EDCT GE2550: DATA SCIENCE IN EDUCATION

Big Data, Learning Analytics & The Information Age

In the news

March 01, 2016

Deloitte adds muscle in higher-ed tech consulting

JOHN JAY COLLEGE OF CRIMINAL JUSTICE

John Jay College of Criminal Justice seeks instructional staff for our **2016 CYBER SECURITY HIGH SCHOOL SUMMER PROGRAM**. The aim of this pilot summer program is to expand the pipeline of college-ready NYC high school students who enter college prepared to pursue a path to become the next generation of cybersecurity experts. The program will run from July 5-28. Instructional staff will participate in a series of professional development sessions in advance of the summer program.

Forbes

QUOTE OF THE DAY

"The rewards in business go to the man who does something with an idea."

- William Benton

New data shows obstacles for city's special education population

How Math Geeks Almost Doxxed Banksy

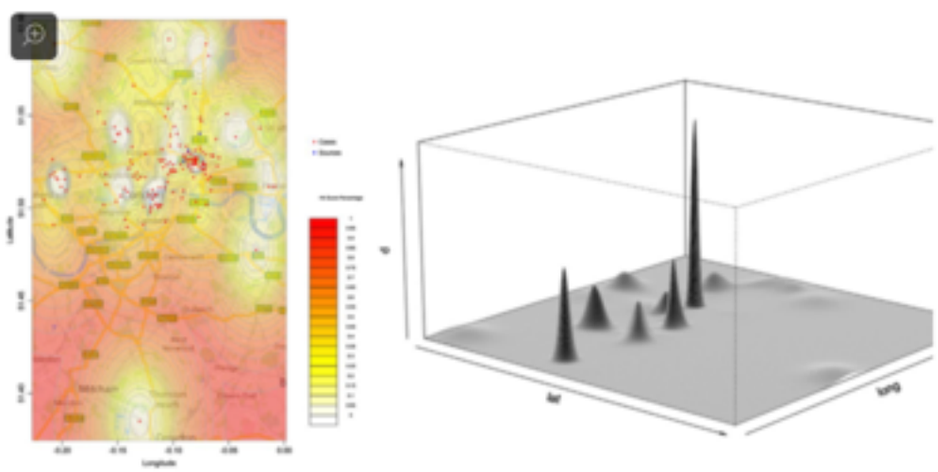
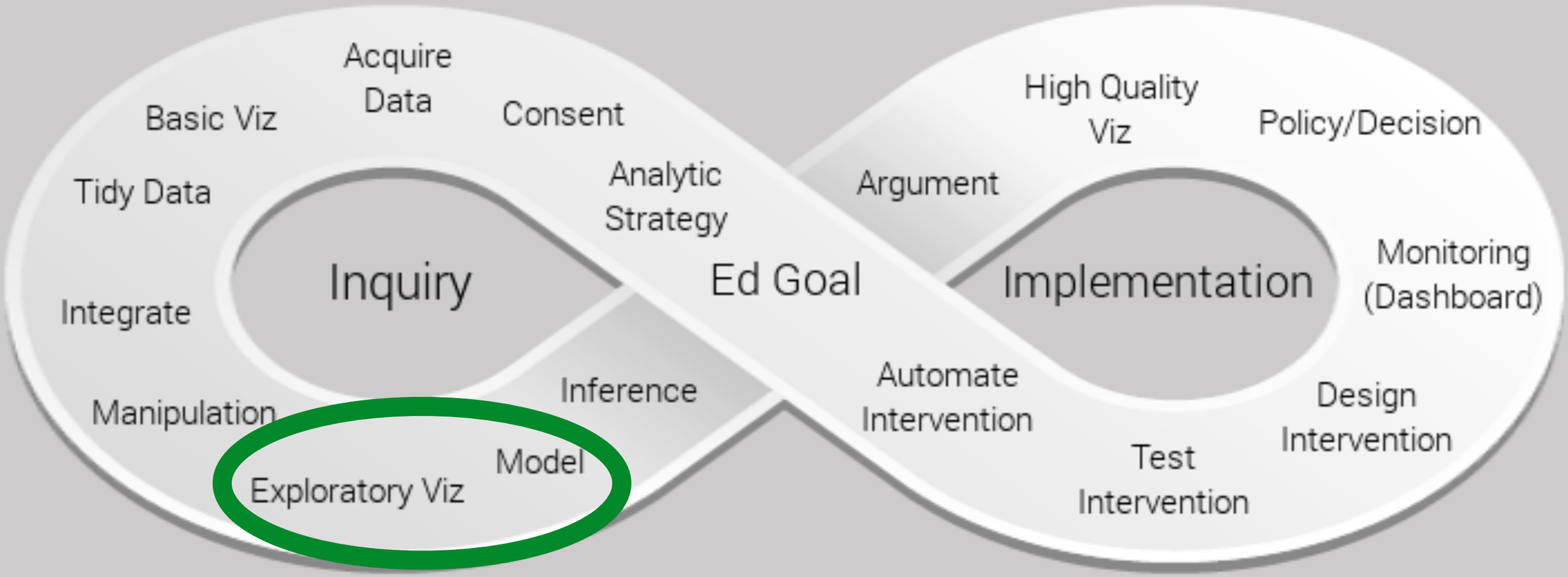


Image: Hauge et al., 2016/Journal of Spatial Science

Today

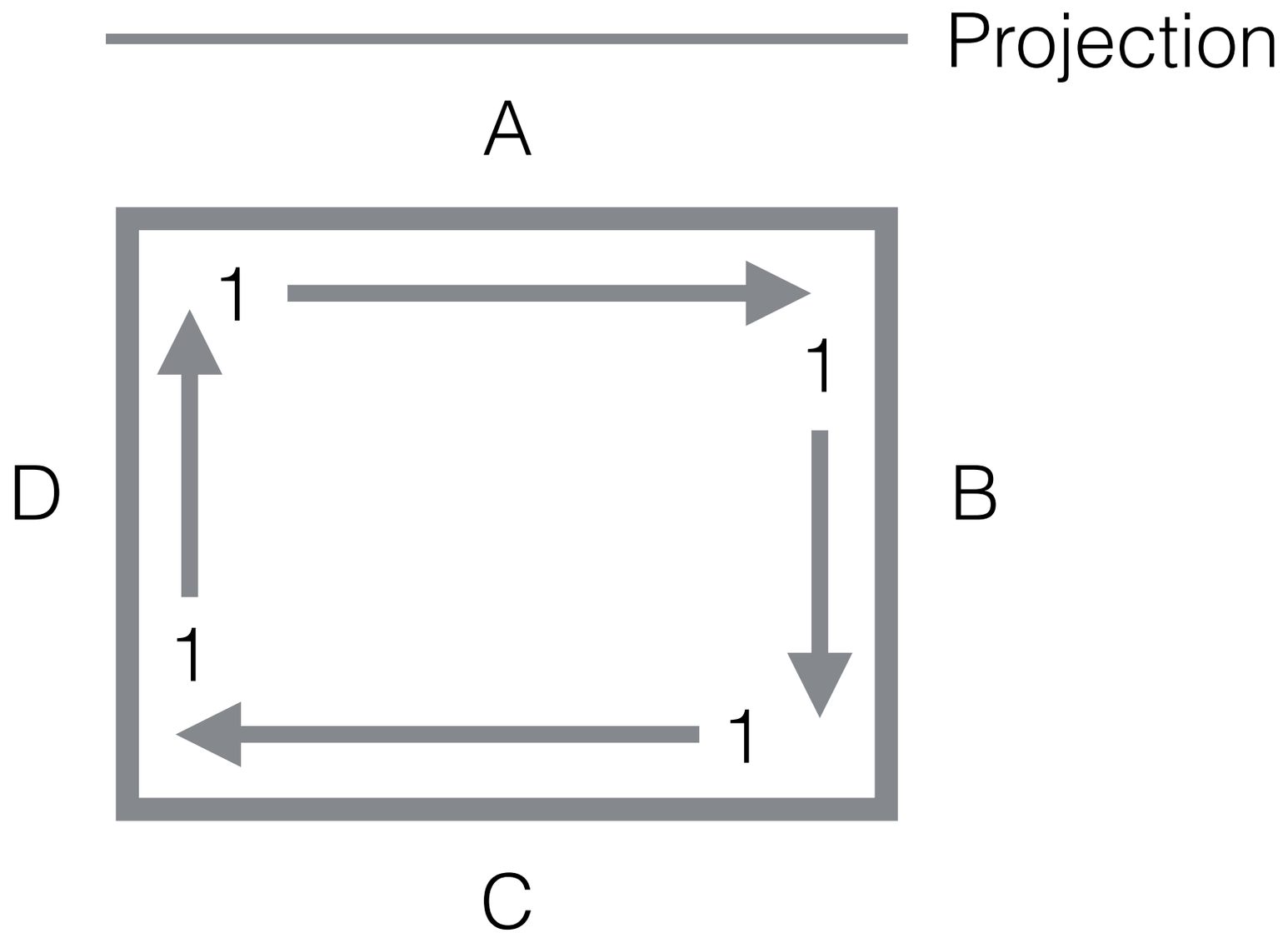
In the news	6:45 - 6:50
Quiz	6:50 - 7:00
Privacy	7:00 - 7:10
Hypothesis Generation	7:10 - 7:40
Social Network Analysis	7:40 - 7:50
Twitter	7:50 - 8:20

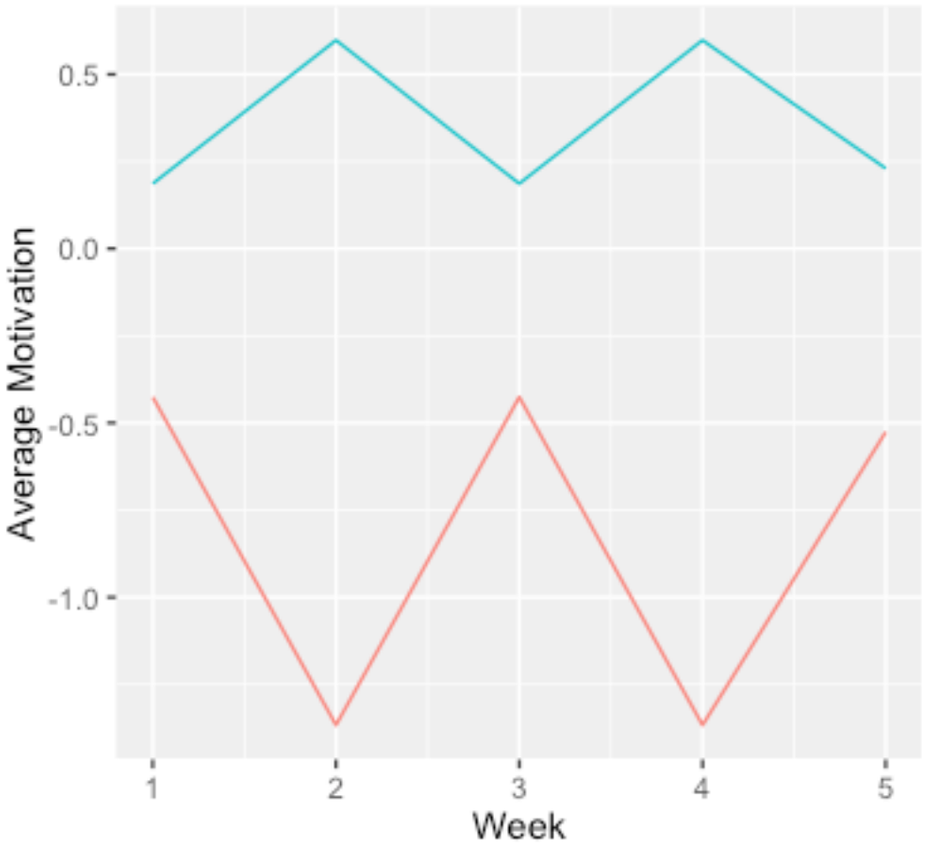
Ed Data Science Cycle



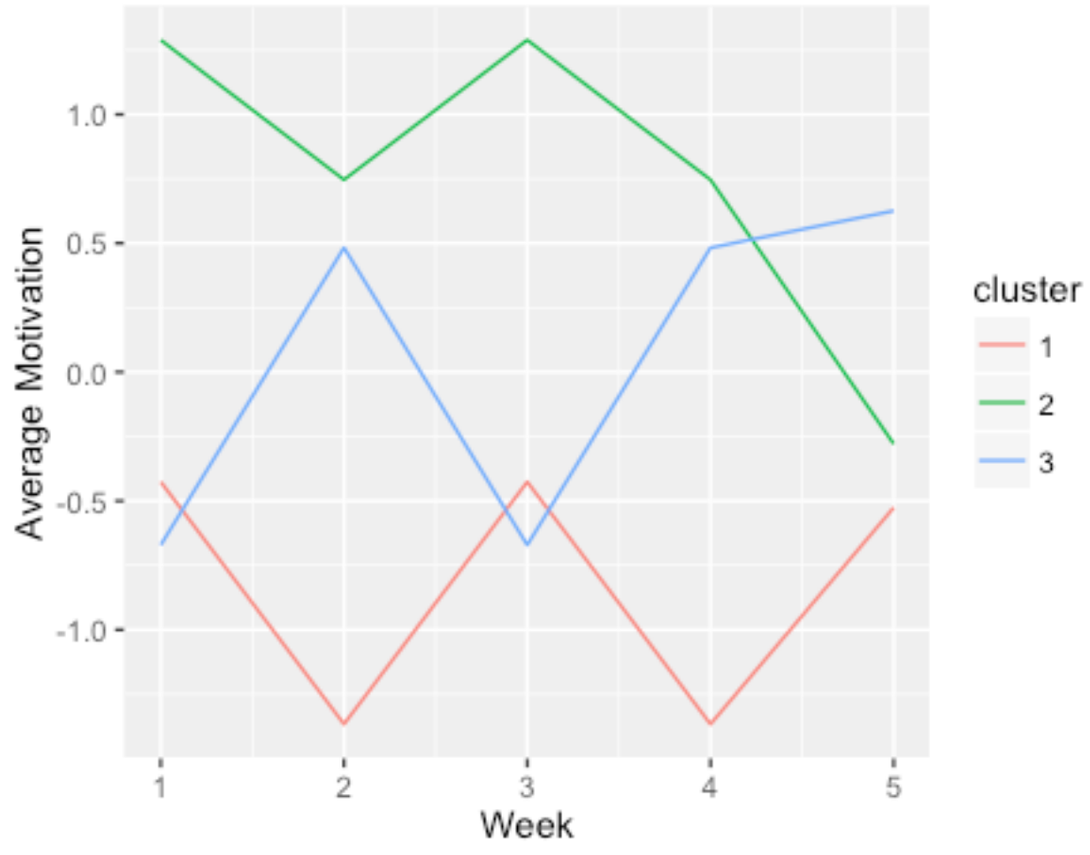
Quiz

bit.ly/1QsU0HK





Is there any information in missingness?



Student Privacy

Privacy Generally

- Companies have, on average, 75,000 data points on each person in the US
- Data is a $300 billion-a-year industry
- Acxiom made $1.1billion last year
- Federal Trade Commission recommended that Congress consider legislation overseeing online privacy, data security and breach notification, and data brokers
- Google plans to “close the data loop” through “Internet of Things”

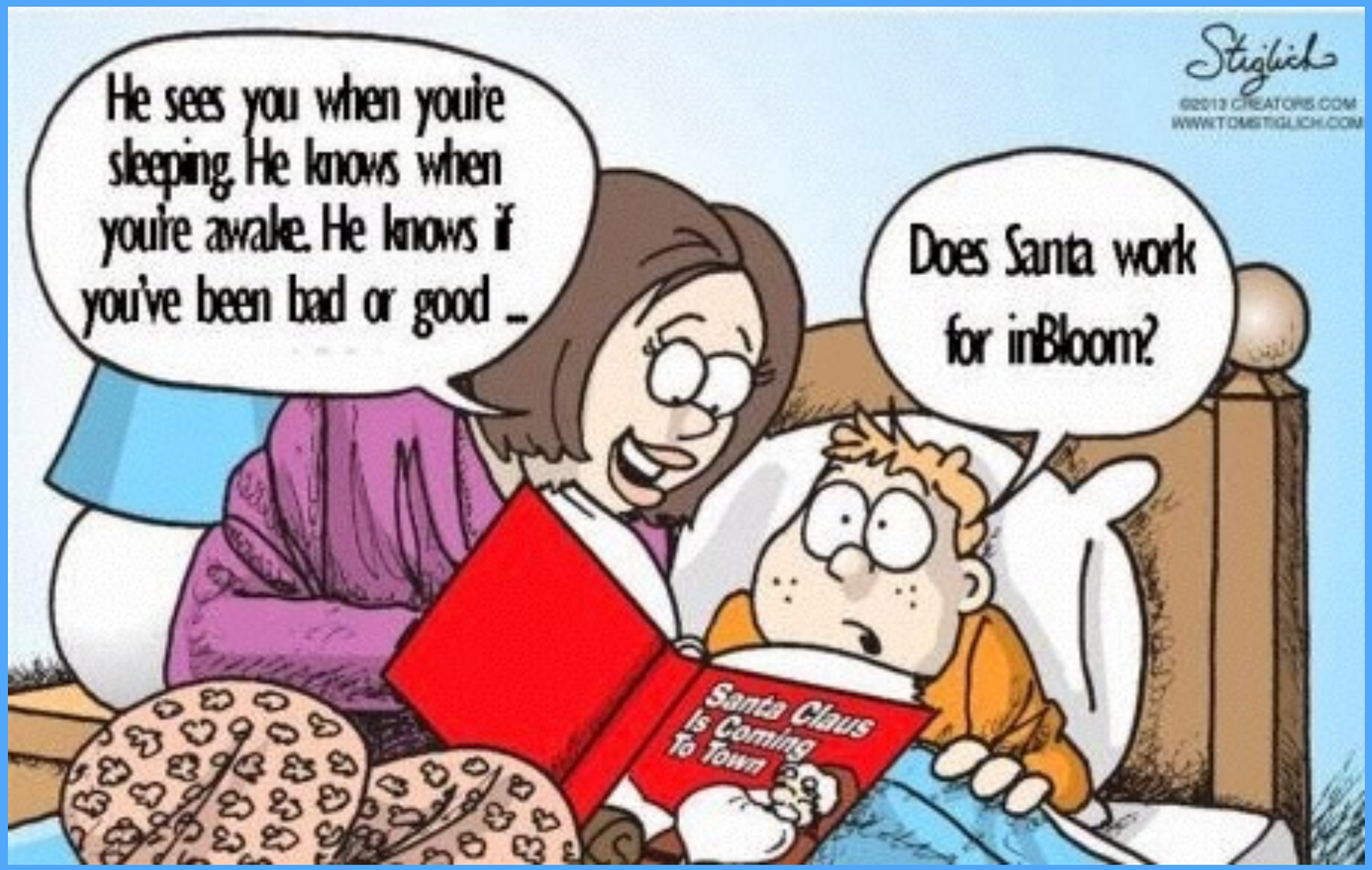
Ed Specific

- Data collection has been happening for a long time in education, but there are barriers to integration

- Family Educational Rights and Privacy Act (“FERPA”) and the Protection of Pupil Rights Amendment (“PPRA”) do not necessarily apply to data collected by companies

- The Software & Information Industry Association represents the industry and is against students being able to **access or amend** data points collected

- Fewer than 7 percent of districts that contract with cloud-service providers restrict the sale or marketing of student information


He sees you when you're sleeping. He knows when you're awake. He knows if you've been bad or good ...
Does Santa work for inBloom?
Santa Claus Is Coming To Town
Stiglich
©2013 CREATORS.COM
WWW.TOMSTIGLICH.COM

inBloom

2011 - April, 2014

- Non-profit
- Open-source infrastructure
- Funded by Gates Foundation, Carnegie Foundation, Carnegie Corporation, Washington Post, New Schools Venture Fund (+$100million)

- Store, clean, and aggregate data
- States and districts would then make the data available to district-approved third parties to develop tools for schools & teachers
- Benefits:
  - Interoperability of school data systems
  - Push the use of personalization

Why didn’t it work?

- Became a political liability
- Parents and teachers opposed its use:
  - 400 “optional” data points per student including SSN, medical history, legal records
  - Tone-deaf leadership, disregard for the concerns of parents & teachers
  - New York pulled out

Consequences?

- Increased political interest in student privacy
- Set back data sharing in the education sector
- Left the door open to private companies
- Changed the tone of Ed Tech relationship with stake holders
- Hard to tell: we just don’t know anything
    - EG - Informed consent

Hypothesis Generation

- We just don't know…anything
- One of the things computers can't do
- Activity:
  - "What might happen to student privacy in the future" write down as many hypotheses as you can

Social Network Analysis

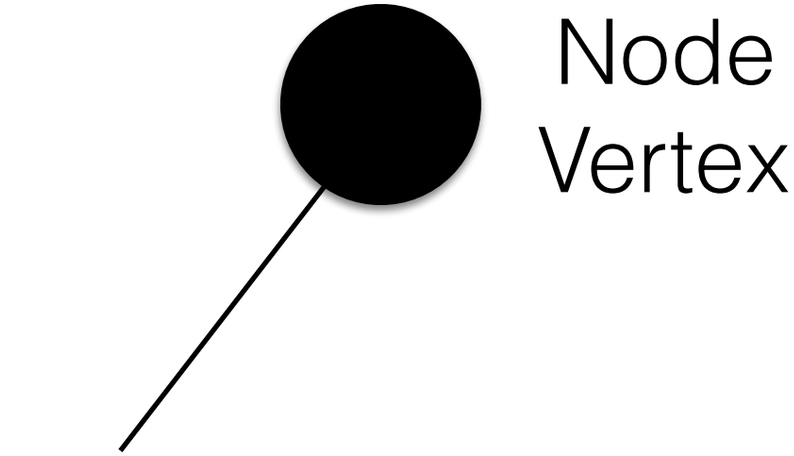
History

- Random network the size of the US (in 1950) would require at most two intermediaries to connect any two people (Kochen)

- Small World Experiments (Miligram)
  - Postcards sent to random people in Kansas
  - People instructed to send their postcard to a target person in Boston or someone that they think might know that person

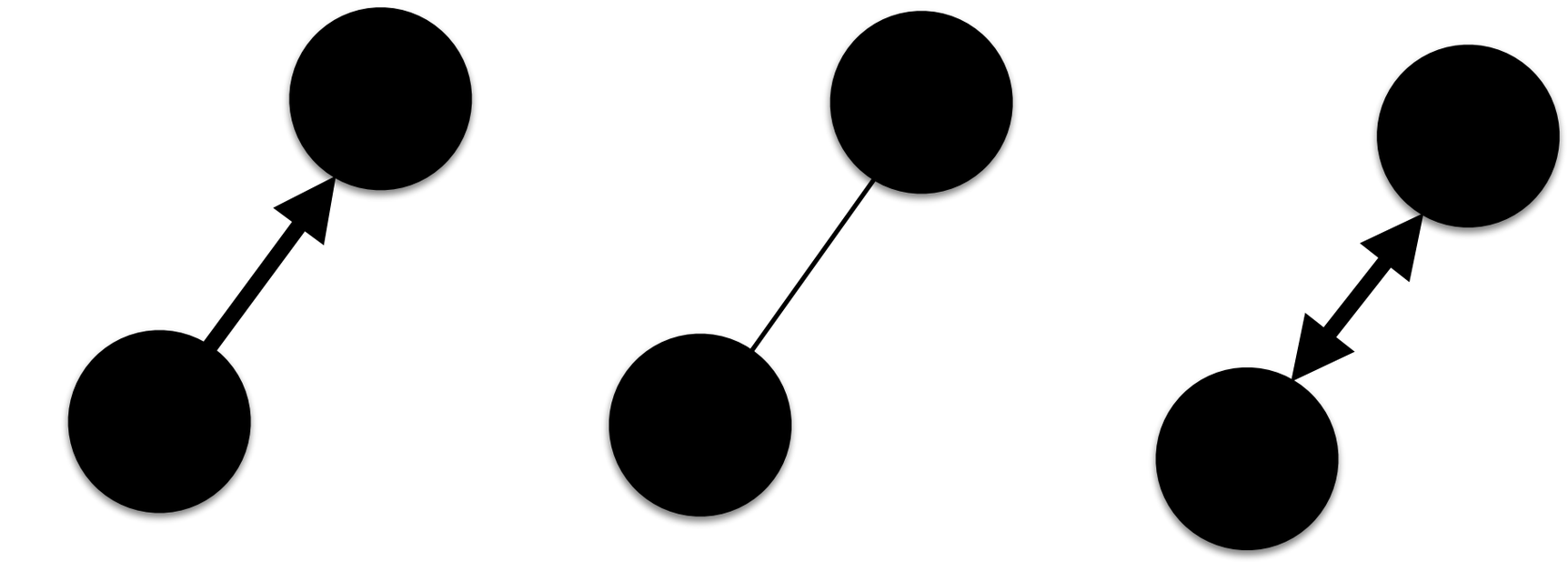
Networks

(Graphs)



Link
Edge
Tie
Arc

Networks

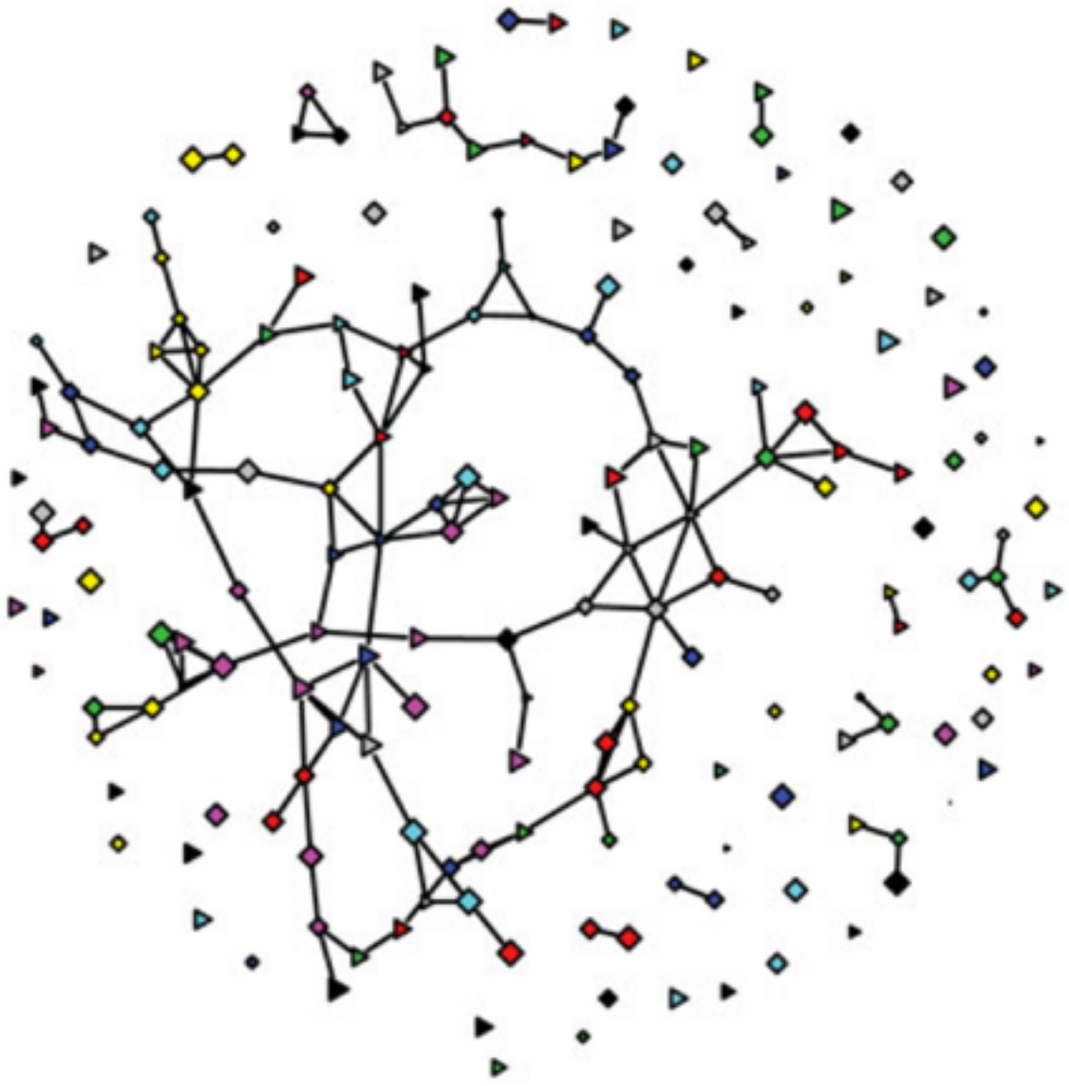


Directed Undirected Reciprocal

Degree

The number of links to other nodes in the network

Undirected



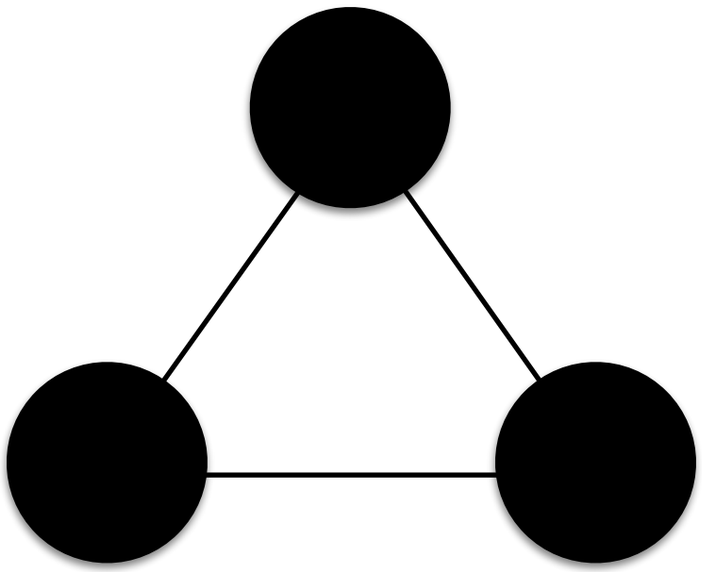
Directed

Indegree = Popularity

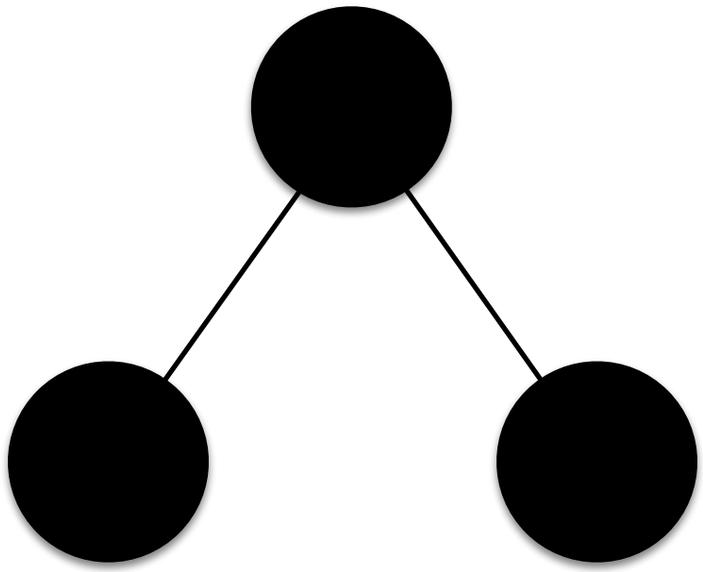
Outdegree = No shame

Density

How close is the graph to the maximal number of links



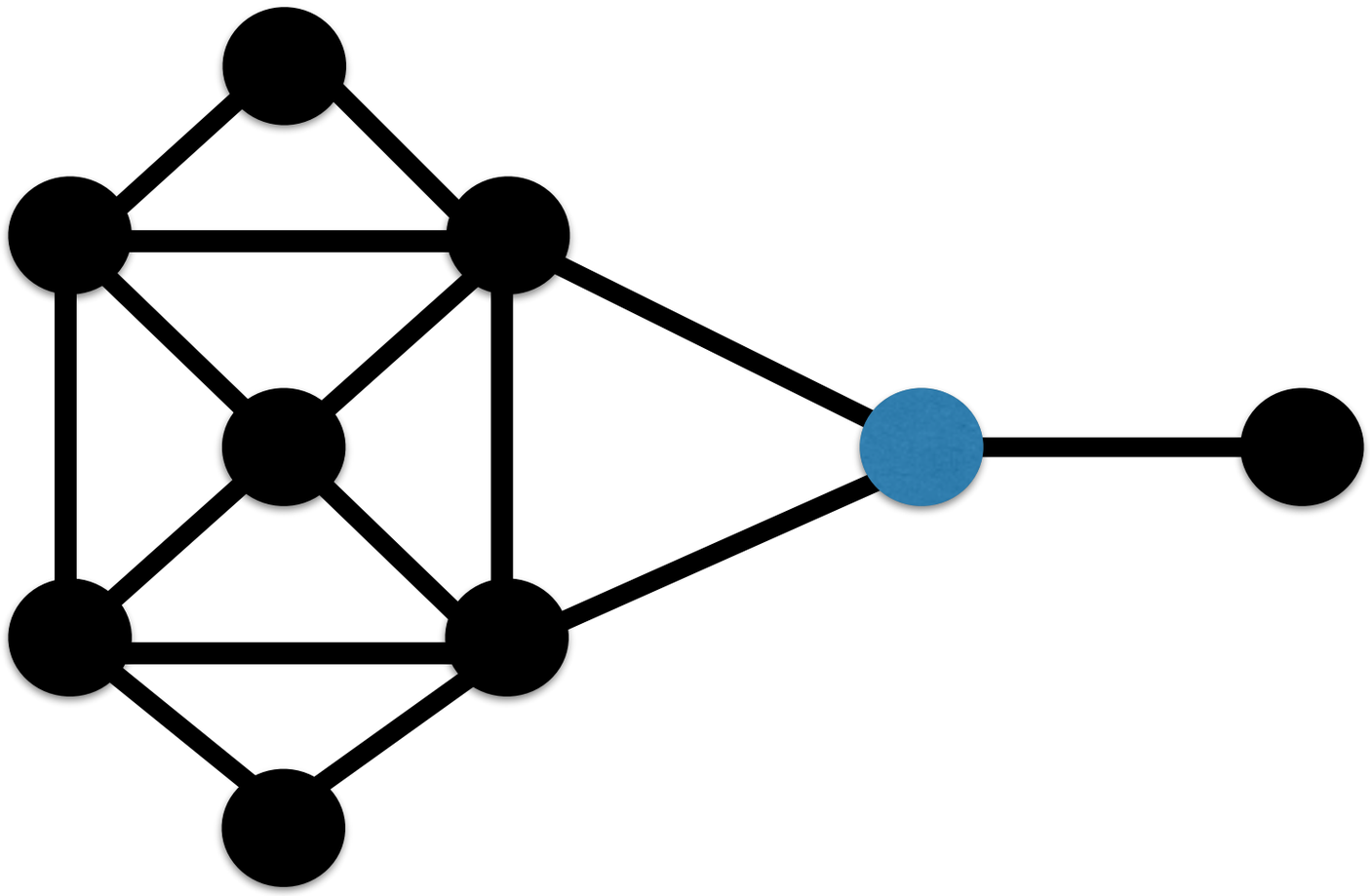
3 actual
3 possible
Density = 1



2 actual
3 possible
Density = 0.67

Betweenness Centrality

The extent to which a node lies between other nodes



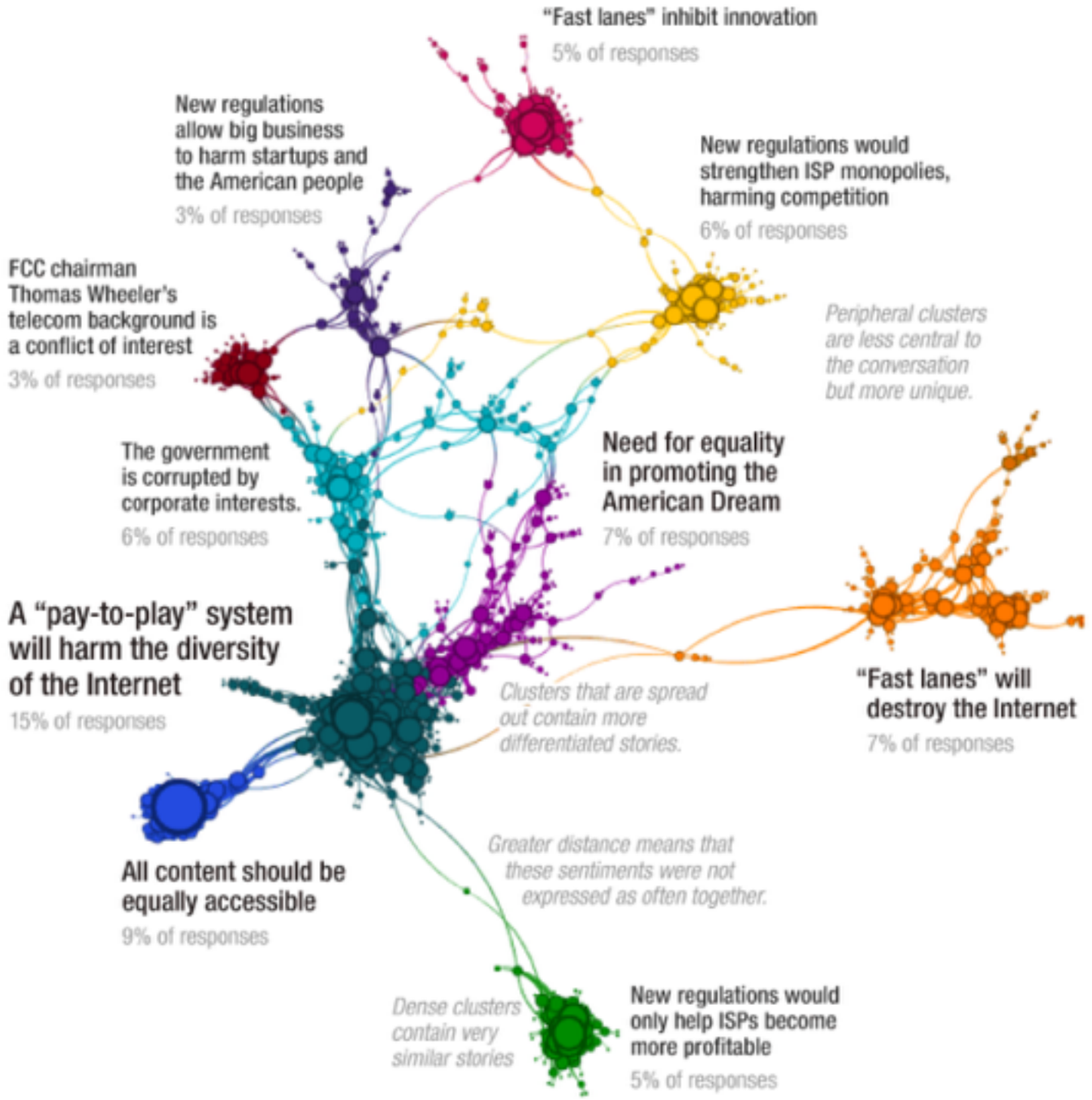
It is equal to the number of shortest paths from all nodes to all others that pass through that node

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?



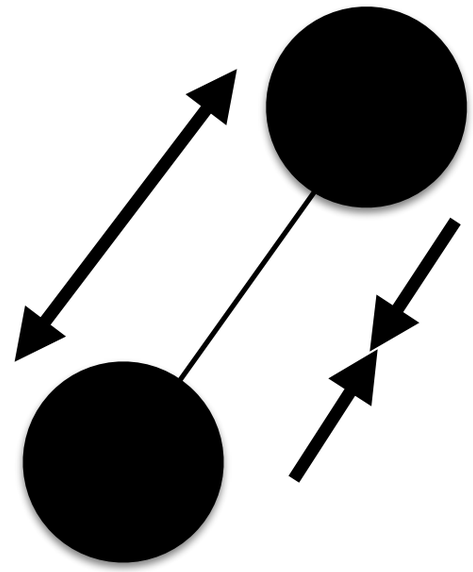
How do we make the network look nice?

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

http://kieranhealy.org/blog/archives/2013/06/09/using-metadata-to-find-paul-revere/