CPSC 534L: Information Networks – Introduction

http://www.cs.ubc.ca/~laks/534l/cpsc534l.html

Laks V.S. Lakshmanan

UBC

Fall 2021

Some Practical Matters 1/3

- Health and Safety first!
  - Masking is required!
  - Vaccination is highly recommended: ”UBC will require COVID-19 testing for all students, faculty and staff, with exemptions provided for those who are vaccinated against COVID-19.”

Some Practical Matters 2/3

- **Meeting time**: Mon/Wed 1:30–3:00 pm
  Pacific Time.
- **Meeting place**: ICCS 246.

- **Class mode**: in-person for now; but we should be prepared to *switch to online if needed.*

Some Practical Matters 2/3

- **Meeting time**: Mon/Wed 1:30–3:00 pm
  Pacific Time.
- **Meeting place**: ICCS 246.

- **Class mode**: in-person for now; but we should be prepared to *switch to online if needed.*

Some Practical Matters 3/3

Formal Prerequisites: None, but ...

a working knowledge of graphs, algorithms, basic theory, basic data mining, ML, basic DB will be assumed.

Online discussion: We will use piazza for online discussion on most course related matters. Sign up at piazza.com/ubc.ca/winterterm12021/cpsc5341

Some Practical Matters 3/3

Formal Prerequisites: None, but ...

a working knowledge of graphs, algorithms, basic theory, basic data mining, ML, basic DB will be assumed.

Online discussion: We will use piazza for online discussion on most course related matters. Sign up at piazza.com/ubc.ca/winterterm12021/cpsc5341

Course Objectives

- Learn about interesting, impactful, and challenging problems in information networks, recommender systems, and knowledge graphs.
  - Modeling, Prediction, & Optimization
  - Finding Communities
  - Information/Influence/Infection Propagation: modeling and optimization.
  - Recommender Systems: models, composite/holistic recs.
  - Knowledge Graph Completion, Question Answering, & Fact Checking.
- Emphasis:
  - Algorithmic Issues, Analysis
  - Fairness
- Focus on research.
- $\Uparrow$ not an exhaustive list!

Course Outline 1/3

- ✓ **Intro.**: different perspectives and terminologies, centrality, what we care about in this course.

- Structural Analysis over Networks:
  - Team Formation.
  - Community detection – General.
  - k-core and k-truss. k-core and k-truss decomposition.
  - k-core/k-truss communities.

Course Outline 1/3

- ✓ **Intro.**: different perspectives and terminologies, centrality, what we care about in this course.

- Structural Analysis over Networks:
  - Team Formation.
  - Community detection – General.
  - k-core and k-truss. k-core and k-truss decomposition.
  - k-core/k-truss communities.

Course Outline 2/3

Fair Optimization

- Infection, Influence & Information Propagation: single and multiple campaigns; revenue maximization, misinformation (fake news) mitigation, filter bubbles.
- Recommender Systems: Content-based vs. Collaborative Filtering; Memory-based vs. Model-based; Recommendations of novel objects; Strategic/Holistic Recommendations.
- Knowledge Graphs: Construction, Augmentation, Completion, and Attacks.

Course Outline 2/3

Fair Optimization

- Infection, Influence & Information Propagation: single and multiple campaigns; revenue maximization, misinformation (fake news) mitigation, filter bubbles.
- Recommender Systems: Content-based vs. Collaborative Filtering; Memory-based vs. Model-based; Recommendations of novel objects; Strategic/Holistic Recommendations.
- Knowledge Graphs: Construction, Augmentation, Completion, and Attacks.

Course Outline 2/3

Fair Optimization

- Infection, Influence & Information Propagation: single and multiple campaigns; revenue maximization, misinformation (fake news) mitigation, filter bubbles.
- Recommender Systems: Content-based vs. Collaborative Filtering; Memory-based vs. Model-based; Recommendations of novel objects; Strategic/Holistic Recommendations.
- Knowledge Graphs: Construction, Augmentation, Completion, and Attacks.

Course Outline 3/3

- Your Talks (interspersed with lectures, per topic).
- Bibliography to choose papers from will be published topic-wise and will be constantly updated as we progress through the course. (Check out class piazza.)
- You're welcome to bring your own idea for a paper presentation: just run it by me to make sure the fit is appropriate.

Marking Scheme

- Homework Assignments: 20%.
- Class Participation: 5%
- Paper Presentation/Discussion: 30%
- Course Project (typically teams of size 2-3): 45%

  $\Rightarrow$ What each of these means for you.

- Some topics — assigned reading.
- Questions?

Plagiarism

- See http://www.cs.ubc.ca/about/policies/collaboration.shtml for department policy.
- Take the time to read it and understand it.
- Be sure to attribute everything that is not your own original idea/contribution, to the source you got it from.
- When in doubt, ask me.

Books, References, and Resources

No text book, but $\exists$ several excellent books:

- David Easley and Jon Kleinberg. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge Univ. Press, 2010.
- Jure Leskovec, Anand Rajaraman and Jeff Ullman. Mining Massive Datasets. 3rd Ed., 2014.
- K. Faust and S. Wasserman. Social Network Analysis. Methods and Applications. Cambridge: Cambridge University Press. 1994.
- Carrington, Peter J. , John Scott and Stanley Wasserman (Eds.). 2005. Models and Methods in Social Network Analysis. New York: Cambridge University Press.
- W. Chen, L., and C. Castillo. Information and Influence Propagation in Social Networks. Morgan & Claypool, October 2013. http://www.morganclaypool.com/doi/abs/10.2200/S00527ED1V01Y201308DTM03
- Xin Huang, L., and Jianliang Xu. Community Search over Big Graphs. Morgan & Claypool, July 2019.

More Resources

- Research Literature (WWW, KDD, ICDM, SDM, ICML, WSDM, VLDB, SIGMOD, ICDE, NeurIPS, SIGIR, RecSys, ACL, EMNLP, COLING, selected journals, ...).
- Course Notes.
- Discussion Report on Papers (for student talks).
- "Network Analytics" Reading Group (NARG): all of you are invited! Send me mail to sign up, if interested.

Tentative Schedule

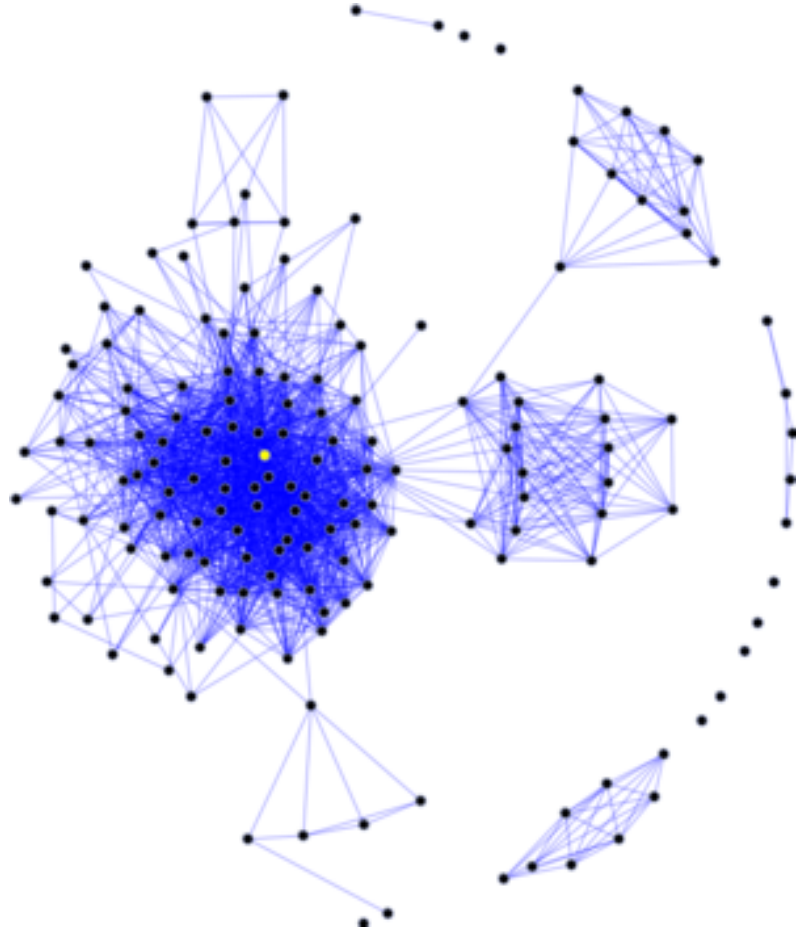
1. Intro, Structural Analysis/Networks – start.
2. Structural Analysis/Networks.
3. Structural Analysis/Networks, IM start
4. Influence Maximization.
5. Influence Maximization.
6. Recommender Systems.
7. Recommender Systems, Knowledge Graphs.
8. Knowledge Graphs.

- Schedule may evolve as we adjust to pace.
- Your talks will be interleaved with lectures.

Brass Tacks: Graphs/Networks Are Everywhere

- *Social networks*: FaceBook, Twitter, LinkedIn, ...
- *Information networks*: e.g., citation networks, collaboration networks, biological (e.g., protein-protein interaction), road networks, ...
- user-product graphs (e.g., in recommender systems).
- *knowledge graphs*.
- Brain ...

Some Example Networks



Social/Information Networks

- Didn't quite start like that!
- What is a SN from a sociologist's perspective?
- Conventional data is tabular:
  e.g., *R(Name, Age, Gender, Salary).*
- rows="cases", "actors","subjects", "observations" (tuples/users) and
- columns = "variables", "measures" (attributes).

Graphs/Networks: different perspectives

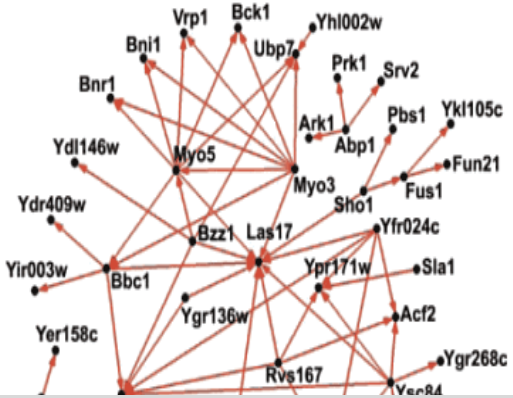
- Sociology (matrix): rows & columns = users[1]
- cell = relationship (tie) between users. (possibly weighted and/or labeled).
- nodes/vertices = "users";
- edges (links) = relationships.
- attributes (properties) may qualify nodes and/or edges.

[1] groups/communities could be nodes too.

Network Types

Social/Information Networks – Intro.

- a network may be *homogeneous*:
  e.g., collaboration graph between scientists; facebook, LinkedIn networks; citation networks, some biological networks (protein-protein interaction) .



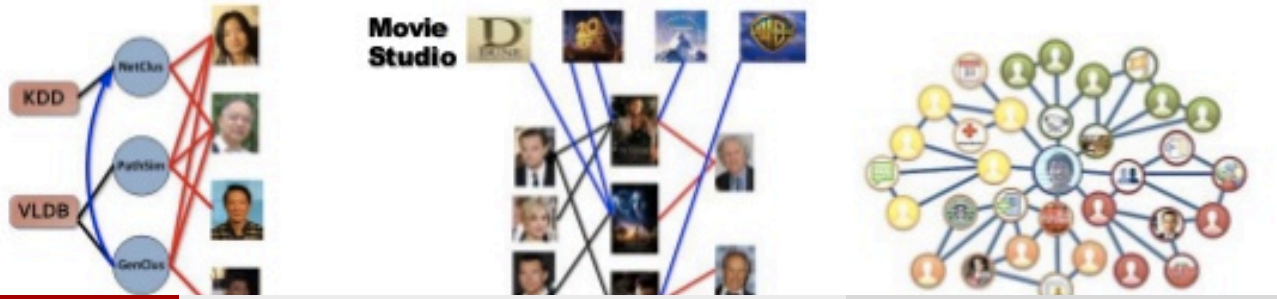
Network Types

Social Networks – Intro.

- Or it may be *heterogeneous*: e.g., publication datasets – pubmed, DBLP, etc.; flickr; last.fm; flixster; IMDB; any recommender system; …

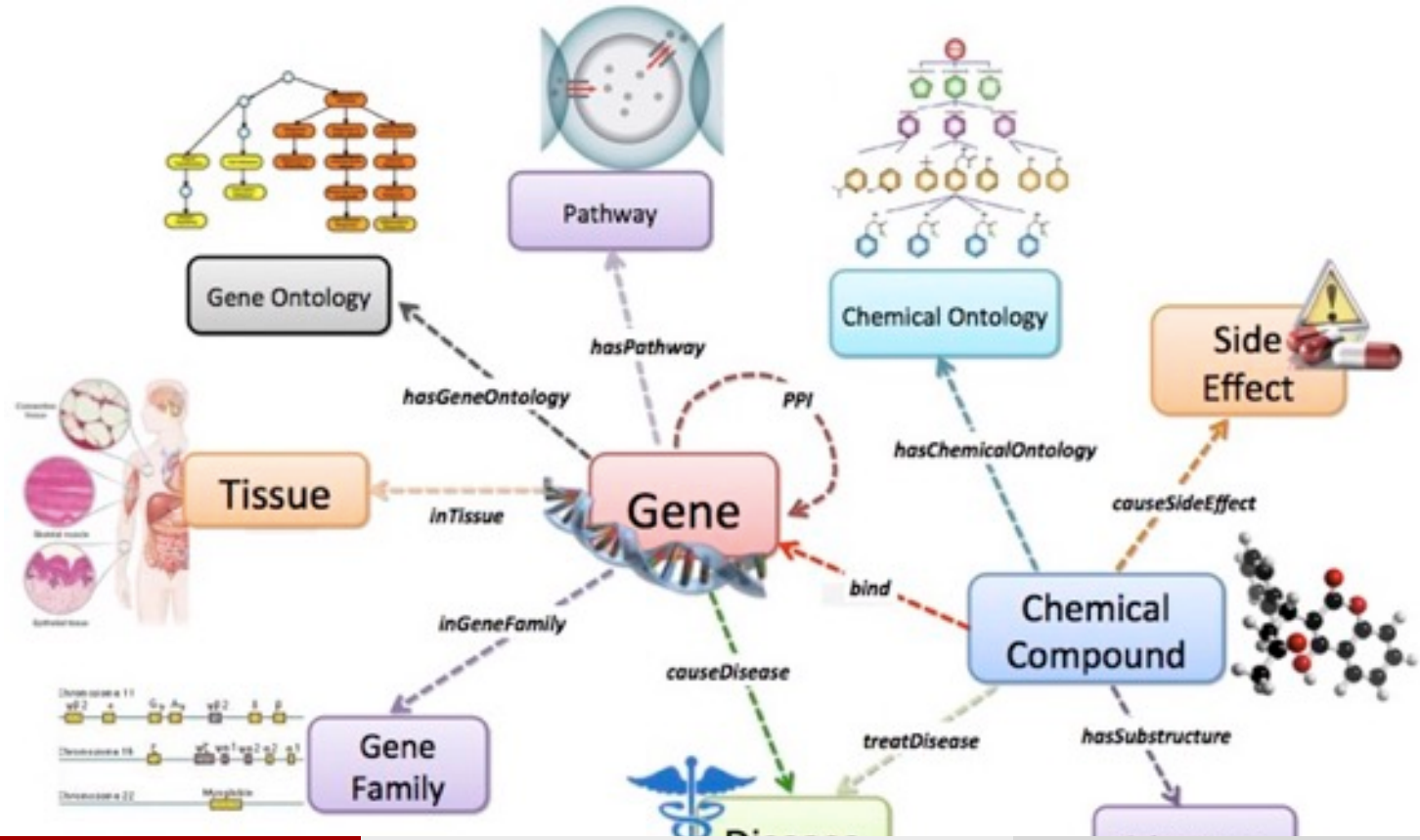
Structures Facilitate Heterogeneous Information Network Analysis

Real-world data: Multiple object types and/or multiple link types



Network Types

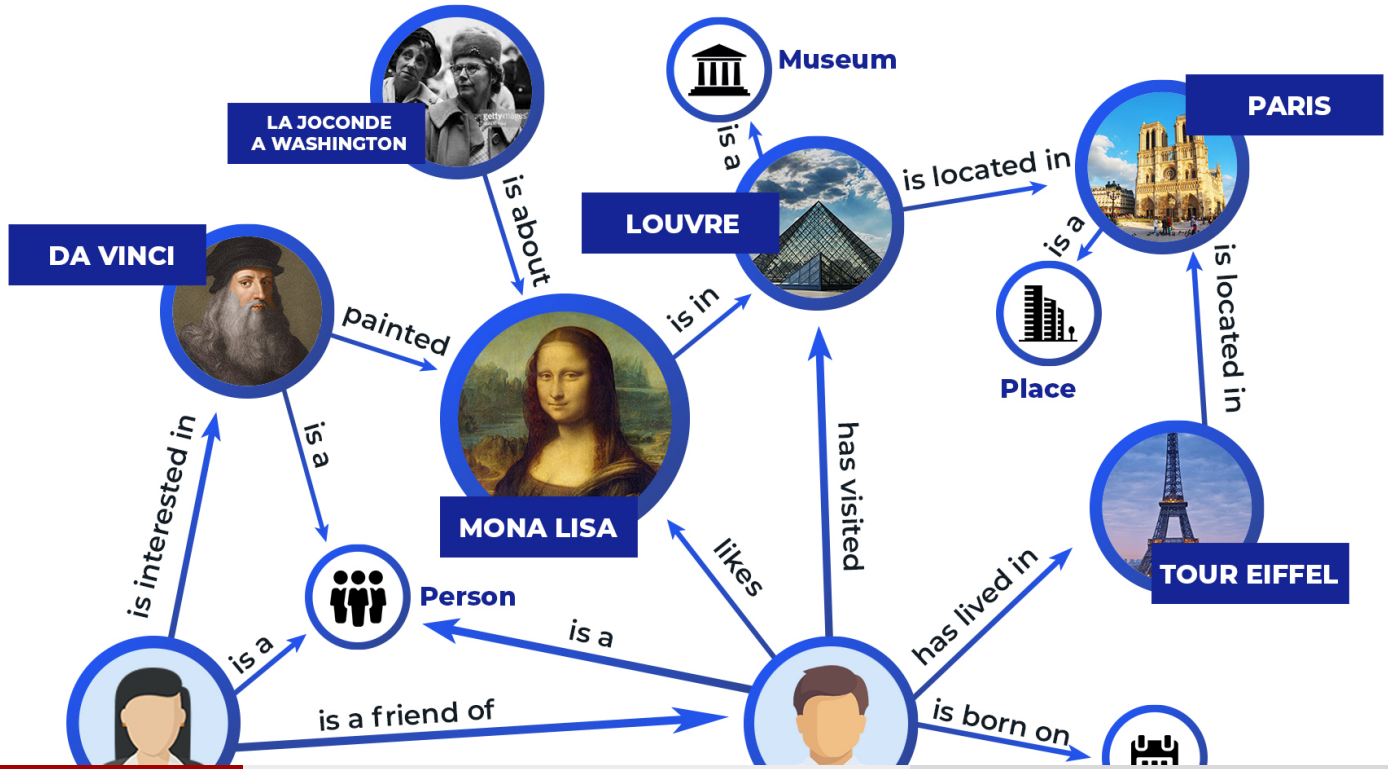
One more heterogeneous example



Network Types

And one more – the last one!

Knowledge Graphs.



Network Analytics – Intro.

- *Similarity*: which subjects (equiv., variables) are similar?
  - similarity analysis is frequently employed in RS (Collaborative Filtering!) [Users $\times$ Items].

- Consider a social network, i.e., a graph of users: similarity between Jack and Jill in terms of their friends.

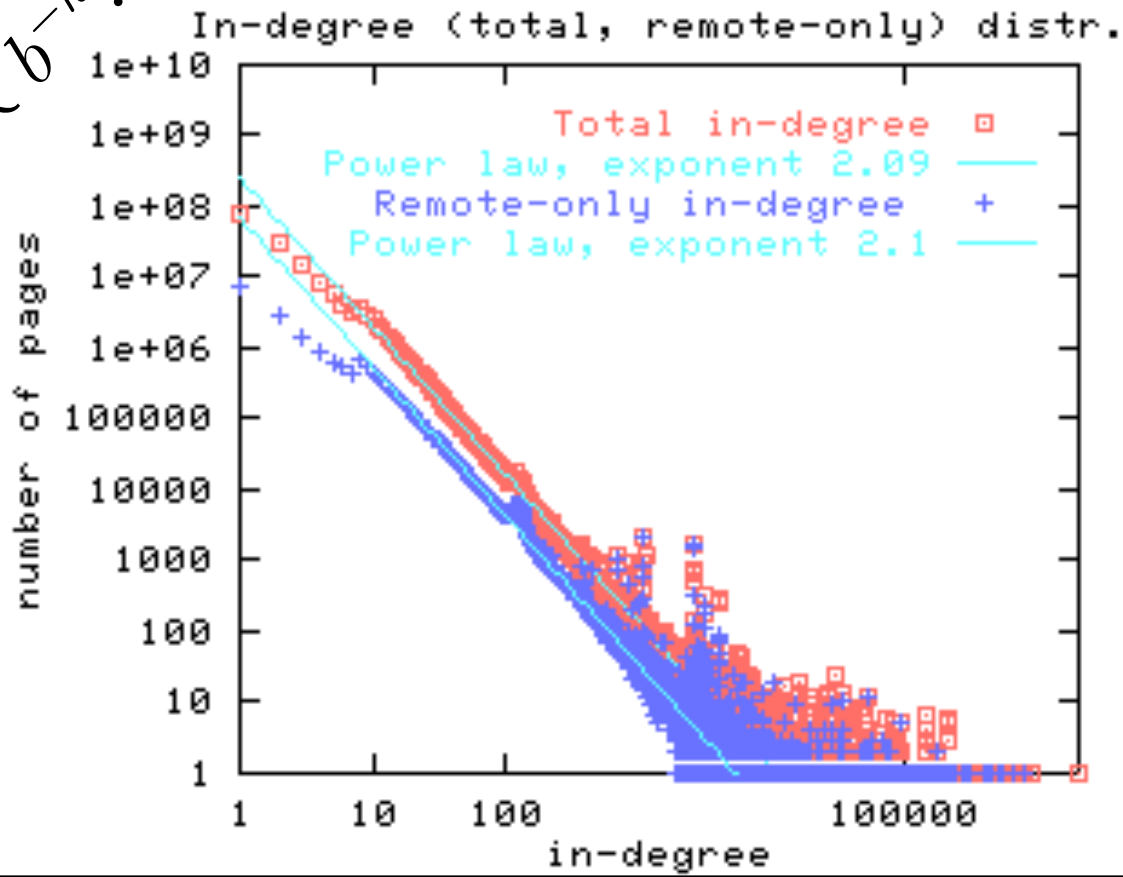
Network Analytics – Intro.

- Similarity in terms of whom you like versus who likes you, in the case of directed graphs.
- Some common (heuristic) measures of similarity: cosine, Jaccard, (Pearson) correlation.
- Notion of local network: friends of user u within h hops; users in a community/group; in a school/class; in an income group; n/w need not be explicitly declared by users.
- "semantic" similarity: graph embeddings!

SN – Intro: some characteristics

- SN often exhibit a power law in a number of ways:

$$\#\text{nodes with degree } k \propto b^{-k}.$$



Taken from: **Networks, Crowds, and Markets: Reasoning About a Highly Connected World By David Easley and Jon Kleinberg Cambridge U. Press 2010.**

Some Characteristics of Info. Networks

- Power law: most users have a very small in-degree; small #users have a large in-degree (observe the long tail).
- When joining the network, users are more likely to connect to popular nodes.
- Most of the blogs are posted by a small #users.
- Most ratings/reviews of movies/songs come from a small #users.
- #downloads of songs; #citations of papers; populations of cities; #copies of genes in genomes; ...

Some Characteristics of Info. Networks

- Network for us is just a graph G=(V,E).
  - Nodes = users/actors/individuals/orgs/entities.
  - can be heterogeneous.
  - Edges = ties/relationships; can be of several types and be complex; can model using labels; can be directed.

- Network Analysis:
  - How do rumors spread (or innovations happen)?
  - How do diseases spread?
  - What is the avg degree of separation of the n/w?

Some Characteristics of Info. Networks

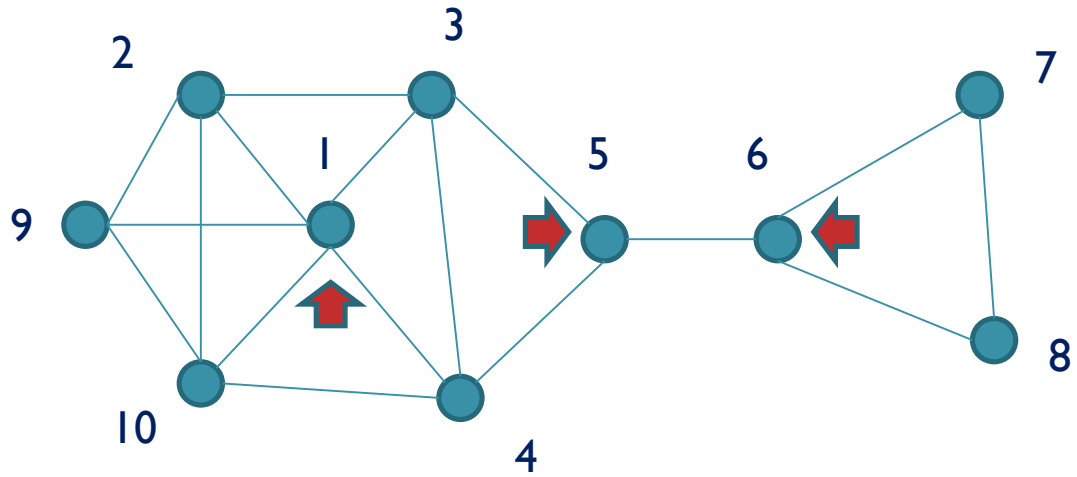
- Millgram's experiment: measure average min. #ties connecting two random people in the US by asking people to forward a mail to their contacts and seeing how many hops it took the mail to reach a certain target starting from random sources. Precursor of modern-day small worlds experiment.

- The so-called six degrees of separation: https://www.youtube.com/watch?v=V2biPHBGm3c

- Leskovec & Horvitz: Microsoft IM – large version of Milgram experiment.

SN – Intro: centrality

- Notions of centrality:
  - (Bidirectional) degree as an indication of info. flow or activity.
  - Betweenness: Boundary spanner between different clusters; how many (shortest) paths pass through me? Potential point of failure.



SN – Intro: centrality

- Closeness: how close am I to other nodes, on avg.?
- E.g. Distance distributions:

Node	Dist=1	Dist=2	Dist=3	Dist=4	Dist=5
1	5	1	1	2	0
3	4	3	2	0	0
5	3	5	1	0	0
6	3	2	3	1	0
7	2	1	2	3	1

Just one possible def.

- closeness(u) = $\sum_{\{v \in G\}} dist(u,v)/|G|$.
- closeness(1) = 18/10; closeness(5) = 16/10;
- closeness(3) = 16/10; closeness(6) = 17/10;
- closeness(7) = 27/10. Any surprises?

Centrality

- so far, local perspective: centrality of a node.
- From a global perspective ...
- How vulnerable is a network to single points of failure?
  - How many (central) nodes/links can fail before n/w is disconnected?
  - global measure.
  - How are the various central nodes situated in the n/w?
- Avg path length in the network.

More Example Networks

- Citation graphs, collaboration graphs.
- SN of corporate members.
- Wikipedia collaboration graph.
- Road and subway networks
- WWW, Internet
- Water distribution networks
- (Romantic) relationships between people
- Online (social) networks like ~~MySpace~~, Friendster, Facebook, Flickr, last.fm, Youtube, MSN IM, ..., LinkedIn, ...