Social Network Analysis

Guide to AI and Data Courses · Part IV: Data

Networks and graphs | Centrality | Community detection | Network models | Diffusion | Neighboring courses

Sercan KÜLCÜ · Giresun University · Computer Engineering

© Sercan Külcü. All rights reserved. · Interactive HTML version of these slides:
https://sercankulcu.github.io/files/ai_courses/slides/Chapter_30_Social_Network_Analysis.html

What Question Does It Ask?

> **How are networks of people and their relationships analyzed?**

Social network analysis studies the **graph** formed by treating people as nodes and relationships as edges. The question is not the traits of individuals but the structure of their ties: who is central, which groups exist, and how a rumor or a disease spreads through the network.

At a glance

APPROACH graph theory and statistics	**INPUT** nodes and links	**OUTPUT** centrality, communities, diffusion
SUCCESS MEASURE modularity, interpretability	**MATH LOAD** medium: graphs, matrices	**DATA NEEDS** network data

A Short History

1736
Euler solves the Königsberg bridges problem; graph theory is born.

1934
Moreno draws relationships in diagrams he calls sociograms.

1967
Milgram's letter experiment popularizes the idea of "six degrees".

1973
Granovetter shows "the strength of weak ties": job leads come from distant acquaintances.

1998
Watts and Strogatz publish the small-world network model.

2008
The Louvain method makes fast community detection possible on networks with millions of nodes.

THE COURSE

Core Concepts

Nodes and edges
Directed or undirected, weighted or unweighted.

Degree distribution
A few highly connected nodes, many with few links.

Centrality
Degree, betweenness, closeness and eigenvector measures.

Clustering coefficient
How many of my friends know each other.

Small world
High clustering but short average paths.

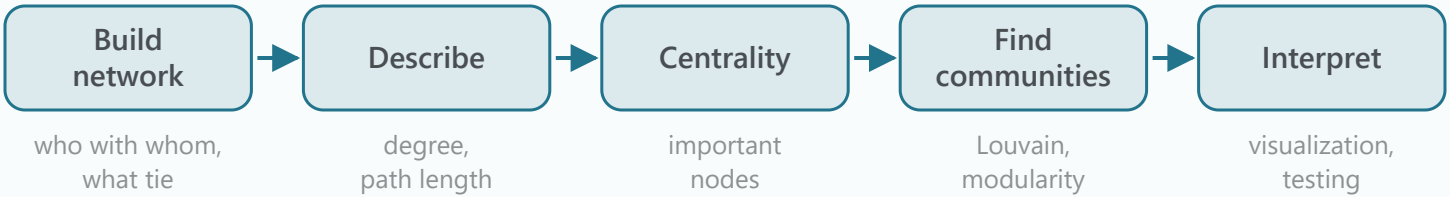
Community
Nodes densely linked inside, sparsely to the outside.

Modularity
How good a partition into communities is.

Diffusion models
Spread of information or disease on a network (SIR).

THE COURSE

A Typical Workflow



Example: A researcher builds a network from a school's friendship survey, looks at the degree distribution, finds "bridge" students with high betweenness, extracts friend groups that cross classes with Louvain, and discusses the drawn result with the teachers.

THE COURSE

What Does the Course Cover?

1. Graph basics: nodes, edges, adjacency matrix
2. Collecting network data: surveys, traces, ego nets
3. Degree distribution, path length and diameter
4. Centrality: degree, betweenness, closeness
5. Eigenvector centrality and PageRank
6. Clustering, triadic closure and weak ties
7. Homophily and assortativity
8. Random graphs and the small-world model
9. Scale-free networks and preferential attachment
10. Community detection: modularity and Louvain
11. Spreading: the SIR model and information cascades
12. Link prediction and network visualization

Skills you will gain

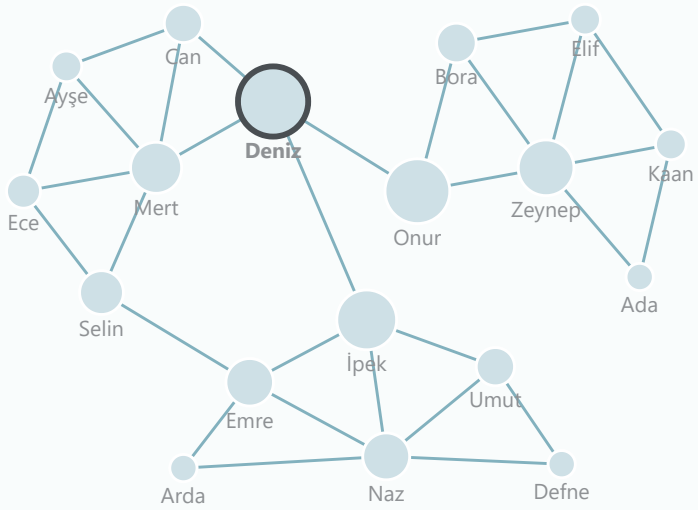
network analysis with NetworkX · network visualization with Gephi · interpreting centrality measures · finding and testing communities · writing a spreading simulation

A typical semester; the order and coverage vary between universities.

TRY IT

Who Should Start the Campaign?

A club will hand its announcement to one person in the friendship network of its 18 members. Click a person to see their centrality, find the communities, then watch the news spread from that person step by step.



Size: Degree **Betweenness**

Find communities | Start spreading

Pass-on probability p = **0.4**

Deniz: degree **4** (rank 4), betweenness **0.58** (rank 1)

Betweenness counts how many shortest paths pass through a person; bridges score high even with few friends.

By degree, people with many friends such as Mert stand out, but the bridge between groups is Deniz (betweenness). Starting the news at a bridge carries it into several communities; this is how influencers are chosen.

CONTEXT

Prerequisites

Courses worth knowing before this one, and why they are needed.

Course	Why it is needed
Linear Algebra	adjacency matrix, spectral methods
Data Structures	adjacency lists and matrices: storing a graph in memory
Algorithms	BFS, DFS and shortest paths: the basis of centrality computations
Discrete Mathematics	graph theory, sets and counting

CONTEXT

Are You Ready for This Course?

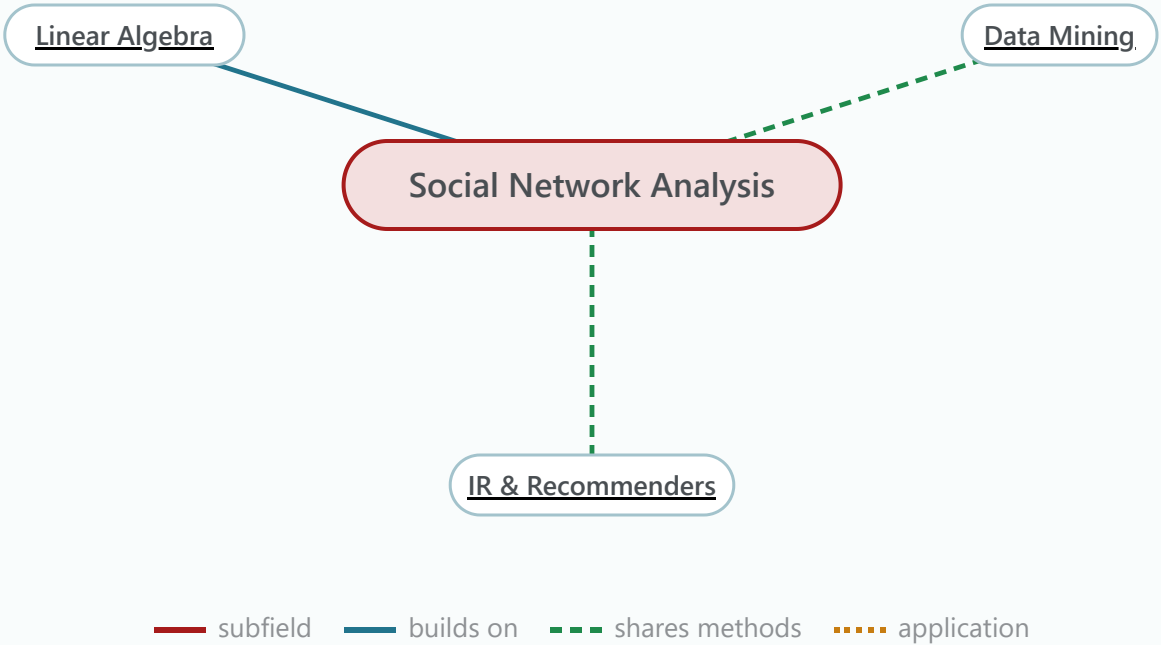
Ask yourself; for every "no", the right-hand column shows where to start.

Ask yourself	If not, start with
Can you find the shortest path between two nodes of a graph with BFS?	Graph search from an algorithms course
Can you explain what an eigenvector of a matrix means?	Linear Algebra
Can you compute the mean and variance of a distribution?	Probability and Statistics
Can you read an edge-list file into a Python dictionary?	An intro programming course, then a short NetworkX tutorial

If you can say "yes" to most of these, you are ready; one or two gaps can be closed during the semester.

CONTEXT

Neighbouring Courses



Overlapping Courses

Relation	Course	Shared ground
builds on	This course builds on Linear Algebra	adjacency matrix, spectral methods
shares methods	Shares methods with Data Mining	community detection
shares methods	Shares methods with Information Retrieval and Recommender Systems	PageRank, link analysis

DIFFERENCES

Where It Differs (1/2)

Criterion	Social Network Analysis	Data Mining	IR & Recommenders
Core question	How are networks of people and their relationships analyzed?	How are hidden, useful patterns found in large piles of data?	How are the most relevant documents or products picked from millions of candidates?
Approach	graph theory and statistics	discovery-driven analysis	matching and ranking
Input	nodes and links	large tables or transaction data	query, documents, user history
Output	centrality, communities, diffusion	rules, clusters, outliers	ranked result list
Success measure	modularity, interpretability	interestingness, support, confidence	precision, recall, NDCG
Math load	medium: graphs, matrices	medium	medium: vector spaces
Data needs	network data	large	lots of interaction data

- **Data Mining:** Data mining looks for patterns in tabular records; this course studies the structure of the links between records.
- **IR & Recommenders:** Information retrieval uses PageRank as a ranking signal; this course asks what centrality says about the network.

DIFFERENCES

Where It Differs (2/2)

Criterion	Social Network Analysis	Linear Algebra	Multi-Agent Systems
Core question	How are networks of people and their relationships analyzed?	How do vectors and matrices represent data and transformations?	How do several autonomous agents decide together or against each other?
Approach	graph theory and statistics	vector spaces and transformations	interacting autonomous agents
Input	nodes and links	vectors, matrices	agents, goals, protocols
Output	centrality, communities, diffusion	solution, decomposition, transformation	joint behavior, equilibrium
Success measure	modularity, interpretability	accuracy, numerical stability	social welfare, stability
Math load	medium: graphs, matrices	the theory itself	medium: game theory
Data needs	network data	none	little; simulation

- **Linear Algebra:** Linear algebra computes the eigenvalues of the adjacency matrix; this course reads them as communities and importance.
- **Multi-Agent Systems:** Multi-agent systems model how agents make decisions; this course analyzes the observed network of ties.

PRACTICE

Applications

Social media
Influencers and echo chambers.

Epidemiology
Disease spread on contact networks.

Fraud
Rings of linked fake accounts.

Science networks
Citation and co-authorship analysis.

Workplace networks
How knowledge flows in a company.

Friend suggestions
Link prediction from mutual friends.

PRACTICE

Risks and Ethics

Inferring what was never shared

A friendship network can reveal political views or sexual orientation that a person never disclosed.

Removing names is not enough

Even with names removed, the pattern of links can identify a person, as structural re-identification attacks showed.

Consent and scraping

In the Cambridge Analytica case, a personality quiz app also collected data on its users' friends without their consent.

Missing ties, wrong answers

Ties forgotten in a survey or invisible on a platform can badly distort centrality rankings and communities.

In depth: Explainable AI and AI Ethics.

PRACTICE

Tools and Datasets

Tools and libraries

NetworkX | igraph | Gephi | graph-tool | Neo4j | PyTorch Geometric

Well-known datasets

Zachary's karate club | Les Misérables character network | Enron e-mails | SNAP network collection | DBLP co-authorship network | Facebook ego networks (SNAP)

PRACTICE

Common Misconceptions

Common belief	In fact
The person with the most friends is the most influential.	Degree is one measure; a bridge joining two groups (betweenness) can matter more.
Looking at a network drawing is enough to understand it.	Large networks draw as a "hairball"; numerical measures and tests are needed.
Community detection has one right answer.	Different methods give different partitions; results are read with domain knowledge.
Connected people have influenced each other.	Similar people choose each other anyway (homophily); influence is hard to separate from selection.

NEXT STEPS

After This Course

Courses that build on or extend this one

This course is mostly an end point: it builds on the others.

Related careers

Network scientist | Data scientist | Social media analyst | Fraud analyst | Computational social scientist

After this course you can move on to learning on networks with graph neural networks, link-based ranking with information retrieval, and diffusion models with epidemiology. Computational social science is an interdisciplinary field that applies these methods to questions about society.

NEXT STEPS

Further Reading

- Wasserman and Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, 1994.
- Easley and Kleinberg, *Networks, Crowds, and Markets*, Cambridge University Press, 2010.
- Newman, *Networks*, 2nd ed., Oxford University Press, 2018.
- Barabási, *Network Science*, Cambridge University Press, 2016.

The End

Chapter 30: Social Network Analysis

← Previous chapter · Map of the field · Next chapter →