Digital Trace Data 2
(Network Analysis)

Soc 121D: People Analytics

Austin van Loon

Network Science

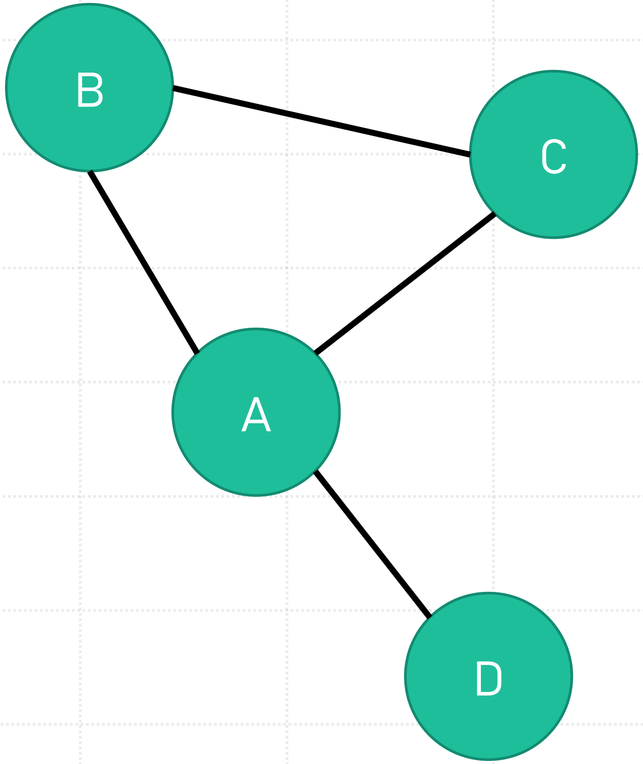
- The analysis of relational data (or data where observations have relationships)
- Highly interdisciplinary (sociologists, economists, computer scientists, psychologists, political scientists, mathematicians, neuroscientists, anthropologists, biologists/ecologists, physicists, chemists, etc.)

What is a “network”?

- A set of “nodes” or “vertices” (in our typical case, people) connected by “ties” or “edges” which represent some relationship between nodes
- You can have nodes and ties of multiple “kinds” (e.g., people and projects; friendship and working together), but we’ll be sticking with simple networks
- Ties can be either directed (e.g., Twitter following) or undirected (e.g., Facebook friends), weighted (e.g., number of text messages) or unweighted (e.g., saved contact)
- Can be represented numerically as an **adjacency matrix**

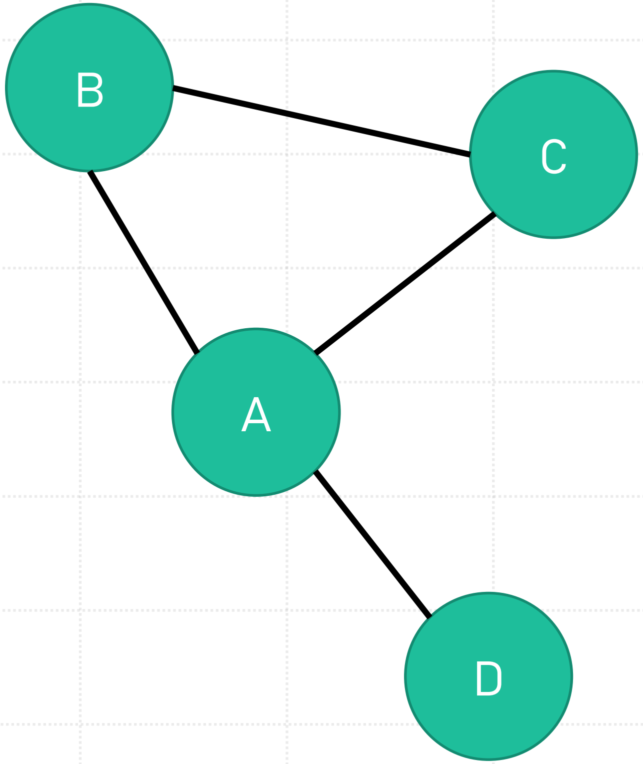
Adjacency Matrix

	A	B	C	D
A	0	?	?	?
B	?	0	?	?
C	?	?	0	?
D	?	?	?	0



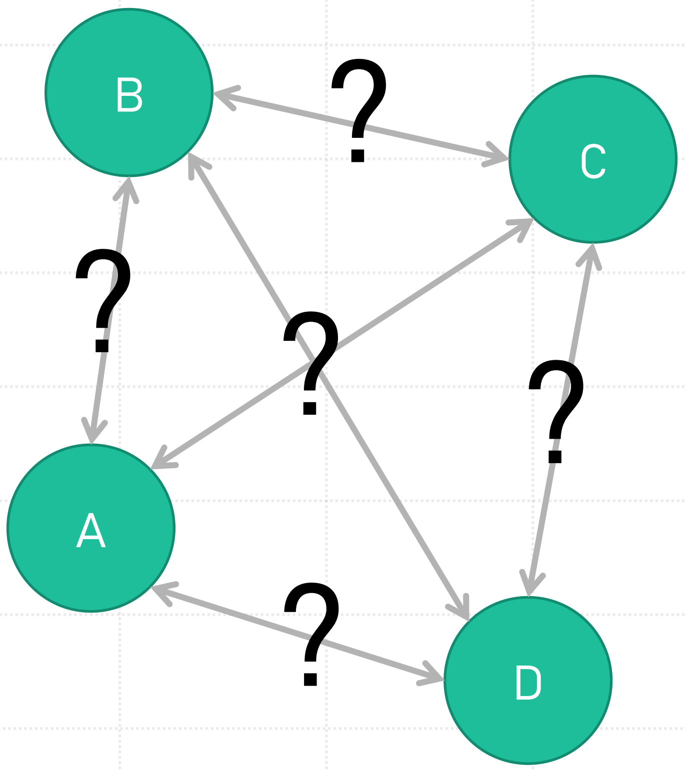
Adjacency Matrix

	A	B	C	D
A	0	1	1	1
B	1	0	1	0
C	1	1	0	0
D	1	0	0	0



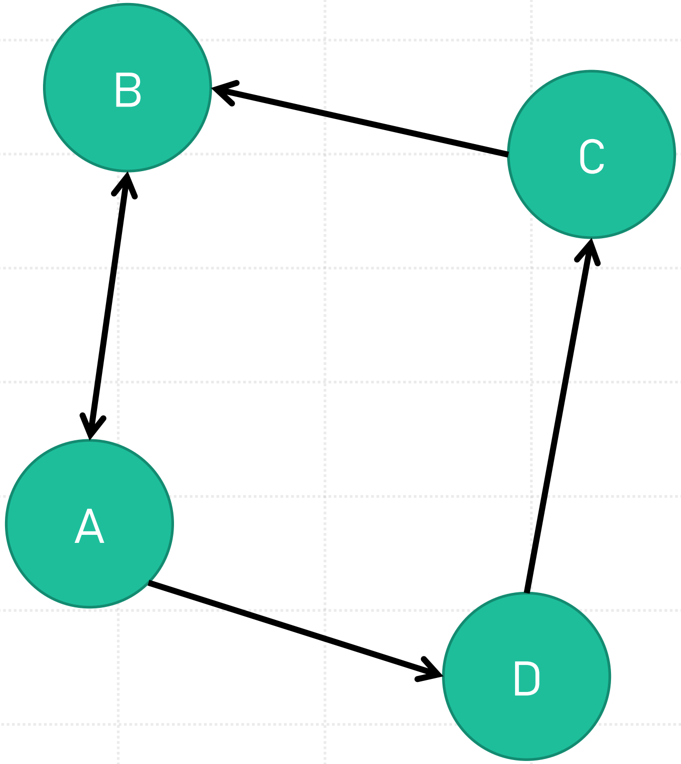
Adjacency Matrix

	> A	> B	> C	> D
A >	0	1	0	1
B >	1	0	0	0
C >	0	1	0	0
D >	0	0	1	0



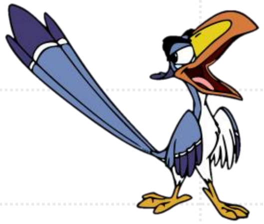
Adjacency Matrix

	> A	> B	> C	> D
A >	0	1	0	1
B >	1	0	0	0
C >	0	1	0	0
D >	0	0	1	0



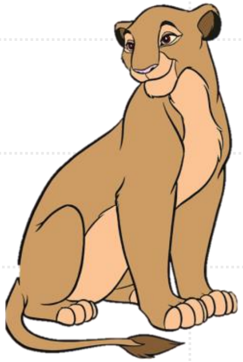







?



Say Hello

First Name Basis

Help with Paper

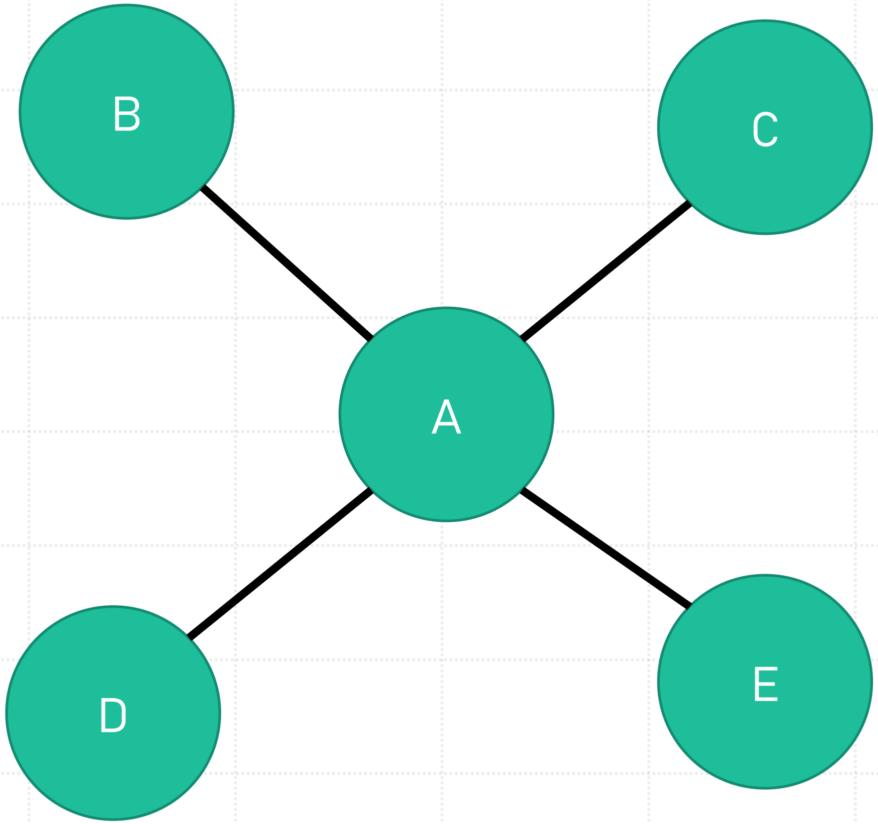
Ask a Favor

Friendship

Three Kinds of Network Analysis

- **Centrality** – The analysis of individuals' position within the network
  - Who is important to the network?
  - Who is positioned to have new/innovative ideas?
- **Community detection** – Detecting group structure, or "clusters" of nodes
  - Where are the boundaries of salient groups?
  - How "closed off" are groups from one another; are the information "silos"?
- **Network-level characteristics** – Measuring tendencies among all nodes in general
  - What are the unwritten rules governing the network structure?
  - What types of connections are (un)likely to form in the future?
  - What types of interventions might help us unlock our workforce's potential?

Centrality



Individuals' Position in Networks

- One's position in a network has various **social consequences**
  - Those who are "brokers" get unique information, granting them unique opportunities and ability to derive unique insights ("good ideas")
  - Those who are "important" to advice networks tend to be granted high prestige and leadership positions
  - Those who are "popular" activate goal-oriented parts of others' brains
- But is also have **psychological origins**
  - Personality is associated with network position
  - Empathy is related to having more connections
  - Attractive people build more advantageous networks

(Some) Measures of Centrality

- **Degree**: how many connections each individual has
- **Closeness**: how “close” are you to everyone on average?
- **PageRank**: if each tie is a vote of confidence, a measure of importance
- **Betweenness**: if each tie means exchange, these folks are essential for the efficient flow over the network
- **Constraint**: measures how much of a “broker” an individual is (i.e., your connections aren’t connected to each other)
- In reality, these all (with maybe the exception of constraint) are highly correlated

PageRank

Number of stops

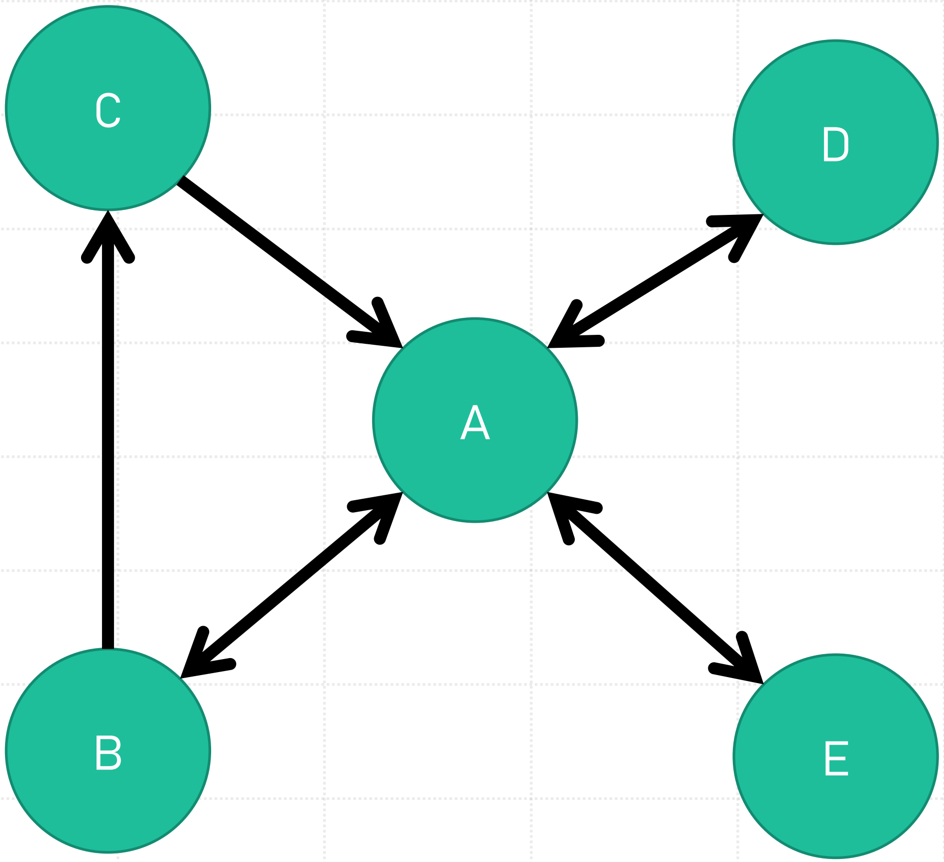
A: 0

B: 0

C: 0

D: 0

E: 0



Number of stops

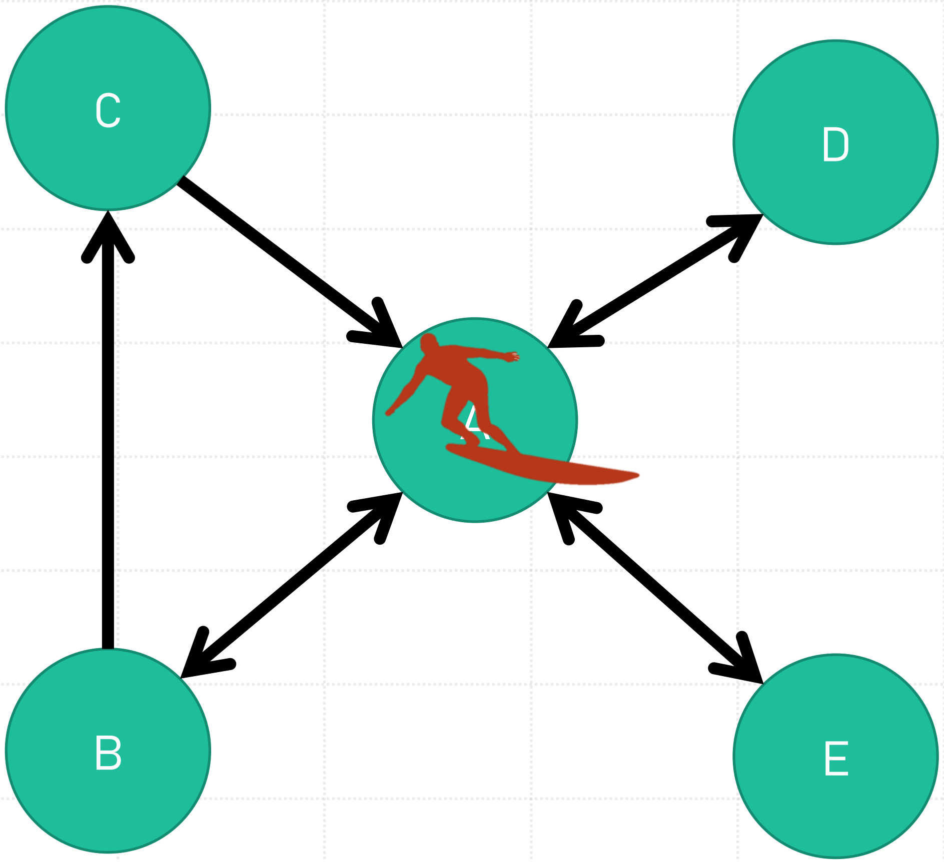
A: 1

B: 0

C: 0

D: 0

E: 0



Number of stops

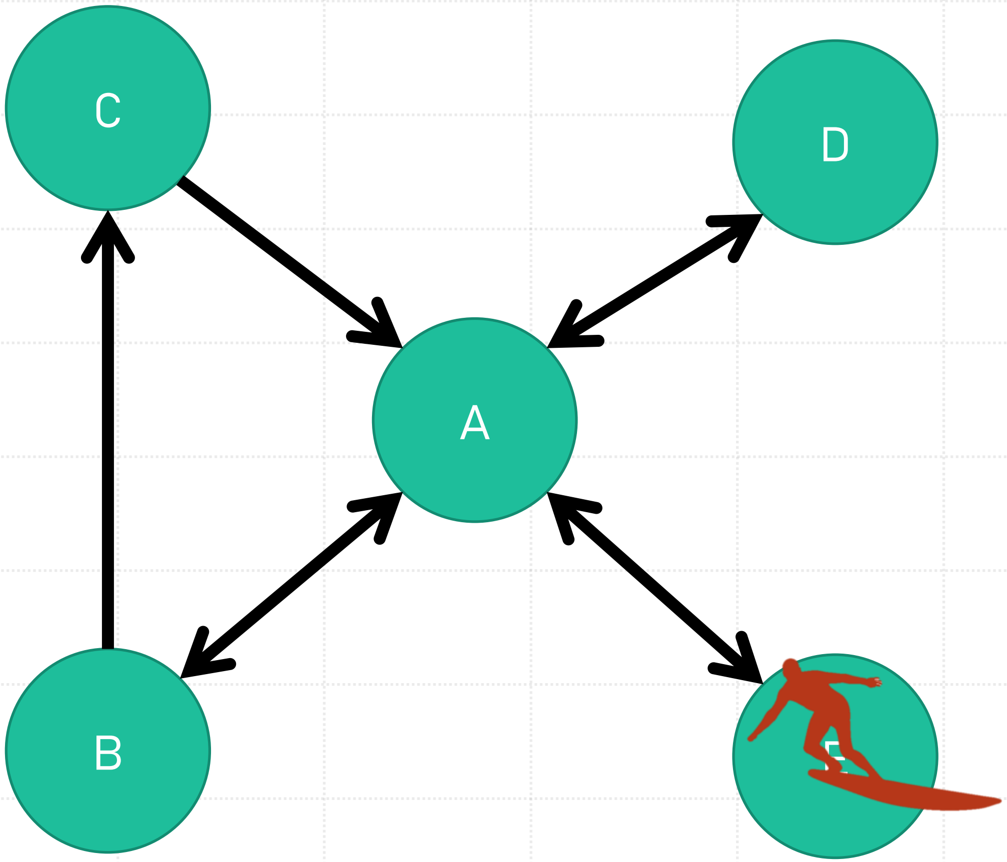
A: 1

B: 0

C: 0

D: 0

E: 1



Number of stops

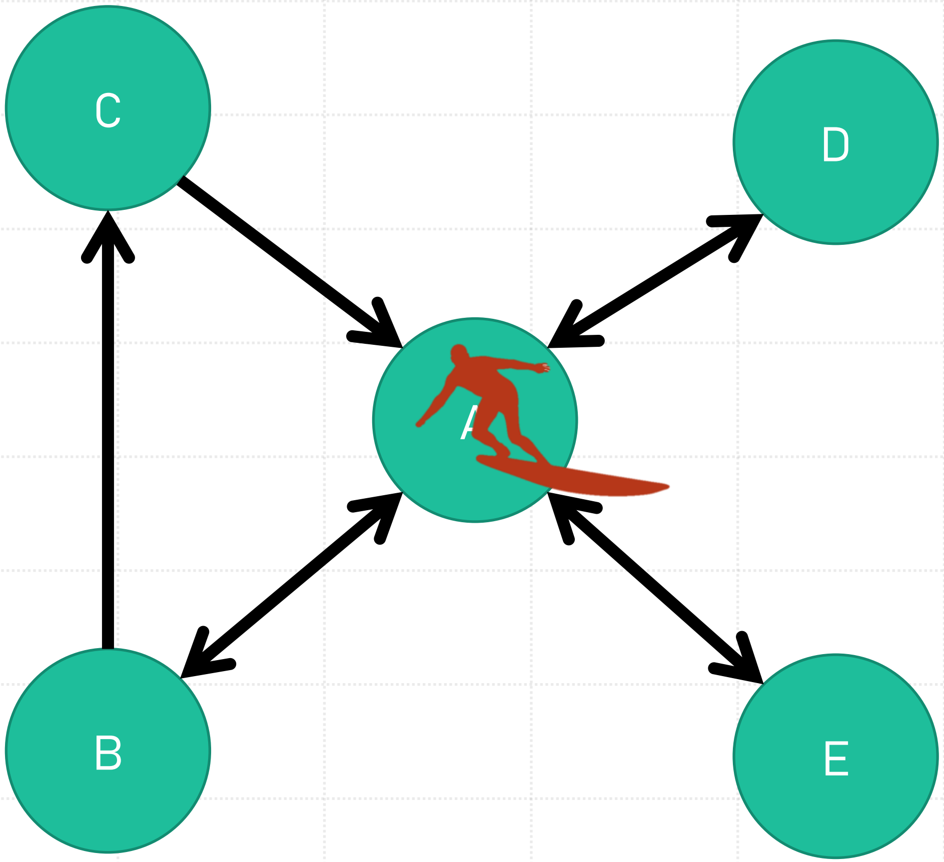
A: 2

B: 0

C: 0

D: 0

E: 1



Number of stops

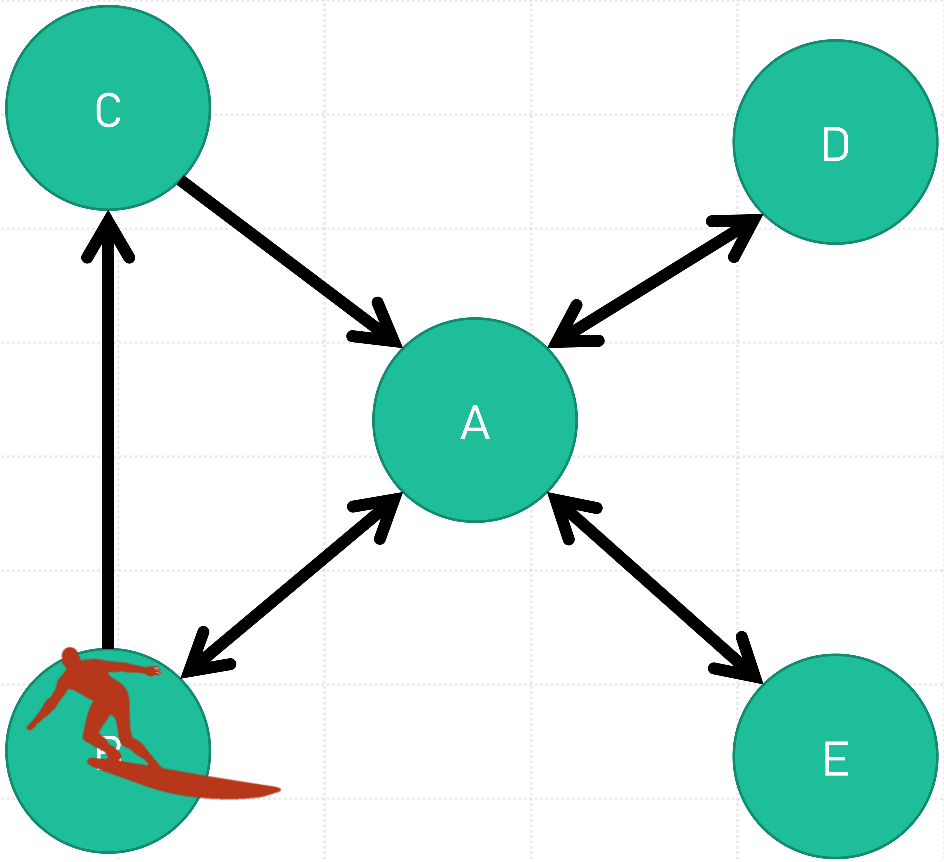
A: 2

B: 1

C: 0

D: 0

E: 1



Number of stops

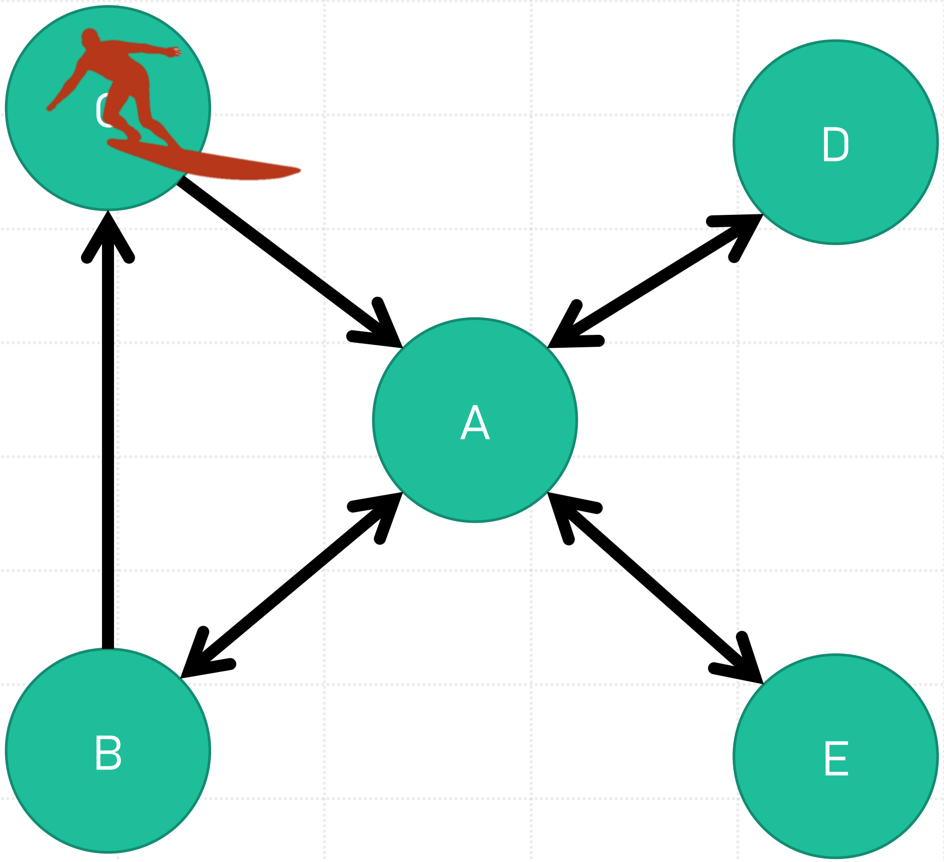
A: 2

B: 1

C: 1

D: 0

E: 1



Number of stops

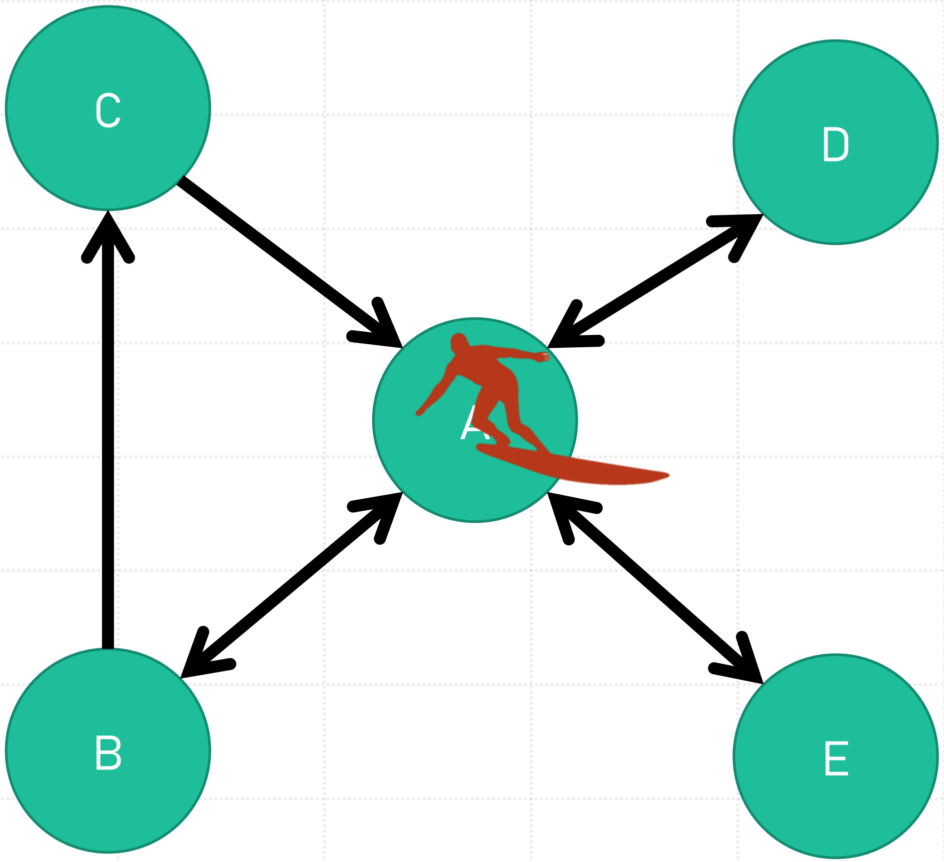
A: 3

B: 1

C: 1

D: 0

E: 1



Number of stops

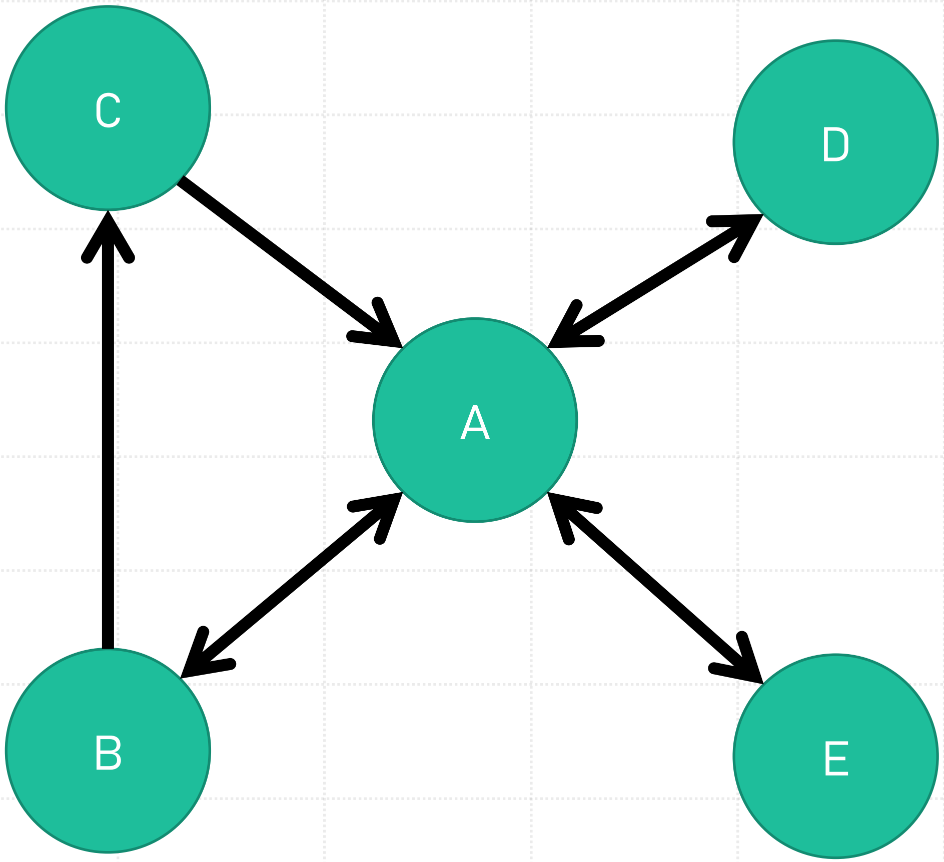
A: 3

B: 1

C: 1

D: 0

E: 1



Number of stops

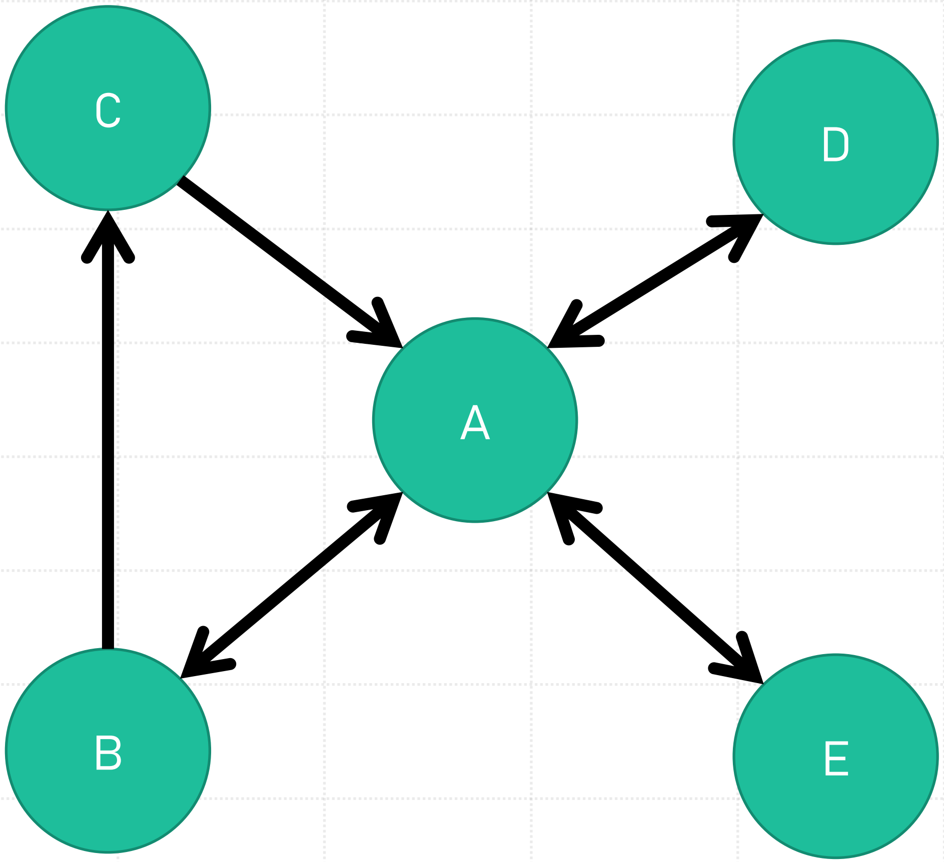
A: 736,955

B: 271,707

C: 177,314

D: 270,869

E: 208,589



Number of stops

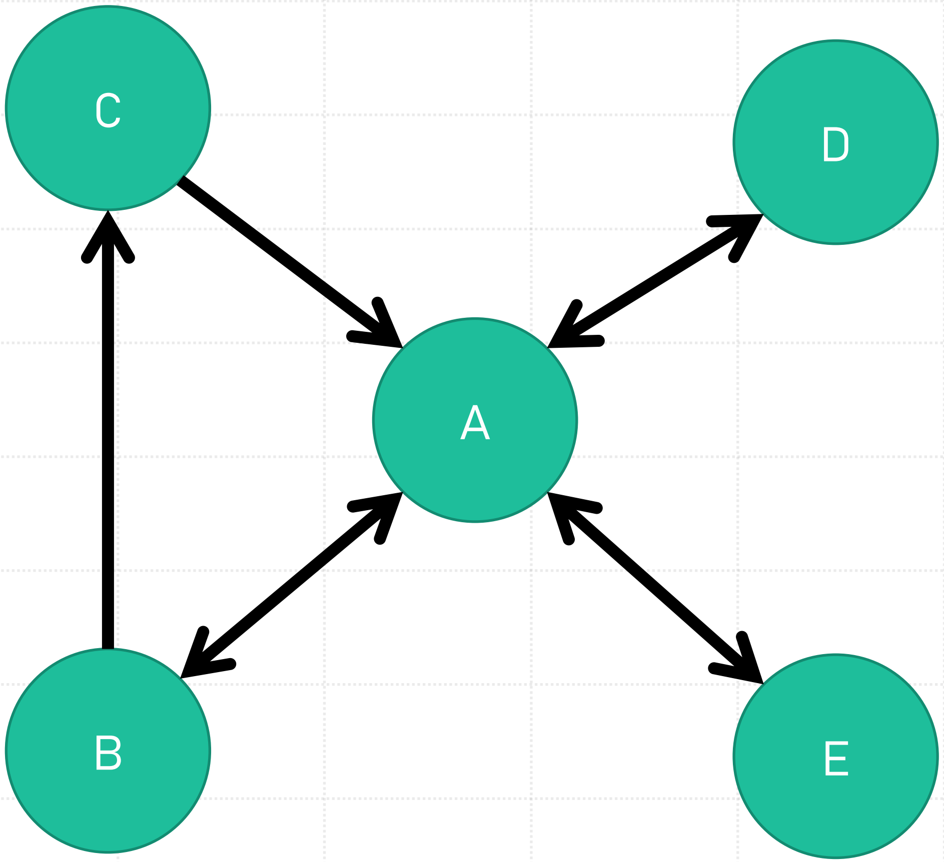
A: 0.44

B: 0.16

C: 0.11

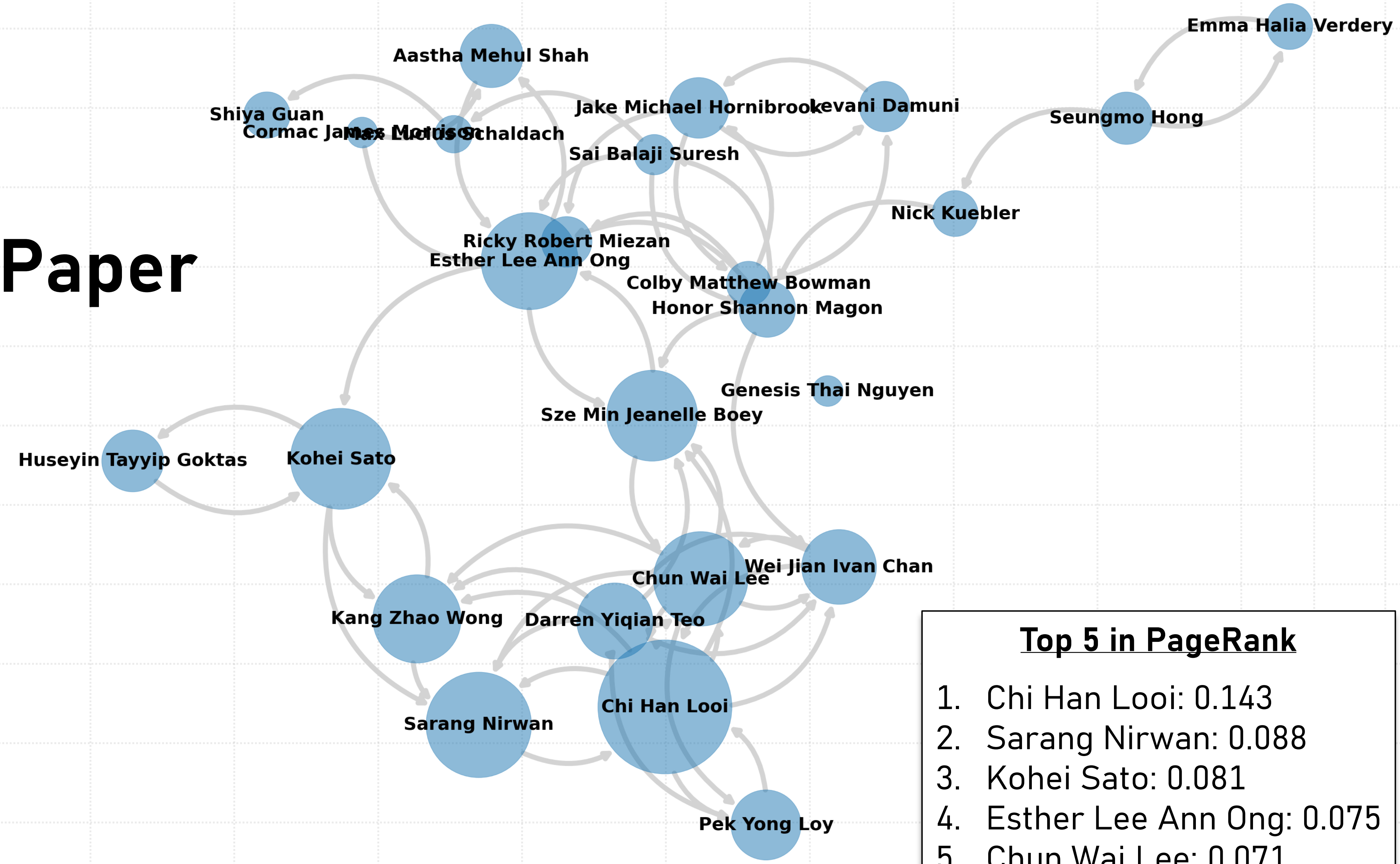
D: 0.16

E: 0.13



Help with Paper

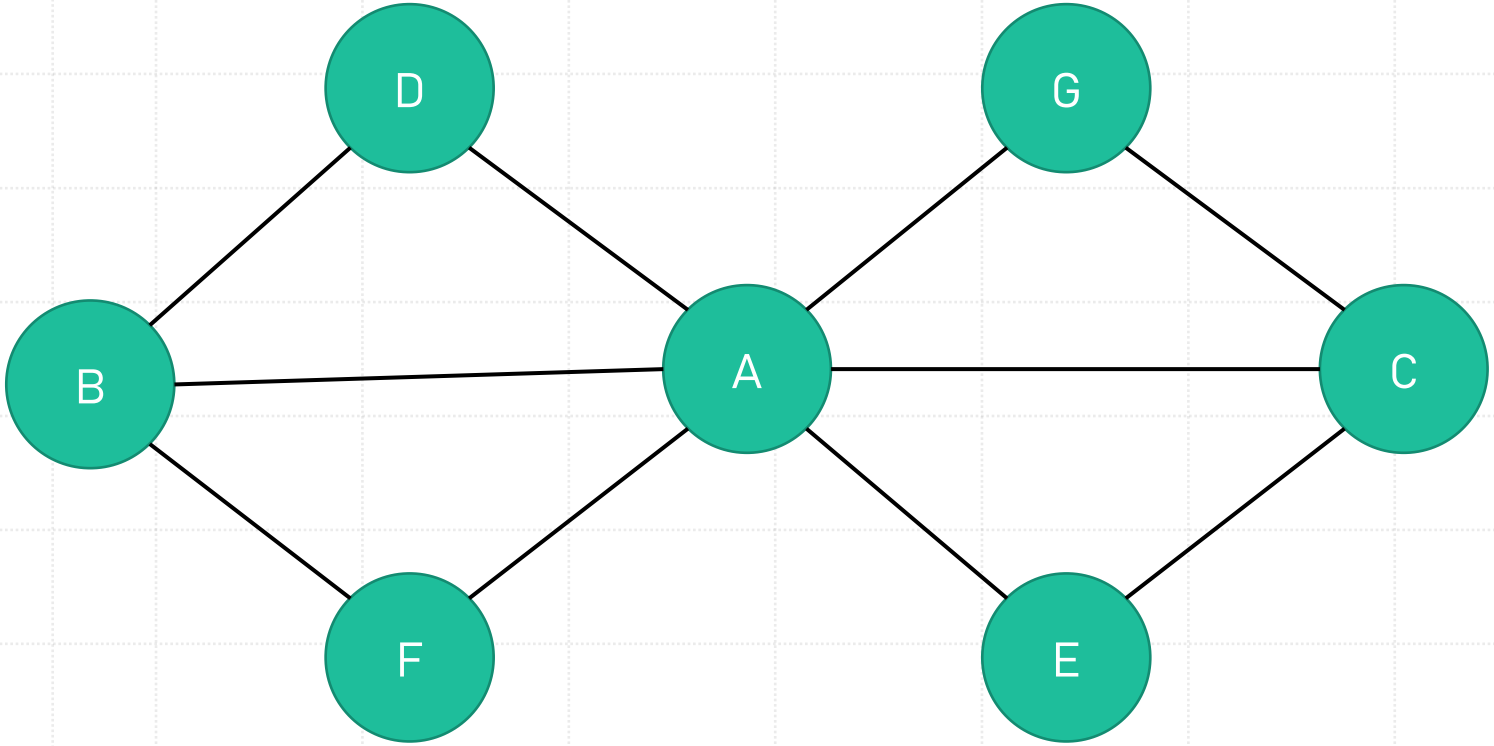
(sized by PageRank)



Top 5 in PageRank

1. Chi Han Looi: 0.143
2. Sarang Nirwan: 0.088
3. Kohei Sato: 0.081
4. Esther Lee Ann Ong: 0.075
5. Chun Wai Lee: 0.071

Betweenness



Number of paths

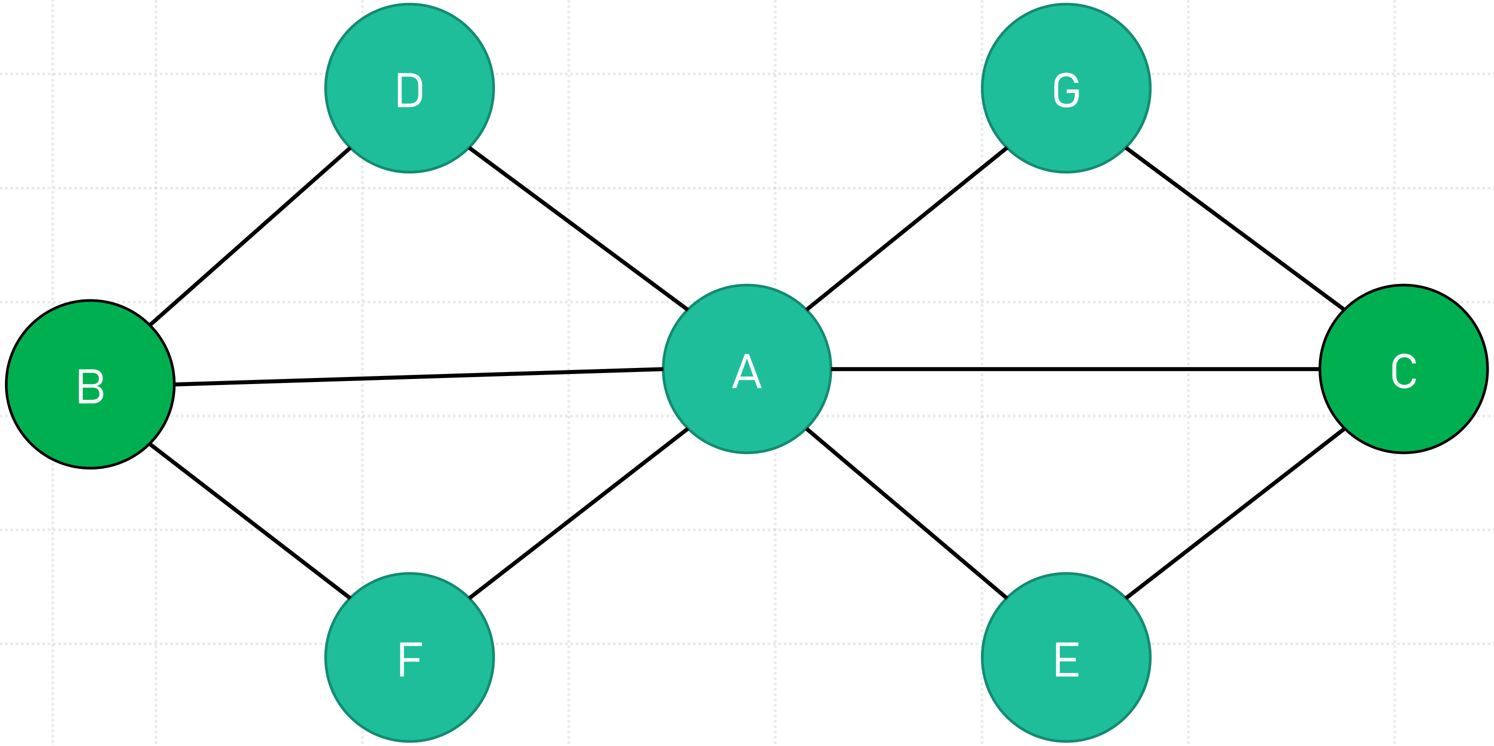
A: 0

B: 0

C: 0

D: 0

E: 0



Number of paths

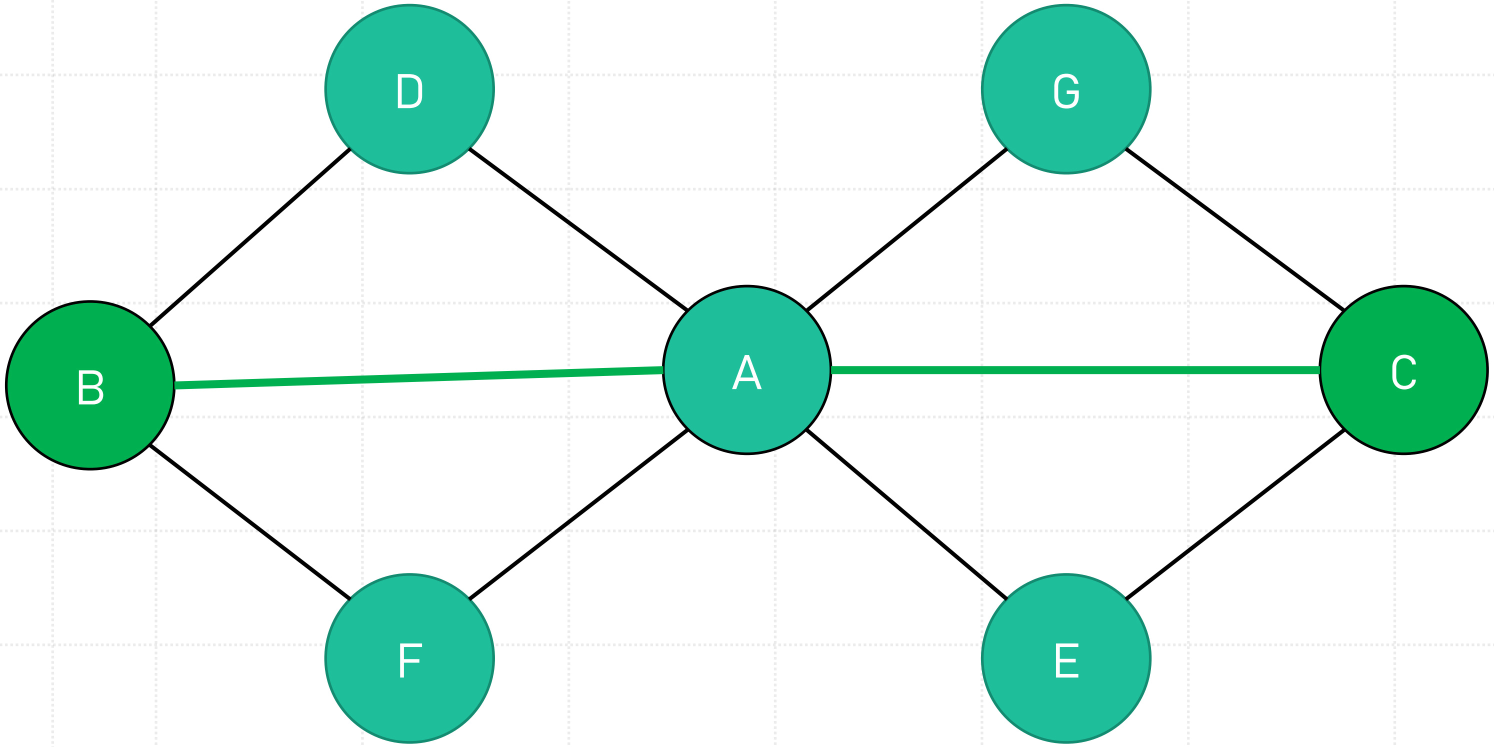
A: 0

B: 0

C: 0

D: 0

E: 0



Number of paths

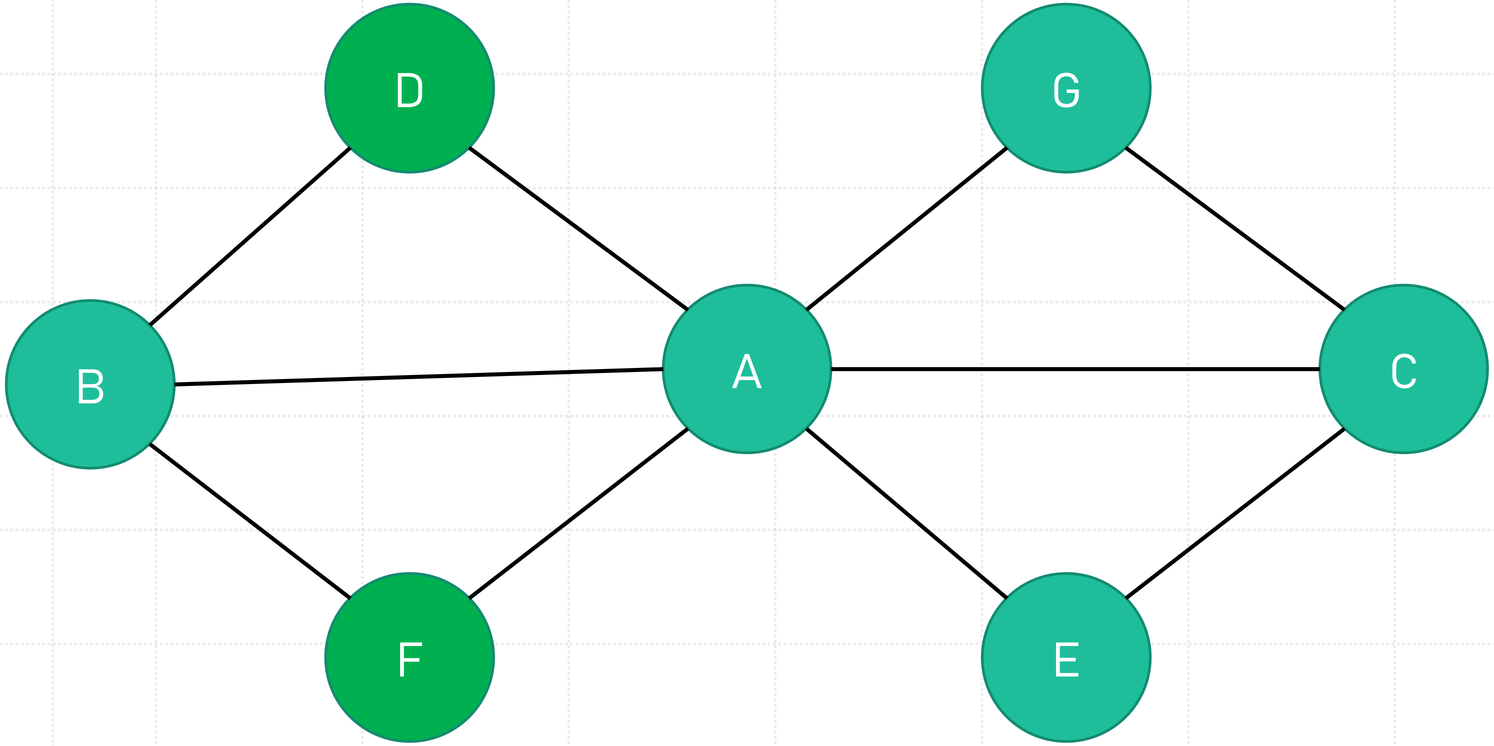
A: 1

B: 0

C: 0

D: 0

E: 0



Number of paths

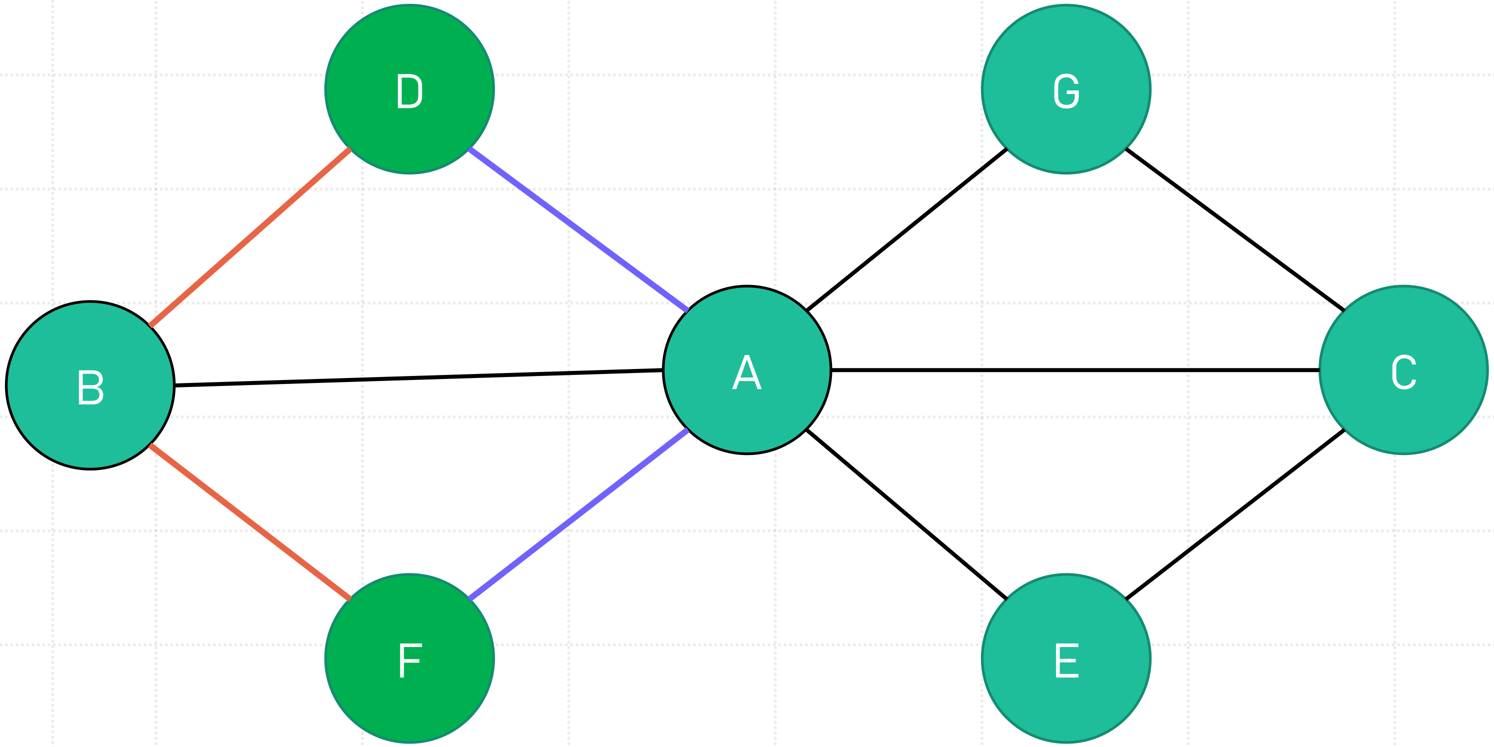
A: 0

B: 0

C: 0

D: 0

E: 0



Number of paths

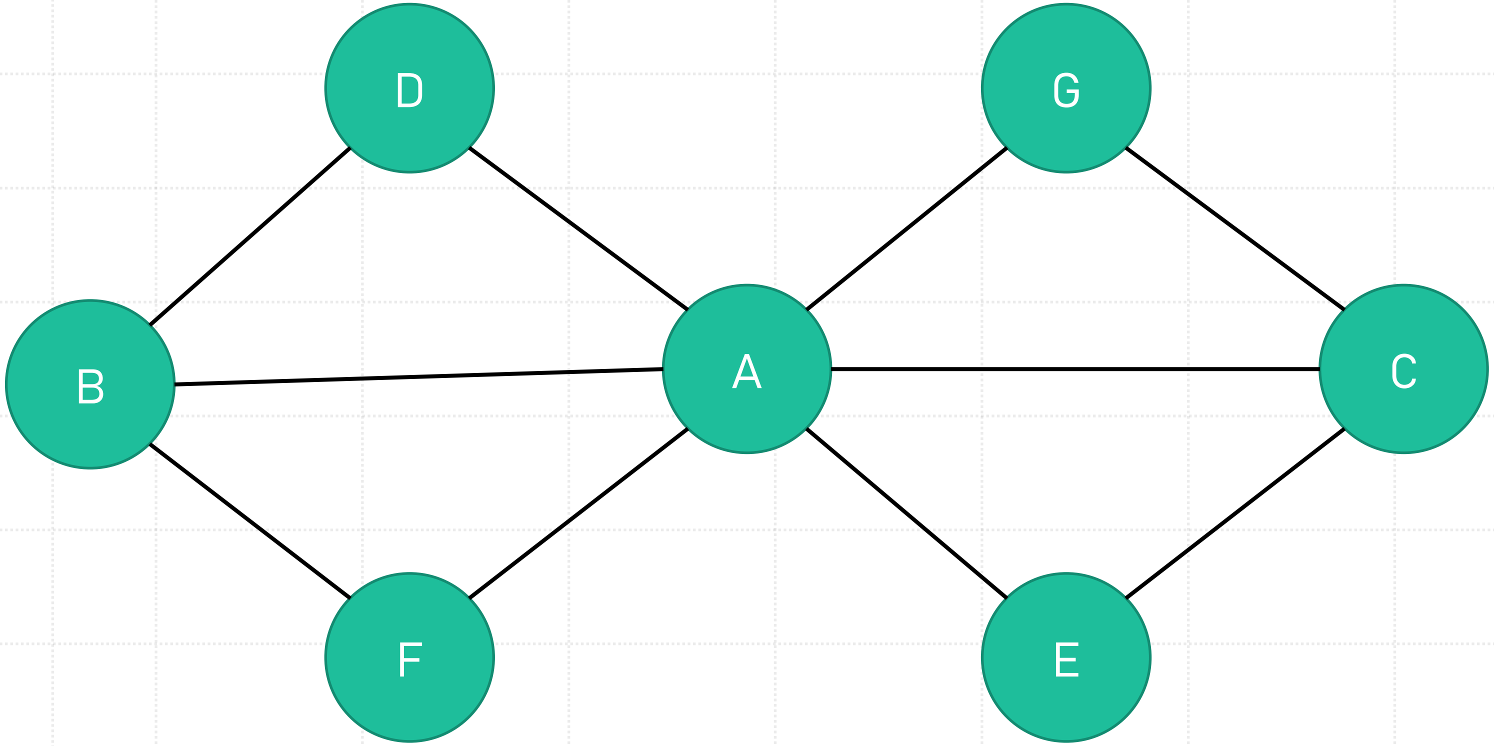
A: 1.5

B: 0.5

C: 0

D: 0

E: 0



Number of paths

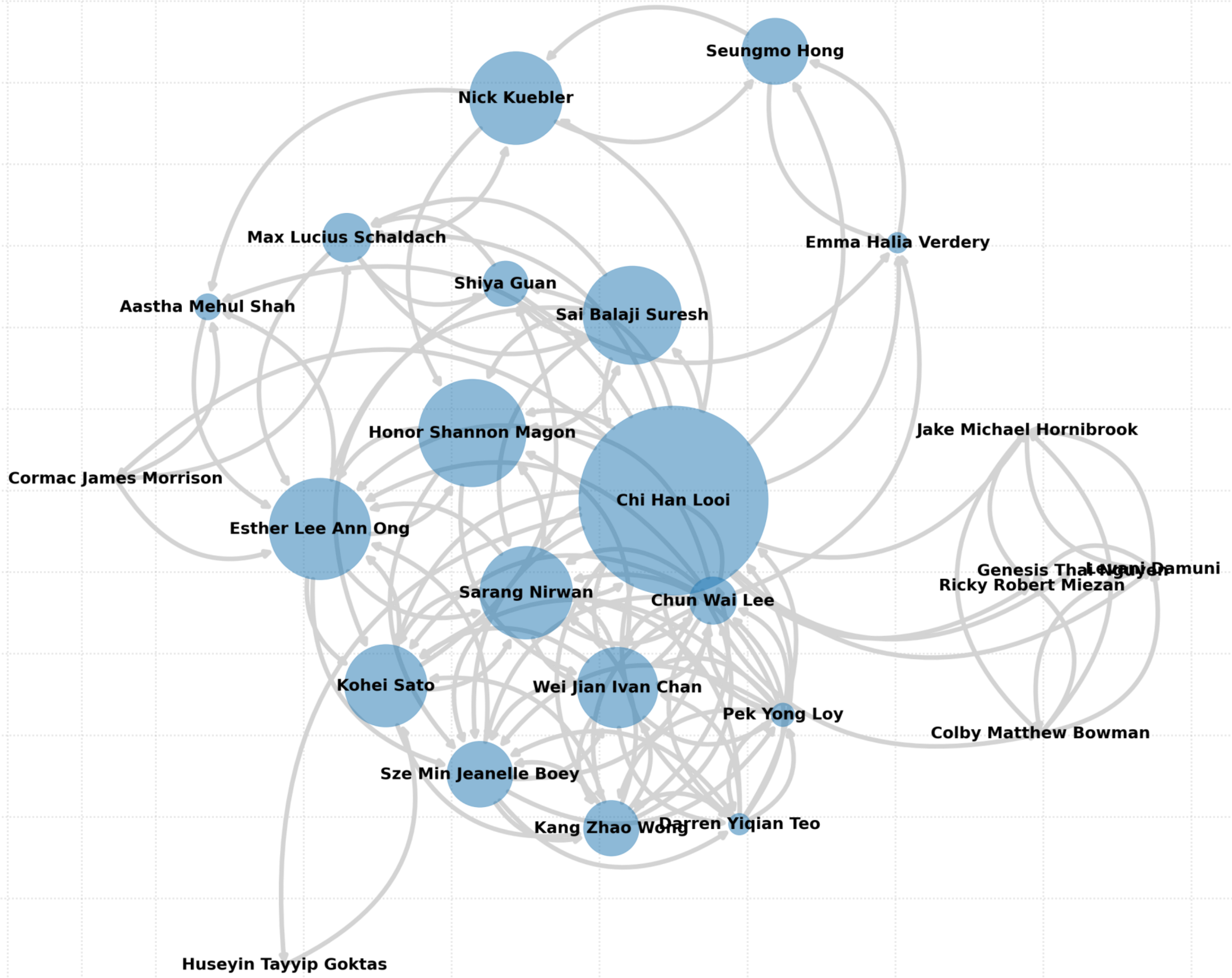
A: 10

B: 0.5

C: 0.5

D: 0

E: 0



Say Hello

(sized by betweenness)

Top 5 in Betweenness

1. Chi Han: 195.5
2. Honor Shannon Magon: 63.3
3. Esther Lee Ann Ong: 56.4
4. Sai Balaji Suresh: 52.9
5. Nick Kuebler: 47.2

Structural Holes and Good Ideas[1]

Ronald S. Burt

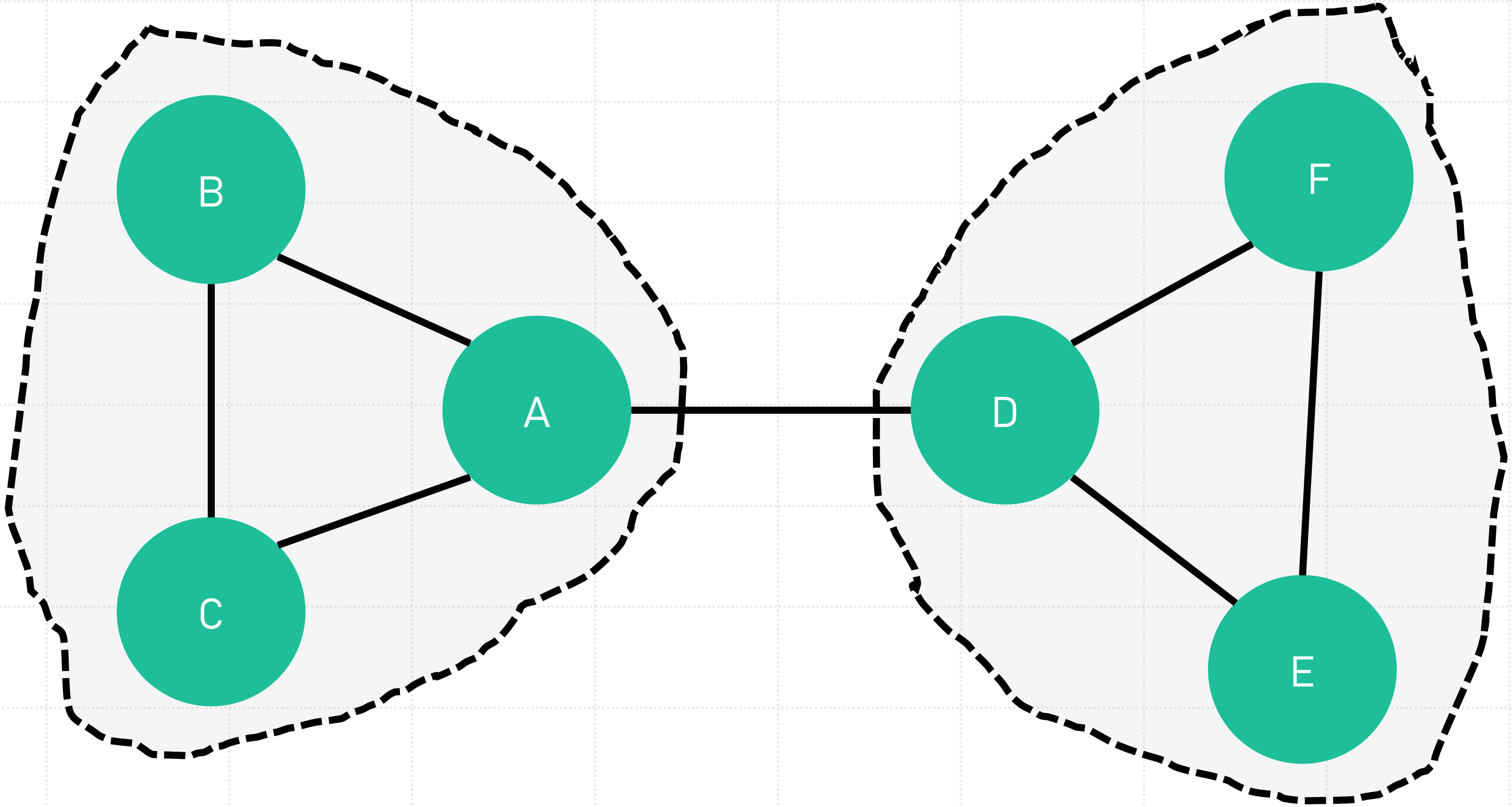
University of Chicago

The Diffusion of Microfinance

ABHIJIT BANERJEE , ARUN G. CHANDRASEKHAR , ESTHER DUFLO , AND , MATTHEW O. JACKSON

SCIENCE • 26 Jul 2013 • Vol 341, Issue 6144 • DOI: 10.1126/science.1236498

Community Detection

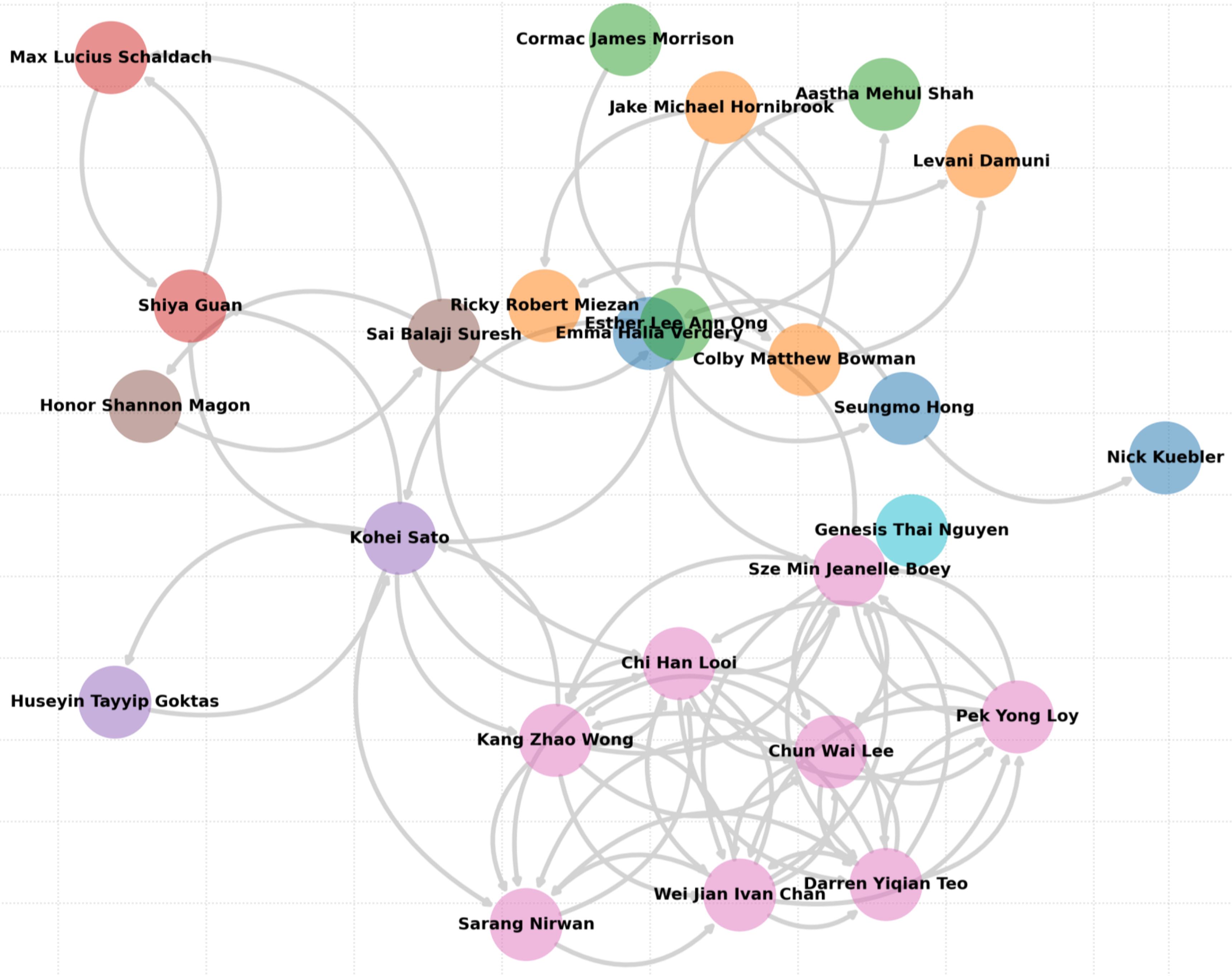


Groups, Communities, and Cliques (oh my!)

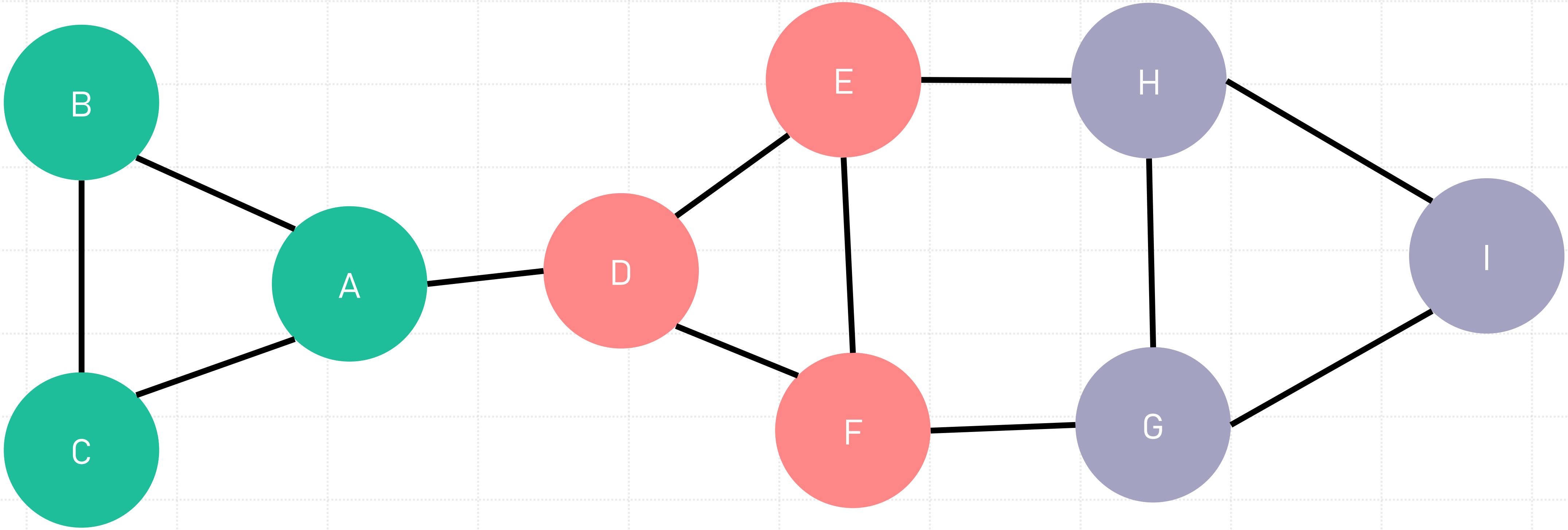
- Human social activity often happens in groups—while most possible ties don't exist, there are dense connections among subsets
  - Fun fact: This gives rise to the **caveman graph** (pardon the gendered language), which is especially efficient at transmitting information (yay!) and disease (aw!)
- To measure this, we typically derive some **quantitative measure** that defines what a "group" is (typically something along the lines of "lots of connections within groups, not as many connections outside of groups")
  - Then we might derive an **algorithmic approach** to finding groups that maximize/minimize this measure
  - Or we might find a clever way to find a **closed-form approximation** through manipulating the adjacency matrix

Friendship

(colored by Louvain community)



Network Characteristics



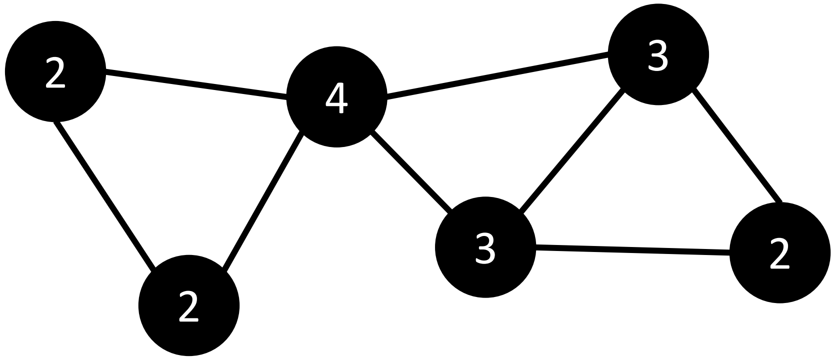
Network Characteristics

- Social networks tend to follow a set of “rules”…
  - **Homophily**: “birds of a feather block together”
  - **Triadic closure**: “the friend of my friend is my friend”
  - **Reciprocity**: “you scratch my back, I scratch yours”
  - **Preferential attachment**: “the rich get richer” (also called “cumulative advantage”)
    - Fun Fact: this gives rise to the “friendship paradox”—for most people, your friends are more popular than you
- Human social networks also vary in important ways:
  - **Density**: “how likely are two random people to be connected?”
  - **Robustness**: “how fragile is this network?”
  - **Centralization**: “how unequal is this network?”
  - **Epidemic threshold**: “how susceptible to disease spread is this network?”

Network Characteristics, cont.

- For each characteristic of a network, we can **measure** it on our network and then **compare** it to what we would expect “by chance”.
- How do we figure out what we would expect “by chance”? We create networks that are in some ways like our network but are otherwise “random”
  - **Erdős-Rényi**: the same number of nodes and edges
  - **Configuration model**: the same “degree distribution”

Observed

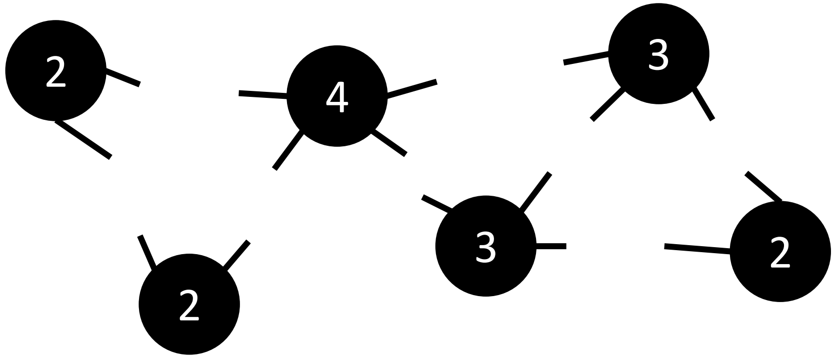


Degree = 2: 3 nodes
Degree = 3: 2 nodes
Degree = 4: 1 nodes

Configuration Model

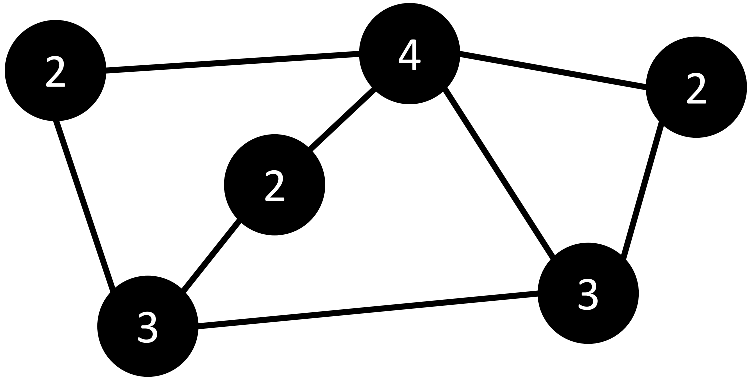
Step 1:

Create "nubs" to replicate degree distribution



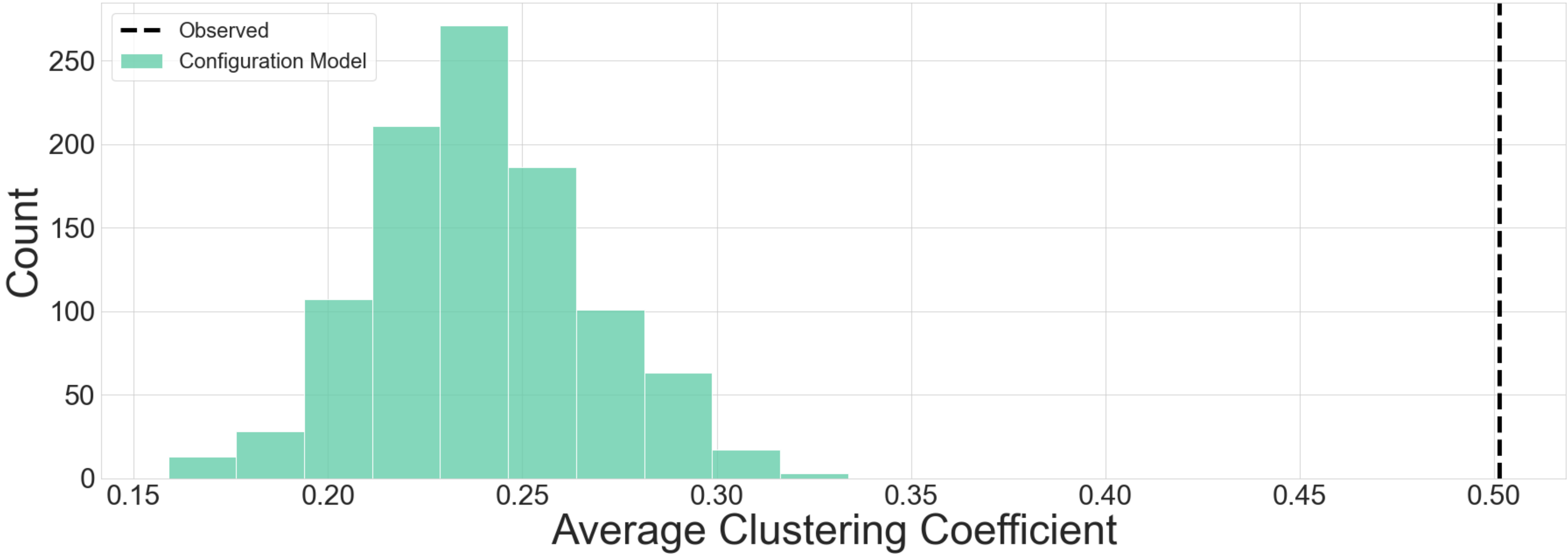
Step 2:

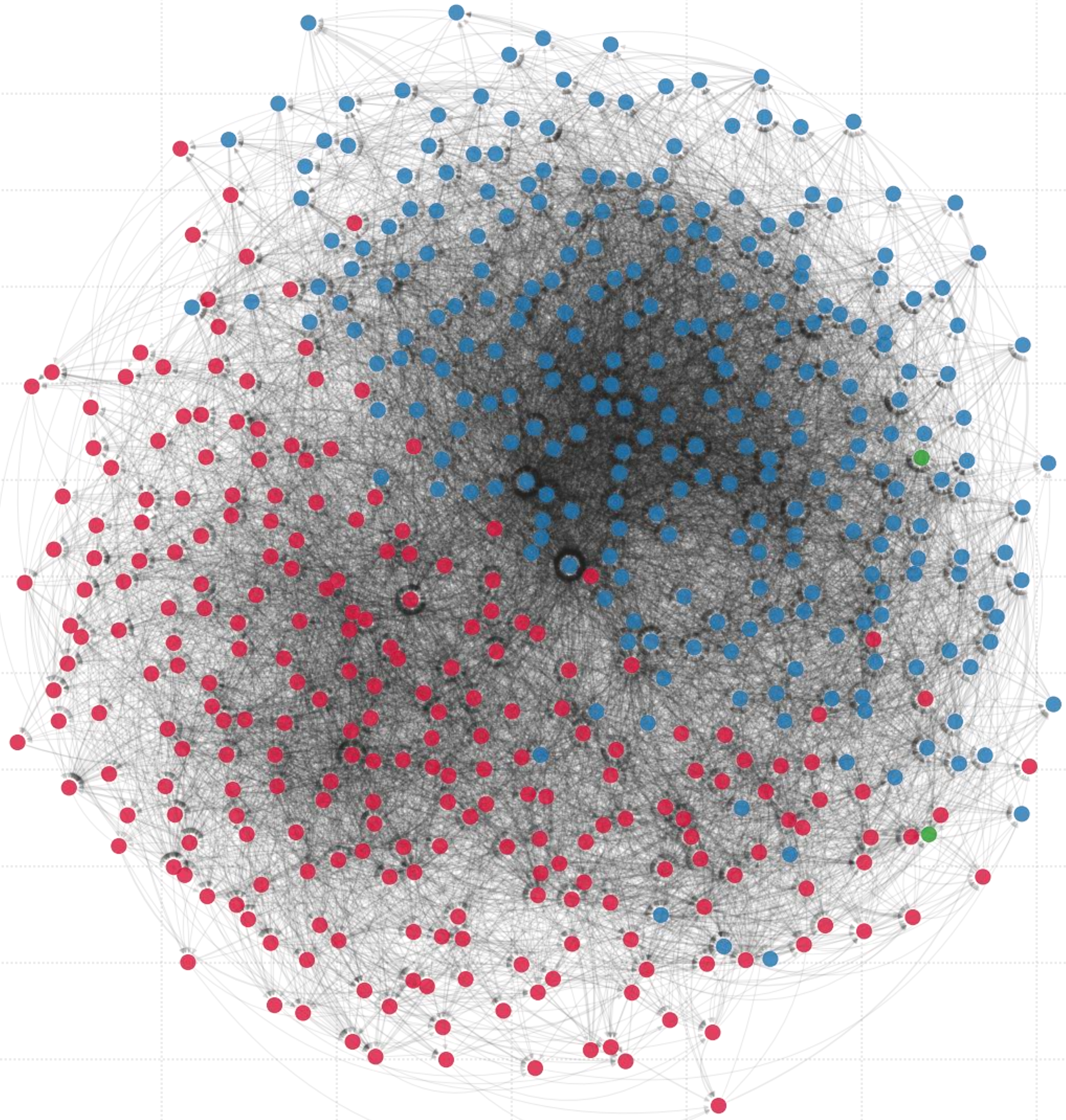
Randomly connect "nubs" to create network



Clustering in First Name Basis

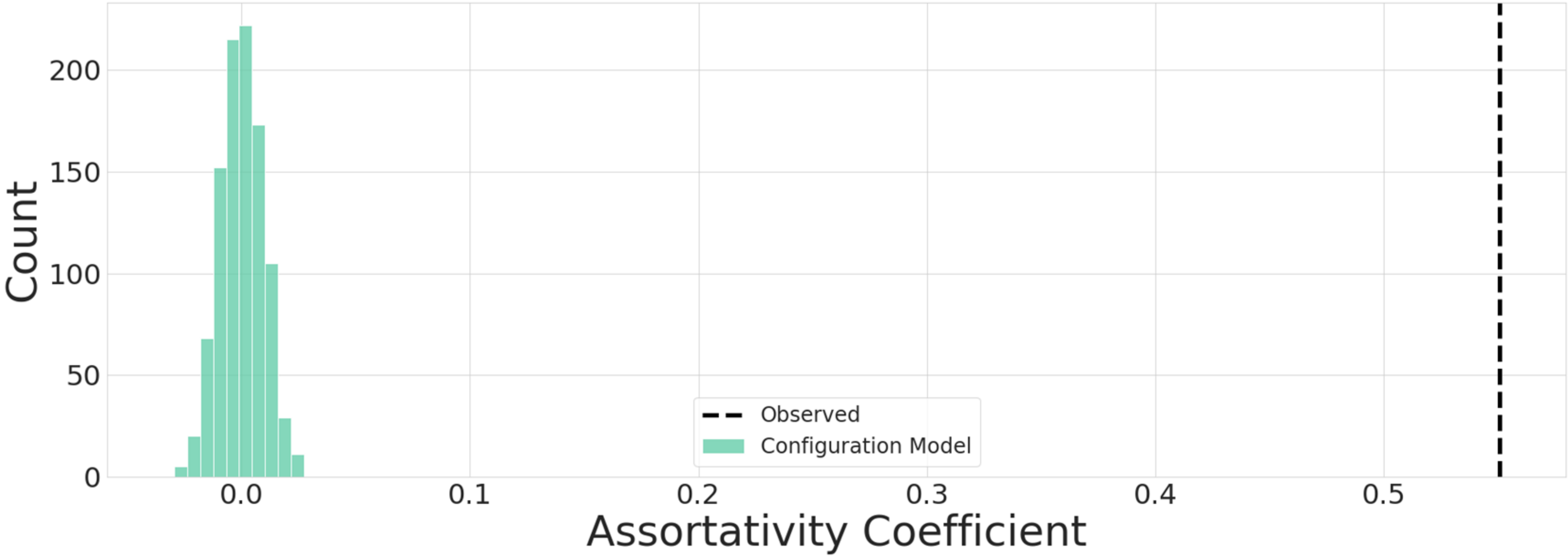
(observed vs. configuration model)





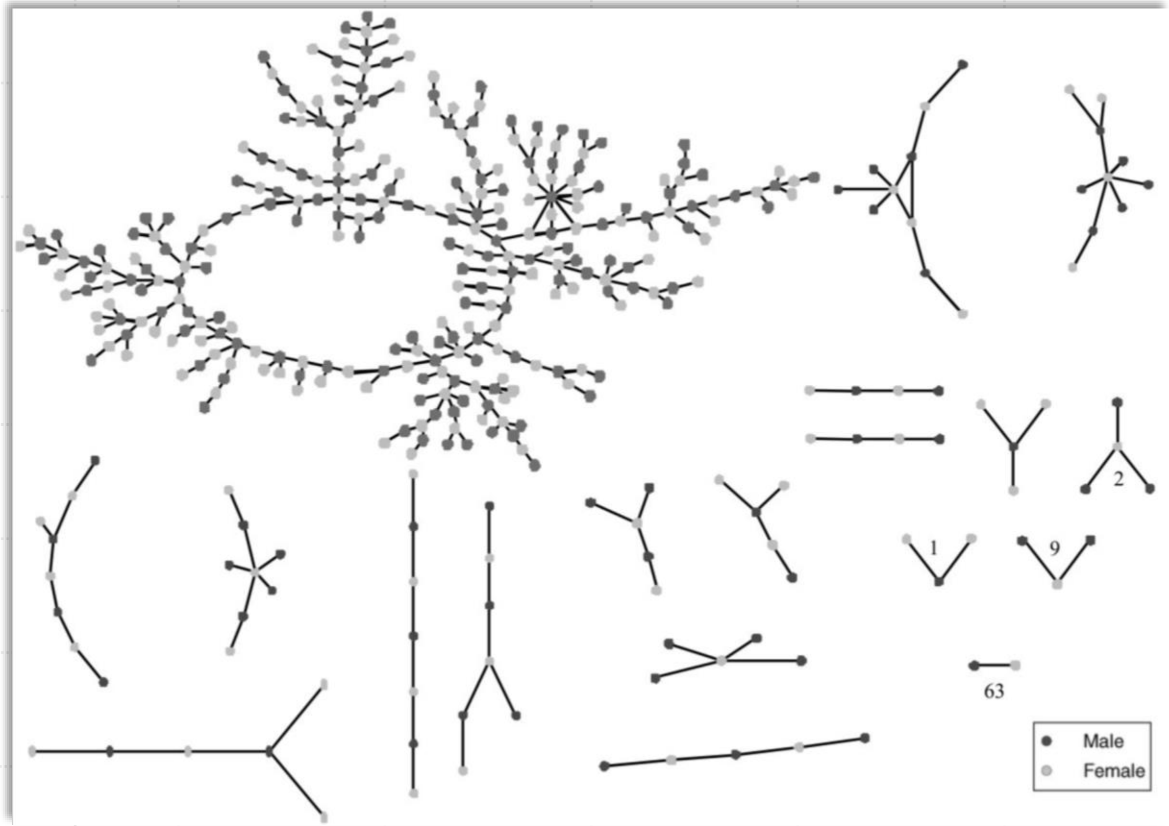
Party Assortativity among MoCs on Twitter

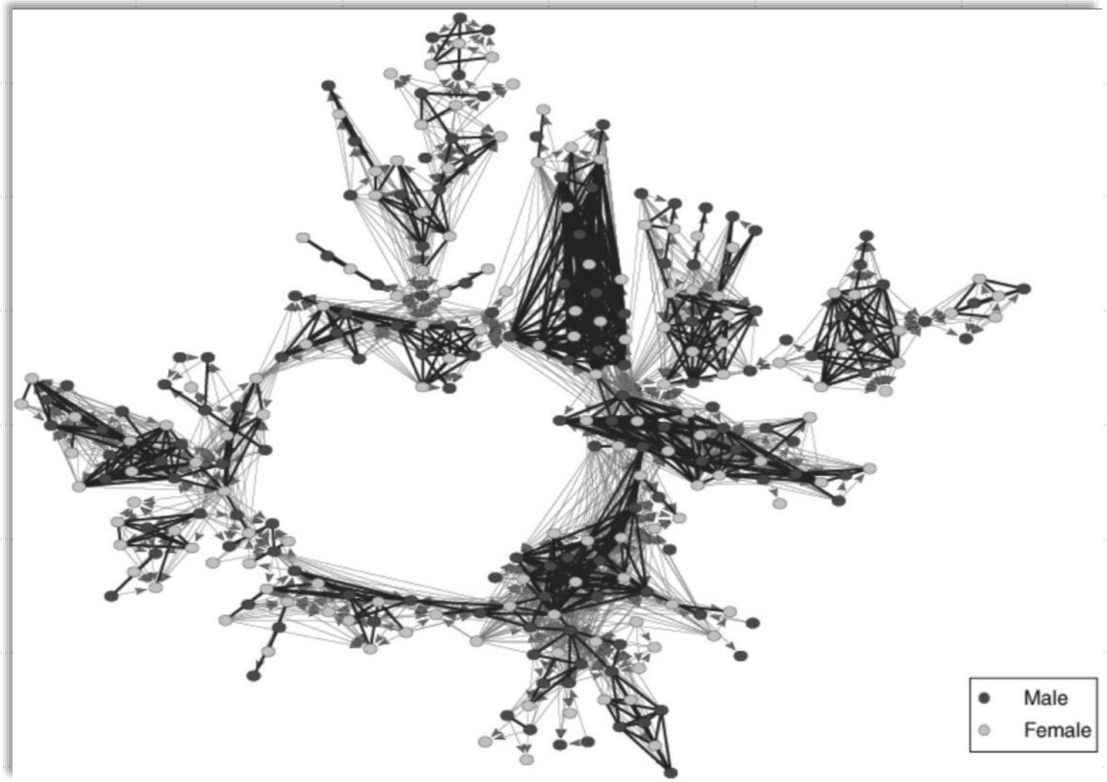
(observed vs. configuration model)



Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks[1]

Peter S. Bearman, James Moody, and Katherine Stovel





Some Applications

- Have a problem but no ideas for a solution? Ask the brokers!
- Need to encourage compliance with a policy change? Convince the influencers!
- Information siloes causing mistakes? Identify the silos (communities) and intervene to increase the flow of information!
- Worker feeling not engaged? Look at their network!
  - Are they stuck in silo? Find ways for them to escape their “cave”!
  - Do they not have a clear community? See if you can integrate them into one or build a community by intervening on their alters!

The Ethics of Digital Trace Data

- Do those being analyzed have a reasonable **expectation of privacy**?
  - Different mediums and environments come with different expectations.
  - It’s important that we consider these expectations when we analyze data
- How will the data be used for or against the people who generated the data?
  - Will the data be **anonymized**, and interventions only on the group level?
  - Will the data be used to allocate **rewards, punishment, or both**?
  - How much do the results of your analysis reflect **conscious decisions**?
  - Can this data be used to **empower those being analyzed**, instead of those ordering the analysis?
- **What do those being analyzed think of your plan?**
  - In my experience, a lot of people (at least will say they) are excited about these kinds of analyses
  - If they mention issues, it often helps make the analysis better!
  - It’s **always** better to be doing analyses that have buy-in from those being analyzed!

Please complete the review survey at:

tinyurl.com/PAweek6survey

Then you’re free to go! Have a great weekend!