Day 4 - Networks

UMN CSS Workshop 2025

Instructor: Alvin Zhou

Learning Goals

- Understand the logic of networks and network analysis
- Learn basic network concepts (node, edge, centrality, closure)
- Explore ways networks are used in social science research
- Distinguish descriptive, predictive, and inferential use of networks
- Review case studies applying network science across disciplines

What is a Network?

- A network is a collection of *nodes* (entities, also called vertices) and *edges* (relationships, also called ties)
  - *Node*: Entity (e.g., person, organization, word)
  - *Edge*: Relationship between nodes (e.g., friendship, citation)
    - Edges can be directed vs. undirected, and weighted (stronger ties = higher weight)
  - *Graph*: set of nodes + edges
- Common examples:
  - Social networks: people connected by friendship
  - Citation networks: papers citing each other
  - Semantic networks: words co-occurring in texts

Adjectives for Networks

Property Type	Description	Example
Directed / Undirected	Are edges one-way or two-way?	Twitter (directed), Facebook (undirected)
Weighted / Unweighted	Do edges have strength/weight?	# of emails sent between people
Static / Dynamic	Does the network change over time?	Temporal networks of contact tracing
Bipartite (Two-mode)	Are there two types of nodes, and edges only run between types?	Users and events, students and classes
Multiplex	Are there multiple types of relationships between nodes?	Friends + coworkers
Signed	Do edges carry positive/negative valence?	Trust/distrust networks
Sparse / Dense	How many edges exist relative to possible?	Most real networks are sparse; thus, we sometimes use network "components" (i.e., maximally connected subgraph)

Network Measures – Centrality

- Degree centrality: # of ties
- Betweenness centrality: How much a node bridges other nodes
- Closeness centrality: how close a node is to others
- Eigenvector centrality: influence via influential neighbors

- Measured per Node

Centrality in Directed or Weighted Networks

- Directed Graphs:
  - In-degree centrality: # of incoming ties (e.g., followers on Twitter)
  - Out-degree centrality: # of outgoing ties (e.g., following on Twitter)
  - Betweenness and Closeness may yield asymmetric values depending on directionality.
- Weighted Graphs:
  - Centralities account for tie strength
  - Example: A node with 2 weak ties (weighted 1 each) is not equal to one with a strong tie (weighted 5).
  - Use weighted degree (we usually call “weighted degree” as “strength”), weighted betweenness, etc.

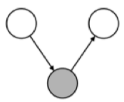
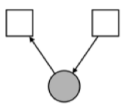
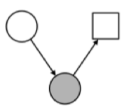
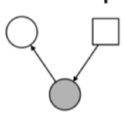
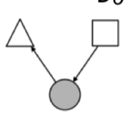
Network Measures – Other Measures

- Local clustering coefficient: Likelihood that a node's neighbors are also connected.
- Constraint (Burt's constraint): How redundant a node's contacts are; lower values suggest brokerage potential.
- Brokerage (Gould-Fernandez Brokerage): Nodes that connect otherwise disconnected communities; you can also calculate scores for different brokerage roles
- K-core: The maximal subgraph in which every vertex has at least degree *K*
- PageRank Centrality: A variant of eigenvector centrality, used by Google's search algorithm.
- Measured per Node

- Structural holes: not a per-node measure, but refers to *gaps* in the network where two alters are not connected; brokers who span these gaps may hold strategic advantage.

Network Measures – Other Measures

Table 1. Movement Spilloverers' Brokerage Roles.

Structural configuration	Brokerage definition (Gould & Fernandez, 1989)	Definition in movement spillover
Coordinator w_I 	Intermediary between two in-group members. $w_{I_j} = \sum_{i}^{N} \sum_{k}^{N} W_I\ (ik)$ where $w_I(ik)$ equals 1 if ijk is true and they have the same membership.	Coordinator served as brokers between two movement participants and then spilled over from the original movement to a subsequent movement.
Itinerant w_O 	Intermediary between two out-group members that have the same membership. $w_{O_j} = \sum_{i}^{N} \sum_{k}^{N} W_O\ (ik)$ where $w_O(ik)$ equals 1 if ijk is true and j has a different membership from i and k.	Itinerants spilled over to a subsequent movement and served as brokers between two participants in the subsequent movement.
Representative b_{IO} 	Intermediary from in-group to out-group members. $b_{IO_j} = \sum_{i}^{N} \sum_{k}^{N} b_{IO}\ (ik)$ where $b_{IO}(ik)$ equals 1 if ijk is true and k has a different membership from i and j.	Representatives received ties from participants in the original movement and then initiated ties with new contacts in a subsequent movement.
Gatekeeper b_{OI} 	Intermediary that mediates relationships from out-group to in-group members. $b_{OI_j} = \sum_{i}^{N} \sum_{k}^{N} b_{OI}\ (ik)$ where $b_{OI}(ik)$ equals 1 if ijk is true and i has a different membership from j and k.	Gatekeepers initiated ties to participants in the original movement and received ties from new contacts in a subsequent movement.
Liaison b_O 	Intermediary between two out-group members that have different memberships. $b_{O_j} = \sum_{i}^{N} \sum_{k}^{N} b_O\ (ik)$ where $b_O(ik)$ equals 1 if ijk is true and i, j, k all have different memberships.	Liaisons served as brokers between two participants from their respective movements. Our analyses, focusing on the spillover between two movements, did not consider this brokerage role.

Note. Movement spilloverers are marked with gray color. Shapes indicate group memberships. Those who spillovered from Movement 1 to Movement 2 are marked with Movement 1 membership.

Network Measures – Other Measures

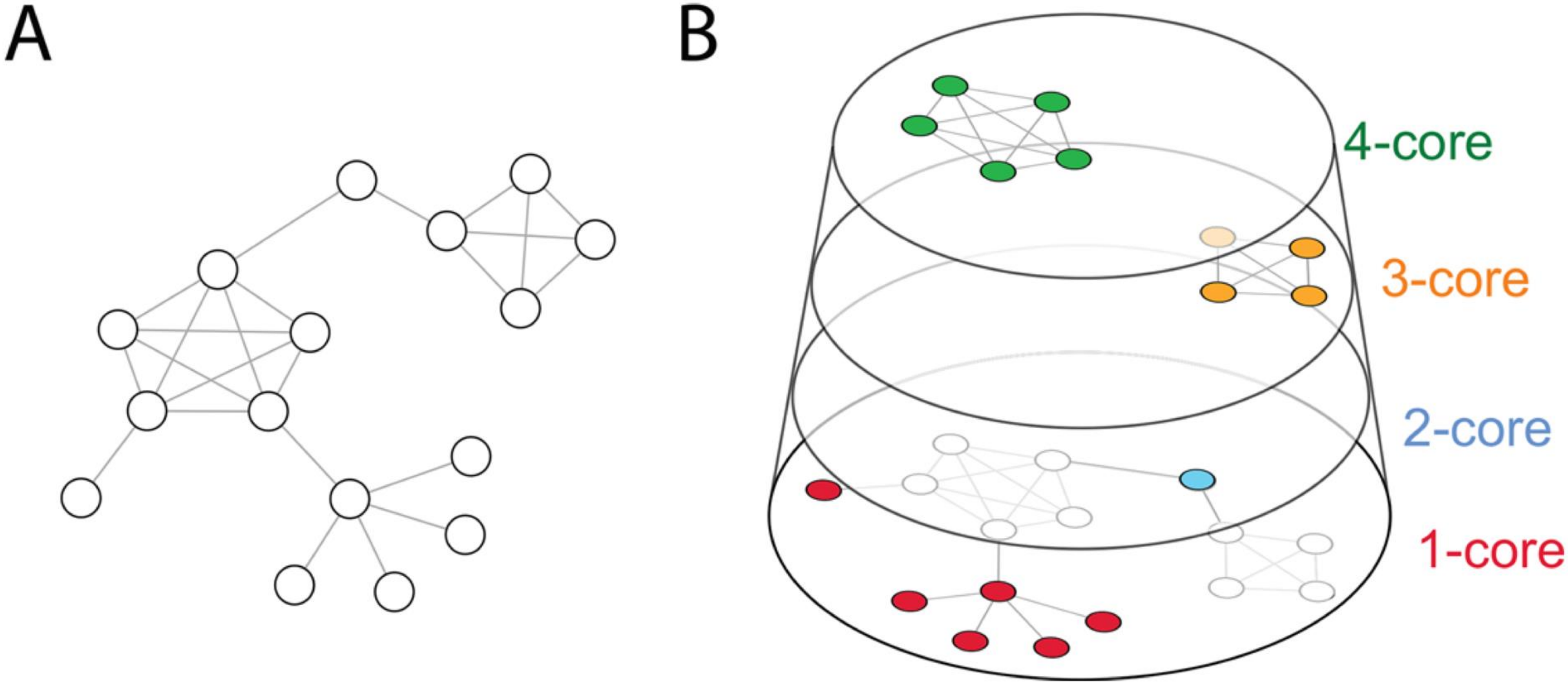


Fig 2. Schematic representation of the *k*-core decomposition for a random network with N = 16 vertices and E = 24 edges. This technique recursively prunes the network to remove nodes with the lowest degree. The coreness of a vertex is *k* if it belongs to the *k*-core but not to the (*k*+1)-core.

Network Measures – Calculation

```
A—B—C
|    |
D    E
```

- Calculate the degree for all nodes.
- Who has the highest betweenness and closeness?

- Open R script <Network_Demo_1.R> and run it ☺

Network Measures – Calculation

```
A—B—C
|   |
D   E
```

- Calculate the degree for all nodes.
- Who has the highest betweenness and closeness?

- Now, imagine the edges are weighted: A–B = 5, B–C = 1. How does that change the measures and your intuition?
- Open R script <Network_Demo_2.R> and run it ☺

Network Measures – Measured per Graph

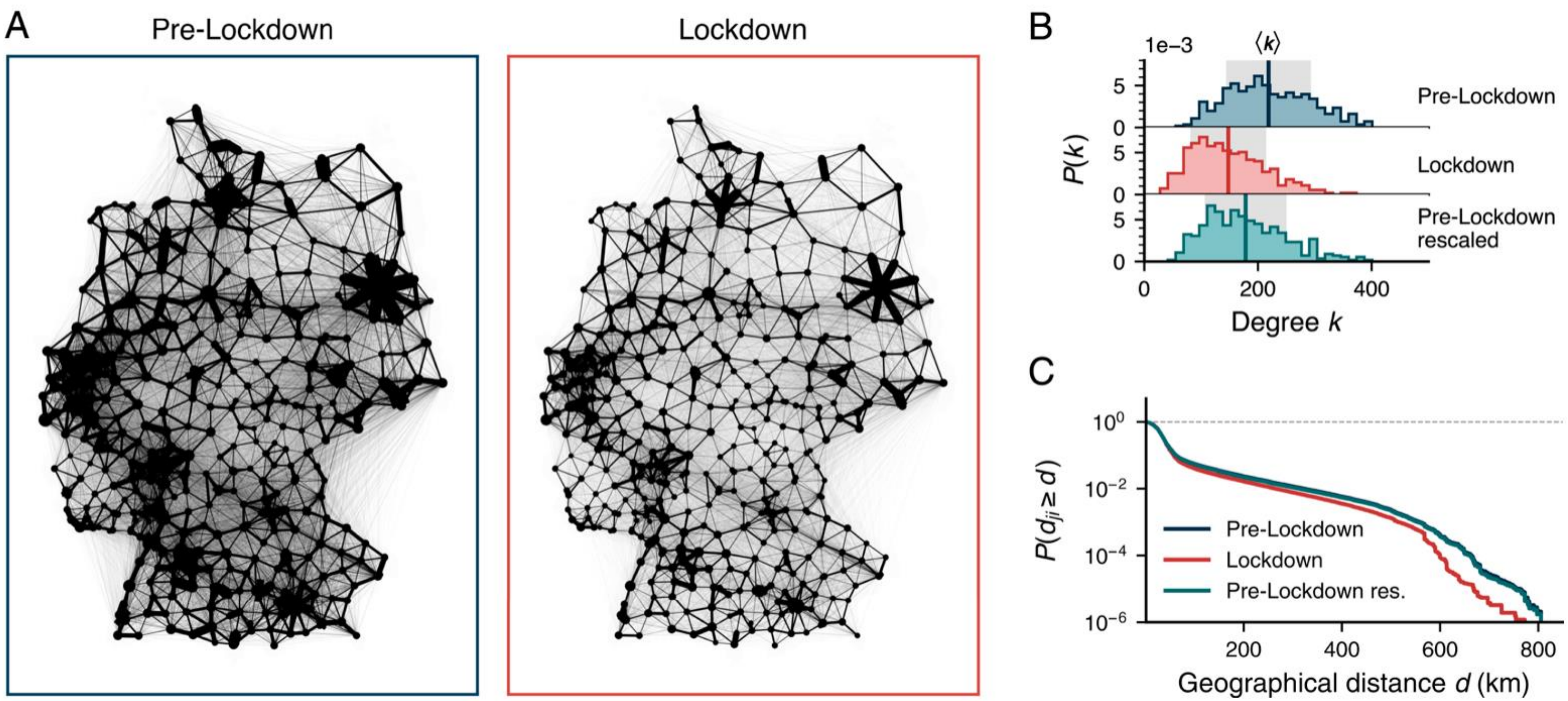
- Density
- Average path length
- Transitivity (a.k.a. clustering coefficient)
  - There are many types of transitivity
  - What we calculated per node earlier is called “local transitivity / local clustering coefficient.”
  - We can also measure the clustering coefficient for the whole graph
- Modularity (community strength)

- Open R script <Network_Demo_3.R> with mock human friendship data and run it ☺
- The script also contains other measures and visualization code

Presentation: Schlosser et al. (2020)

Presentation: Schlosser et al. (2020)

Fig. 3.



How to use networks in SoSci research

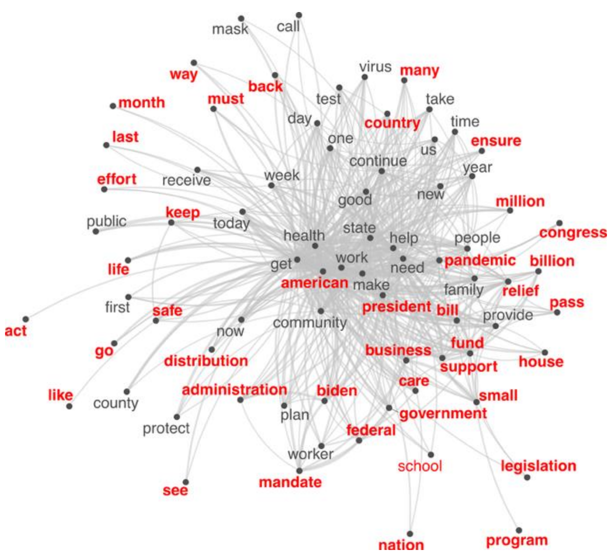
- Descriptive use
  - Semantic networks
  - Network visualization
- Predictive use (network as IVs)
  - Centrality/brokerage/… serve as IVs in regressions
- Inferential modeling (network as DV)
  - ERGM and its variants
  - Network simulation and other generative network models
- Networks as embedded structures for inquiry
  - Network experiment

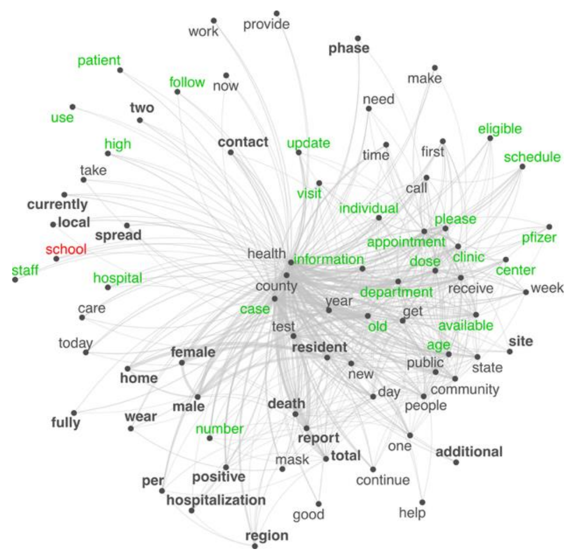
How to use networks in SoSci research

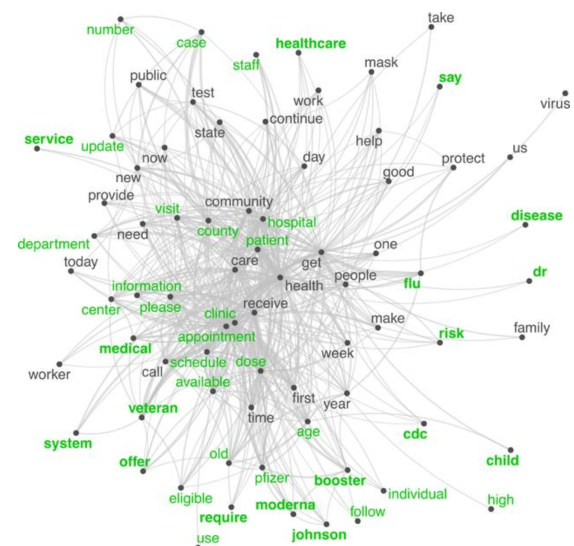
- Descriptive use
  - Semantic networks
  - Network visualization
- Predictive use (network as IVs)
  - Centrality/brokerage/… serve as IVs in regressions
- Inferential modeling (network as DV)
  - ERGM and its variants
  - Network simulation and other generative network models
- Networks as embedded structures for inquiry
  - Network experiment

Semantic Networks

- Words/terms as nodes
- Co-appearance as ties – (could be) weighted, undirected
- Easily represent textual information

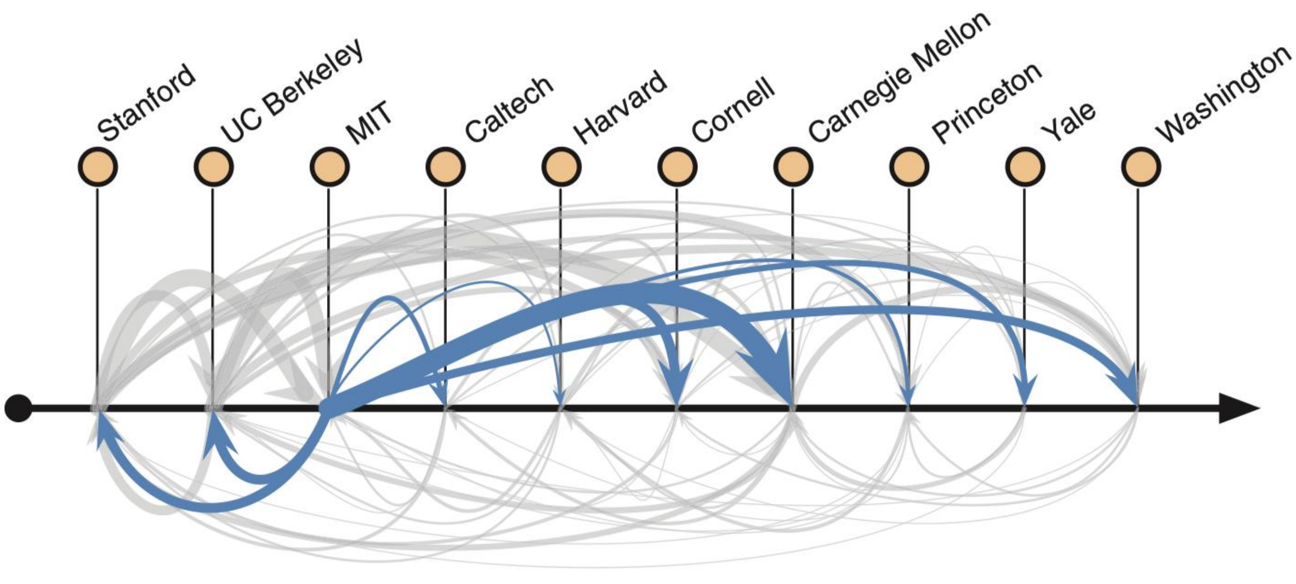






Network Visualization

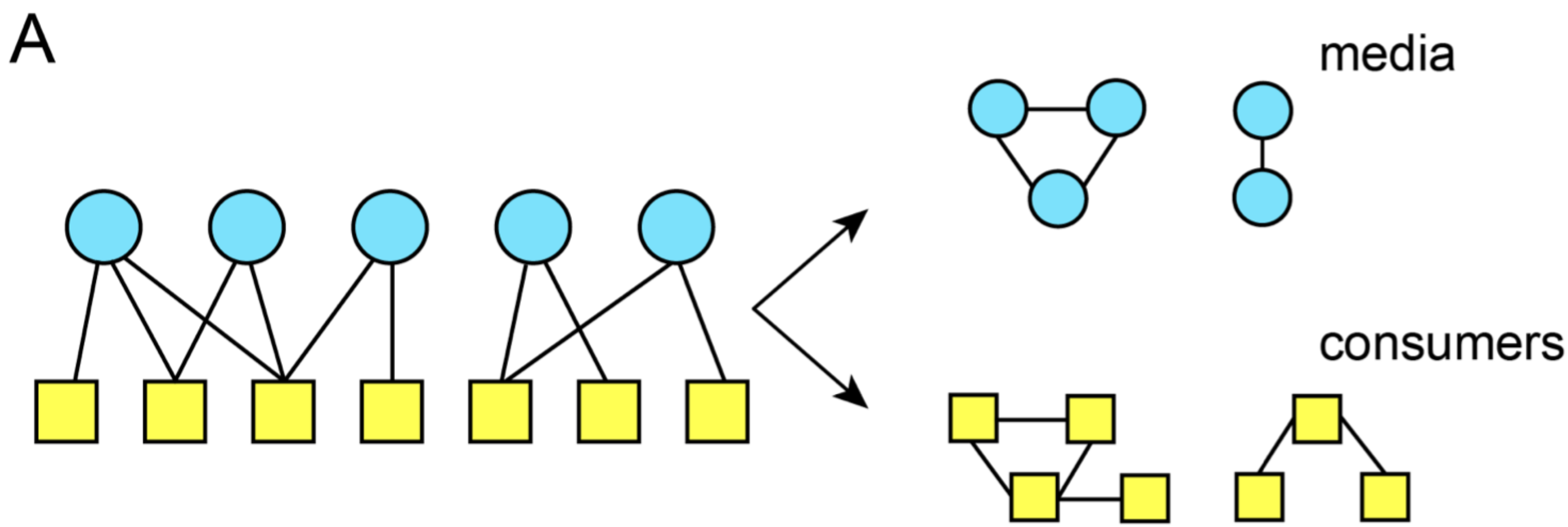
- Convey arguments on networking phenomena.
- Hiring networks among prestigious CS PhD programs.
  - Clauset, A., Arbesman, S., & Larremore, D. B. (2015). Systematic inequality and hierarchy in faculty hiring networks. *Science Advances, 1*(1), e1400005.



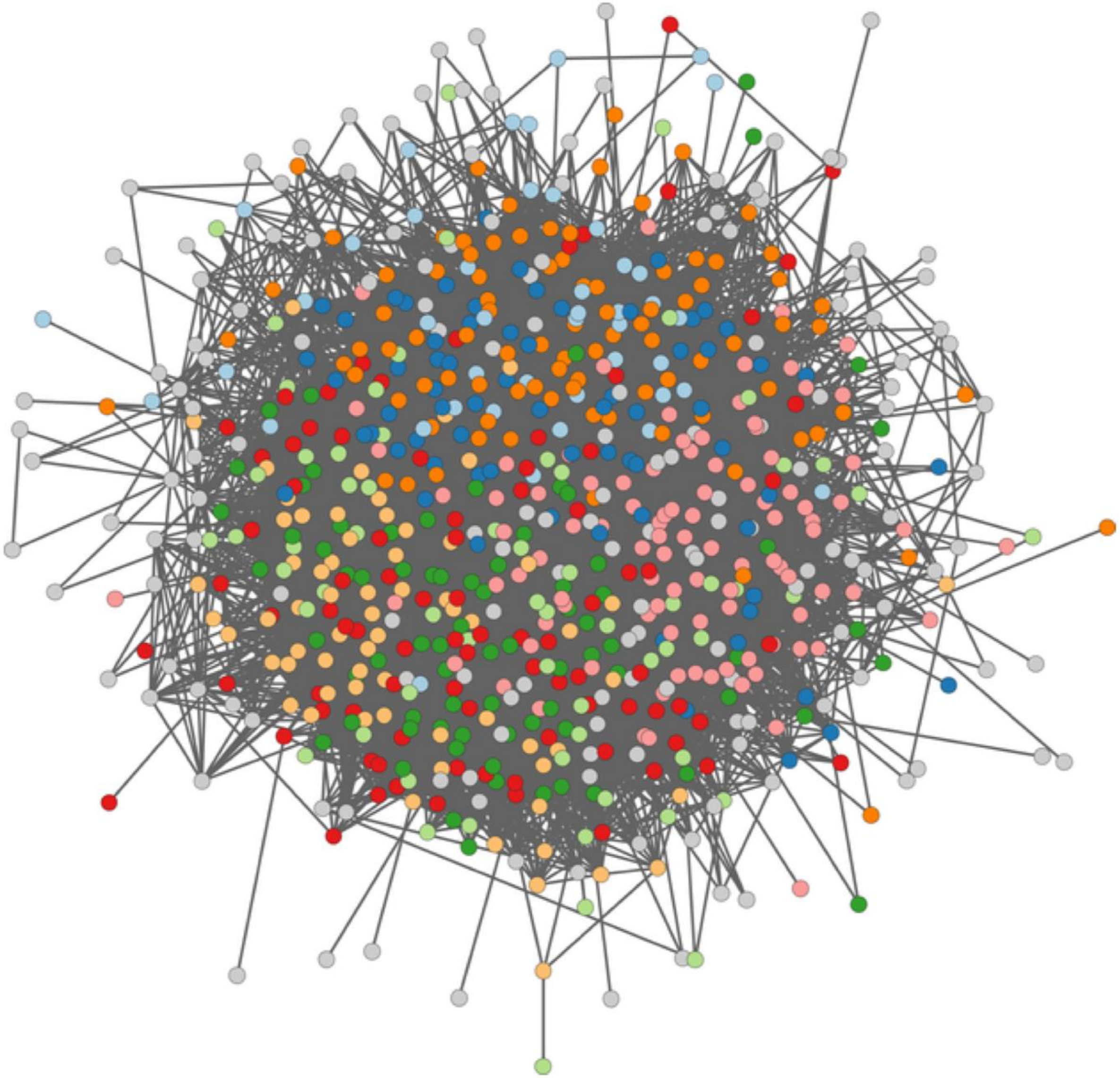
Network Visualization

- Convey arguments on networking phenomena.
- Dual-projection networks.
  - Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018). Networks of audience overlap in the consumption of digital news. *Journal of Communication, 68*(1), 26–50.

Figure 1. The Construction of Audience Overlap Networks



Network Visualization

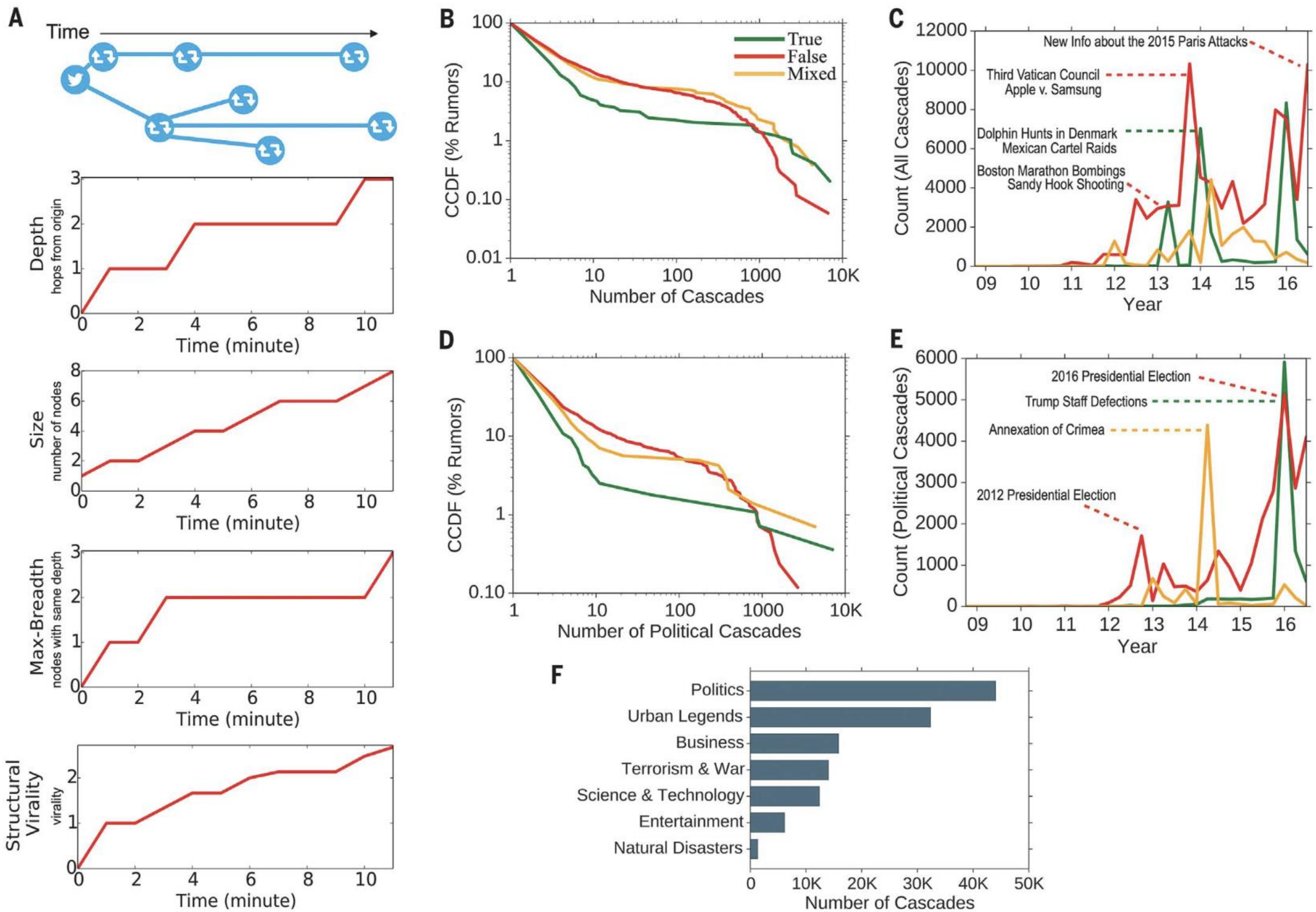


Network Visualization

- Convey arguments on networking phenomena.
- **Avoid hairball visualizations that convey little information.**
  - Venturini, T., Jacomy, M., & Jensen, P. (2021). What do we see when we look at networks: Visual network analysis, relational ambiguity, and force-directed layouts. *Big Data & Society, 8*(1), 205395172110184.
  - Nocaj, A., Ortmann, M., & Brandes, U. (2015). Untangling the hairballs of multi-centered, small-world online social media networks. *Journal of Graph Algorithms and Applications, 19*(2), 595–618.
- Use **filtering** techniques (e.g., backbone extraction) / color / size / force-directed layouts, etc

Presentation: Vosoughi et al. (2018)

Presentation: Vosoughi et al. (2018)



How to use networks in SoSci research

- Descriptive use
  - Semantic networks
  - Network visualization
- Predictive use (network as IVs)
  - Centrality/brokerage/… serve as IVs in regressions
- Inferential modeling (network as DV)
  - ERGM and its variants
  - Network simulation and other generative network models
- Networks as embedded structures for inquiry
  - Network experiment

Network measures as variables

- Network position predicts / associates with future outcomes.
- Centrality, brokerage, and other network-based metrics act as IVs in regression, panel models, or other methods.
- You can also compare groups' network measures to tell a story

Network measures as variables

- Yang, A., & Saffer, A. J. (2021). Standing out in a networked communication context: Toward a network contingency model of public attention. *New Media & Society, 23*(10), 2902–2925.

- **Context:** Public attention to organizations during protests.
- **Network IVs:**
  - **Network constraint** (Burt): Measures redundancy of connections (low = brokerage).
  - **Centrality Measures**
- **DV:** Volume of public engagement.
- **Method:** Regression Trees – How IVs help predict DV
- **Finding:** Less constrained (i.e., more broker-like) accounts received more attention.

Network measures as variables

Table 2. Factors that influence the frequency of mentions in a 2016 refugee discussion network.

Root

Split	Change in DV	# of cases at the point	Deviance
/	0.23	21,500	2007.99

Average neighbor in-degree

Split	Change in DV	# of cases at the pc
<0.01	0.18	18,
⩾0.01	0.54	31

Number of followers

Split	Change in DV	# of cases at the point	Deviance
<13,848.5	.07	11872	312.84
⩾13,848.5	.40	966	139.36

Structural hole effective network size

Split	Change in DV	# of cases at the pc
⩾1.0	0.63	3,15
<1.03	0.37*	20
<1.5	0.34*	48
⩾1.5	0.70 *	55
<5.40	0.44	28
⩾5.40	1.29	53

Structural hole constraint

Split	Change in DV	# of cases at the point	Deviance
<0.19	1.11 *	58	11.18
⩾0.19	.35*	908	96.56

In-degree centrality

Split	Change in DV	# of cases at the pc
<1.24e–6	1.039 *	23
⩾1.24e-6	1.777*	11

$M = 0.23$, complexity param$= 0.17$, $N = 21{,}500$ observations.
Nodes indicated by * are terminal or lead nodes in the decision tree, which means that the results are statistically significant.

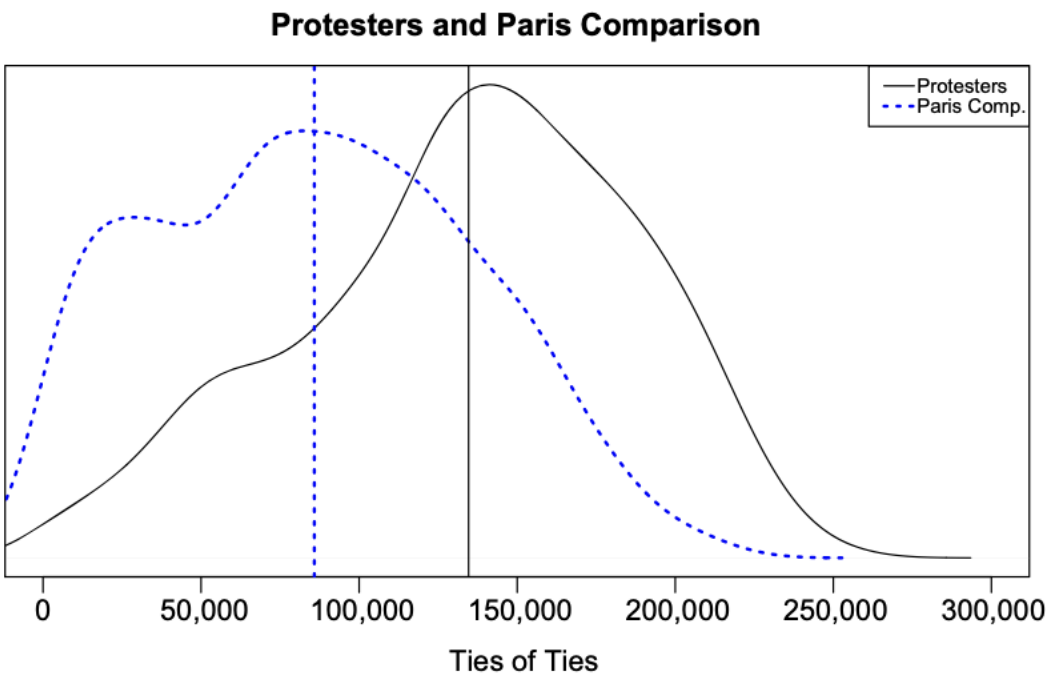
Network measures as variables

- Larson, J. M., Nagler, J., Ronen, J., & Tucker, J. A. (2019). Social networks and protest participation: Evidence from 130 million Twitter users. *American Journal of Political Science, 63*(3), 690–705.

- **Context:** Protest participation during political events on Twitter.
- **Pair Comparison:**
  - **Participants**: Those who were confirmed in attendance at the protest
  - **Potential Participants**: Those who could have participated but did not
- **DV:** Network measures (e.g., ego-network density; tie strength)
- **Finding:** Users central to protest communities were more likely to participate.

Network measures as variables

FIGURE 2 Distribution of the Number of Ties of Ties per User in the Set of Protesters and the Paris Comparison Set



Note: Vertical lines indicate the distributions' means.

Network measures as variables

- Network position predicts / associates with future outcomes.

- IVs = Position, Brokerage, Exposure
- Can be used in panel, event-history, standard regression, or other kinds of advanced methods.
- Network provides **added explanatory power** beyond demographics or content.

How to use networks in SoSci research

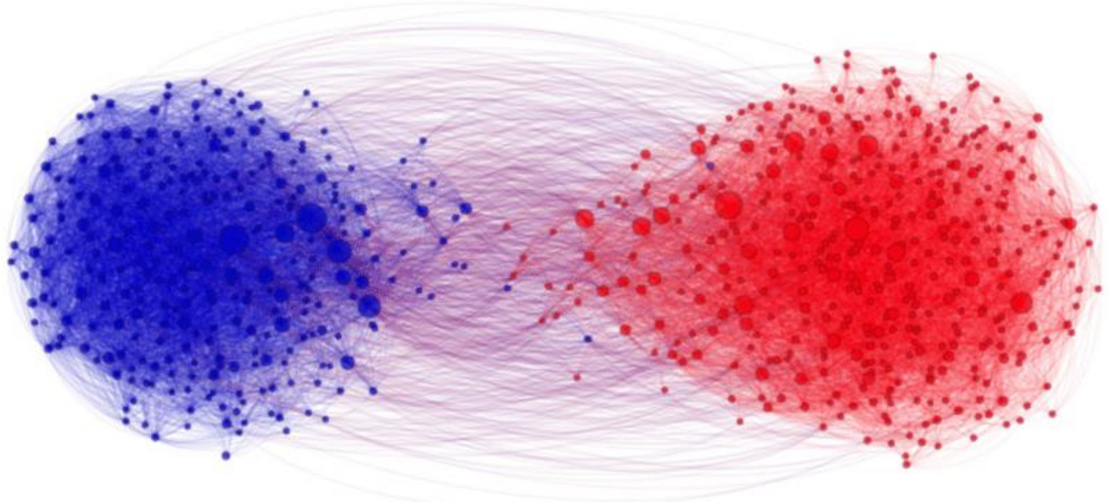
- Descriptive use
  - Semantic networks
  - Network visualization
- Predictive use (network as IVs)
  - Centrality/brokerage/… serve as IVs in regressions
- Inferential modeling (network as DV)
  - ERGM and its variants
  - Network simulation and other generative network models
- Networks as embedded structures for inquiry
  - Network experiment

Network as DV to be predicted

- We observed a network, then we ask:
  - What nodal attributes, between-node forces, and network tendencies might have existed that contributed to the observed network?

Network as DV to be predicted

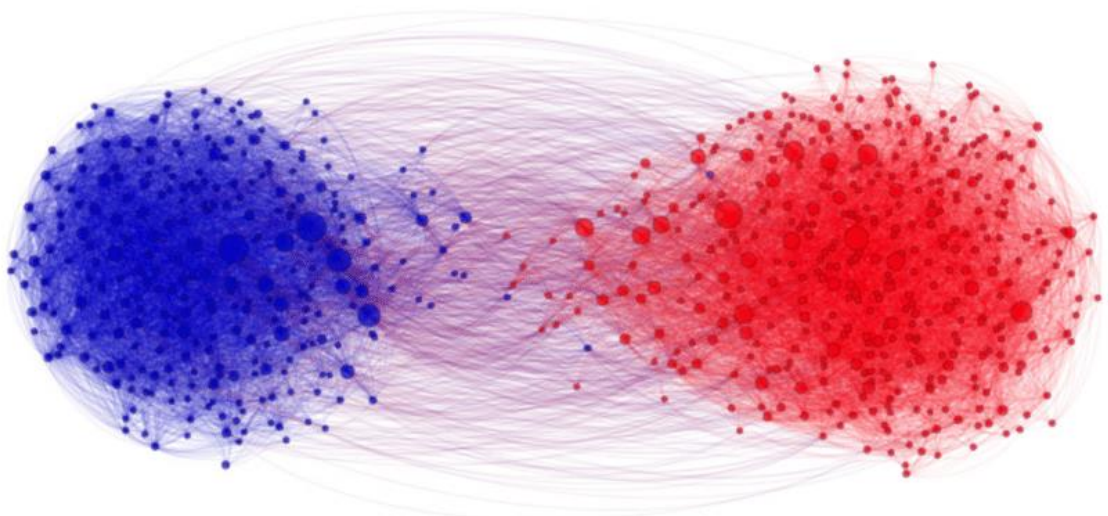
- We observed a network, then we ask:
  - What nodal attributes, between-node forces, and network tendencies might have existed that contributed to the observed network?
- For example:
  - See this network as a DV
  - What predictor might have been significant in its emergence?



Network as DV to be predicted

- We observed a network, then we ask:
  - What nodal attributes, between-node forces, and network tendencies might have existed that contributed to the observed network?

- How can we statistically show this network comes from political polarization (Democrats and Republicans don't connect with each other) though?



Network as DV to be predicted

- Inferential Network Analysis:
  - Statistical models that treat network ties as outcomes and aim to infer the mechanisms behind observed network structure; similar to logistic regression, it predicts the probability that a pair of nodes in a network will have a tie between them, with a set of IVs
  - Latent Space Models – infer hidden proximity that explains ties
  - TERGM, SAOM (e.g., SIENA) – extend to temporal dynamics
  - ERGMs

- ERGMs (can be applied to directed, undirected, valued, unvalued, and bipartite networks, resulting in numerous variants):
  - Control variables: edges (constant term, i.e., probability of tie formation overall)
  - Nodal attributes: organization's industry, person's political affiliation, etc.
  - Homophily tendencies: same gender, age difference, same hashtags, <u>**same party affiliation**</u>, <u>**different party affiliation**</u>, etc.
  - Structural parameters: reciprocity, transitivity, i.e., 2-star, 3-star, triangle, etc.
  - DV: 1/0 (whether or not a tie exists)

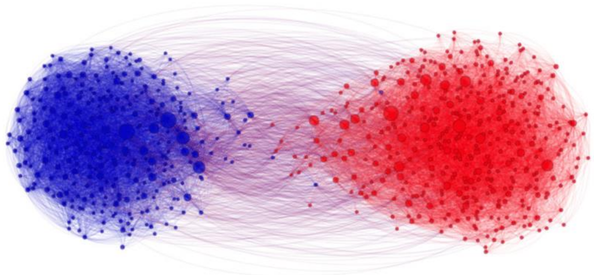
Network as DV to be predicted

- Inferential Network Analysis
  - Suffer from model specification complexity similar to regressions
    - Which nodal covariates, dyadic covariates, and structural parameters to include
    - Whether to model homophily, reciprocity, transitivity, shared popularity, etc.
    - Like regressions, overfitting, underfitting, and omitted variable bias are risks
    - With TERGM/SAOM, you now also deal with temporal dynamics, which means even more decisions (e.g., memory terms, time lags)
  - Computationally heavy
    - All of these models rely on simulation-based estimation techniques like MCMC or GMM, which are slow and can be unstable
    - Can't apply to enormous networks, because you are modeling *N*(N-1)/2* ties
    - Your PC will crash with CPU/Memory problems, and in the end, you may receive a "models do not converge" warning.

Network as DV to be predicted

- Beyond ERGMs – Other Ways to Explain Network Patterns
  - Simulation-based Hypothesis Testing
    - Examples: Permutation tests, edge rewiring, attribute shuffling
    - Null: What would the network look like if attributes/ties were random?
    - Based on the observed network, not model-driven
    - You keep the structure fixed, but shuffle things around
  - Theory-informed generative models
    - Examples: SBM, Degree-Corrected SBM, Barabási–Albert, Latent Space Model
    - Null: What structure would have emerged from these mathematical models?
    - Model-driven, often probabilistic
    - You generate networks from scratch (or from parameterized theory)
- In other words: ERGMs *infer*; simulations and generative models ask: *Could this structure have emerged from this process?*

Network as DV to be predicted



- Simulation-based Hypothesis Testing (permutation test)
  - Shuffle party labels to generate 100 new networks and calculate modularity each time.
  - How extreme is the observed modularity compared to the 100 networks?
  - **Keep structure, randomize labels → "Is party relevant?"**
- Theory-informed generative models (stochastic block models)
  - Assign nodes to latent groups (e.g., Democrat, Republican)
  - Specify high within-group tie probabilities, low between-group tie probabilities
  - Generate 100 simulated networks under this group-based structure
  - Do they show similar modularity, assortativity, or community structure?
  - **Keep labels, simulate structure → "Is group-based tie formation sufficient to (re)produce this structure?"**

Common Generative Network Models

Category	Model	Key Idea	Characteristics	Famous Example
1. Random Graphs	**Erdős–Rényi (ER)**	Randomly connect node pairs	Poisson degree distribution, few hubs	Baseline for randomness
	Configuration Model	Fix degree sequence, randomly connect stubs	Preserves degree distribution	Null model with preserved structure
2. Small-World & Clustering Models	**Watts–Strogatz (WS)**	Start with a ring lattice, rewire with probability	High clustering + short average path length	Human friendship networks
3. Power-Law Models	**Barabási–Albert (BA)**	New nodes attach to high-degree nodes	Scale-free, power-law degree, hubs	Web links, citation networks
4. Community Detection / Group-Based	**Stochastic Block Model (SBM)**	Partition nodes into groups; connect based on group probabilities	Captures community/cluster structure	Political blogosphere
	Degree-Corrected SBM	Like SBM, but adjusts for degree heterogeneity	Realistic with uneven group sizes	Social affiliation networks
5. Statistical Network Models	**Exponential Random Graph Models (ERGM)**	Estimate tie probability based on structural features (e.g., reciprocity, triads)	Flexible and interpretable	Face-to-face friendship networks
	Latent Space Model (LSM)	Nodes exist in a latent space; closer nodes more likely to connect	Captures proximity effects	Online dating or collaboration networks
6. Dynamic/Temporal Models	**Actor-Oriented Models (SIENA)**	Model evolution of ties based on actor-level decisions	Micro-level changes over time	Teen friendship evolution
	Relational Event Models (REM)	Model sequences of interaction events	Time-stamped dyadic data	Email, chat logs

How to use networks in SoSci research

- Descriptive use
  - Semantic networks
  - Network visualization
- Predictive use (network as IVs)
  - Centrality/brokerage/… serve as IVs in regressions
- Inferential modeling (network as DV)
  - ERGM and its variants
  - Network simulation and other generative network models
- Networks as embedded structures for inquiry
  - Network experiment

Networks as Basic Structure for Inquiries

- Not just analyzing the network, we are using it as infrastructure
- We define networks and determine who interacts with whom, when, and how often, and ask how structure affects outcomes
- Examples:
  - Network experiment: Manipulate network structure (e.g., add influencers, remove ties).
    - Control: No Influencer; Treatment: With an influencer who communicates more
    - Control: Observed communication network; Treatment: All long-range tie removed
  - Message experiment within a network: Deliver different messages to central vs peripheral nodes.
    - Same network, deliver message to different nodes to start with
  - Etc.

Networks as Basic Structure for Inquiries

- Centola (2010, Science): Built artificial online health networks
- Randomly assigned participants to:
  - Clustered-lattice networks vs random networks
  - Tracked how health behavior (e.g., signing up for an exercise program) spread
- Found: Clustered networks led to more stable behavioral adoption than random ones.
- Takeaway: Even with the same messages and people, the network structure changed the outcome.

Networks as Basic Structure for Inquiries

- We can also think of a field experiment design that uses existing/observed networks

- Field experiment on Twitter: Inject political messages into different types of nodes:
  - High-degree (popular) vs low-degree users
  - Echo chambers vs bridge nodes
- Outcome: Which messages travel farther? Which reach across party lines?
- Lesson: Network position + content jointly affect political diffusion and polarization.

How to use networks in SoSci research

- Descriptive use
  - Semantic networks
  - Network visualization
- Predictive use (network as IVs)
  - Centrality/brokerage/… serve as IVs in regressions
- Inferential modeling (network as DV)
  - ERGM and its variants
  - Network simulation and other generative network models
- Networks as embedded structures for inquiry
  - Network experiment

Words of Caution

Caution – Interpretation

Metrics are constructed and must be interpreted in context, for example, which centrality measure captures “influencer”?

- Central ≠ Important: A node with high degree might be popular, but not influential in diffusion
- Unit of analysis matters: Actor-level stats must align with research questions
- Complementary metrics: Use multiple indicators to triangulate

- Ask if a measure is substantively meaningful

Caution – Network for Network's Sake

- Some papers try to use network analysis to fluff up papers methodologically, but in fact, there is no need for network at all
  - E.g., a network of text messages between people
    - Nodes as persons; weighted ties as messages
    - The indegree of a (person) node is basically the number of messages they receive. Do you need a network for that?
  - E.g., seductive hairball visualization while analytically empty

- Use networks only when…
  - Your theory is relational, or
  - You're interested in structure (e.g., brokerage, redundancy), or
  - You're modeling processes (e.g., diffusion, competition, influence)
- Use networks only when you are studying networked phenomena

Network - Rarely Used in Isolation

- Network structure reveals who is connected to whom, but not why, how, or what it means.
- Other methods complement network analysis:
  - Text Analysis / NLP: for uncovering meaning or themes behind the nodes/edges.
  - Statistical Modeling: to predict influence, centrality, diffusion.
  - Visualization: for pattern recognition and communicative clarity.
  - Temporal Models: to understand how networks evolve over time.
- In excellent CSS studies, network analysis is rarely used in isolation --- it usually comes with a strategic combination of other novel CSS methods.

Presentation: Zhao et al. (2022)

Presentation: Zhao et al. (2022)

- Understand the social network and influence of ancient Chinese poets.
- Combined methods:
  - Text mining + Named Entity Recognition: Identify poets and their references to one another.
  - Network Analysis: Build directed influence networks among poets.
  - Network Simulation: Susceptible–Infected (SI) Model
  - Temporal Slicing: Examine changing influence across historical periods.
  - Centrality Measures + Topic Modeling: Identify influential poets and thematic shifts.

Lab Preview

- Short Lecture <What Networks Could Be>

- Semantic network of <presidential_speeches.csv>
- Visualization of semantic networks in R and Gephi

- ERGM for social networks
- Other than ERGM

- **YouTube API**
  - Go to https://developers.google.com/youtube/v3/getting-started
  - Apply for the API
  - We will use it in the next lab