HUDK 4051: LEARNING ANALYTICS: PROCESS & THEORY

2/2/17 11:47 AM

In the news

The New York Times

Here Is How to Fend Off a Hijacking of Home Devices

Will You Graduate? Ask Big Data

Assessing the Legacy of inBloom — Data&Society

Our Schools are Drowning in Data — SANTA FE REPORTER

IN TRUMP, TECH FINDS A TROLL IT CAN'T IGNORE — WIRED

DATA SCIENCE ASSOCIATION

Why Machine Learning Algorithms Fall Short

Knowledge to Practice Receives $6.5 Million in Series A Financing for Expansion into Continuing Medical Education Specialty Areas — YAHOO! FINANCE

Department of Education's Higher Ed Tech Plan Puts Focus on Data Analytics and Online Education — EdTech Focus On Higher Education

TYTO ONLINE

http://www.tytoonline.com/

LearnLaunch & Rocky Hill School Pioneer Bringing EdTech Development into Schools; Building New Innovation Center for Entrepreneurs, Students & Faculty

Penn State team using artificial intelligence to improve education experience — CENTRE DAILY TIMES

THE WALL STREET JOURNAL.

What's Better in the Classroom—Teacher or Machine?

'Edtech' is attracting investors, but some parents prefer traditional teaching methods

SmartDataCollective

Challenges of Big Data in Education

ENVIRONMENTAL DATA & GOVERNANCE INITIATIVE

INSIDE HIGHER ED

Data Fears and Downloads .... Data Rescue NYC

Direct-to-Student Data

DOMINO DATA SCIENCE POPUP

https://popup.dominodatalab.com/

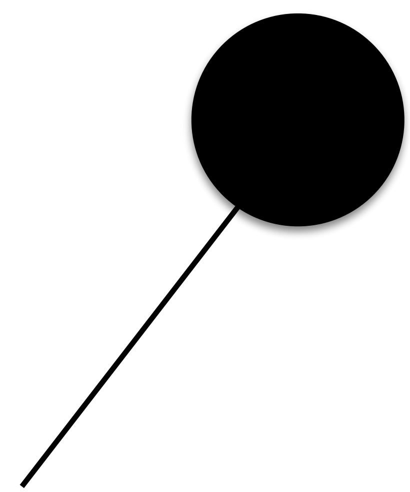
Xiaoting & LAKPOD

Soundcloud*

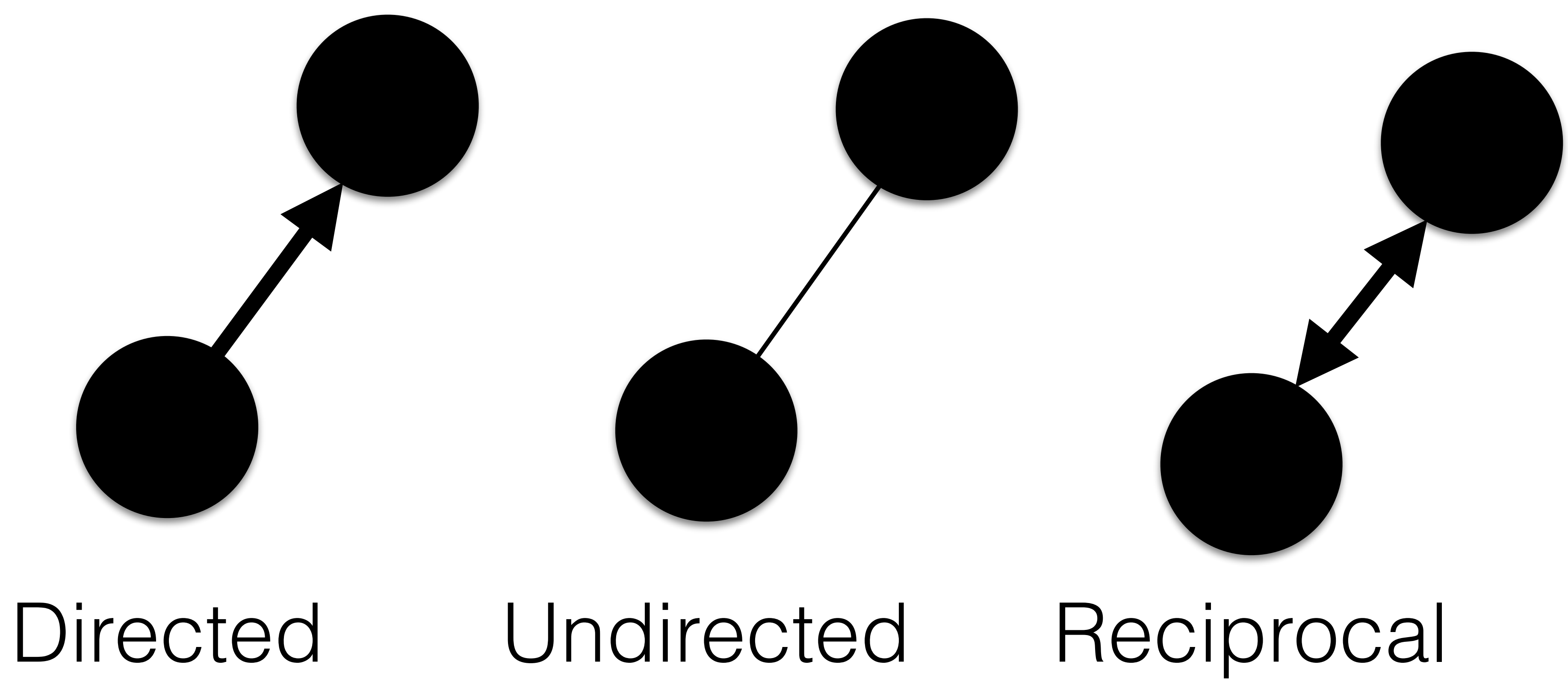
Social Network Analysis Recap

Networks

(Graphs)



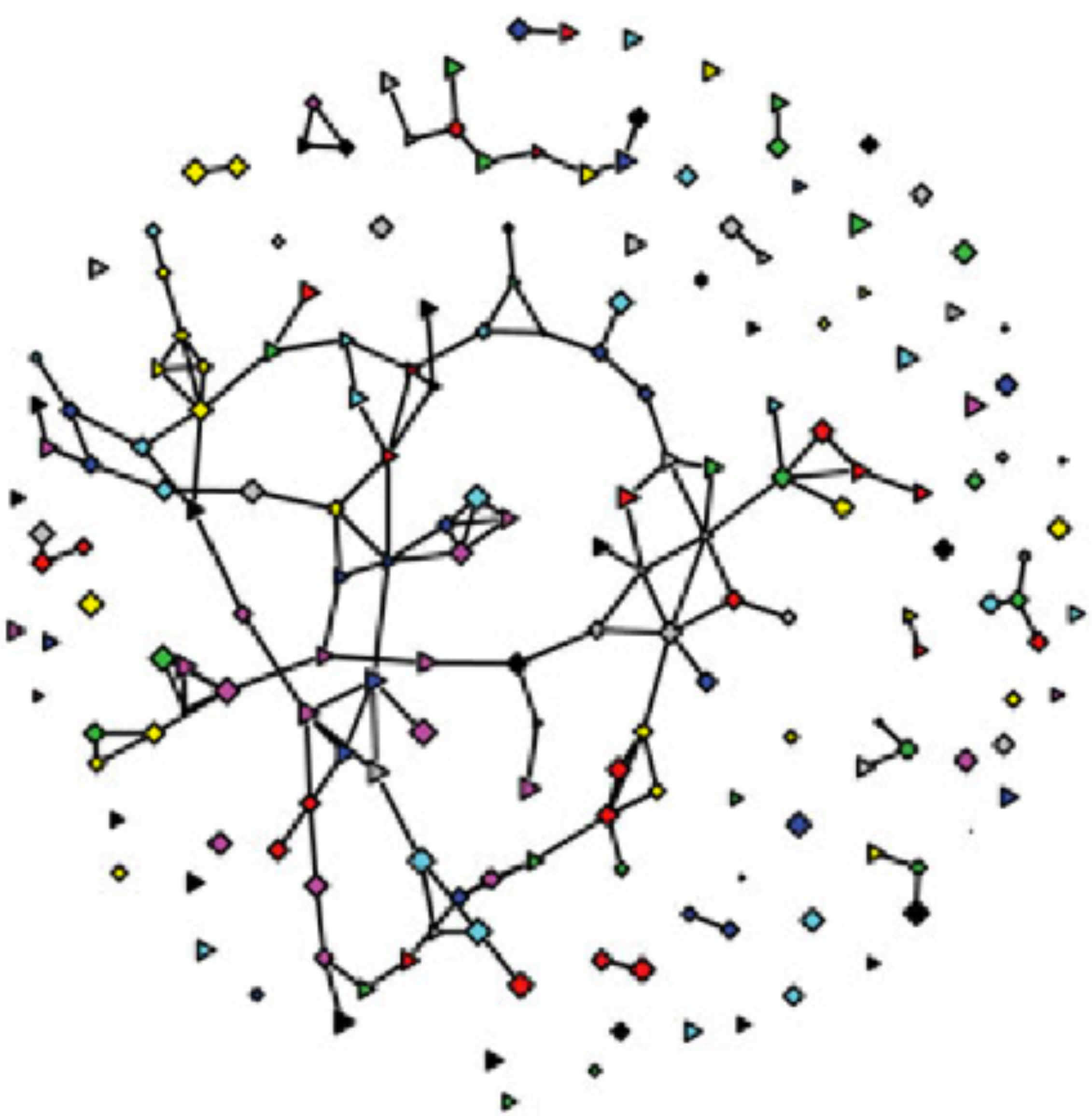
Networks



Degree

The number of links to other nodes in the network

Undirected

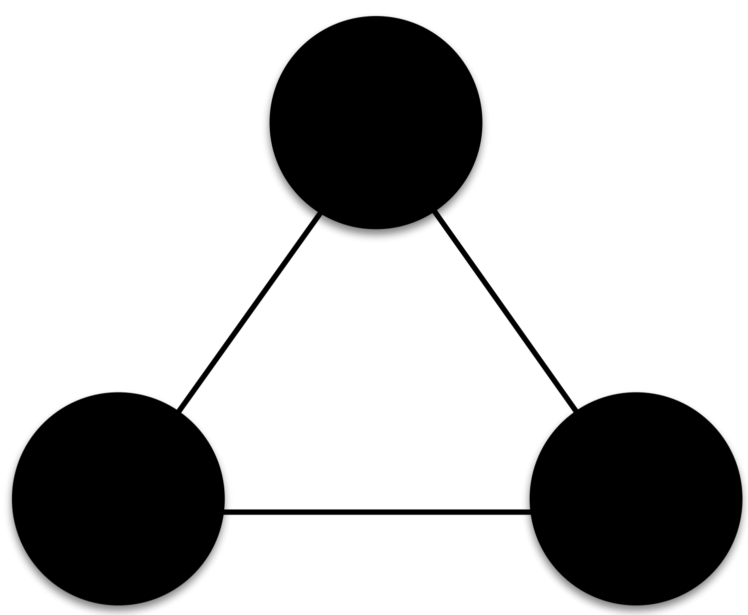


Directed

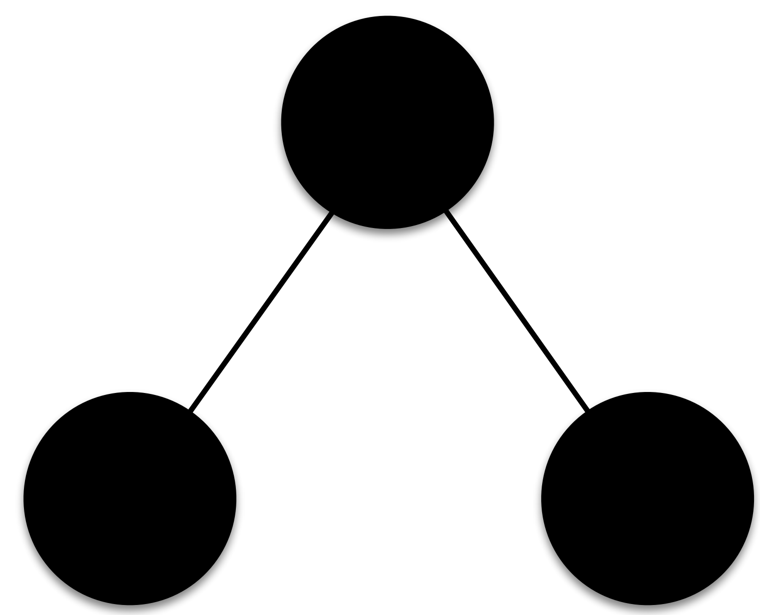
Indegree = Popularity
Outdegree = No shame

Density

How close is the graph to the maximal number of links



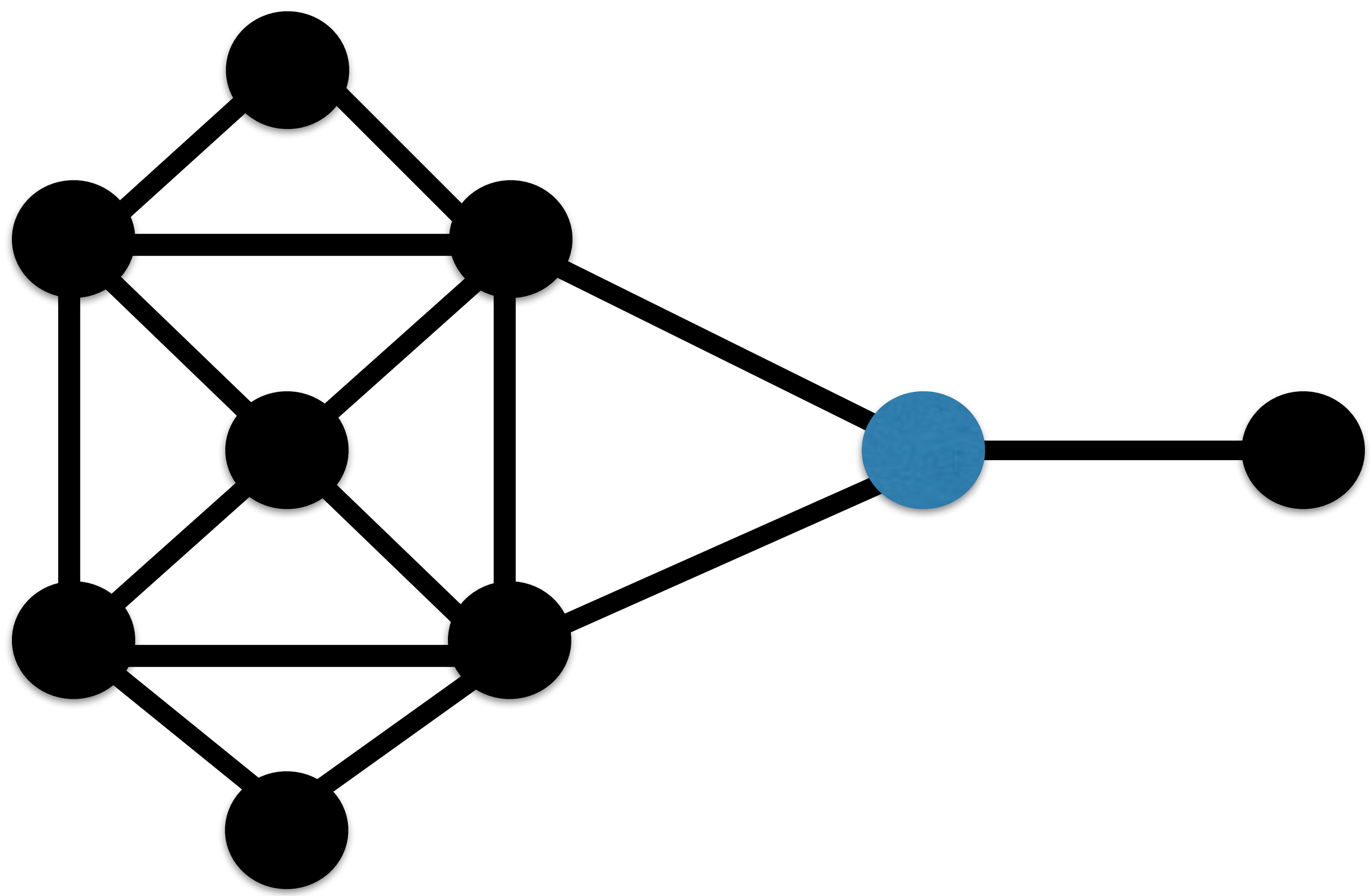
3 actual
3 possible
Density = 1



2 actual
3 possible
Density = 0.67

Betweenness Centrality

The extent to which a node lies between other nodes



It is equal to the number of shortest paths from all nodes to all others that pass through that node

Modularity

The fraction of the edges that fall within the given groups minus the expected such fraction if edges were distributed at random

$$
\begin{aligned}
Q_S &= \frac{1}{2\bar{w}} \sum_i \sum_j \left(\bar{w}_{ij} - \frac{\bar{w}_i \bar{w}_j}{2\bar{w}} \right) \delta(C_i, C_j) \\
&= \frac{1}{4w} \sum_i \sum_j \left(w_{ij} + w_{ji} - \frac{(w_i^{\text{out}} + w_i^{\text{in}})(w_j^{\text{out}} + w_j^{\text{in}})}{4w} \right) \delta(C_i, C_j) \\
&= \frac{1}{4w} \sum_i \sum_j \left[\left(w_{ij} - \frac{w_i^{\text{out}} w_j^{\text{in}}}{2w} \right) + \left(w_{ji} - \frac{w_i^{\text{in}} w_j^{\text{out}}}{2w} \right) \right] \delta(C_i, C_j) \\
&= -\frac{1}{(4w)^2} \sum_i \sum_j (w_i^{\text{out}} - w_i^{\text{in}})(w_j^{\text{out}} - w_j^{\text{in}}) \delta(C_i, C_j) \\
&= Q_D - \frac{1}{(4w)^2} \sum_i \sum_j (w_i^{\text{out}} - w_i^{\text{in}})(w_j^{\text{out}} - w_j^{\text{in}}) \delta(C_i, C_j).
\end{aligned}
$$

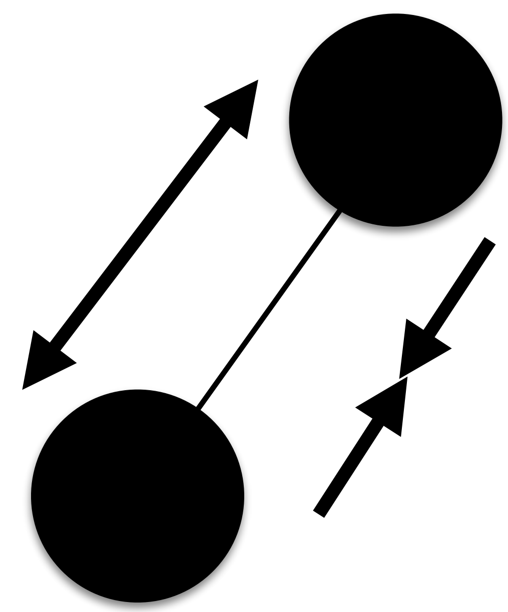
How do we make the network look nice?

Force directed graphing

- Attractive forces

  Springs

  Hooke’s Law: $F = kX$



- Repulsive forces

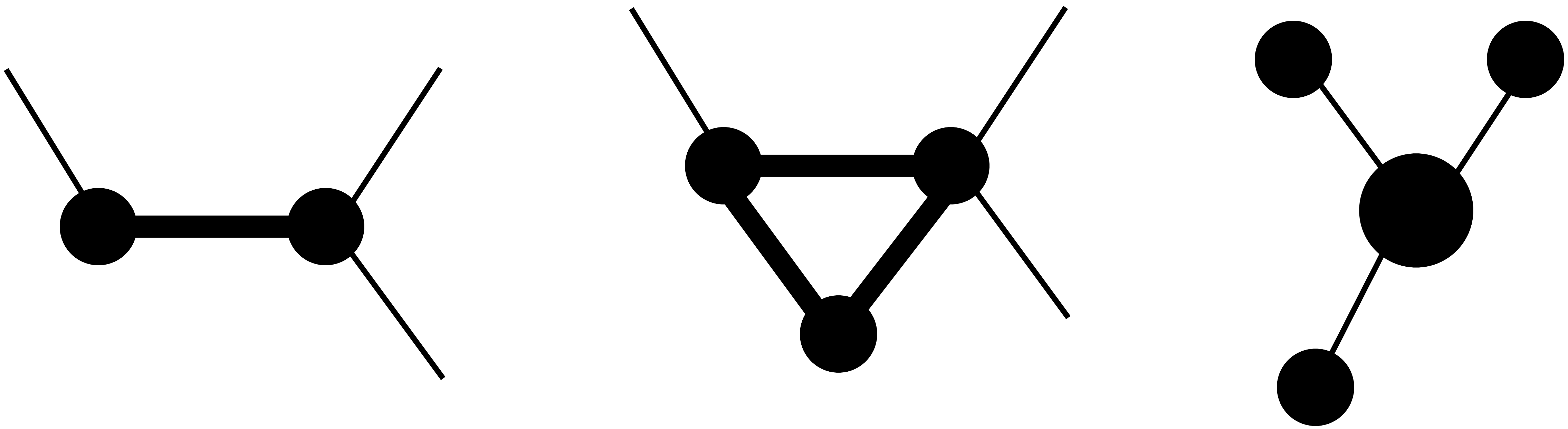
  Electrons

  Coulomb’s Law: $|\mathbf{F}| = k_e \frac{|q_1 q_2|}{r^2}$

https://youtu.be/YGDvR6CRwEc

SNA: Cliques & Subnetworks

Dyads, Triads, Ego-Centric Networks



- Ego - node of interest
- Neighborhood - nodes connected to ego at some path length (usually 1)
- Neighborhood can in In/Out, weak/strong, N-step

Larger Structures

- Need to make an argument that nodes *belong together*

- Maximal clique = a clique that is not included in a larger clique

- Lots of ways to do this (density, Strength of ties, etc)

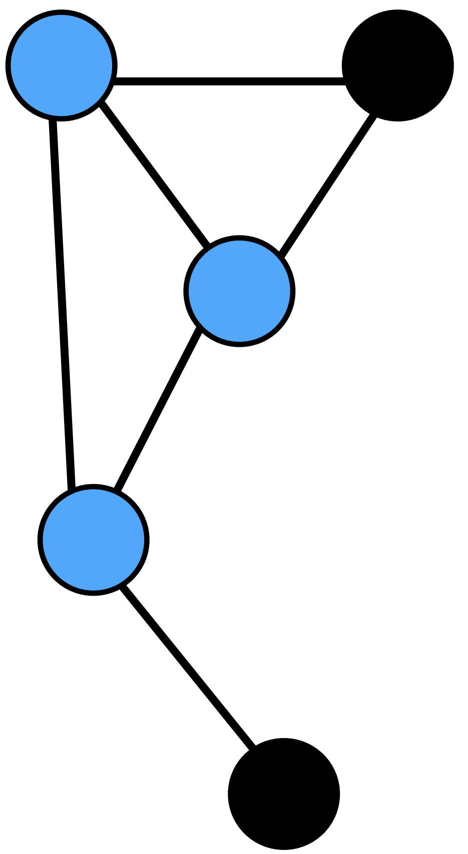
- Two main groups of methods: Bottom Up or Top Down

Bottom Up

- How does the macro emerge from the micro?
- Dyad = simplest group
- Clique/Complete Subgraph = maximum number of actors who have all possible ties present among themselves (all members are tied to all other members in the group)
- *Build* the network up from the bottom, comparing cliques sizes and relationships with each other

Algorithms

- <u>Greedy algorithm</u>: makes the locally best decision at each stage
    - Start with arbitrary node
    - Examine each node that it is connected to
    - If it is connected to every other node, keep, otherwise discard
    - Gotcha: Might not find the maximal clique only a local minima



Algorithms

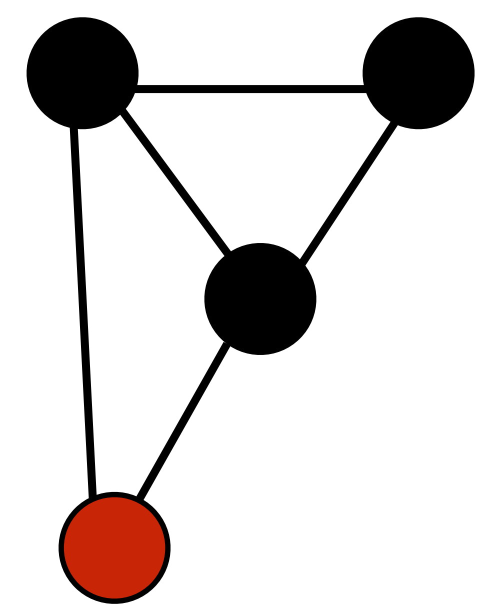
- Brute force algorithm: enumerate every possible combination of nodes
- Usually used to identify a clique of certain size (*k*)

Larger than cliques

- Co-membership of cliques (overlap)

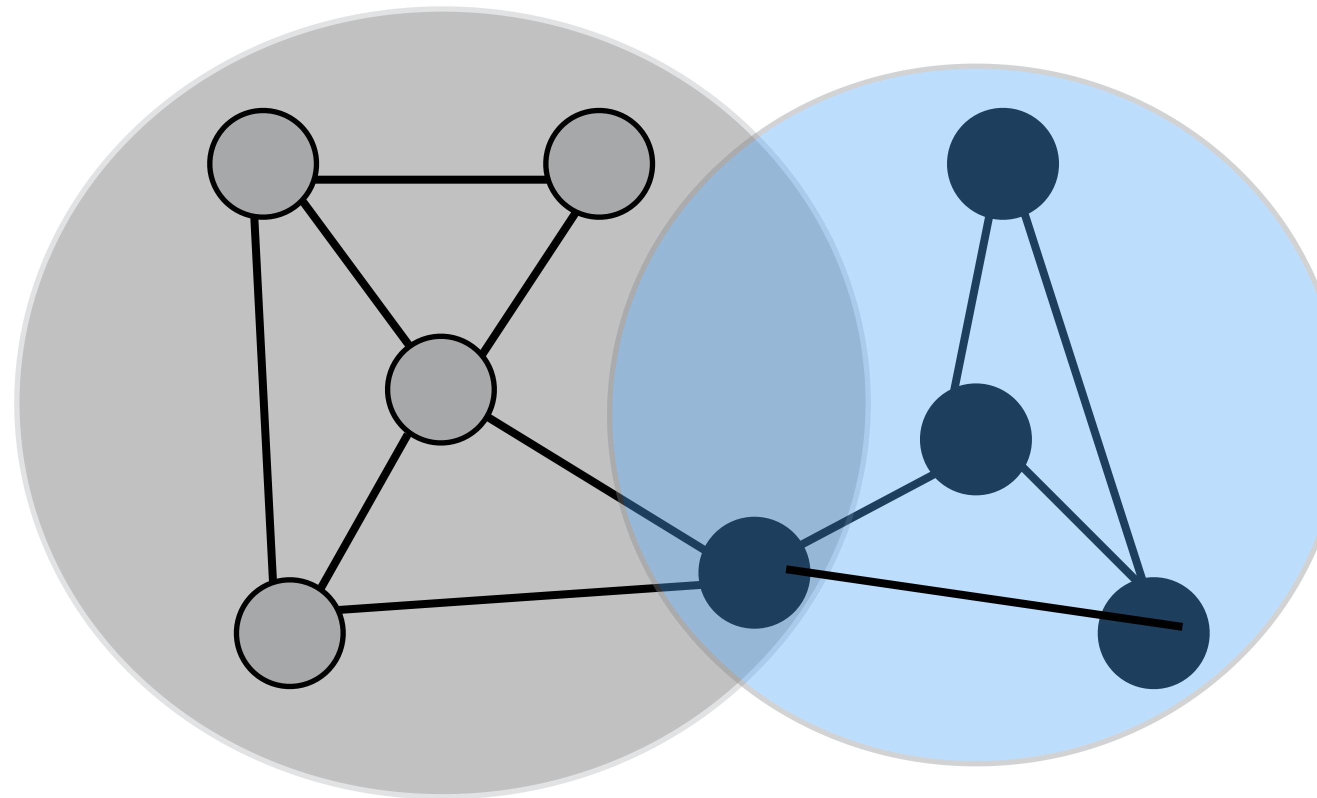
- N-cliques: Friend of a friend clique (can produce “stringy” groups)

- N-clans: Requires that all the ties among actors occur through other members of the group



Larger than Clans

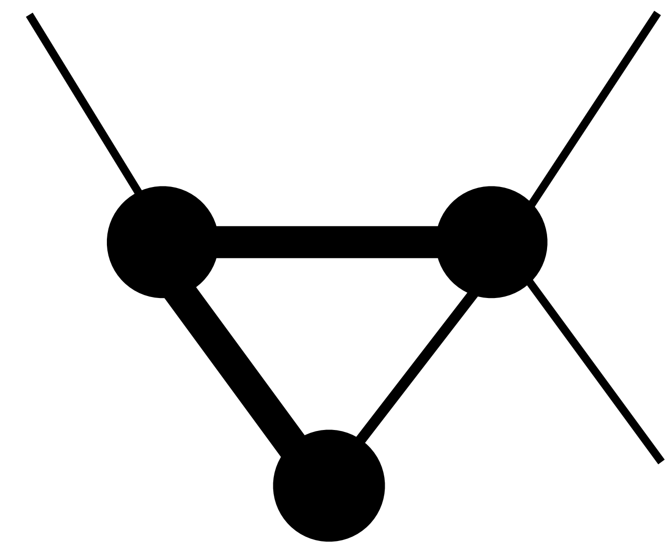
- <u>K-plexes</u>: A node is a member of a clique of size *n* if it has direct ties to *n*-*k* members of that clique
- <u>K-cores</u>: To be included in a core, an actor must be tied to all but *k* other actors in the group



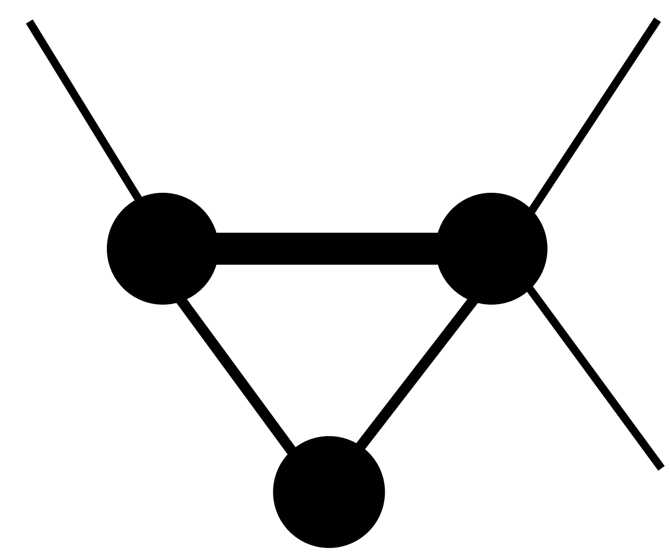
n >= 4, k = 2

F-Groups

- Incorporate strength of connection into grouping
- Equates grouping with balanced triads



weakly transitive



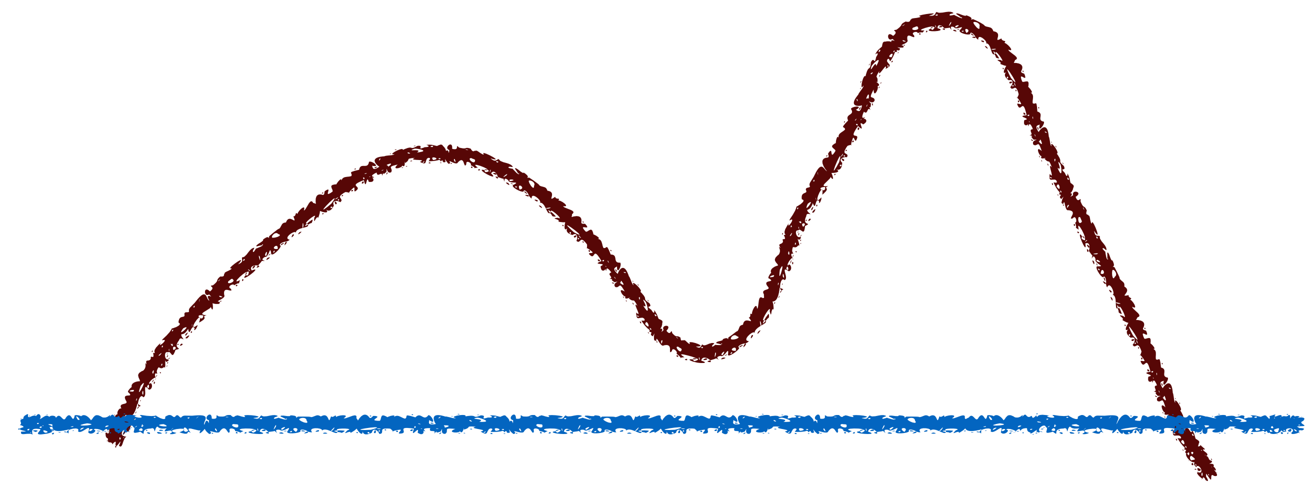
strongly transitive

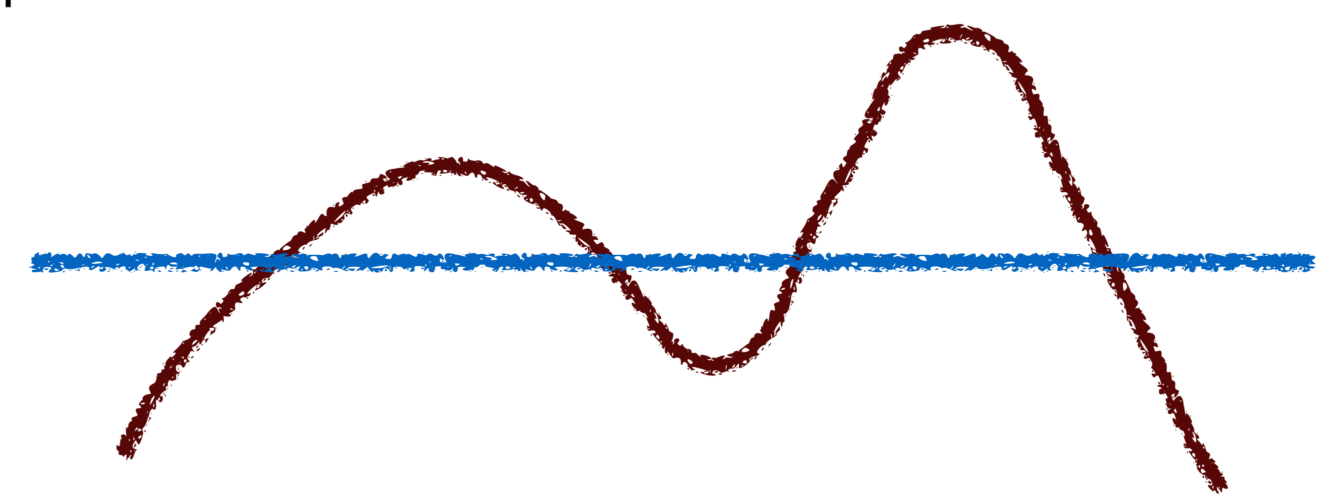
Top Down

- Looking for "holes" or "vulnerabilities" or "weak spots" in the overall structure
- Focuses on the constraints under which the network has developed
- Asks questions about overall system dynamics

Top Down Groupings

- Components: Disconnected parts of a graph
- Island Method: Take a measure of connectivity, remove nodes below a certain threshold. As if you are raising the water level around an island. What is left are the subgraphs
- Could also use connection weight or direction also



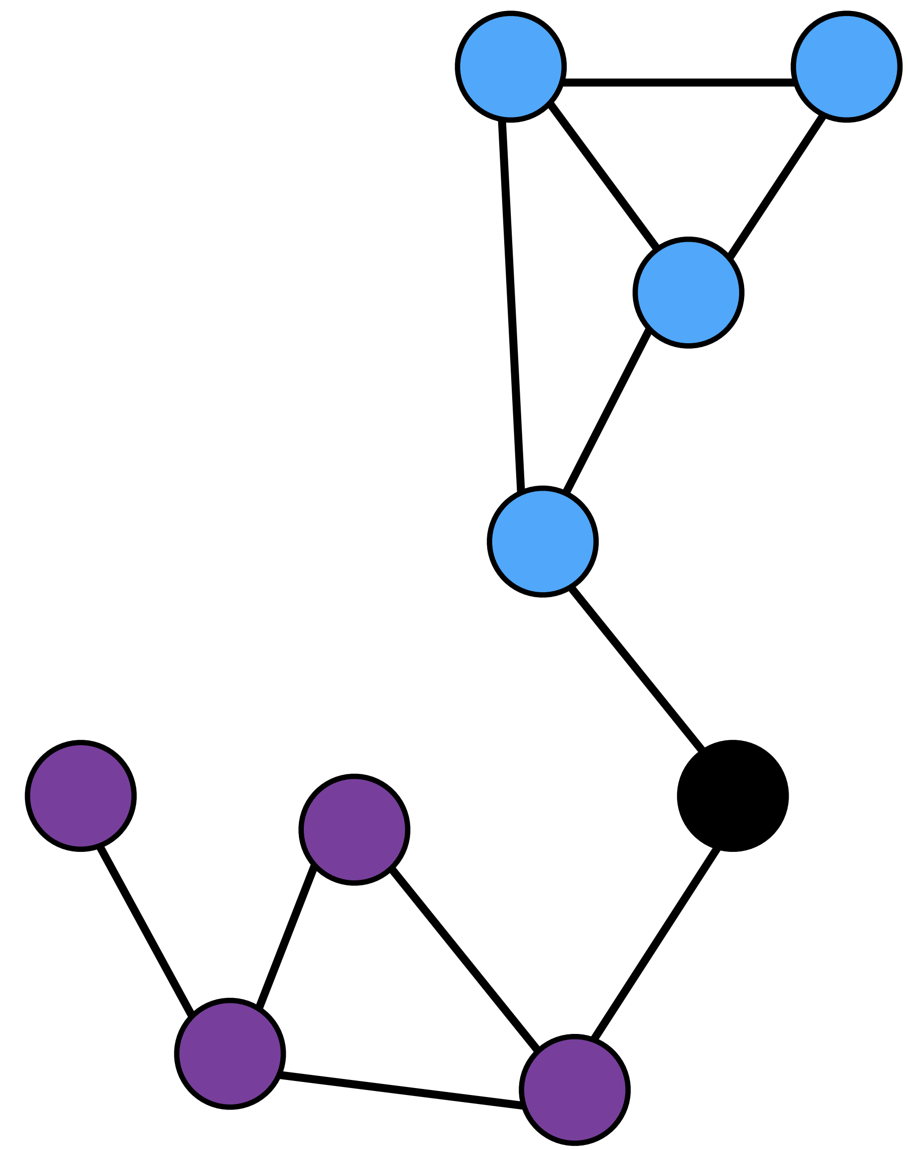


Blocks & Cut/Articulation points

- <u>Cutpoint</u>: If node is removed it creates two separate graphs

- <u>Block</u>: group without cut-points (non-separable)

- Could also use connection weight or direction



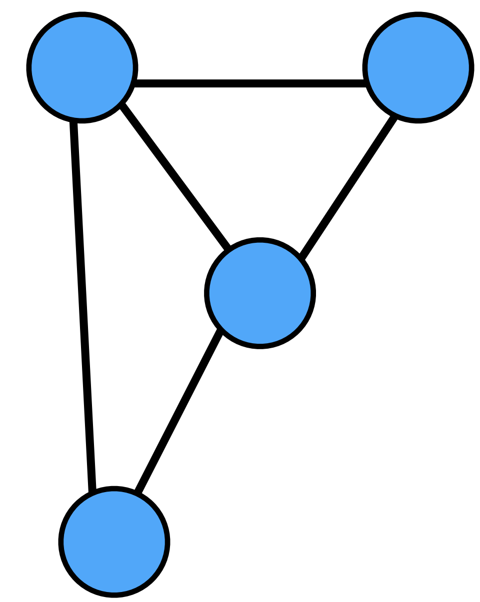
Bridges & Lambda Sets

- Bridge: If edge is removed it creates two separate graphs

- Lambda Sets: Measures the flow through the network

- Rank sets with respect to importance (amount of flow)

- Highlights points at which the fabric of connection is most vulnerable to disruption



Factions

- Imagine an idealized graph made up of three unconnected groups
- These are factions
- We can compare this imaginary graph to an actual graph to get a sense of how factionalized a network is
- Algorithm compares adjacency matrices to find the arrangement of nodes within the real data that most closely matches a theoretical factionalized network

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