Great Ideas in Computing

University of Toronto CSC196
Winter/Spring 2019

Week 7: October 26-Oct 30 (2020)

Announcements

- We did a quick poll to see how many students felt that Q1 was a fair quiz? If you thought the quiz was unfair, please do email me (preferably) or post an annoymous message on piazza. You can also discuss the quiz with Marta on Monday, November 2. The quiz grades will be released today.
- A2 has been submitted by everyone. Start of A3 will be posted today.
- On Monday, Professor Roger Grosse led a discussion on neural nets and deep learning, which is part of machine learning (ML) which in turn is part of AI. The success of ML in many applications is one of the main reasons for the increasing interest in computer science. As indicated by Roger Grosse, the University of Toronto has been a leader in AI and ML research and specifically deep learning for many years.
- Your questions with regard to neural nets and deep learning were great and significantly added to the discussion.
- However, many students do not actively participate. That's a hint reminding you that 18% of your final grade depends on active particpation (which, of course includes attending class but that is not sufficient).

Agenda for next couple of lectures

- Our next topic is an overview of network applications and basic graph theory concepts. I will do this quickly today and Marta will follow up on Monday. What I am doing today is terminology and motivation.
- In particular, we will then discuss social networks.
- Question 1 in Assignment A3 refers to matrix multiplction. Marta can go over matrix multiplication in tutorial if necessary. Here is the definition: Suppose A is a $p \times n$ matrix and B is a $n \times q$ matrix, then matrix multiplication $C = A * B$ is defined as the $p \times q$ matrix $C = C[i, j]$ where the i, j entry of C is $c_{i,j} = \sum_{k=1}^{n} a_{i,k} \cdot b_{k,j}$.
- Here I am using $*$ to denote matrix multiplcication and $a \cdot b$ to denote scalar multiplication. We assume the elements of the matrix belong to a *ring* (e..g., $\mathbb{Z}$, the ring of integers) or a *field* (e.g., $\mathbb{Q}$, the field of rattional numbers) where $+$ and $\cdot$ and defined and satisfy some natural properties.

New topic: Networks and Social Networks

As mentioned earlier, I view search engines, social networks, and email to be arguably the three most influential killer applications in making computing a necessary commodity.

But before discussing social networks, we need to introduce a few concepts about networks in general. I am using slides from the first week of my spring course CSC303 (Social and Information Networks).

What we will see from these slides is that *networks* (called *graphs* in mathematics) play a fundamental role in the way we think about many diverse real world systems. In fact, in our discussion of search engines we have been making implicit and explicit references to networks. A neural net is also a network.

The web is a network of documents connected by hyperlinks. The figure on slide 15 of the Week 6 slides is an example of a special type of graph called a *bipartite graph* where the *links* (also called *edges*) between the nodes (also called *vertices*) are restricted to be between Ad nodes and Query nodes. Trees (e.g., search trees) are another special type of graph.

What's in a name? Graphs or Networks?

Networks are graphs with (for some people) different terminology where graphs have vertices connected by edges, and networks have nodes connected by links. I do not worry about this "convention", to the extent it is really a vague convention without any real significance.

Here is one explanation for the different terminology: We use networks for settings where we think of links transmitting or transporting "things" (e.g. information, physical objects, friendship).

Many different types of networks

- Social networks
- Information networks
- Transportation networks
- Communication networks
- Biological networks (e.g., protein interactions)
- Neural networks

Visualizing Networks

- nodes: entities (people, countries, companies, organizations, . . .)
- links (may be directed or weighted): relationship between entities
  - friendship, classmates, did business together, viewed the same web pages, . . .
  - membership in a club, class, political party, . . .

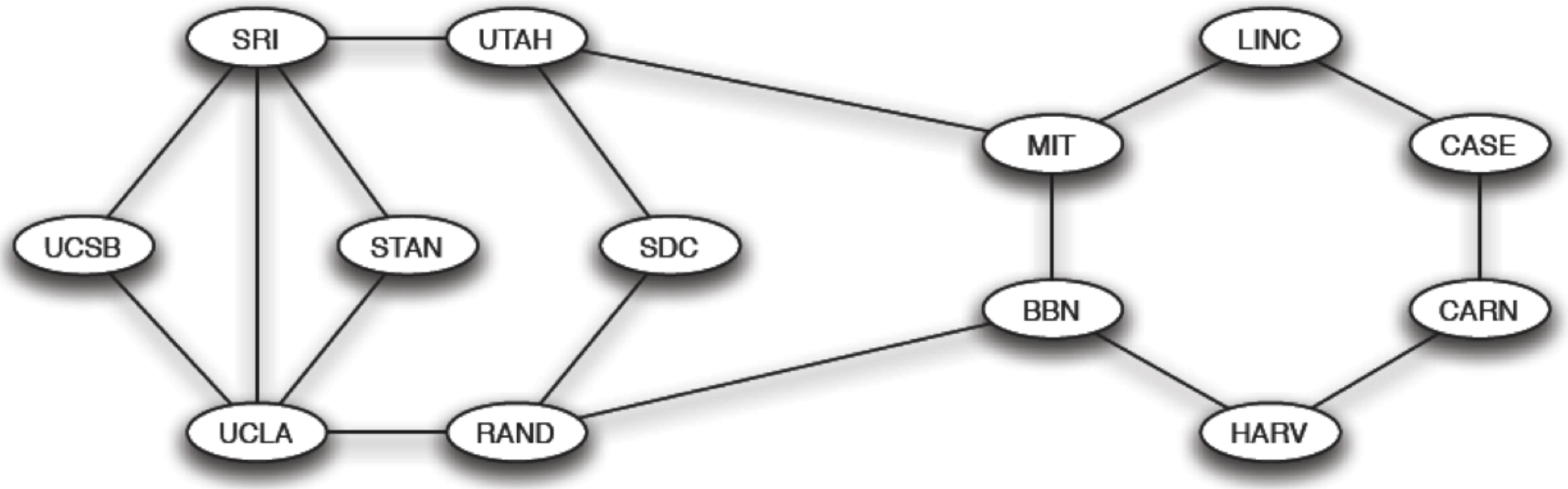
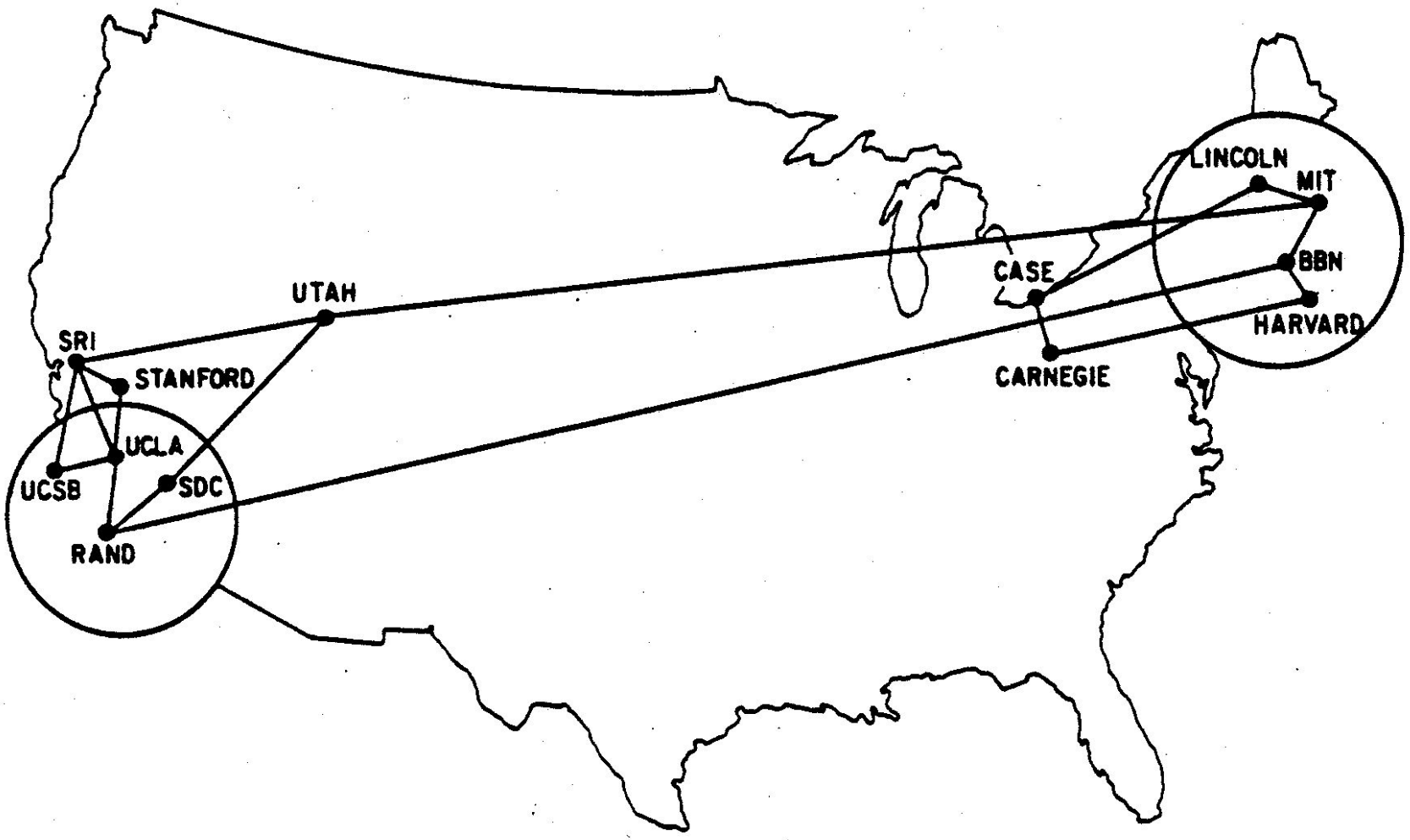


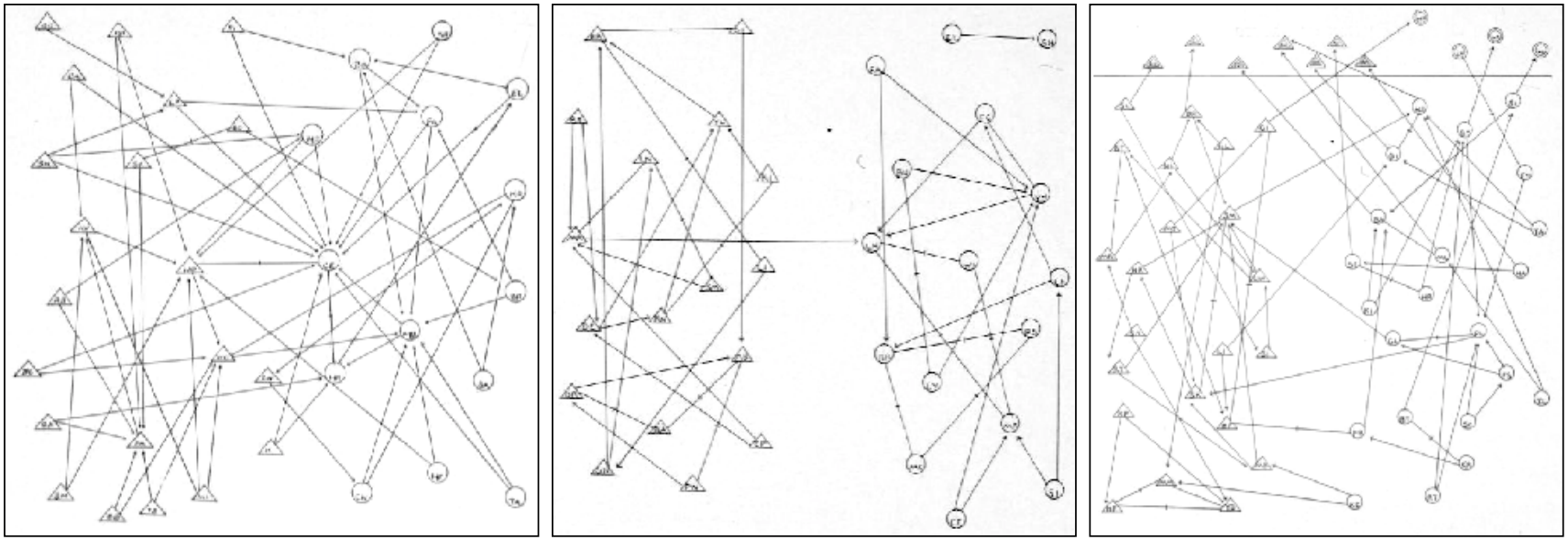
Figure: Initial internet: Dec. 1970 [E&K, Ch.2]

December 1970 internet visualized geographically [Heart et al 1978]



The first social network analysis

In his **1934** book *Who Shall Survive: A New Approach to the Problem of Human Interrelations*, Jacob Moreno (Romanian-US psychiatrist) introduced *sociograms* and used these graphs/networks to understand relationships. In one study (that was repeated to test changes) he asked each child in various elementary grades at a public school to choose two children to sit next to in class. He used this to study inter-gender relationships (and other relationships). Here boys are depicted by triangles and girls by circles.



1st grade 4th grade 8th grade

A closer look at grade 1 in Moreno sociogram

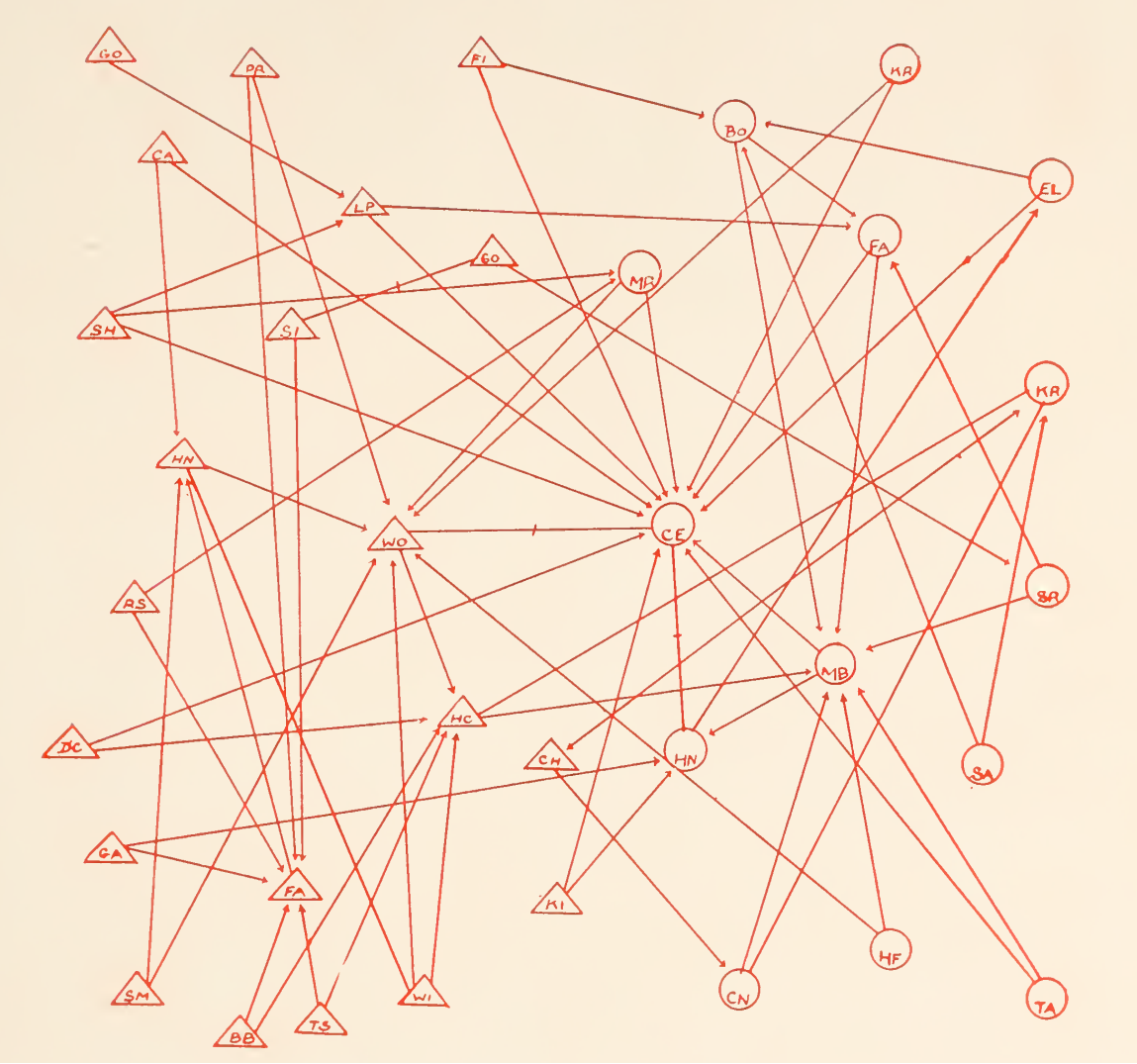


Figure: 21 boys, 14 girls. Directed graph. Every node has out-degree 2. 18 unchosen having in-degree 0. Note also that there are some "stars" with high in-degree.

A closer look at grade 4 in Moreno sociogram

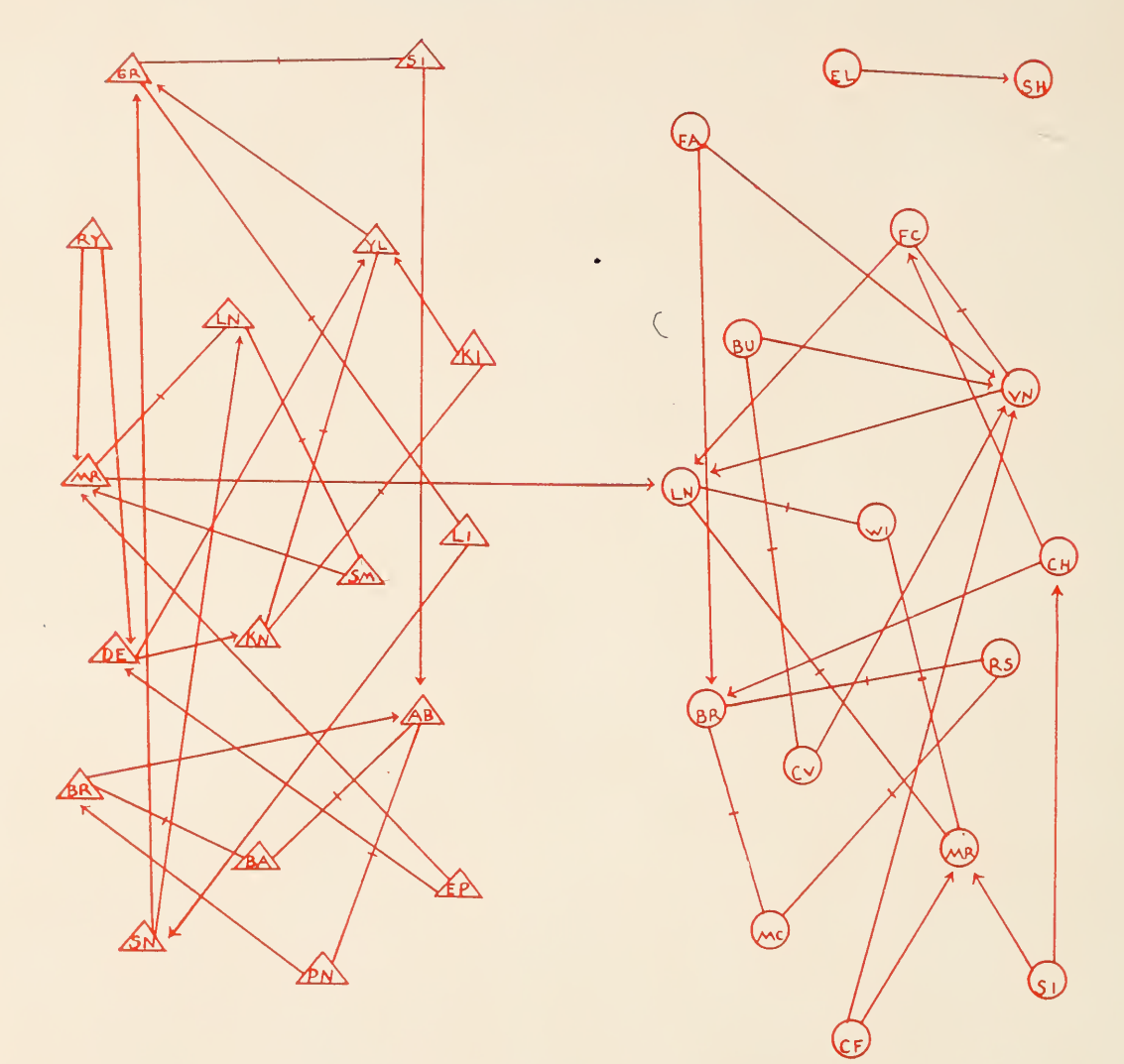


Figure: 17 boys, 16 girls. Directed graph with 6 unchosen having in-degree 0. Moreno depicted his graphs to emphasize inter-gender relations. Note only one edge from a boy to a girl.

A closer look at grade 8 in Moreno sociogram

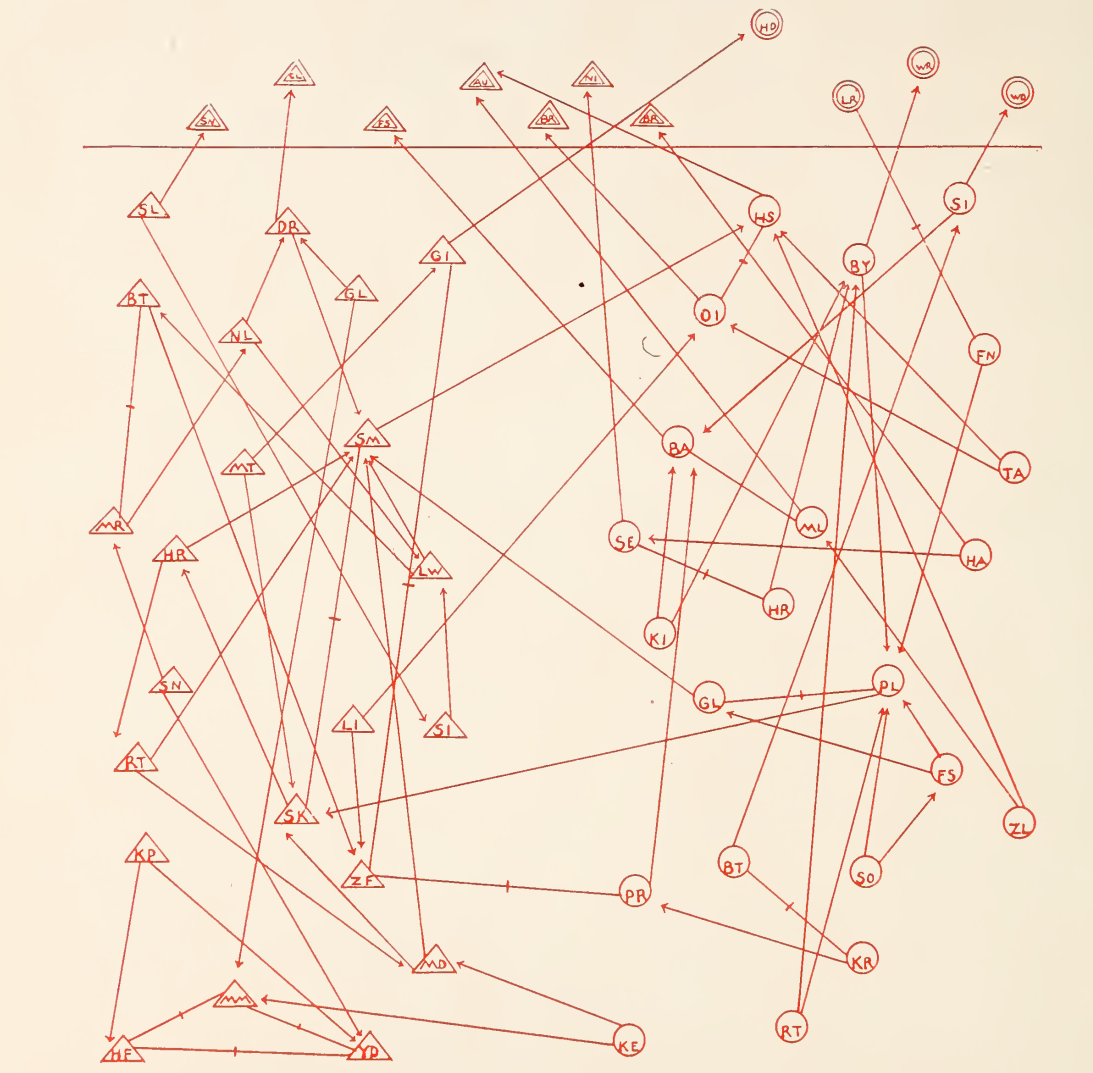


Figure: 22 boys, 22 girls. Directed graph with 12 unchosen having in-degree 0. Some increase in inter-gender relations. Double stars and circles above line indicte different "groups".

Romantic Relationships [Bearman et al, 2004]

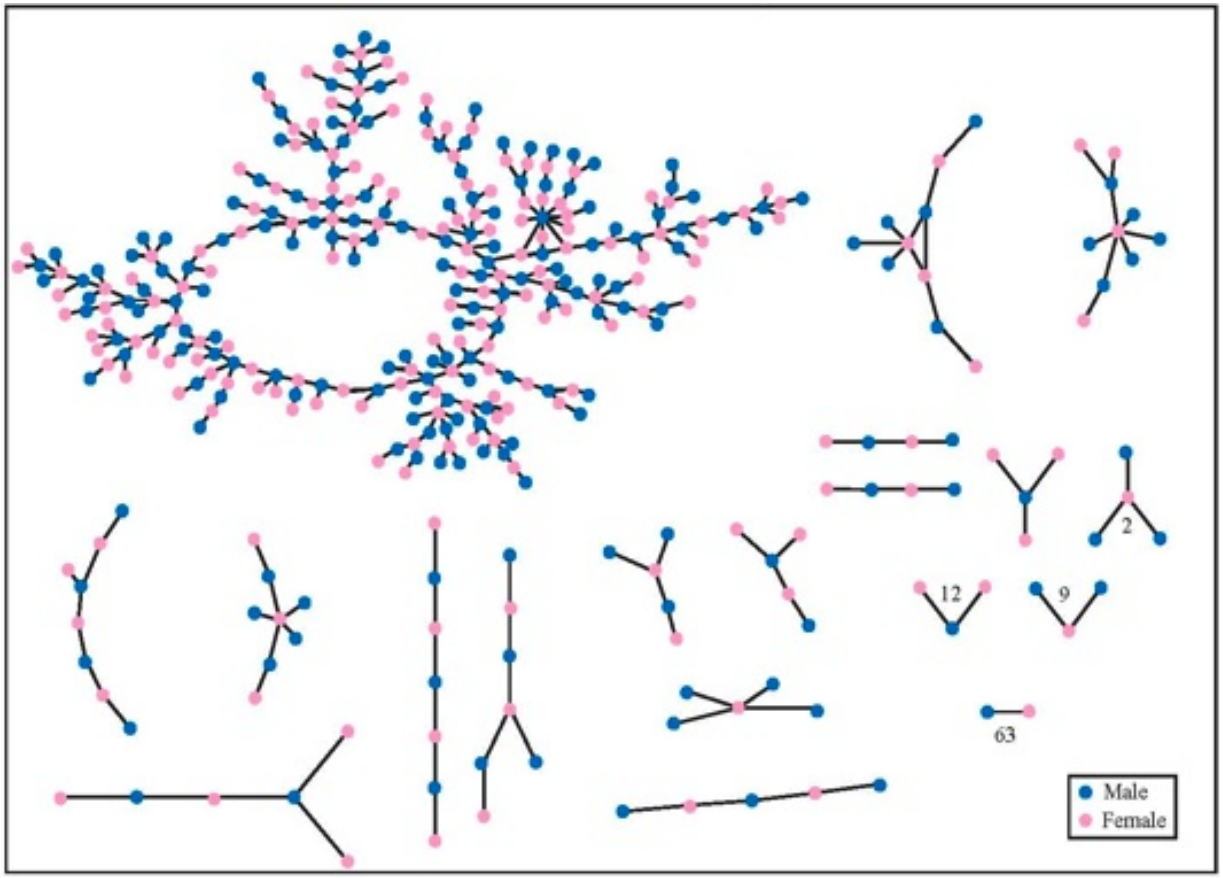


Figure: Dating network in US high school over 18 months.

- Illustrates common "structural" properties of many networks
- What is the benefit of understanding this network structure?

Kidney Exchange: Swap Chains

- Waiting list for kidney donation: approximately 100K in US and growing (i.e., new patients added but many deaths while waiting). The wait for a deceased donor could be 5 years and longer.
- Live kidney donations becoming somewhat more common in N.A. to get around waiting list problems: requires donor-recipient pairs
- Exchange: supports willing pairs who are incompatible
  1. allows multiway-exchange
  2. supported by sophisticated algorithms to find matches

Kidney Exchange: Swap Chains

- Waiting list for kidney donation: approximately 100K in US and growing (i.e., new patients added but many deaths while waiting). The wait for a deceased donor could be 5 years and longer.
- Live kidney donations becoming somewhat more common in N.A. to get around waiting list problems: requires donor-recipient pairs
- Exchange: supports willing pairs who are incompatible
  1. allows multiway-exchange
  2. supported by sophisticated algorithms to find matches
- But what if someone renegs? $\Rightarrow$ Cyclyes require simultaneous transplantation; Paths require altruisitic an donor!

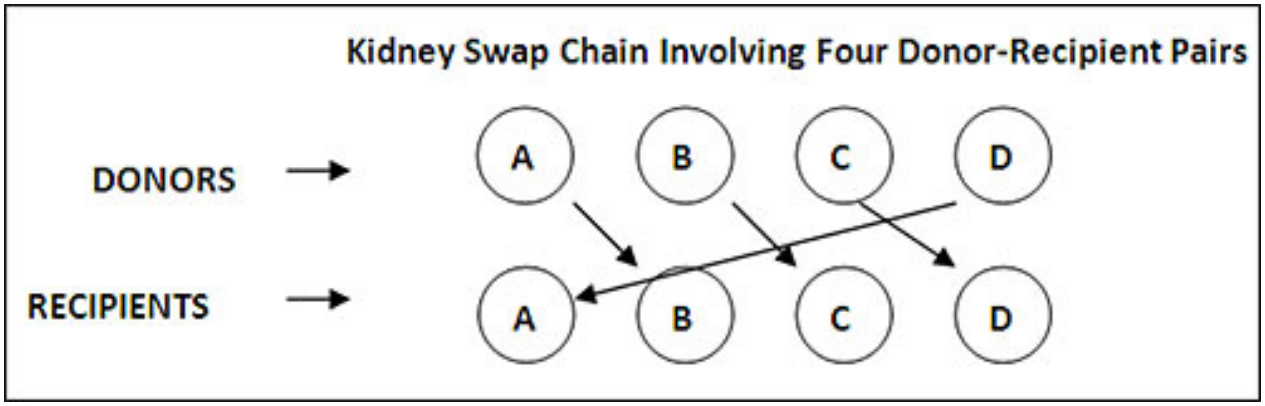
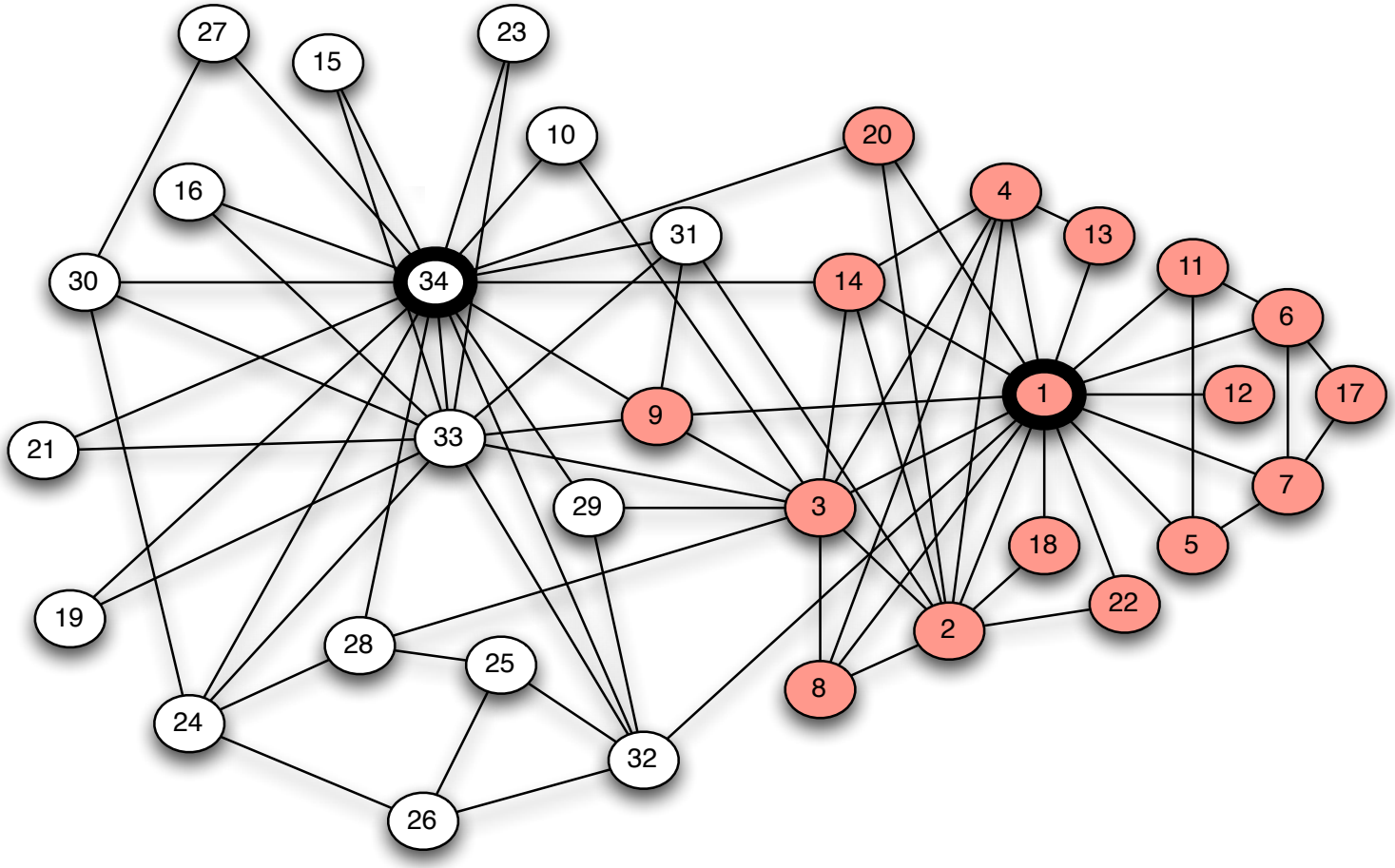


Figure: Dartmouth-Hitchcock Medical Center, NH, 2010

Communities: Karate club division



Karate Club social network, Zachary 1977

Figure: Karate club splis into two clubs (or *communities*)

Communities: 2004 Political blogsphere

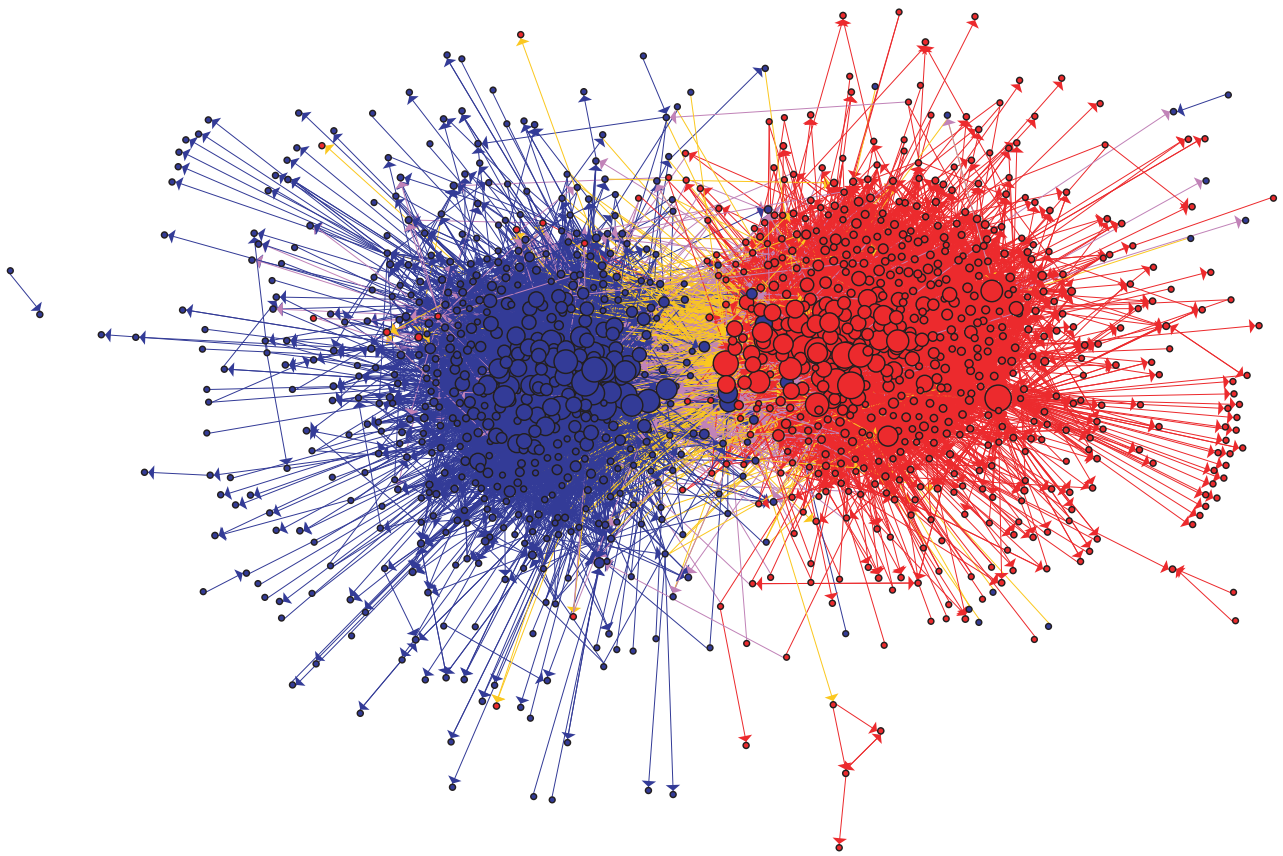


Figure 1: Community structure of political blogs (expanded set), shown using utilizing a GEM layout [11] in the GUESS[3] visualization and analysis tool. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

Communities: 2017 Twitter online discourse regarding Black Lives Matter

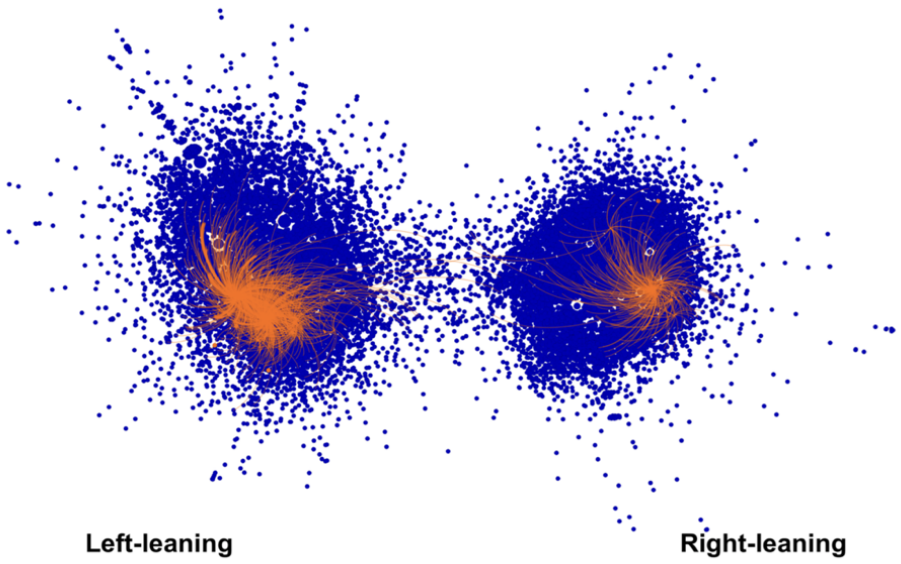


Fig. 1. Retweet Network Graph: RU-IRA Agents in #BlackLivesMatter Discourse. The graph (originally published [3]) shows accounts active in Twitter conversations about #BlackLivesMatter and shooting events in 2016. Each node is an account. Accounts are closer together when one account retweeted another account. The structural graph shows two distinct communities (pro-BlackLivesMatter on the left; anti-BlackLivesMatter on the right).

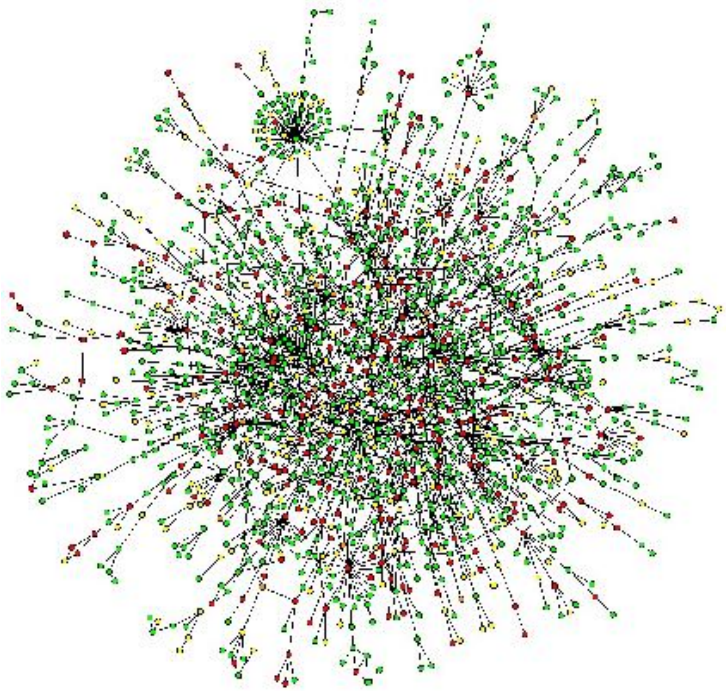
Accounts colored orange were determined by Twitter to have been operated by Russia's Internet Research Agency. Orange lines represent retweets of those account, showing how their content echoed across the different communities.
The graph shows IRA agents active in both "sides" of that discourse.

Figure: From Starbird et al [2017, 2019]

Communities and hierarchical structure: Email communication

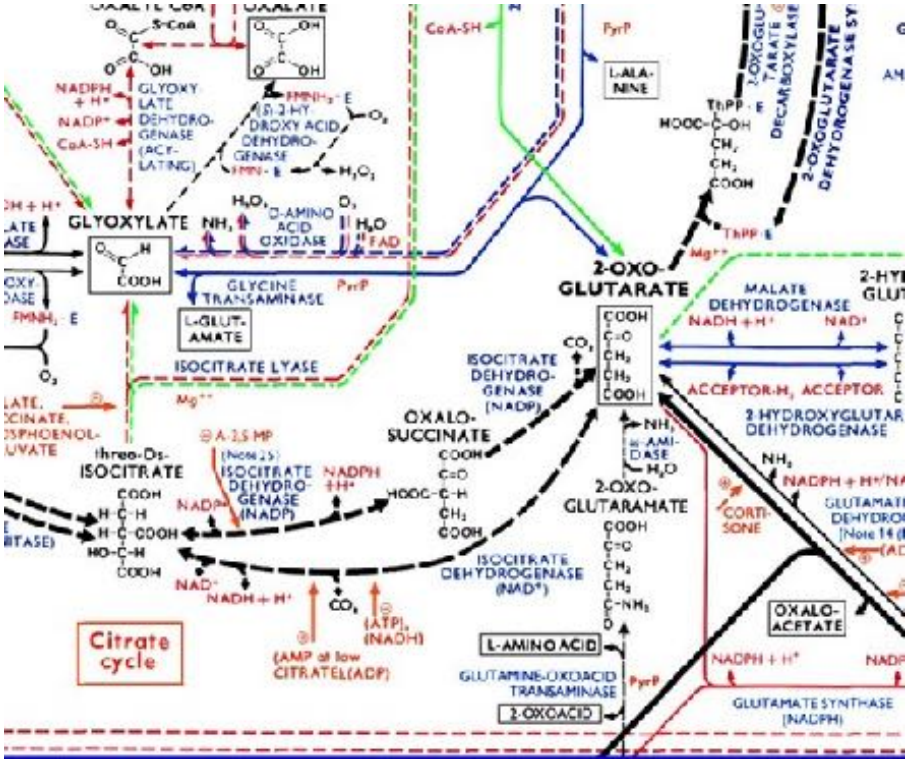
Figure: Email communication amongst 436 employees of Hewlett Packard Research Lab, superimposed on the Lab organizational hierarchy

Protein-protein interaction network



Protein-Protein Interaction Networks
Nodes: Proteins
Edges: 'physical' interactions

Metabolic network



Metabolic networks

Nodes: Metabolites and enzymes

Edges: Chemical reactions

The web as a directed graph of hyperlinks

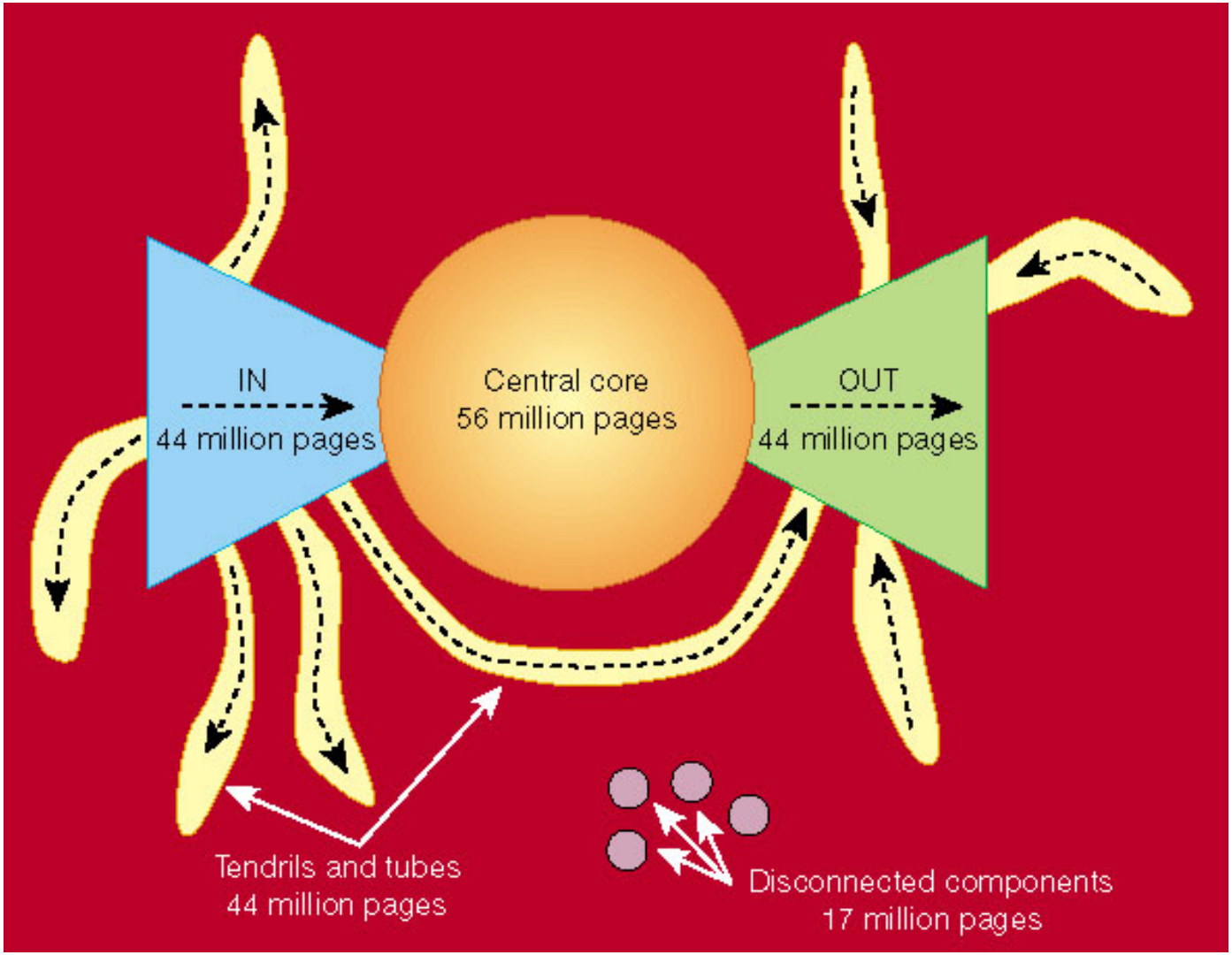


Figure: A schematic picture of the bow tie structure of the 1999 Web. Although the numbers are outdated, the structure has persisted. [Fig 13.7, EK textbook]

The current interest in networks

- Clearly there are complex systems and networks that we are in contact with daily.
- The population of the world can be thought of as social network of approximately 7.75 billion people. AS OF January 2020, The people on Facebook are a *subnetwork* of approximatley 2.45 billion active monthly users of which 1.62 billion are daily users. AS OF THIRD QUARTER 2019.
- The language of networks and graph analysis provides a common language and framework to study systems in diverse disciplines. Moreover, networks relating to diverse disciplines may sometimes share common features and analysis.
- The ability to store and process massive amounts of data, makes computational aspects of networks essential.
- The current impact of social and information networks will almost surely continue to escalate (even if Facebook and other social networks are under increasing presure to protect privacy and eliminate "bad actors").

What can one accomplish by studying networks

We use networks as **a model** of real systems. As such, we always have to keep in mind the goals of any model which neceessarily simplifies things to make analysis possible.
In studying social and information networks we can hopefully

- Discover interesting phenomena and statistical properties of the network and the system it attempts to model.
- Formulate hypotheses as to say how networks form and evolve over time
- Predict behaviour for the system being modeled.

And how do we accomplish stated goals

Much of what people do in this field is empirical analysis. We formulate our network model, hypotheses and predictions and then compare against real world (or sometimes synthetically generated) data.

Sometimes we can theoretically analyze properties of a network and then again compare to real or synthetic data.

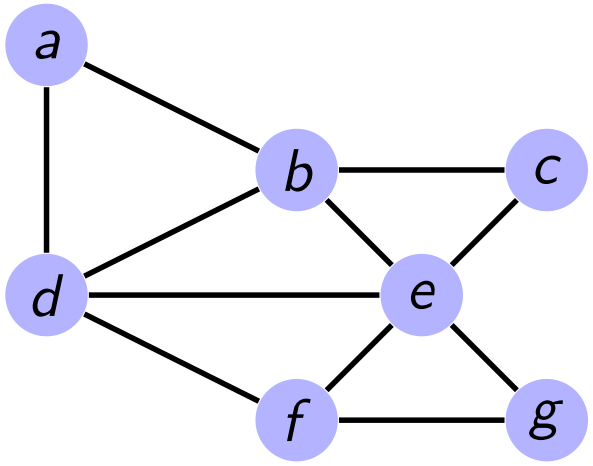
What are the challenges?

- Real world data is sometimes hard to obtain. Like search enginess, social networks treat much of what they do as proprietary.
- Many graph theory problems are known to be computationally difficult (i.e., *NP* hard) and given the size of many networks, results can often only be approximated and even then this may require a significant amount of specialized heuristics and approaches to help overcome (to some extent) computational limitations.
- And we are always faced with the difficulty of bridging the simplification of a model with that of the many real world details that are lost in the abstraction.

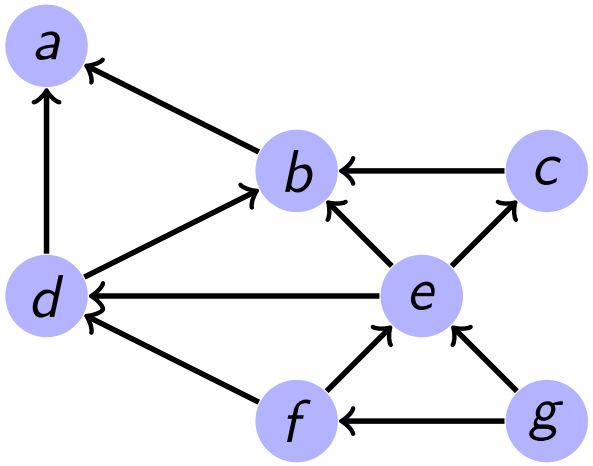
Start of network (graph) concepts

Graphs come in two varieties

1. undirected graphs ("graph" usually means an undirected graph.)



2. directed graphs (often called di-graphs).



Visualizing Networks as Graphs

- nodes: entities (people, countries, companies, organizations, . . .)
- links (may be directed or weighted): relationship between entities
  - friendship, classmates, did business together, viewed the same web pages, . . .
  - membership in a club, class, political party, . . .

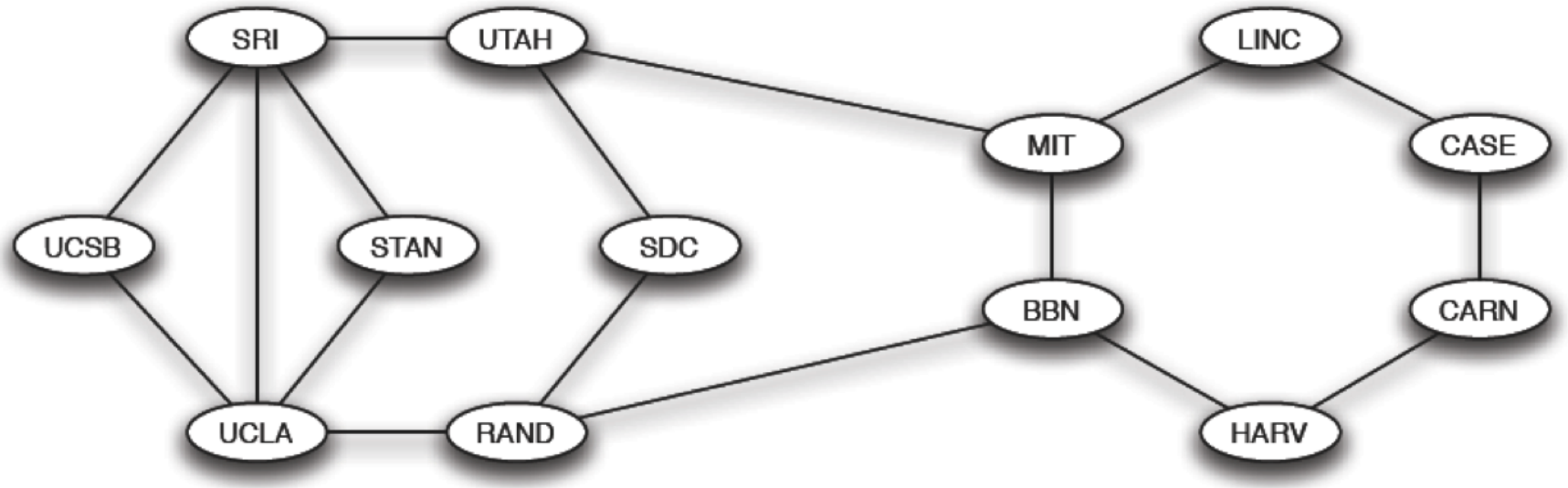


Figure: Internet: Dec. 1970 [E&K, Ch.2]

Adjacency matrix for graph induced by eastern sites) in 1970 internet graph: another way to represent a graph

$$A(G) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

- This node induced subgraph (for the sites MIT = 1, LINC = 2, CASE = 3, CARN = 4, HARV = 5, BBN = 6) is a 6 node regular graph of degree 2. It is a simple graph in that there are no self-loops or multiple edges.
- Note that the adjacency matrix of an (undirected) simple graph is a symmetric matrix (i.e. $A_{i,j} = A_{j,i}$) with $\{0,1\}$ entries.
- To specify distances, we would need to give weights to the edges to represent the distances.

The matrix A^2 where $A = A(G)$

Consider squaring the previous matrix $A = A(G)$. That is, $A^2 = A * A$.

$$A^2 = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Draw a visualization of the graph represented by A^2. If we let $c_{i,j}$ be the i, j entry in A^2, can you desribe the meaning of $c_{i,j}$?

The matrix $B = A + I$

Consider the 6×6 identity matrix $I = (\iota_{i,j})$. That is, $\iota_{i,i} = 1$ for $1 \leq i \leq 6$ and $\iota_{i,j} = 0$ for $1 \leq i, j \leq 6$ and $i \neq j$.

Let $B = A + I$ (as above). That is, $b_{i,j} = a_{i,j} + \iota_{i,j}$ for all i, j. We have

$$B(G) = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Note that now the matrix B has self loops and hence is not a simple graph.

Kidney Exchange: Swap Cycles

- Live kidney donation common in N.A. to get around waiting list problems: donor-recipient pairs are nodes and links are directed.
- Exchange: supports willing pairs who are incompatible
  1. allows multiway-exchange
  2. supported by sophisticated algorithms to find matches

Kidney Exchange: Swap Cycles

- Live kidney donation common in N.A. to get around waiting list problems: donor-recipient pairs are nodes and links are directed.
- Exchange: supports willing pairs who are incompatible
  1. allows multiway-exchange
  2. supported by sophisticated algorithms to find matches
- But what if someone reneges? $\Rightarrow$ require simultaneous transplantation! Non-cyclic paths can be started by an altruistic donor!

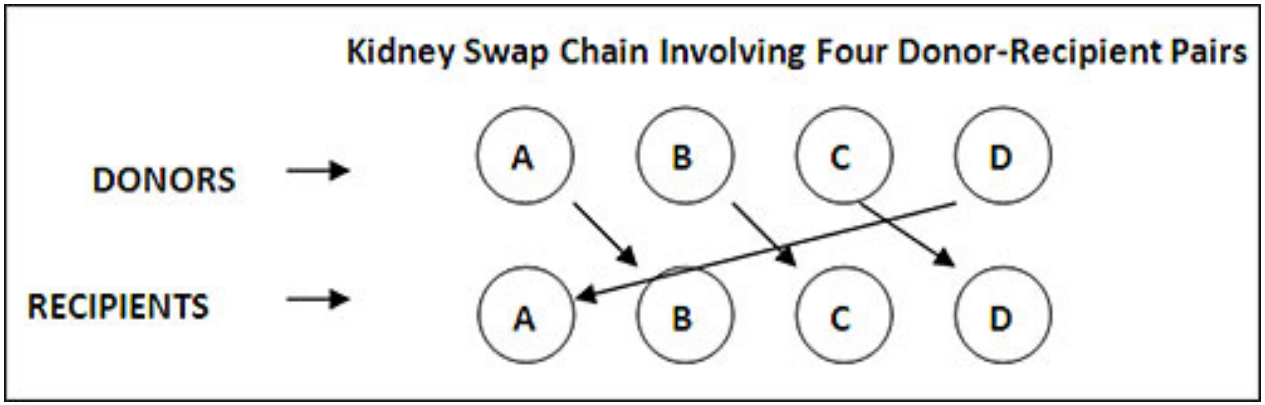


Figure: Dartmouth-Hitchcock Medical Center, NH, 2010

End of Week 7

We will continue with network definitions and terminology next week in the tutorial on Monday and in our regular Wed, Friday classes.
With regard to social networks, we will discuss influence spread in a social network and also how just the network stucture by itself can yield infomation about people in a social network.