Introduction to Computational Social Science

Session 5: Network data and analysis

Maximilian Haag Constantin Kaplaner

Geschwister-Scholl-Institute for Political Science
LMU Munich

12.12.2022
Room B U103, Tue 14:00–18:00 (bi-weekly)

Today's session

Lecture

1. Basic concepts
2. Data collection and storage
3. Network measures
4. Applications in Political Science / CSS

Lab

1. Network basics in R
2. Twitter friendships of German MPs
3. Visualization

Before we begin: Advent of Code

Advent of Code is an annual set of computer programming challenges that have a Christmas-themed Advent calendar format.

→ It's great for learning a new programming language!

Try it out if you like. Feel free to ask for help!

Derek Sollberger has created tutorial videos of how to solve the 2022 Advent of code in R.

Before we begin: ChatGPT i

ChatGPT

is a chatbot that responds to prompts and instructions developed by OpenAI (Brown et al., 2020; Ouyang et al., 2022).

Before we begin: ChatGPT ii

Features / uses (examples)

- Give explanations
- Write text
- Write code(!)

Technology
Generative transformer based (deep learning) text model (GPT-3.5) combined with Reinforcement Learning from Human Feedback

Limitations

- No independent reasoning
- Massive in size and probably not be free to use for long
- No open source code

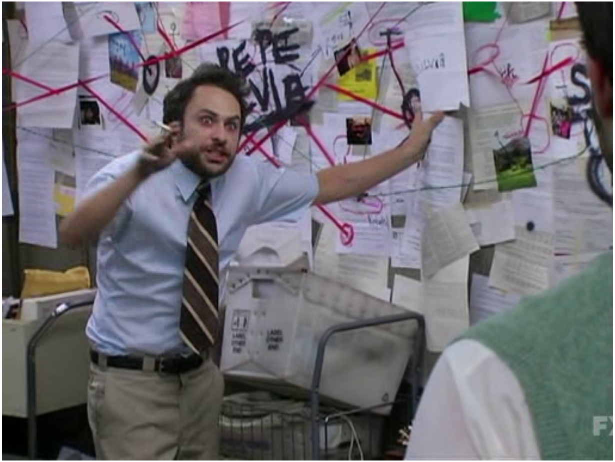
See also https://openai.com/blog/chatgpt/

Introduction & basic concepts

Networks around us i

Source: https://de.m.wikipedia.org/wiki/Datei:Netzplan_U-Bahn_M%C3%BCnchen.svg

Networks around us ii



Source: It's Always Sunny in Philadelphia (Season 4 Episode 10)

Networks around us iii

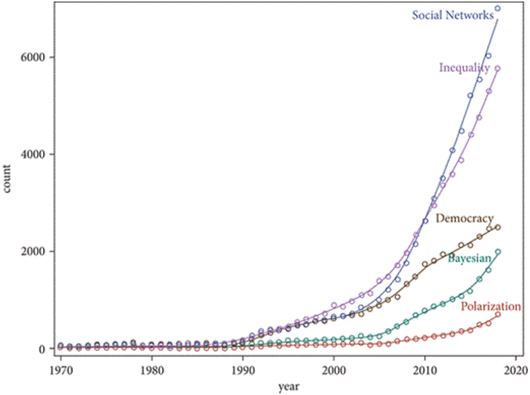
Introduction to networks in the Social Sciences

- Networks have a long tradition in science, mostly in Physics and Biology

Introduction to networks in the Social Sciences

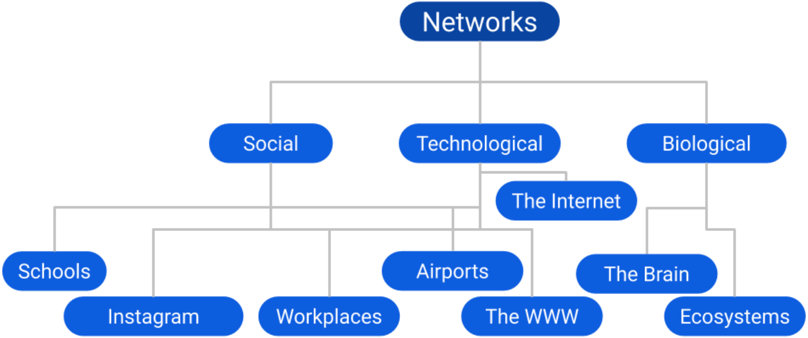
- Networks have a long tradition in science, mostly in Physics and Biology
- In the *Social Sciences*, we are often interested in *social* networks

Shy study networks?



Trend in articles on “Social Network*” topics among all papers indexed in Web of Science Social Science Citation Index, with other keywords for comparison. Source: Light & Moody (2021)

Example overview of types of networked systems



Source: Jilbert (n.d.)

Research questions and goals of network analysis (Borgatti, 2013)

Consequences of networks → independent variable

Example

- How the position of an actor in a network shapes the actor's power
- How a characteristic of a network influences actor behavior within the network

Research questions and goals of network analysis (Borgatti, 2013)

Consequences of networks → independent variable

Example

- How the position of an actor in a network shapes the actor's power
- How a characteristic of a network influences actor behavior within the network

Prediciton of network characteristics → dependent variable

Example

- How actor preferences influence the actor's position in the network
- How one type of connection in a network (e.g. business ties) can influence the emergence of another type of connection (e.g. friendship)

Example

Six degrees of separation

refers to the the separation of people by social connections
(e.g. friends of friends).

It can be shown that in a random network of friendships between 6 billion people, two people are on average separated by roughly 6 'chained' friendships.

Example

Six degrees of separation

refers to the the separation of people by social connections (e.g. friends of friends).

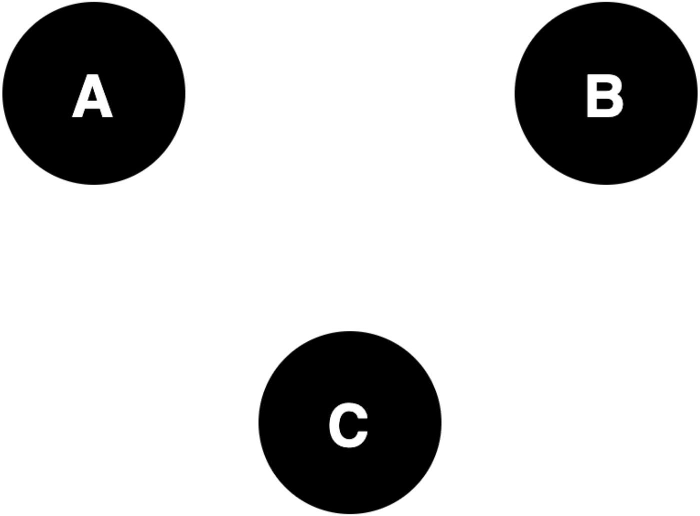
It can be shown that in a random network of friendships between 6 billion people, two people are on average separated by roughly 6 ‘chained’ friendships.

See also

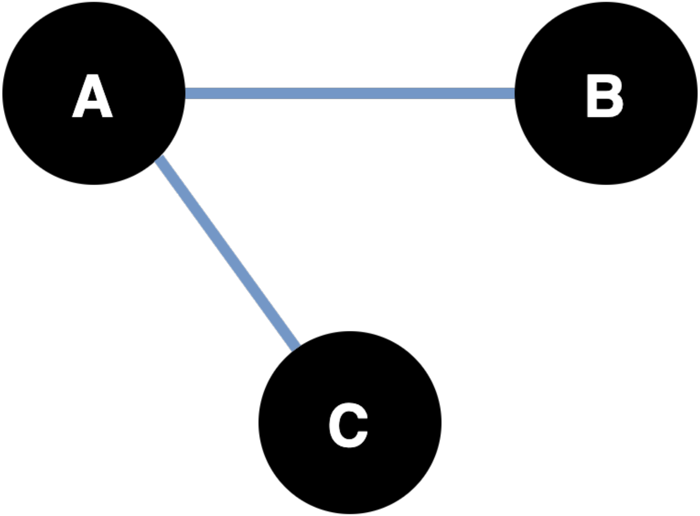
Six degrees of Wikipedia

Six degrees of Kevin Bacon

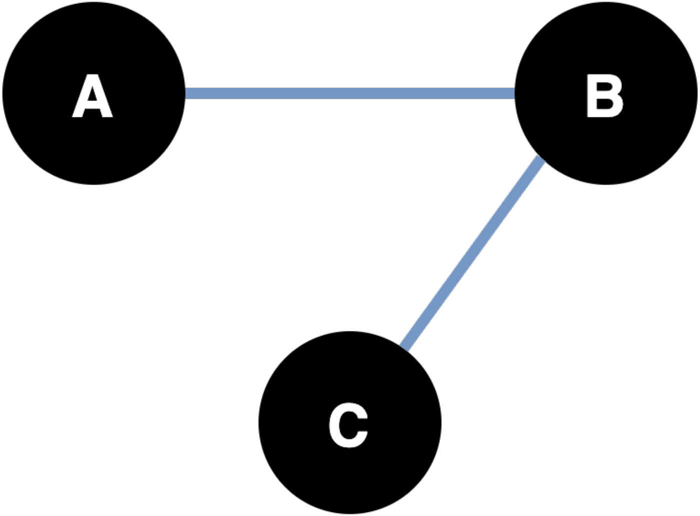
A simple network i



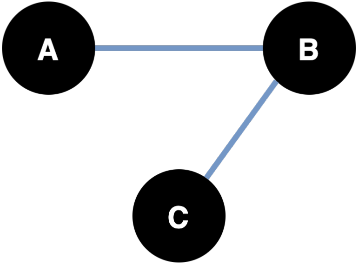
A simple network ii



A simple network iii

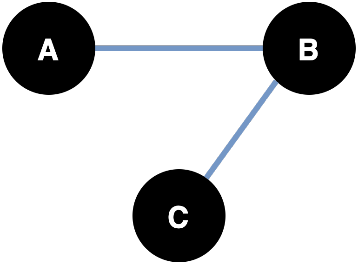


Nodes and edges



Nodes (also vertices) Entities or units in a network

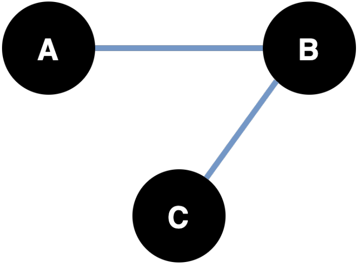
Nodes and edges



Nodes (also vertices) Entities or units in a network

Edges (also ties, links) Connections between nodes

Nodes and edges

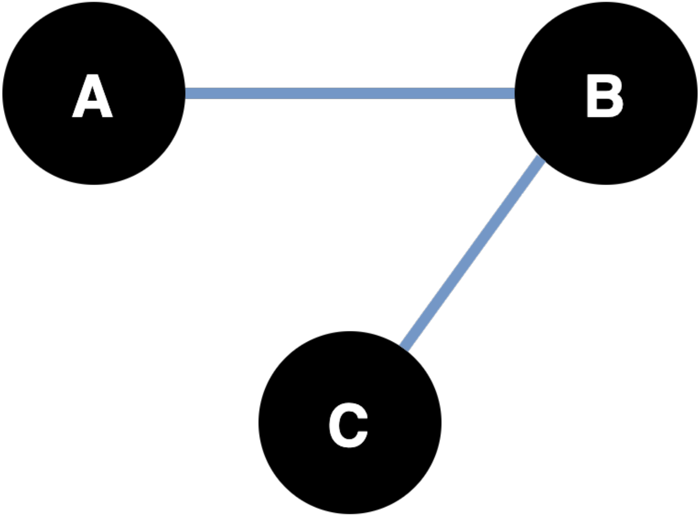


Nodes (also vertices) Entities or units in a network

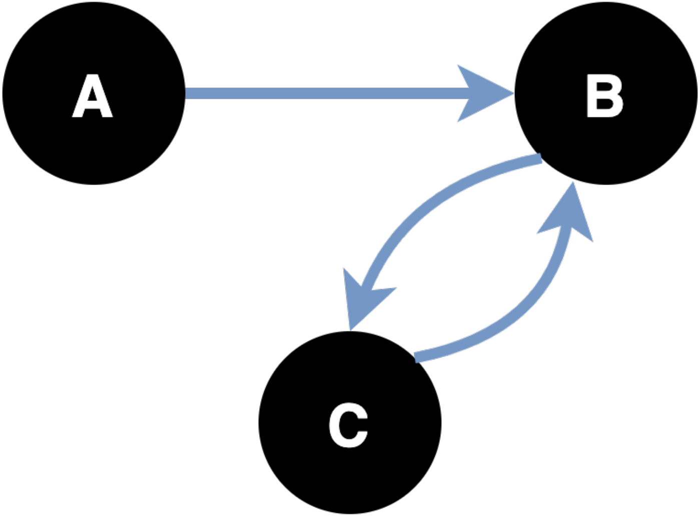
Edges (also ties, links) Connections between nodes

→ **Network**

Undirected edges

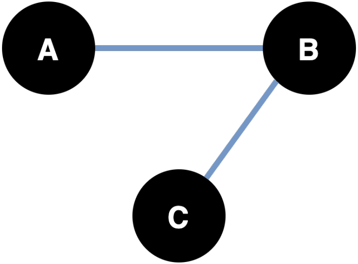


Directed edges



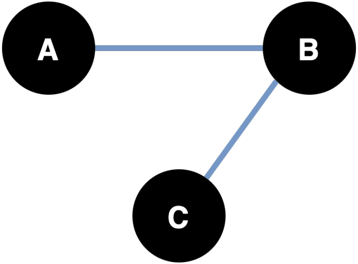
Symmetric and asymmetric relations

Undirected, symmetric relation

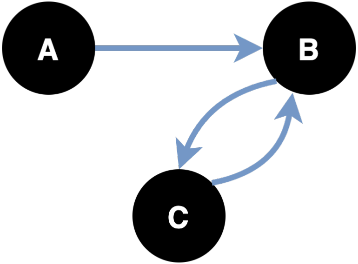


Symmetric and asymmetric relations

Undirected, symmetric relation



Directed, asymmetric relation



Typology of relations between nodes

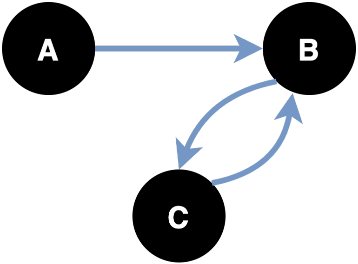
Similarities			Social Relations				Interactions	Flows
Location	Membership	Attribute	Kinship	Other role	Affective	Cognitive	e.g.,	e.g.,
e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	Sex with	Information
Same spatial and temporal space	Same clubs	Same gender	Mother of	Friend of	Likes	Knows	Talked to	Beliefs
	Same events	Same attitude	Sibling of	Boss of	Hates	Knows about	Advice to	Personnel
	etc.	etc.		Student of	etc.	Sees as happy	Helped	Resources
				Competitor of		etc.	Harmed	etc.
							etc.	

A typology of ties studied in social network analysis by Borgatti et al. (2009)

One and two-mode networks

Text messaging among friends

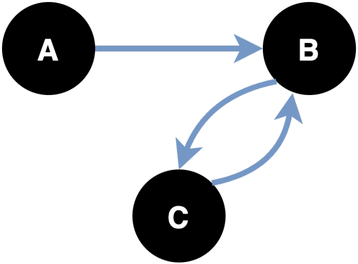
→ one-mode network



One and two-mode networks

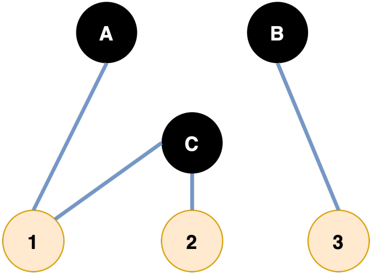
Text messaging among friends

→ one-mode network

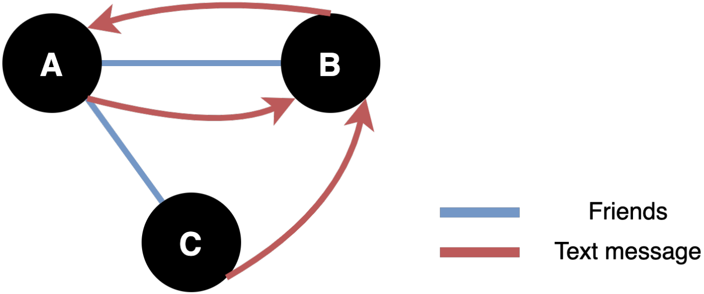


Hobbies among friends

→ two-mode network

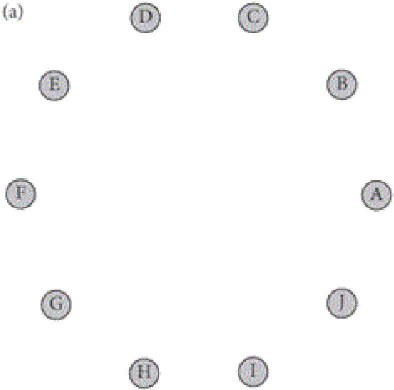


Multiplexity



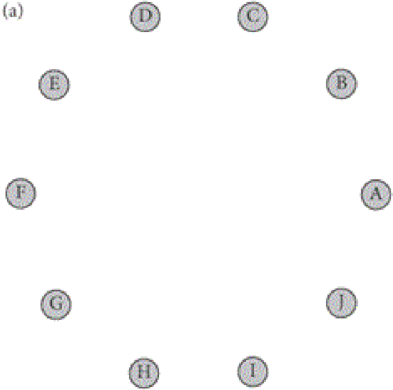
Unconnected and maximally connected networks

Unconnected network

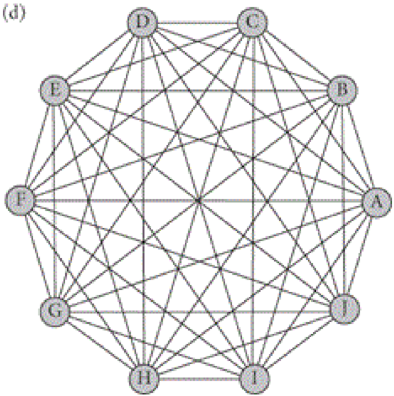


Unconnected and maximally connected networks

Unconnected network



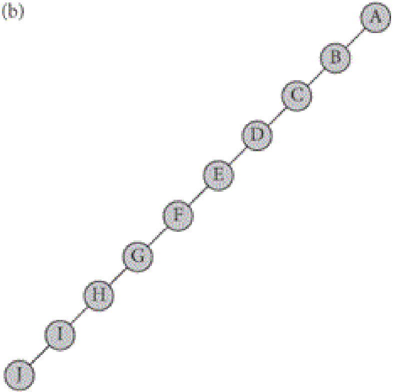
Maximally connected network



Chains and trees

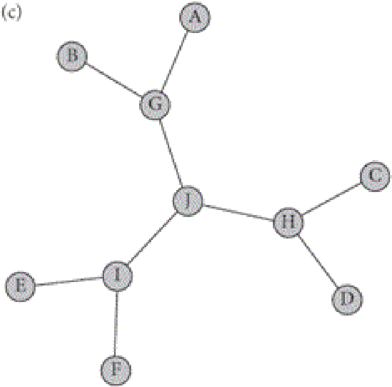
Chain

(b)

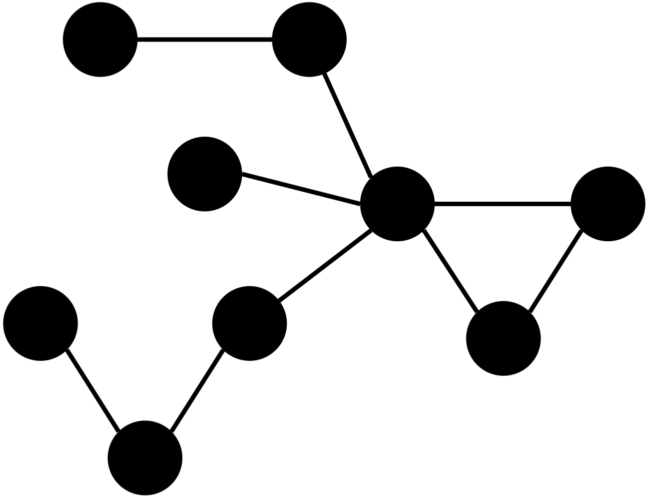


Tree

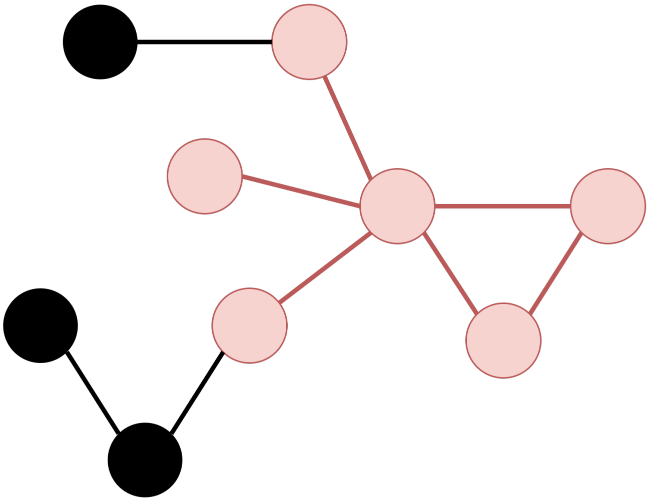
(c)



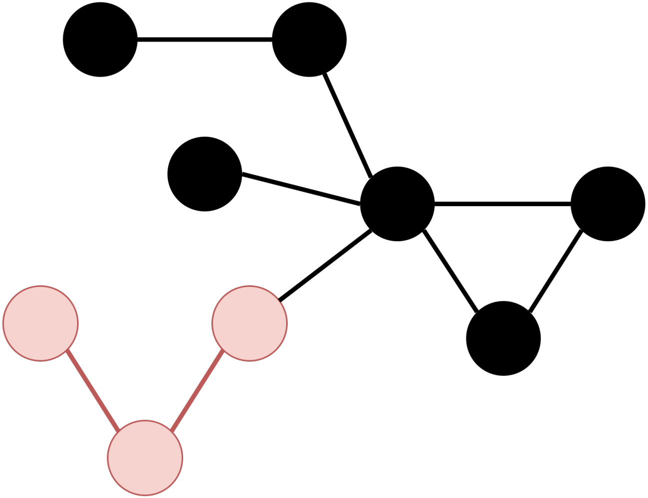
Levels of analysis: network



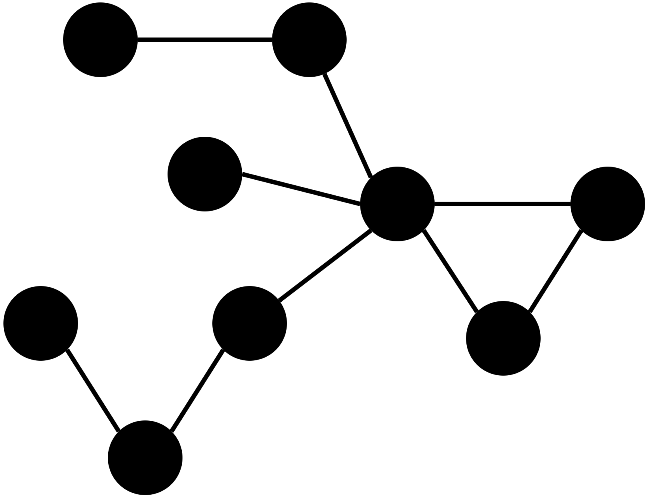
Levels of analysis: subgroup



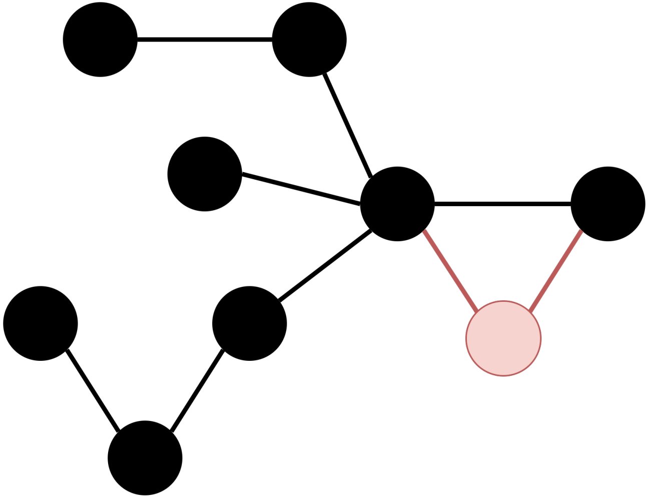
Levels of analysis: triad



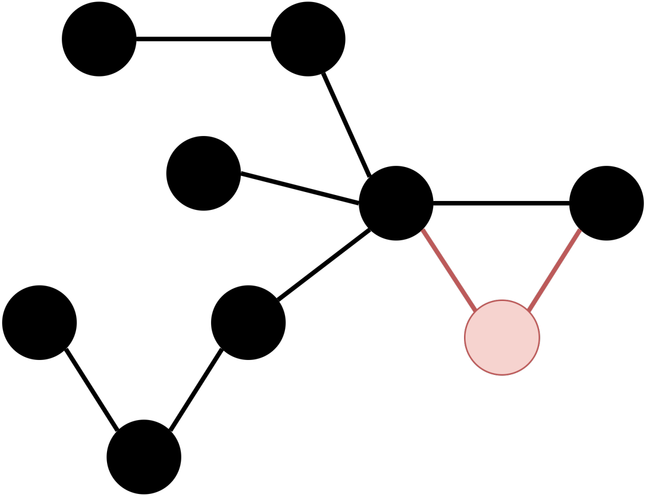
Levels of analysis: dyad



Levels of analysis: ego



Levels of analysis: ego and alter



Network data collection and storage

What do we want to know? (Light & Moody, 2021)

Ties (Borgatti et al., 2009)

- social relationships (e.g. friendship, kinship)
- interactions (e.g. sending and receiving messages, sharing resources)
- flows (spread of, e.g. ideas, diseases) → between nodes connected by relation or interaction ties

What do we want to know? (Light & Moody, 2021)

Ties (Borgatti et al., 2009)

- social relationships (e.g. friendship, kinship)
- interactions (e.g. sending and receiving messages, sharing resources)
- flows (spread of, e.g. ideas, diseases) → between nodes connected by relation or interaction ties

Scope

- Local networks (convenience samples)
- Complete / global networks

What do we want to know? (Light & Moody, 2021)

Ties (Borgatti et al., 2009)

- social relationships (e.g. friendship, kinship)
- interactions (e.g. sending and receiving messages, sharing resources)
- flows (spread of, e.g. ideas, diseases) → between nodes connected by relation or interaction ties

Scope

- Local networks (convenience samples)
- Complete / global networks

Complete networks are often unobtainable → *Sampling*

- Respondent-driven sampling → gain insights in network of hidden populations
- Network scale-up method → estimate the size hidden populations

Boundary specification

Boundary specification problem "*In social networks boundaries can often logically extend to every human on the planet.*" (Light & Moody, 2021)

→ Researchers need to specify the boundaries of the network they want to analyze

How can we collect netowrk data? (Adams et al., 2021)

Collection via

Primary sources

- Survey / Interviews

Secondary sources

- Information on individuals (e.g. archival sources)

Hybrid sources

- Sensor data / digital trace data / metadata

- text-corpora → Text networks

Surveys / Interviews i

In surveys / interviews, we can ask respondents about their connections and relations to other entities. We can

- ask *open* or *closed* format questions

Surveys / Interviews i

In surveys / interviews, we can ask respondents about their connections and relations to other entities. We can

- ask *open* or *closed* format questions
- ask whether or not a relation exists

Surveys / Interviews i

In surveys / interviews, we can ask respondents about their connections and relations to other entities. We can

- ask *open* or *closed* format questions
- ask whether or not a relation exists
- ask about the strength of relations

Surveys / Interviews i

In surveys / interviews, we can ask respondents about their connections and relations to other entities. We can

- ask *open* or *closed* format questions
- ask whether or not a relation exists
- ask about the strength of relations

Surveys / Interviews ii

Name generators

- Example: “*Who do you contact if you get stuck with your work?*”

Surveys / Interviews ii

Name generators

- Example: "*Who do you contact if you get stuck with your work?*"
- General Social Survey: "*From time to time, most people discuss important matters with other people. Looking back over the last six months, who are the people with whom you discussed matters important to you?*" → can be used to identify alters of an ego

Surveys / Interviews ii

Name generators

- Example: "*Who do you contact if you get stuck with your work?*"
- General Social Survey: "*From time to time, most people discuss important matters with other people. Looking back over the last six months, who are the people with whom you discussed matters important to you?*" → can be used to identify alters of an ego

→ The exact name generator is determined by the research question

Surveys / Interviews iii (Adams et al., 2021)

Name interpreters

Ask additional questions on alters:

- attributes of the nominated alters
- details of the relationship
- strength or frequency of relationship

Digital trace data / metadata

Information on networks from

- Monitoring of behavior (e.g. sensor data, app use log)
- Digital trace data (e.g. interaction on social media)
- Metadata (e.g. who cites whom, who follows whom)

Ethics (Adams et al., 2021)

Data collection

- Anonymity / Confidentiality → Problem: anonymized data can be used to deduct identities
- Informed consent
- Active consent / opt-out in population studies

Ethics (Adams et al., 2021)

Data collection

- Anonymity / Confidentiality → Problem: anonymized data can be used to deduct identities
- Informed consent
- Active consent / opt-out in population studies

Analysis

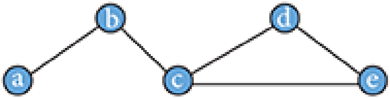
- Network visualization can reveal identities
- Balance between benefit and harm (e.g. infectious disease tracing)
- Network analysis use in business and management to assess performance

How to store network data?

Adjacency matrices

Undirected

(a)



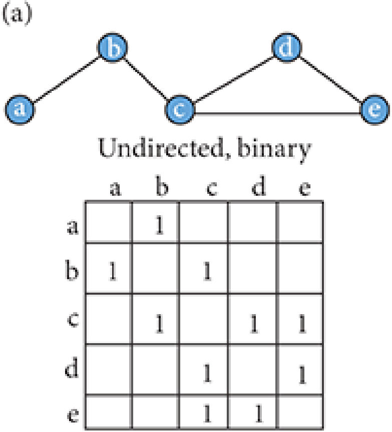
Undirected, binary

	a	b	c	d	e
a		1			
b	1		1		
c		1		1	1
d			1		1
e			1	1	

How to store network data?

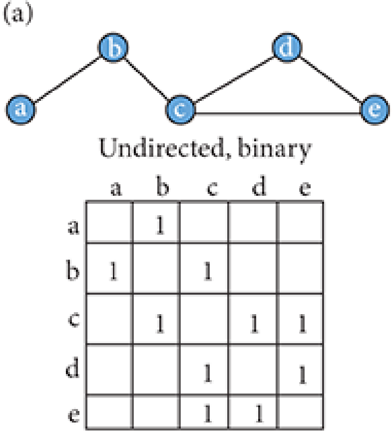
Adjacency matrices

Undirected



	a	b	c	d	e
a		1			
b	1		1		
c		1		1	1
d			1		1
e			1	1	

Directed



	a	b	c	d	e
a		1			
b	1		1		
c		1		1	1
d			1		1
e			1	1	

Source (both): Light & Moody (2021)

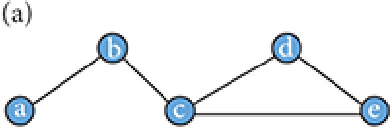
Network measures and models

Social Network Analysis (SNA)

Social Network Analysis (SNA) is the use of graph-theoretic and matrix algebraic techniques to study ***social structure*** *and* ***social relationships****. (Jilbert, n.d.)*

Graphs and matrices

Graph



Matrix

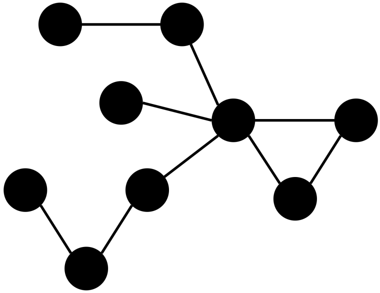
	a	b	c	d	e
a		1			
b	1		1		
c		1		1	1
d			1		1
e			1	1	

→ Graphs and matrices are *mathematical* structures that can be used to model and make inferences about a network.

Describing a graph: order

Order

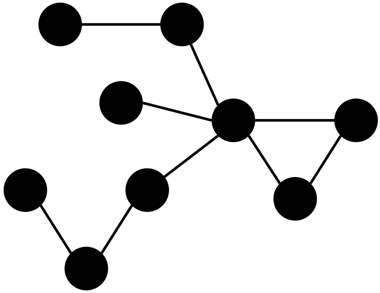
→ the number of nodes in a graph



Describing a graph: order

Order

→ the number of nodes in a graph

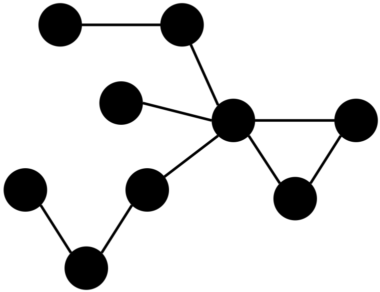


$$n = |V| = 9$$

Describing a graph: density

Density

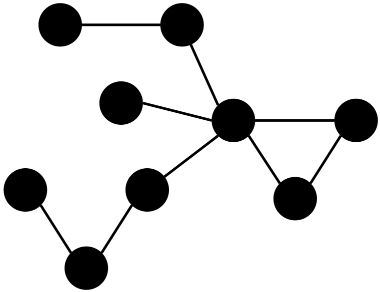
→ the relation of the number of *actual* edges in a graph to the number of *possible* edges in a graph



Describing a graph: density

Density

→ the relation of the number of *actual* edges in a graph to the number of *possible* edges in a graph

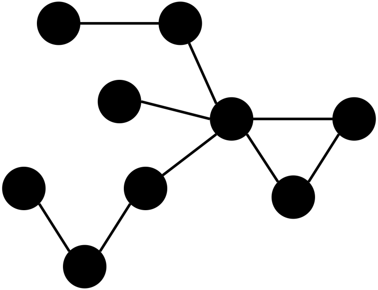


$$n = |V| = 9$$

Describing a graph: density

Density

→ the relation of the number of *actual* edges in a graph to the number of *possible* edges in a graph

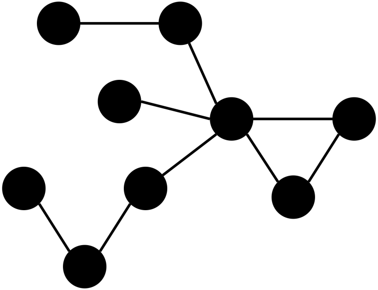


$$n = |V| = 9$$
$$e = |E| = 9$$

Describing a graph: density

Density

→ the relation of the number of *actual* edges in a graph to the number of *possible* edges in a graph

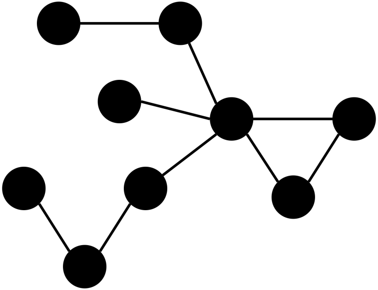


$$n = |V| = 9$$
$$e = |E| = 9$$
$$\frac{e}{n(n-1)/2} = \frac{9}{36} = 0.25$$

Describing a graph: density

Density

→ the relation of the number of *actual* edges in a graph to the number of *possible* edges in a graph



$$n = |V| = 9$$
$$e = |E| = 9$$
$$\frac{e}{n(n-1)/2} = \frac{9}{36} = 0.25$$

Note: for directed graphs maximum number of possible edges is given by $n(n-1)/1$

Note

Notation and terms are not always used uniformly across different authors or disciplines.

Overview: levels and measures

Depending on the level of analysis, we might be interested in different properties of our network / nodes / relations.

Overview: levels and measures

Depending on the level of analysis, we might be interested in different properties of our network / nodes / relations.

- **Dyad** and **Triad** level → *structure* of the relations and whether or not a relation between nodes is *reciprocal*

Overview: levels and measures

Depending on the level of analysis, we might be interested in different properties of our network / nodes / relations.

- **Dyad** and **Triad** level → *structure* of the relations and whether or not a relation between nodes is *reciprocal*
- Single (**Ego**) nodes → position within the network, such as it's *centrality*

Overview: levels and measures

Depending on the level of analysis, we might be interested in different properties of our network / nodes / relations.

- **Dyad** and **Triad** level → *structure* of the relations and whether or not a relation between nodes is *reciprocal*
- Single (**Ego**) nodes → position within the network, such as it's *centrality*
- **Subnetwork** level → *distance* between nodes

Overview: levels and measures

Depending on the level of analysis, we might be interested in different properties of our network / nodes / relations.

- **Dyad** and **Triad** level → *structure* of the relations and whether or not a relation between nodes is *reciprocal*
- Single (**Ego**) nodes → position within the network, such as it's *centrality*
- **Subnetwork** level → *distance* between nodes
- **Network level**
  - → groups within the network
  - → comparisons between networks (network properties, averages)

Overview: levels and measures

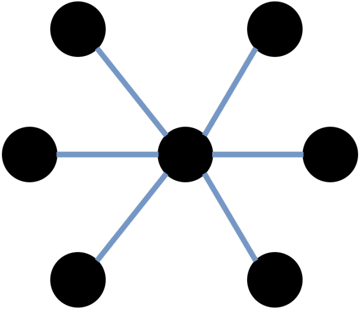
Depending on the level of analysis, we might be interested in different properties of our network / nodes / relations.

- **Dyad** and **Triad** level → *structure* of the relations and whether or not a relation between nodes is *reciprocal*
- Single (**Ego**) nodes → position within the network, such as it's *centrality*
- **Subnetwork** level → *distance* between nodes
- **Network level**
  - → groups within the network
  - → comparisons between networks (network properties, averages)

Note: levels / measures are not mutually exclusive

Structural holes

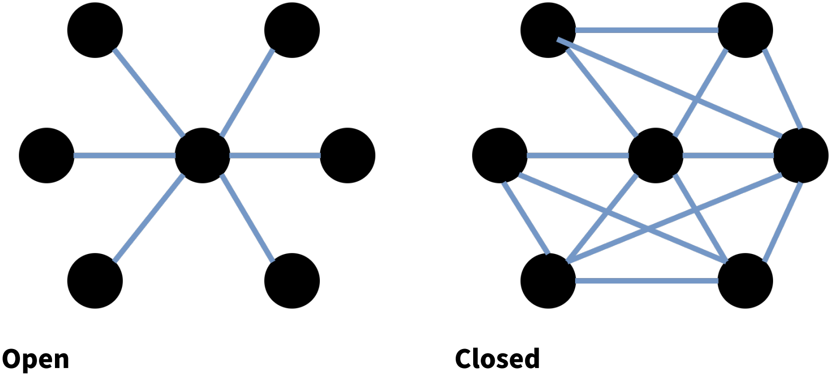
Structural holes describe the absence of connections between two nodes that are connected through other nodes



Open

Structural holes

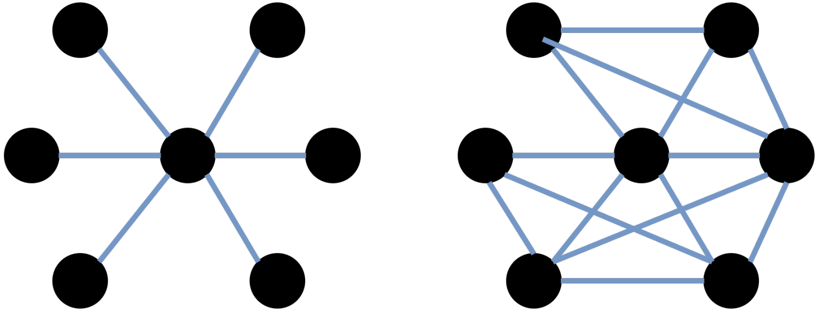
Structural holes describe the absence of connections between two nodes that are connected through other nodes



based on Fig. 4 from Borgatti et al. (2009)

Structural holes

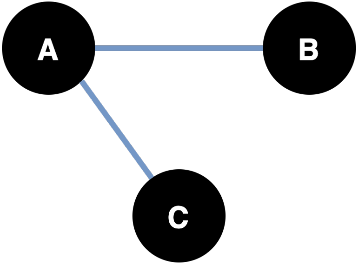
Structural holes describe the absence of connections between two nodes that are connected through other nodes



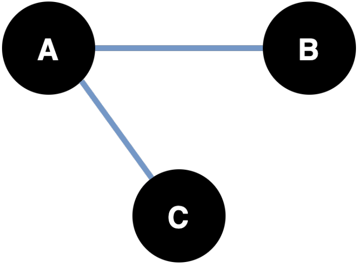
Open **Closed**

based on Fig. 4 from Borgatti et al. (2009)

Paths and distance (Borgatti & Everett, 2021)

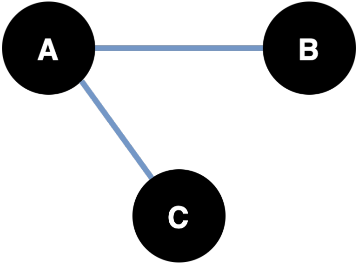


Paths and distance (Borgatti & Everett, 2021)



Path refers to the ‘way’ in which two nodes are connected in a network → only adjacent nodes can be considered!

Paths and distance (Borgatti & Everett, 2021)



→ .e.g. useful to assess *centrality*

Path refers to the 'way' in which two nodes are connected in a network → only adjacent nodes can be considered!

Distance refers to the length (number of edges) of the shortest *path* between two nodes.

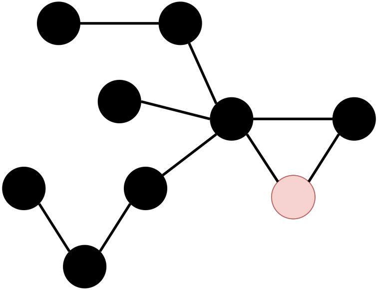
Centrality overview (Borgatti & Everett, 2021)

Centrality refers to the ‘importance’ or ‘advantage’ of the position of a node in a network.

There are multiple ways to define and approach centrality and, hence, different measures of centrality, e.g.

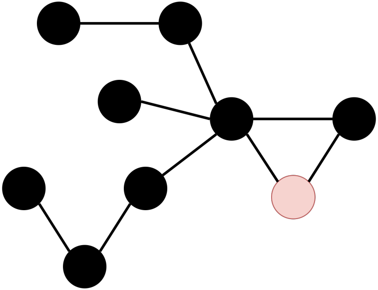
- *Degree centrality*
- *Betweenness centrality*
- *Closeness centrality*
- Eigenvector centrality
- PageRank

Degree centrality



Degree centrality *refers to the number of edges a node has.*

Degree centrality

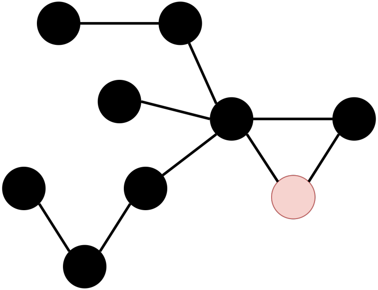


Degree centrality *refers to the number of edges a node has.*

Variants

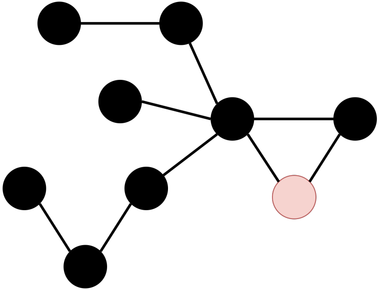
- *In- and outdegree centrality* (for directed networks)
- *Avg. degree* (comparison between networks)

Betweenness centrality



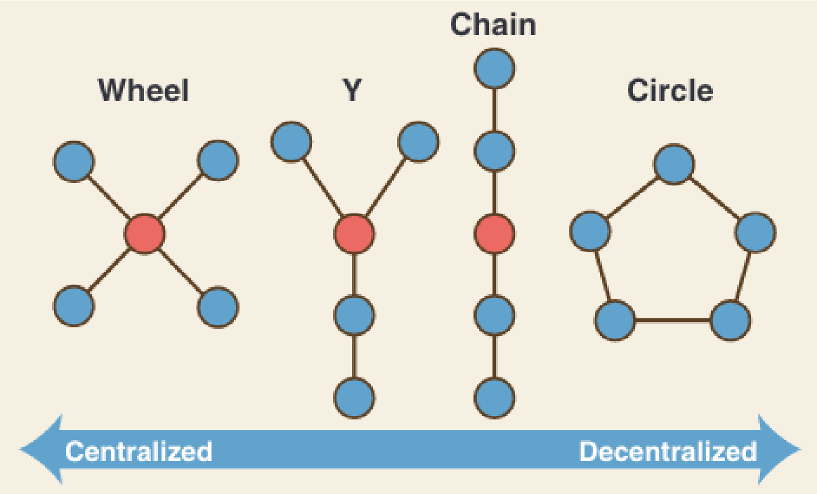
Betweenness centrality *refers to the number of times a node is a 'bridge' on the shortest path between two other nodes.*

Closeness centrality



Closeness centrality *refers to the sum of the shortest paths to all other nodes in the network.*

Centrality and network structures



Source: Borgatti et al. (2009)

Networks, Political Science and CSS

Naidu et al. (2021)

Social Origins of Dictatorships: Elite Networks and Political Transitions in Haiti

How do social networks influence the organization of resistance to democracy?

Theory

- Social networks within groups of elites important for coordinating activities or spreading information
- Network position of individuals create variation in the amount of influence individuals can exert over others
- The higher the centrality of an actor, the higher the incentive for a coup

Network

- Network of individuals based on firm-level data, business ownership data, and genealogical data
- Measure: Network centrality of individuals

Results

The higher the centrality, the more likely to participate in Coup attempts

Wang (2022)

Blood is Thicker Than Water: Elite Kinship Networks and State Building in Imperial China

Under which conditions are kinship-based institutions compatible with state building?

Theory

- Geographically dispersed kinship networks cross-cut local cleavages and incentivize elites to unite in pursuit of national, rather than sectarian, goals
- Elites embedded in such dispersed networks can benefit from a strong central state, which generates scale economies in providing protection and justice throughout a large territory

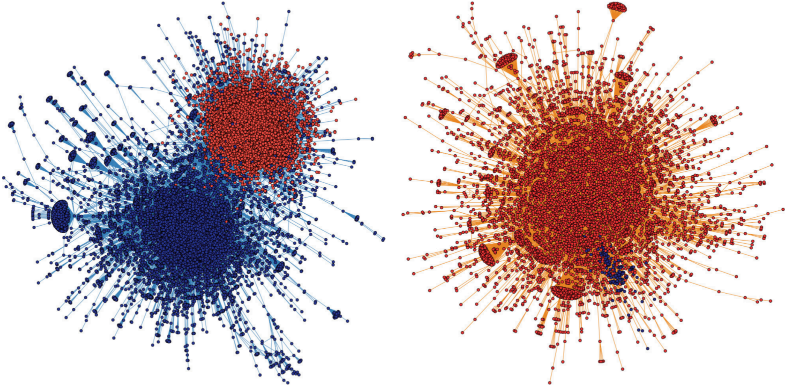
Network

- (Geographic) politician kinship network based on tomb epitaphs from 11th century China
- Measure: Concentration of the network

Results

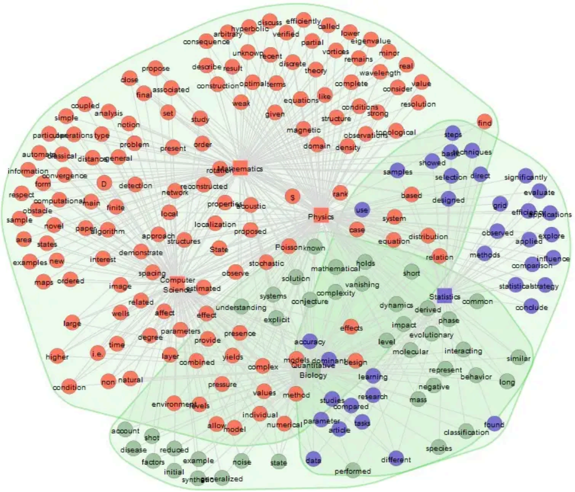
The lower concentration (more dispersed) a network, the higher support for state building

Conover et al. (2011)



Political retweet (left) and mention (right) networks on Twitter in the 6 weeks leading up to the US 2010 presidential election, laid out using a force-directed algorithm. Source: Conover et al. (2011)

Text networks



Source:
https://towardsdatascience.com/text-network-analysis-theory-and-practice-223ac81c5f07

Outlook

There are many concepts and aspects of (social) network analysis not covered in today's lecture. For example:

- Dynamic (temporal) networks
- Exponential random graphs
- Text networks
- Network experiments

Reading

Today's reading (Borgatti et al., 2009) can be found in the literature folder.

Next session

Next session 10 Jan 2023: Geo-spatial data

Guest lecture: Franziska Quoß – The impact of political business cycles on the environment

Lab

References i

Adams, jimi, Santos, T., & Williams, V. N. (2021). Strategies for Collecting Social Network Data: Overview, Assessment, and Ethics. In R. Light & J. Moody (Eds.), *The Oxford Handbook of Social Networks* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190251765.013.10

Borgatti, S. P. (2013). *Analyzing social networks* (1. publ.). Sage.

Borgatti, S. P., & Everett, M. G. (2021). Three Perspectives on Centrality. In R. Light & J. Moody (Eds.), *The Oxford Handbook of Social Networks* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190251765.013.22

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, *323*(5916), 892–895. https://doi.org/10.1126/science.1165821

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are Few-Shot Learners* (No. arXiv:2005.14165). arXiv. https://doi.org/10.48550/arXiv.2005.14165

References ii

Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political Polarization on Twitter. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*.

Jilbert, O. L. and I. (n.d.). *Social Networks: An Introduction*. Retrieved December 8, 2022, from https://bookdown.org/omarlizardo/_main/

Light, R., & Moody, J. (2021). Introduction. In R. Light & J. Moody (Eds.), *The Oxford Handbook of Social Networks* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190251765.013.1

Naidu, S., Robinson, J. A., & Young, L. E. (2021). Social Origins of Dictatorships: Elite Networks and Political Transitions in Haiti. *American Political Science Review*, *115*(3), 900–916. https://doi.org/10.1017/S0003055421000289

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback*.

References iii

Wang, Y. (2022). Blood is Thicker Than Water: Elite Kinship Networks and State Building in Imperial China. *American Political Science Review*, *116*(3), 896–910. https://doi.org/10.1017/S0003055421001490

Appendix i