Recap: Network Centrality and Clustering

DS701 Session 24 — Mon Nov 30, 2026

Today’s plan

Knowledge-Check Review

KC 1: Three measures, three notions

*The lecture defines three centrality measures — closeness, betweenness, and eigenvector centrality. For each one, state in a single phrase what makes a node score **high**, and name the karate-club member (0, the instructor, or 33, the president) that each measure ranked first.*

KC 2: Why the min cut is useless here

The global minimum cut of the karate-club graph has cost 1: it slices off node 11, which has a single edge. Explain why this "best" cut is useless for partitioning the club, and how the isoperimetric ratio $\alpha = E(U, V \setminus U) / \min(|U|, |V \setminus U|)$ changes the goal so that a cut like Zachary's two factions wins instead.

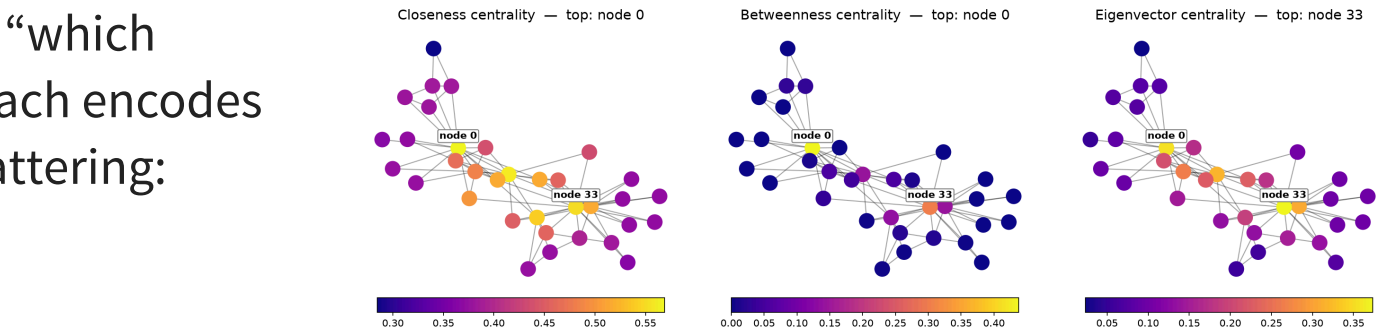
KC 3: Spectral clustering — which eigenvectors, and why they are coordinates

*Spectral clustering embeds each node as a point in $\mathbb{R}^d$ using eigenvectors of the Laplacian $L = D - A$, then runs k-means on those points. (a) Which eigenvectors are used, and why is the eigenvector for the smallest eigenvalue **excluded**? (b) Name the two earlier tools from this course this procedure combines, and say in one sentence why the Laplacian eigenvectors are a sensible set of coordinates for the nodes.*

Highlights

1. Centrality — “important” according to *whom*?

Every measure answers “which nodes matter?” — but each encodes a **different notion** of mattering:



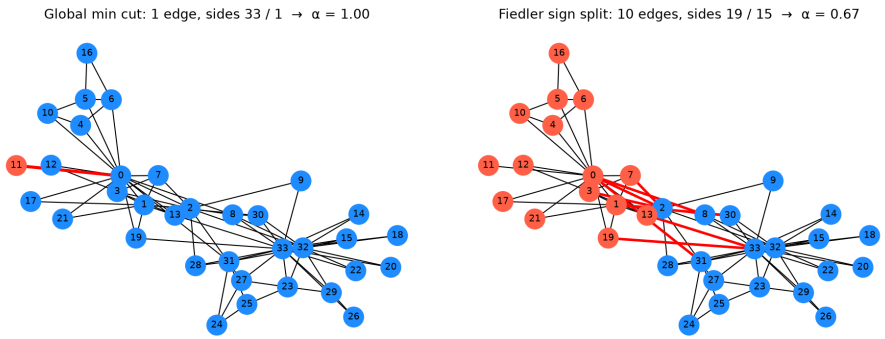
2. Eigenvector centrality: importance is recursive

$$x_i = \frac{1}{\lambda} \sum_j A_{ij}\, x_j \qquad \Longleftrightarrow \qquad A\mathbf{x} = \lambda \mathbf{x}$$

3. From eigenvector centrality to PageRank

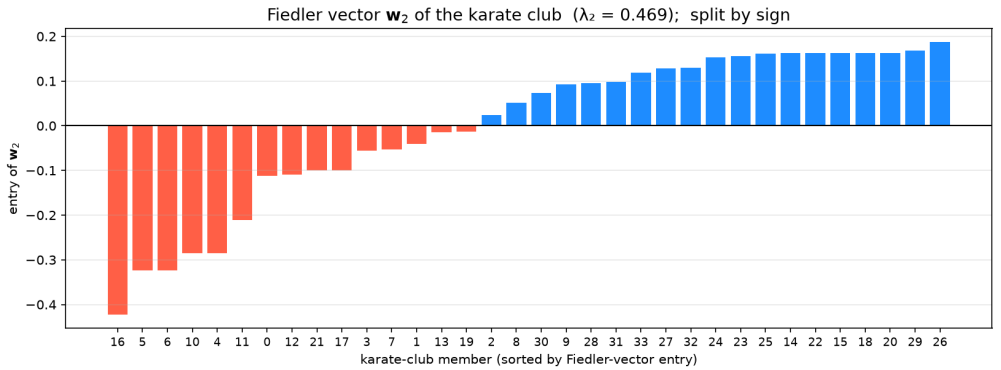
The lecture stops at $A\mathbf{x} = \lambda\mathbf{x}$; the Web needed two fixes. **A random surfer** on a page follows a random link with probability d, and jumps to a uniformly random page with probability $1 - d$ (damping, $d = 0.85$).

4. Cuts: from “cheapest” to “balanced”



5. The Laplacian, and why its second eigenvector cuts the graph

$$L = D - A, \qquad \mathbf{x}^T L \mathbf{x} = \sum_{(i,j)\in E} (x_i - x_j)^2 \ \geq 0$$



6. Spectral clustering *is* community detection

7. The measures disagree — so the question picks the measure

you are asking …	report
who can reach everyone fastest (spread a rumour, place a depot)?	**closeness**
whose removal breaks the most shortest paths (broker, bottleneck, gatekeeper)?	**betweenness**
who is connected to the well-connected (prestige, status)?	**eigenvector**
the same, but each node's vote is diluted by how many it gives (links, citations)?	**PageRank**
who simply has the most contacts?	**degree**

In-Class Activity

Activity: centralities by hand, PageRank from scratch, and a graph that disagrees

Goal: compute closeness, betweenness and eigenvector centrality on a real network (power iteration for the last), build PageRank with damping $d = 0.85$ from scratch and check it against NetworkX, spectrally partition the same network with the Fiedler vector — then take a planted graph whose five centrality measures crown different nodes, tabulate and correlate the rankings, and match measures to questions.

- Work in groups of 2–3. Open **your section's** notebook (A1: karate club, B1: *Les Misérables*).
- Parts marked **(autograded)** are submitted to Gradescope; the rest is participation.
- Staff will circulate — be ready to explain any part of your work.
- Dependencies: `numpy`, `pandas`, `matplotlib`, `networkx`.

Going deeper

Full lecture notes: Network Centrality and Clustering

- Closeness Centrality and Betweenness Centrality — definitions, the dependency ratio, the karate colourings
- Eigenvector Centrality Definition, Matrix Form and Key Points — $A\mathbf{x} = \lambda\mathbf{x}$ and Perron–Frobenius
- Centrality Comparison: Karate Example
- Min s–t cut, Minimum Cuts and Balanced Cuts — Zachary's prediction; why min cut peels a leaf
- Spectral Graph Theory, the Laplacian Quadratic Form and the Fiedler Vector
- Spectral Partitioning — sign / bisection / gap / ratio-cut rules, and the karate split
- Spectral Clustering and its football example with ARI vs. dimension
- Last session's foundations: Networks I — degree, paths, clustering coefficient, layouts

Before you go: quiz 3 is Wednesday

Wed Dec 2, in class — closed notes, paper and pencil, covering lectures 14–21 (causal inference, neural networks, NLP, recommenders, networks). Today was the last content session.

- The **reference sheet** for the unit was published last Wednesday: quiz3-reference.pdf. The printed copy handed out on quiz day is identical, and it is the only sheet allowed in the room.
- Today's ideas that live on the sheet: the three centralities, $A\mathbf{x} = \lambda\mathbf{x}$, cuts and α, $L = D - A$, the Fiedler vector, the spectral-clustering steps.
- The homework sets and knowledge checks are the practice; the recap decks are the answer keys.