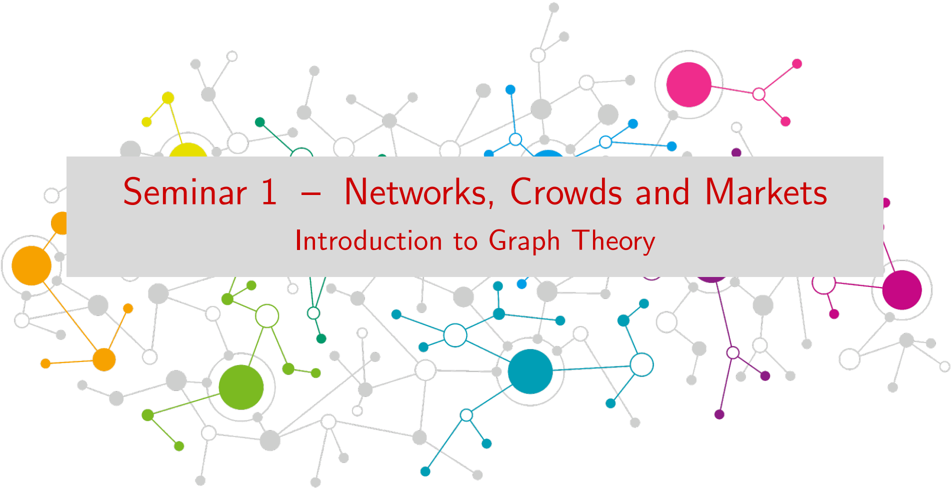
Seminar 1 – Networks, Crowds and Markets

Introduction to Graph Theory



Exercise 1. Adjacency and complements

Let B be the adjacency matrix of a simple *directed* graph G, and A the adjacency matrix of its *underlying undirected* graph (ignore arrow directions; collapse parallel arcs to a single edge). Decide if each statement is true or false and briefly justify.

(a) $a_{ij} = a_{ji}$ for $i \neq j$

(b) $b_{ij} \leq a_{ij}$ for $i \neq j$

(c) $b_{ij} + b_{ji} \leq 2$ for $i \neq j$

(d) $a_{ij} = b_{ij}$ for $i \neq j$

(e) In any simple *undirected* graph: $A(G^c) = J - I - A(G)$.

Solution to Exercise 1

(a) **True.** In the underlying undirected graph, edges have no direction, so its adjacency matrix is symmetric: $a_{ij} = a_{ji}$ for $i \neq j$.

(b) **True.** If $b_{ij} = 1$ then there is a directed edge $i \to j$, so in the underlying undirected graph there is an (undirected) edge $\{i, j\}$ and hence $a_{ij} = 1$. If $b_{ij} = 0$, the inequality holds trivially.

(c) **True.** In a simple directed graph there is at most one arc in each direction. Thus $b_{ij}, b_{ji} \in \{0, 1\}$ and $b_{ij} + b_{ji} \leq 2$ always.

(d) **False.** It may happen that $b_{ij} = 0$ but $b_{ji} = 1$, in which case $a_{ij} = 1$ while $b_{ij} = 0$. So a_{ij} and b_{ij} need not be equal.

(e) **True.** For a simple undirected graph, the complement G^c has edges exactly where G has none, except on the diagonal (no self-loops). This is encoded by

$$A(G^c) = J - I - A(G),$$

where J is the all-ones matrix and I is the identity.

Exercise 2. Quick degree-sequence checks

Consider the multiset

$$\{0, 1, 1, 1, 2, 2, 2, 3, 6\}.$$

(a) Can this be a degree sequence of a simple undirected graph on $N = 9$ vertices? Justify briefly by running the parity and sum check.

(b) If yes, how many edges L does such a graph have? What is $\overline{deg}$?

(c) Intend to draw this graph.

Solution to Exercise 2

Parity / sum check. The sum of degrees is

$$0 + 1 + 1 + 1 + 2 + 2 + 2 + 3 + 6 = 18,$$

which is even, so it passes the basic parity test. Also the maximum degree $6 \leq N - 1 = 8$.

Edges and average degree.

$$\sum_i d_i = 2L = 18 \quad \Rightarrow \quad L = 9.$$

The average degree is

$$\overline{\text{deg}} = \frac{2L}{N} = \frac{18}{9} = 2.$$

Try to plot this graph. You can safely ignore the degree 9 node and then the degree 6 nodes needs to be connected to almost everyone...

Exercise 3: Paths and Connectedness

Show that if a graph (simple undirected) is not connected then its complement is. Provide an example of a four node network that is connected and such that its complement is also connected.

Hint: What happens in $\overline{G}$ is u, v lie in two different components of G? What happens if they lie in the same component?

Solution to Exercise 3

Let G be a simple undirected graph that is *not* connected. We show that its complement $\overline{G}$ is connected.

Take any two distinct vertices u, v.

- If u and v lie in *different* components of G, then there is no edge between them in G, so there *is* an edge uv in $\overline{G}$.
- If u and v lie in the *same* component of G, pick any vertex w in a different component of G (possible since G is not connected). Then there are no edges uw or vw in G, hence both uw and vw are edges in $\overline{G}$. Thus u and v are connected in $\overline{G}$ by the path $u - w - v$.

Since any two vertices in $\overline{G}$ are connected by a path of length 1 or 2, $\overline{G}$ is connected.

Example with 4 nodes. Take the path P_4: $1 - 2 - 3 - 4$. This is connected. Its complement has edges between all nonadjacent pairs:

$$\{1,3\}, \{1,4\}, \{2,4\},$$

which also form a connected graph. So both G and $\overline{G}$ are connected.

Exercise 4. Degrees and handshaking

A simple undirected graph G has $N = 1000$ nodes and $L = 800$ edges.

(a) Compute the average degree $\overline{\deg}(G)$.

(b) You observe the degree sequence has N_0 isolated vertices and N_1 vertices of degree 1. Give two independent checks that your N_0, N_1 are consistent with N, L.
(Hint: handshaking and nonnegativity.)
- e.g. why $N_0 = 960$ is impossible?
- why $N_0 = 0$, $N_1 = 399$ is impossible?

Solution to Exercise 4

(a) By the handshaking lemma,

$$\overline{\deg}(G) = \frac{2L}{N} = \frac{2 \cdot 800}{1000} = \frac{1600}{1000} = 1.6.$$

(b) Let N_0 be the number of isolated vertices and N_1 the number of degree-1 vertices. Then:

$$\sum_{v} \deg(v) = 0 \cdot N_0 + 1 \cdot N_1 + \sum_{\deg(v) \geq 2} \deg(v) = 2L = 1600.$$

Also,

$$N_0 + N_1 + \#\{v : \deg(v) \geq 2\} = N = 1000.$$

Check 1: $N_0 = 960$ **impossible.**
If $N_0 = 960$, then only 40 vertices are non-isolated. The maximum number of edges among 40 vertices is

$$\binom{40}{2} = 780,$$

so the maximum possible number of edges in the entire graph is $780 < 800 = L$. Contradiction. Hence $N_0 = 960$ cannot occur.

Check 2: $N_0 = 0$, $N_1 = 399$ **impossible.**
If there are no isolated vertices and 399 vertices of degree 1, then the remaining 601 vertices must each have degree at least 2. Therefore the sum of degrees satisfies

$$\sum_{v} \deg(v) \geq 399 \cdot 1 + 601 \cdot 2 = 399 + 1202 = 1601 > 1600,$$

which is impossible since $2L = 1600$. So $N_0 = 0$, $N_1 = 399$ cannot happen.

Exercise 5: Comparing networks

Imagine two different undirected networks, each with the same number of nodes and links.

Questions:

1. Must both networks have the same maximum and minimum degree? Explain why or why not.
2. Must they have the same mean degree? Explain why or why not.

Solution to Exercise 5

Let both networks have N nodes and L links.

1. **Maximum/minimum degree.** They do *not* need to have the same maximum or minimum degree. The same N and L only fix the *sum* of all degrees, not how degrees are distributed. One network may have a hub with very high degree and many low-degree nodes; another network may spread degrees more evenly. **Give a simple example on four nodes.**

2. **Mean degree.** They *must* have the same mean degree, because

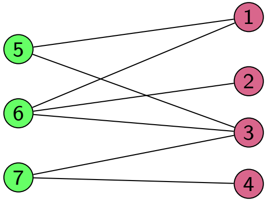
$$\overline{\deg} = \frac{1}{N}\sum_{v}\deg(v) = \frac{2L}{N}.$$

If N and L are the same, then $\overline{\deg}$ is identical for both networks.

Exercise 6: Bipartite Networks

Consider the following bipartite network:

1. Construct its adjacency matrix A. Why the block diagonal blocks are zero?
2. Calculate the average degree of the purple nodes and the average degree of the green nodes in the bipartite network.
3. Without computing A^2 explicitly, what are its $(5,1)$ and $(5,6)$ entries?



Bipartite network with 6 purple and 5 green nodes, 10 links.

Solution to Exercise 6

Label the vertices as purple $1, 2, 3, 4$ and green $5, 6, 7$. From the picture we have edges:

$$\{5,1\}, \{5,3\}, \{6,1\}, \{6,2\}, \{6,3\}, \{7,3\}, \{7,4\}.$$

(a) Adjacency matrix and zero blocks. With ordering $(1, 2, 3, 4, 5, 6, 7)$, the adjacency matrix A (symmetric) has nonzero entries only between purple and green nodes. Thus the top-left 4×4 and bottom-right 3×3 blocks (within purple / within green) are zero, since bipartite graphs have no edges inside each part.

(b) Average degrees. Degrees of purple nodes:

$$\deg(1) = 2\ (5,6), \quad \deg(2) = 1\ (6), \quad \deg(3) = 3\ (5,6,7), \quad \deg(4) = 1\ (7).$$

So the average purple degree is

$$\overline{\deg}_{\text{purple}} = \frac{2+1+3+1}{4} = \frac{7}{4} = 1.75.$$

Degrees of green nodes:

$$\deg(5) = 2\ (1,3), \quad \deg(6) = 3\ (1,2,3), \quad \deg(7) = 2\ (3,4),$$

so

$$\overline{\deg}_{\text{green}} = \frac{2+3+2}{3} = \frac{7}{3} \approx 2.33.$$

(c) Entries of A^2.
Recall $(A^2)_{ij}$ counts the number of walks of length 2 from i to j.

- $(5,1)$: walks $5 \to * \to 1$. Neighbours of 5 are 1 and 3. The only valid middle vertex is 3 (since there is no loop at 1). So we have $5 - 3 - 1$ as the unique length-2 walk, hence $(A^2)_{5,1} = 1$.
- $(5,6)$: walks $5 \to * \to 6$. Neighbours of 5 are 1 and 3. Both are connected to 6:

  $$5 - 1 - 6, \quad 5 - 3 - 6.$$

  Thus $(A^2)_{5,6} = 2$.

Exercise 7. Bipartite graphs

(a) Let a bipartite graph have parts of sizes m and n. Show that the maximum number of edges is mn, attained by $K_{m,n}$.

(b) Among all bipartite graphs on N vertices (where the partition is free to choose), show that the maximum number of edges is

$$L_{\max} = \left\lfloor \frac{N^2}{4} \right\rfloor,$$

attained by $K_{\lfloor N/2 \rfloor, \lceil N/2 \rceil}$.

Solution to Exercise 7

(a) In a bipartite graph with parts of sizes m and n, every edge connects one vertex in the first part to one vertex in the second part. The maximum number of such edges occurs when all possible cross-part pairs are present, i.e. the complete bipartite graph $K_{m,n}$. This has

$$mn$$

edges, and no bipartite graph on the same parts can have more.

(b) Now we are free to choose the bipartition sizes m, n subject to $m + n = N$. The maximum possible number of edges is

$$\max_{m+n=N} mn.$$

Treat m as real: $f(m) = m(N - m)$ is a concave quadratic with maximum at $m = N/2$, giving $f(N/2) = N^2/4$. For integer m, n, the maximum is therefore

$$L_{\max} = \left\lfloor \frac{N^2}{4} \right\rfloor,$$

attained by the most balanced split $m = \lfloor N/2 \rfloor$, $n = \lceil N/2 \rceil$, i.e. by $K_{\lfloor N/2 \rfloor, \lceil N/2 \rceil}$.

Exercise 8: Netflix bipartite network

Netflix keeps data on customer preferences using a big bipartite network connecting users to titles they have watched and/or rated. Say, Netflix's movie library contains approximately 20,000 titles.

In the first quarter of 2025, Netflix reported having about 300 million users. Assume the average user's degree in this network is 100.

Questions:

1. Approximately how many links are in this network?
2. Would you consider this network sparse or dense? Explain.

Solution to Exercise 8

Let U be the number of users and T the number of titles.

- $U \approx 300{,}000{,}000$,
- $T \approx 20{,}000$,
- average user degree (number of titles per user) is 100.

(a) Number of links.
Each of the U users has degree (on average) 100, so the total number of links (user–title pairs) is approximately

$$L \approx U \times 100 \approx 300{,}000{,}000 \times 100 = 30{,}000{,}000{,}000 = 3 \times 10^{10}.$$

(b) Sparse or dense?
The total number of *possible* links in the bipartite graph is

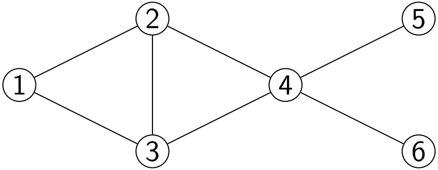
$$U \times T \approx 300{,}000{,}000 \times 20{,}000 \approx 6 \times 10^{12}.$$

The edge density is roughly

$$\frac{L}{UT} \approx \frac{3 \times 10^{10}}{6 \times 10^{12}} \approx 0.005 = 0.5\%.$$

Since only about 0.5% of all possible user–title pairs are present, this network is very *sparse*.

Exercise 9. BFS levels and shortest paths (unweighted)



(a) Run BFS from node 1 by hand: write the level sets $L_0, L_1, L_2, \ldots$.

(b) Give a shortest path from 1 to 6. How many distinct shortest paths from 1 to 6 exist?

(c) Which vertices are at eccentricity equal to the graph diameter?

Solution to Exercise 9 (BFS)

(a) BFS levels from node 1.

$$L_0 = \{1\}, \quad L_1 = \{2, 3\}, \quad L_2 = \{4\}, \quad L_3 = \{5, 6\}.$$

(b) Shortest paths from 1 to 6.
Any shortest path from 1 to 6 must go through 4 (since 4 is the unique node in L_2). There are two distinct shortest paths of length 3:

$$1 \to 2 \to 4 \to 6, \qquad 1 \to 3 \to 4 \to 6.$$

(c) Vertices with eccentricity equal to the diameter.
Compute distances:

- From 1: $\text{dist}(1, \cdot) = (0, 1, 1, 2, 3, 3)$, so $\mathrm{ecc}(1) = 3$.
- From 5: $5 \to 4 \to 2$ or 3 or 1 etc.; the farthest node is at distance 3 (e.g. $5 \to 4 \to 2 \to 1$), so $\mathrm{ecc}(5) = 3$.
- From 6: by symmetry with 5, $\mathrm{ecc}(6) = 3$.

No vertex has eccentricity larger than 3, so the diameter is 3 and the vertices whose eccentricity equals the diameter are

$$\{1, 5, 6\}.$$

Exercise 9: Diameter and Degree

Consider a sequence of networks such that each network in the sequence is connected and involves more nodes than the previous network. Show that if the diameter of the networks is bounded, then the maximal degree of the networks is unbounded. That is, show that if there exists a finite number M such that the diameter of every network in the sequence is less than M, then for any integer K there exists a network in the sequence and a node in that network that has more than K neighbors.

Hint: Use the Moore bound.

Solution to Exercise 9 (Diameter and Degree)

Let G be a connected graph with maximum degree Δ and diameter at most M. The *Moore bound* says that the number of vertices N satisfies

$$N \;\leq\; 1 + \Delta \sum_{i=0}^{M-1} (\Delta - 1)^i$$

for $\Delta \geq 2$. For fixed M, the right-hand side grows on the order of $(\Delta - 1)^M$ as Δ increases.

Now consider sequence of connected graphs (G_n) with strictly increasing number of nodes N_n and all diameters bounded by the same M.

Suppose, for contradiction, that the maximum degree is bounded by some constant K across the whole sequence: $\Delta(G_n) \leq K$ for all n. Then the Moore bound implies that each G_n has at most

$$1 + K \sum_{i=0}^{M-1} (K - 1)^i$$

vertices, a quantity depending only on K and M, not on n. This would bound the sizes N_n uniformly, contradicting that N_n increases without bound.

Hence our assumption is false: the sequence must contain graphs whose maximum degree exceeds any given integer K. In other words, if diameters stay bounded while $N_n \to \infty$, then the maximum degree must be unbounded.

Exercise 10: Facts about Trees

Recall that a tree is a connected graph with no cycles.

Show the following:

1. The minimal number of edges of a connected graph is $N - 1$.
2. A connected network is a tree if and only if it has $N - 1$ links.
3. A tree has at least two leaves, where leaves are nodes that have exactly one link.
4. In a tree, there is a unique path between any two nodes.

Solution to Exercise 10

Let G be a connected simple graph on N vertices.

1. Minimal number of edges is $N-1$. Take any spanning tree T of G (it exists because G is connected). A tree on N vertices always has $N-1$ edges, and any connected graph must have at least as many edges as its spanning tree, so a connected graph has at least $N-1$ edges.

2. Connected and $N-1$ edges $\Leftrightarrow$ tree. ($\Rightarrow$) If G is connected with $N-1$ edges, any cycle would allow us to delete one edge while keeping the graph connected, giving a connected graph with fewer than $N-1$ edges, impossible by part (1). So G has no cycles and is therefore a tree. ($\Leftarrow$) If G is a tree, it is by definition connected and acyclic. A standard property of trees is that they have exactly $N-1$ edges (can be proved by induction on N).

3. A tree has at least two leaves. Consider a longest simple path in the tree, say $v_1, \ldots, v_k$. If v_1 had degree at least 2, then there would be a neighbour of v_1 other than v_2, which would extend the path and contradict maximality. Hence v_1 has degree 1, and similarly v_k has degree 1. So there are at least two leaves.

4. Unique path between any two nodes. In a tree, there is at least one simple path between any two vertices (since it is connected). If there were two distinct simple paths between the same pair, their union would contain a cycle, contradicting acyclicity. Hence the path between any two vertices is unique.

Optional take-home: NetworkX mini-lab

Paste this into Colab:

```
# 1) Setup
!pip -q install networkx matplotlib
import networkx as nx
import matplotlib.pyplot as plt

# 2) Build a small graph by hand
G = nx.Graph()
G.add_edges_from([(1,2),(1,3),(2,4),(3,4),(4,5),(5,6)])

# 3) Basic stats
N, L = G.number_of_nodes(), G.number_of_edges()
print(f"Nodes: {N}, Edges: {L}, Avg degree: {2*L/N:.2f}")

# 4) Degree sequence
deg = [d for _, d in G.degree()]
print("Degree sequence:", sorted(deg))

# 5) Shortest paths from node 1 (BFS)
lengths = nx.single_source_shortest_path_length(G, 1)
print("Distances from 1:", dict(lengths))

# 6) Draw
pos = nx.spring_layout(G, seed=7)
nx.draw(G, pos, with_labels=True, node_color="lightblue",
        node_size=700, edge_color="gray")
plt.show()
```

Stretch goals: change the edge set, recompute degree sequence and distances; add a node with many edges and see how the layout and degree stats change.