Great Ideas in Computing

University of Toronto CSC196
Fall 2021

Week 11 (part 1): November 29-December 3

Announcements

Announcements

- I hope everyone did well on the second quiz.
- I repeat that you have one week from the time that any graded material is returned to request a regrade. The request must be specific indicating why you believe the grade is not a proper assessment of your solution.
- I do ask you to understand that grading most questions can be subjective to some extent. If you received say a 6/10 and think you should have received a 7/10 for a particular question, it is very unlikely to change your final grade. But we do want you to get proper credit for reasonable solutions and to undertsatnd why your solution may or may not be clear. So that is why we entertain grade change requests.
- The final assignment is due December 3 at 8AM.

This weeks agenda

- We will first finish up complexity based cryptograph and then try to get to social networks quickly.
- The social network question on Assignment A4 is mainly a thought question so you should be able to provide reasonable answers based on your own experience and the indicated W11 slides.
- We begin a discussion of graphs/networks in general and social networks in particular
- Given that we only have two weeks left in the term and I have a question on social networks in Assignment 4, I am providing an "appendix" of graph/networks definitions with examples starting on slide 41. I will call attetion to any concepts or definitions needed in the discussion of social-networks. If you have any question about social networks raise them in class, or on piazza.
- The undergraduate Easley and Kleinberg textbook "Networks, Crowds, and Markets: Reasoning about a Highly Connected World" is an excellent text for understanding the importance of network concepts and applications. We teach an undergraduate course CSC303 devoted to social networks.

The basic idea of public key encryption

Public key encryption was introduced by Diffie and Hellman, and a particular method (RSA) was created by Rivest, Shamir and Adelman.

The basic idea is that in order for Alice (or anyone) to send Bob a message, Bob is going to create two related keys, a public key allowing Alice to send an encrypted mesasage to Bob, and a private key that allows Bob to decrypt Alice's message.

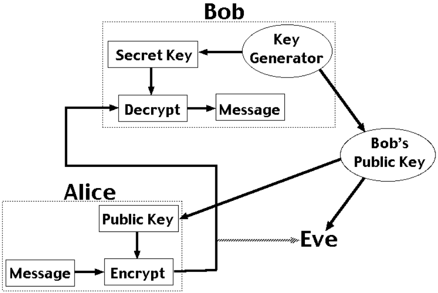


Figure: Diagram of public key encryption. Figure taken from Paul Johnston notes

The RSA method

Bob wants to generate two keys, a public key e, N and a private key d. The claim is that it is hard on average to find d given e and N. Bob chooses $N = p \cdot q$ for two large primes p, q (which for defining "on average" may satisfy some constraint).

Bob will choose the public e such that $gcd(e, \phi(N)) = 1$ where $\phi(N) = \phi(pq) = (p-1)(q-1)$. $\phi(N)$ is called the Euler totient function which is equal to the number integers less than N that are relatively prime to N. $gcd(a, b) = 1$ means that a and b are relatively prime (i.e. have no common proper factors).

Alice encodes a message M by computing $M^e \bmod N$.

Hiding some mathematics, BOB can compute a d such that $de = 1 \bmod (p-1)(q-1)$ since Bob knows p and q. But without knowing p, q, finding d becomes computationally difficult.

Hiding some more mathematics, it will follow that $M^{de} = M \pmod{N}$ for any message M. That is, Bob decrypts a cypher text C by the function $C^d \bmod N$.

What mathematical facts do we need to know.

The main mathematical facts are :

1. There are sufficiently many prime numbers in any range so one can just randomly try to diffent numbers and test if they are prime.
2. $a^{\phi(N)} = 1 \bmod N$ for any a such that $gcd(a, N) = 1$ As a special case, $a^{p-1} = 1 \bmod p$ for any prime p and a not a multiple of p. So we have $M^{(p-1)(q-1)} = 1 \bmod N$.
3. If $gcd(a, b) = 1$ then there exists s and t such that $sa + tb = 1$. In the RSA algorithm, we can let $a = e$ and $b = (p-1)(q-1)$. Then s will become the d we need for decryption. That is $de + t(p-1)(q-1) = 1$.
4. It follows then that $M^{de} = M^{1-t(p-1)(q-1)} = M \cdot M^{-t(p-1)(q=1)} = M \bmod (p-1)(q-1)$.

What computational facts do we need to know?

1. The extended Euclidean algorithm can efficiently compute an s and t such that $sa + tb = gcd(a, b)$
2. $a^k \bmod N$ can be computed efficiently for any a, k, N.
3. We can efficiently determine if a number p is prime.

In practice, public keys e are chosen to be reasonably small so that encryption can be made more efficient.

Note that we have been assuming that an adversary EVE (i.e., is just eavesdropping) and not changing messages. That is, EVE just wants to learn the message or something about the message. If EVE could change messages then EVE could pretend to be BOB. So one needs some sort of a public key infrastructure.

Note that if EVE knows that the message M was one a few possibilities, then EVE can try each of the possibilities; that is compute $M^e modN$ for each possible M to see what message was being sent. So here is where randomness can be used. We can pad or interspers random bits in the plain text M so that the message being sent becomes some one of many random messages M'.

WARNING: Real world cryptography is sophisticated

Complexity based cryptography requires careful consideration of the definitions and what precise assumptions are being made.

Complexity based cryptography has led to many important practical protocols and there are a number of theorems. Fortunatley, many complexity assumptions turn out to be equivalent.
In the Rackoff notes, the following theorem is stated as the fundamental theorem of cryptography. (To make this result precise, one needs precise definitions which we are omitting.)
Theorem: The following are equivalent:

- It is possible to do “computationally secure sessions”
- There exists pseudo-random generators; that is, create strings that comoutationally look random)
- There exist one way functions f; that is functions such that $f(x)$ is easy to compute but given $f(x)$ it is hard to find a z such that $f(z) = f(x)$. Here “hard to find” means not computable in polynomial time.
- There exist computationally secure digital signature schemes.

The discrete log function

RSA is based on the assumed difficulty of factoring. Another assumption that is widely used in cryptography is the discrete log function. Again, we need some facts from number theory.

Let p be a large prime.

- $\mathbb{Z}_p^*$ denotes the set of integers $\{1, 2, \ldots, p-1\}$ under the operations of $+, -, \cdot \mod p$ is a *field*. In particular, for every $a \in \mathbb{Z}_p^*$, there exists a $b \in \mathbb{Z}_p^*$ such that $a \cdot b = 1$; i.e., $b = a^{-1} \mod p$.
- Moroever, $\mathbb{Z}_p^*$ is *cyclic*. That is, there exists a $g \in \mathbb{Z}_p^*$ such that $\{1, g, g^2, g^3, \ldots g^{p-2}\} \mod p = \mathbb{Z}_p^*$. Recall, as a special case of the Euler totient function, $a^{p-1} = 1 \mod p$.

The assumption is that given $(g, p, g^x \bmod p)$, it is computationally difficult to find x. This is another example (factoring can also be an example) of a *one-way function*. In fact the discrete log function is a *one-way permutation*.

A pseudo random generator

We started off our discussion of complexity based cryptpgraphy by noting that randomness is essential. We have also noted that it is not clear (or at what cost) one can obtain strings that "look like" truly random strings.

A pseudo random generator G is a *deterministic* function $G : \{0,1\}^k \rightarrow \{0,1\}^\ell$ for $\ell > k$. When ℓ is exponential in k, G is called a pseudo random function generator. For now, lets even see how to be able to have $\ell = k + 1$.

The random input string $s \in \{0,1\}^k$ is called the seed and the goal is that $r = G(s)$ should be "computationally indistinguishable" from a truly random string in $t = \{0,1\}^\ell$. This means that no polynomial time algorithm can distinguish between r and t with probability better than $\frac{1}{2} + \epsilon$ for any $\epsilon > 0$. (Here I am being sloppy about the quantification but hopefully the idea is clear.)

A pseudo random generator continued

On the previous slide there was a claim that having a pseudo random generator is equivalent to having a one-way function.

How can we use (for example, the assumption that the discrete log function is a one-way function) to construct a pseudo random generator with $\ell = k + 1$.

The Blum-Micali generator. Assumming the discrete log function is a one-way function then the following is a pseudo random generator:

Let x_0 be a random seed in $\mathbb{Z}_p^*$ by interpeting $(s_1, \ldots, s_k)_2$ as a binary number mod p. Let $x_{k+1} = g^{x_k} \bmod p$. Define $s_{k+1} = 1$ if $x_k \leq \frac{p-1}{2}$.

Manual Blum won the Turing award for his contributions to cryptography and Silvio Micali (along with Shafira Goldwasser) won the Turing award for *interactive zero knowledge proofs*. (Note: The authors on the seminal zero knowledge paper are Goldwasser, Micali, and Rackoff where I am noting that Charlie Rackoff is a UT DCS Professor Emeritus.)

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, . . .

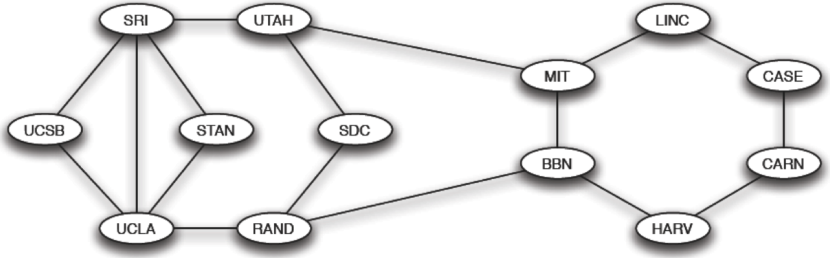
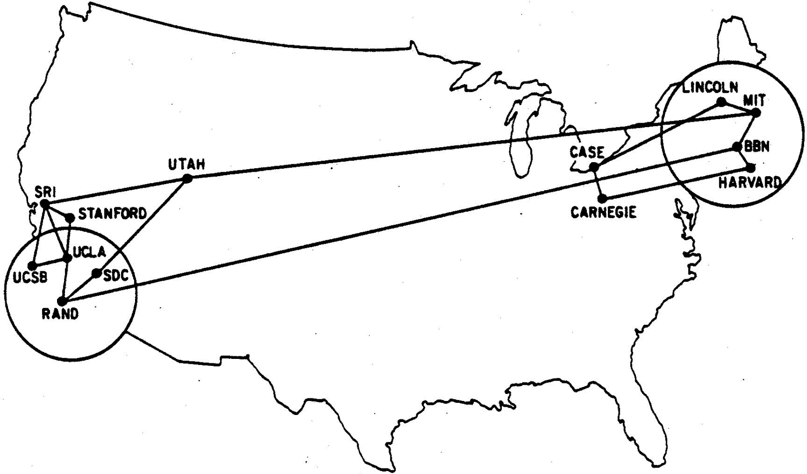


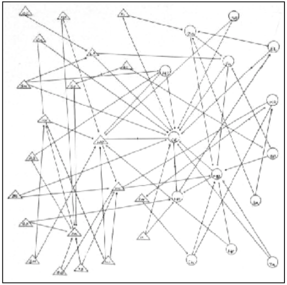
Figure: Initial internet: Dec. 1970 [E&K, Ch.2]

December 1970 internet visualized geographically [Heart et al 1978]

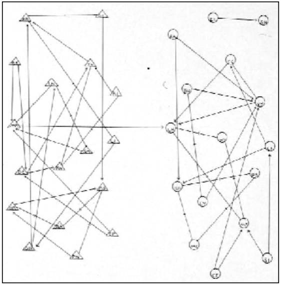


The first social network analysis

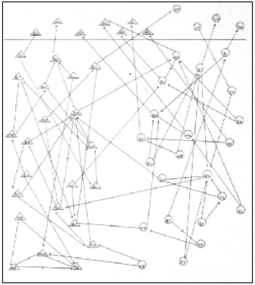
In his **1934** book *Who Shall Survive: A New Approach to the Problem of Human Interrelations*, Jacob Moreno (Romanian-US psychiatrist) introduced *sociograms* and used these graphs/networks to understand relationships. In one study (that was repeated to test changes) he asked each child in various 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

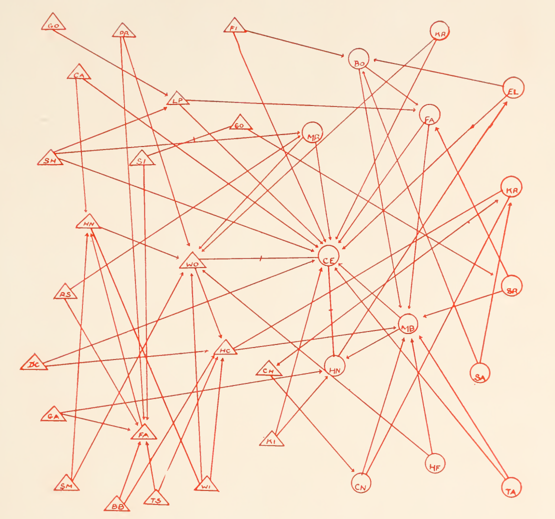


Figure: 21 boys, 14 girls. Directed graph. Every node has out-degree 2. 18 unchosen having in-degree 0. Note also that there are some “stars” with high in-degree.

A closer look at grade 4 in Moreno sociogram

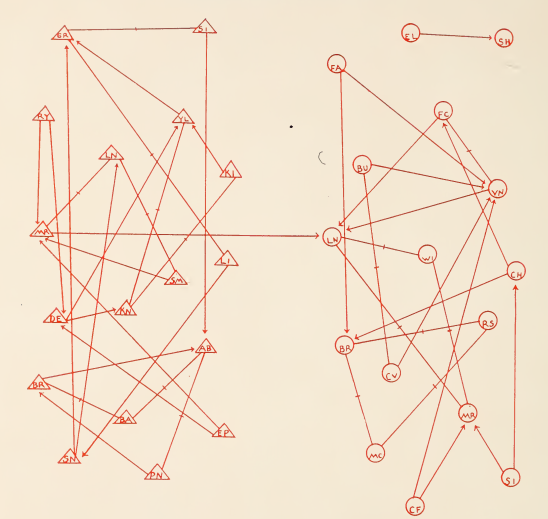


Figure: 17 boys, 16 girls. Directed graph with 6 unchosen having in-degree 0. Moreno depicted his graphs to emphasize inter-gender relations. Note only one edge from a boy to a girl.

A closer look at grade 8 in Moreno sociogram

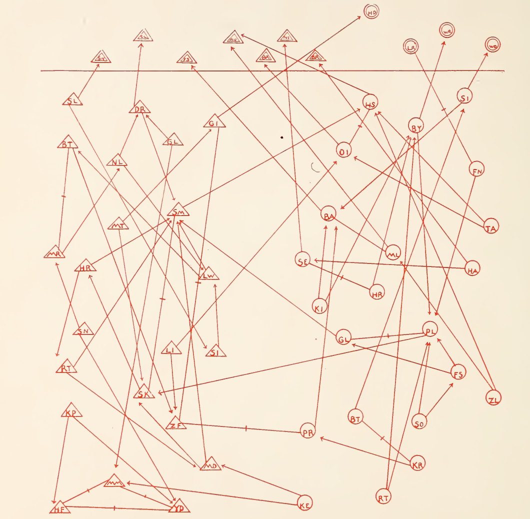


Figure: 22 boys, 22 girls. Directed graph with 12 unchosen having in-degree 0. Some increase in inter-gender relations. Double stars and circles above line indicte different "groups".

Romantic Relationships [Bearman et al, 2004]

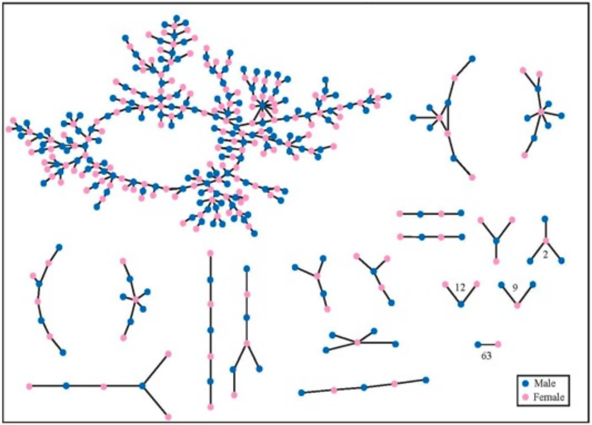


Figure: Dating network in US high school over 18 months.

- 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? ⇒ Cyclyes require simultaneous transplantation; Paths require altruisitic an donor!

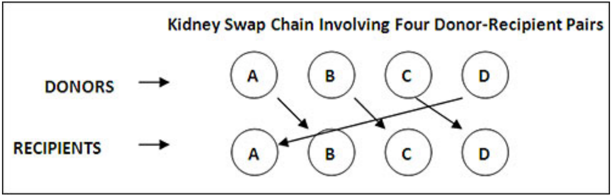
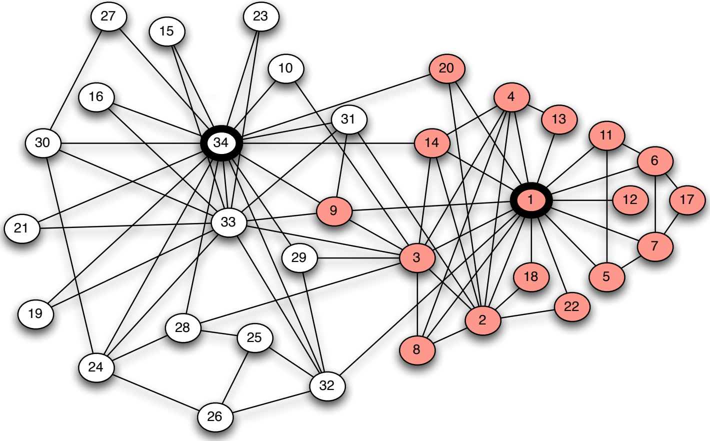


Figure: Dartmouth-Hitchcock Medical Center, NH, 2010

Communities: Karate club division



Karate Club social network, Zachary 1977

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

Communities: 2004 Political blogsphere

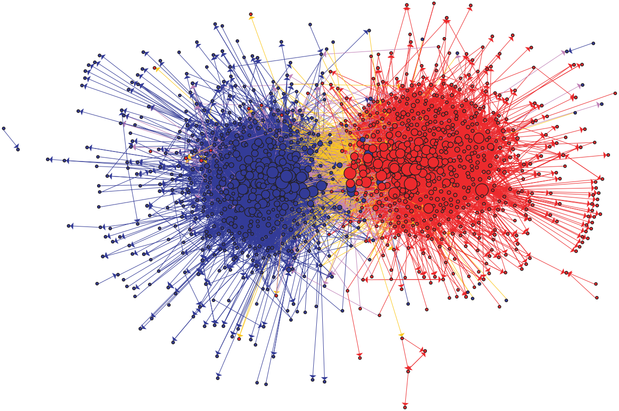


Figure 1: Community structure of political blogs (expanded set), shown using utilizing a GEM layout [11] in the GUESS[3] visualization and analysis tool. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones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

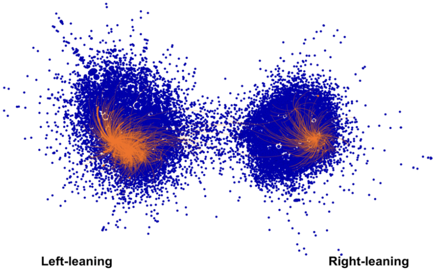


Fig. 1. Retweet Network Graph: RU-IRA Agents in #BlackLivesMatter Discourse. The graph (originally published [3]) shows accounts active in Twitter conversations about #BlackLivesMatter and shooting events in 2016. Each node is an account. Accounts are closer together when one account retweeted another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

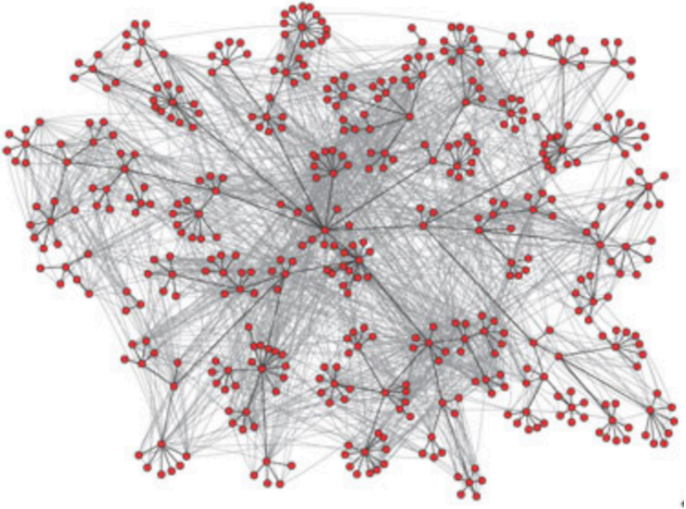
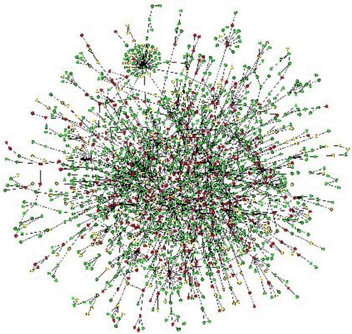


Figure: Email communication amongst 436 employees of Hewlett Packard Research Lab, superimposed on the Lab organizational hierarchy

Protein-protein interaction network

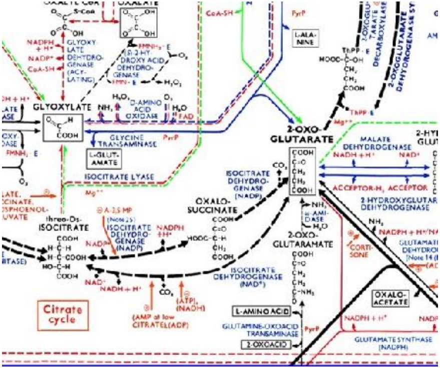


Protein-Protein Interaction Networks

Nodes: Proteins

Edges: 'physical' interactions

Metabolic network


GLYOXYLATE
GLYOXYLATE DEHYDROGENASE (ACYLATING)
D-AMINO ACID OXIDASE
GLYCINE TRANSAMINASE
L-GLUTAMATE
ISOCITRATE LYASE
L-ALANINE
2-OXOGLUTARATE DECARBOXYLASE
2-OXOGLUTARATE DEHYDROGENASE SYSTEM
2-OXO-GLUTARATE
MALATE DEHYDROGENASE
ISOCITRATE DEHYDROGENASE (NADP)
OXALOSUCCINATE
threo-Ds-ISOCITRATE
ISOCITRATE DEHYDROGENASE (NADP)
ISOCITRATE DEHYDROGENASE (NAD+)
2-OXO-GLUTARAMATE
2-HYDROXYGLUTARATE DEHYDROGENASE
ACCEPTOR-H, ACCEPTOR
GLUTAMATE DEHYDROGENASE
CORTISONE
OXALOACETATE
L-AMINO ACID
GLUTAMINE-OXOACID TRANSAMINASE
2-OXOACID
GLUTAMATE SYNTHASE (NADPH)
Citrate cycle
(AMP at low CITRATE)(ADP)
(ATP), (NADH)

Metabolic networks

Nodes: Metabolites and enzymes

Edges: Chemical reactions

The web as a directed graph of hyperlinks

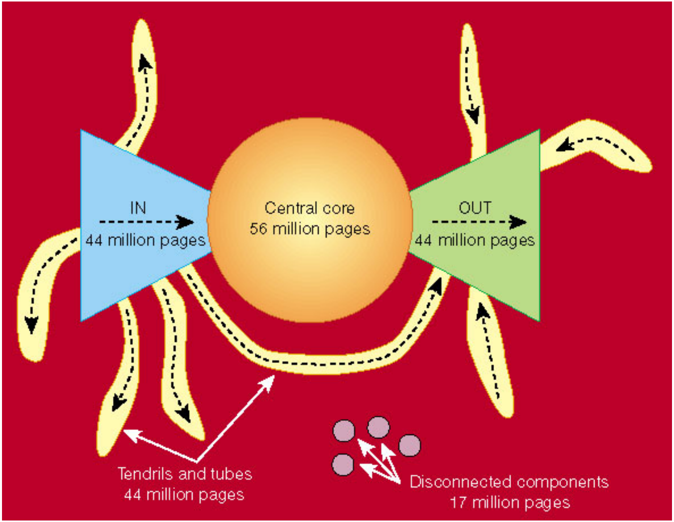


Figure: A schematic picture of the bow tie structure of the 1999 Web. Although the numbers are outdated, the structure has persisted. [Fig 13.7, EK textbook]

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 7.8 billion people. AS of January 2020, The people on Facebook are a *subnetwork* of approximatley 2.9 billion active monthly users of which 1.6 billion are daily users. *Different numbers are reported in different sites.)
- 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, eliminate "bad actors", and eliminate "divisive policies").

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.
- Understand how special interests can target information and misinformation to selected "communities"

And how do we accomplish stated goals

Much of what people do in this research field is empirical analysis. Researchers formulate network models, hypotheses and predictions and then compare against the 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.

Social networks

A social network is a network $G = (V, E)$ where the nodes in V are people or organizations. Social networks can be undirected or directed networks.

The edges can be relations between people (e.g. friendship) or membership of an individual in an organization.

Social networks can be of any size (e.g., a small network like the Karate Club on slide 14 in the week 7 slides) or enormous networks like Facebook and Twitter. We usually think of Facebook as an undirected graph (where *friendship* is an undirected edge) and Twitter as a directed graph (i.e., where *follows* is a directed edge).

Understanding how networks evolve, the resulting structure of social networks, and computational aspects for dealing with large networks is an active field of study in CS as well as in sociology, political science, economics, epidemiology, and any field that studies human behaviour. J. Kleinberg's 2000 analysis with regard to the six degrees of separation phenomena is an early result that sparked interest in algorithmic aspects of social networks.

End of Monday, November 29 class

We ended as we just began discussing the challenges of stuyding large social networks. On Wednesday we we continue this discussion.

If time permits, I will breifly discuss some perhaps surprising studies that give evidence as to how much information about relations (i.e., the edges) between people (i.e. the nodes) can be extracted from just the graph structure. This discussion will utilize graph concepts coming primarily from social networks.

The computational challenge presented by super large networks

The size of some modern networks such as the web and social networks such as Facebook are at an unprecedented scale.

As of November, 2021, Facebook has roughly 2.85 billion monthly active users worldwide. The average facebook user has 155 friends which then implies about $2.85 \cdot \frac{155}{2} \approx 200$ billion edges. It is interesting to note that 90% of daily active users are outside USA and Canada. See https://www.omnicoreagency.com/facebook-statistics/ // for lots of interesting demographic and other facts about Facebook.

What does this imply for the complexity of algorithms involving such super large networks?

Linear is the new exponential

In complexity theory (e.g. in the P vs NP issue that we will be discussing) we say (as an abstraction) that polynomial time algorithms are "efficient" and "exponential time" is infeasible. There are, of course, exceptions but as an abstraction this has led to invaluable fundamental insights.

As problem instances have grown, there was a common saying that "quadratic (time) is the new exponential".

But with the emergence of networks such as the web graph and the Facebook network, we might now say that "linear is the new exponential" when it comes to extracting even the most basic facts about these networks. For example, how do we even estimate the average node degree in a giant network?

There are many facts about large networks that we would like to extract from the network. For example, how do we find "influential" or "interesting nodes" in a social network?

Sublinear time algorithms

What is sublinear time?

In general when we measure complexity, we do so as a funtion of the input/output size. For graphs $G = (V, E)$, the size of the input is usually the number of edges E. (An exception is that when the graph is presented say as an adjacency matrix, the size is n^2 where $n = |V|$.)

Since our interest is in massive information and social networks, we consider sparse graphs (e.g. average constant degree) so that $|E| = O(|V|)$ and hence we will mean sublinear time as a function of n. The desired goal will be time bounds of the form $O(n^{\alpha})$ with $\alpha < 1$ and in some cases maybe even $O(\log n)$ or $polylog(n)$.

Given that optimal algorithms for almost any graph property will depend on the entire graph, we will have to settle for approximations to an optimum solution. Furthermore, we will need to sample the graph so as to avoid having to consider all nodes and edges. And we will need a way to efficiently access these massive graphs,

Coping with massive social graphs continued

One way to help coping with massive networks is to hope to utilize some substantial amount of parallelism. There is an area of current research concerning massive parallel computation (MPC) models where (in principle) we can achieve sublinear time by distributing computation amongst a large (i.e., non constant) number of processors.

But even if we could muster and organize thousands of machines, we will still need random samplng, approximation, and have highly efficient "local information algorithms".

Finally, in addition to random sampling and parallelism, we will have to hope that social networks have some nice structural properties that can be exploited to as to avoid complexity barriers that exist for arbitrary (sparse) graphs. These complexity barriers are hopefully clear from our discussion of complexity theory, *NP completeness* and *NP hardness*.

Preferential attachment models

Preferential attachnment models (also called "rich get richer" models) are probabilistic generative models explaining how various networks can be generated. Namely, after starting with some small graph, when we add a new node u, we create a number of links between u to some number m of randomly chosen nodes $v_1, v_2, \ldots, v_m$. The probability of choosing a v_i is proportional to the current degree of v_i. More generally, the probability of choosing a node v_i can be an increasing function of the degree,

These models have been used to help explain the structure of the web as well as social networks. Furthermore, networks generated by such a process have some nice structural properties allowing for substantially more efficient algorithms than one can obtain for arbitrary graphs.

For such models, there are both provable analytic results as well as experimental evidence on synthetic and real networks that support provable results that follow from the model. (Remember, a model is just a model and is not "reality"; as models are implifications of real networks, they may not account for many aspects in a real network. For example, in this basic model, all the edges for a new node are set upon arrival.

Consequences for networks generated by a preferential attachment process

There are many properties, believed and sometimees proven. about preferential attachment network models that do not hold for uniformly generated random graphs (e.g., create sparse graphs with constant average degree by choosing each possible edge with say probability proportional to $\frac{1}{n}$).

One of the most interesting and consequential proerties is that vertex degrees satisfy a *power law distribution* in expectation. Specifically, the expectation fraction $P(d)$ of nodes whose degree is d is proportional to $d^{-\gamma}$ for some $\gamma \geq 1$. Such a distribution is said to have a *fat tail*.

In a uniformaly random sparse graph (with average degree d_{avg}), with high probability , the fraction of nodes having a large degree $d > d_{avg}$ is proportional to c^{-d} for some $c > 1$.

The Barabasi and Albert preferential model

Barabasi and Albert [1999] specified a particular preferential attachment model and conjectured that the vertex degrees satisfy a power law in which the fraction of nodes having degree d is proportional to d^{-3}.

They obtained $\gamma \approx 2.9$ by experiments and gave a simple heuristic argument suggesting that $\gamma = 3$. That is, $P(d)$ is proportional to d^{-3}

Bollobas et al [2001] prove a result corresponding to this conjectured power law. Namely, they show for all $d \leq n^{1/15}$ that the *expected* degree distribution is a power law distribution with $\gamma = 3$ asymptotically (with n) where n is the number of vertices.

Note: It is known that an actual realized distribution may be far from its expectation, However, for small degree values, the degree distribution is close to expectation.

When we say that a distribution $P(d)$ is a power law distribtion this is often meant to be a "with high probaility" whereas results for networks generated by a preferential attachment process the power law is usually only in expectation.

Proven or observed properties of nodes in a social network generated by preferential attachment models

In addition to the power law phenomena suggesting many nodes with high degree, other properies of social networks have been obseerved such as a relatively large number of nodes u having some or all of properties such as the following: .

- high clustering coefficient defined as : $\frac{(u,v),(u,w),(v,w) \in E}{(u,v),(u,w) \in E}$. That is, mutual friend of u are likely to be friends.
- high centrality ; e,g, nodes on many pairs of shortest paths.

Brautbar and Kearns refer to such nodes (as above) as “interesting indiviudals” and these individuals might be candidates for being “highly influential individuals”. Bonato et al [2015] refers to such nodes as the *elites* of a social network.

Other proven or observed properties of networks generated by preferentical attachment models

- correlation between the degree of a node *u* and the degrees of the neighboring nodes.
- graph has small diameter; suggesting "6 degrees of separation phenomena"
- relatively large dense subgraph communities.
- rapid mixing (for random walks to approach stationary distribution)
- relatively small (almost) *dominating sets* .

On my spring 2020 CSC303 web page, I posted a paper by Avin et al (2018) that shows that preferential attachment is the *only* "rational choice" for players (people) playing a simple natural network formation game. It is the rational choice in the sense that the strategy of the players will lead to a unique equilibrium (i.e. no player will want to deviate assumming other players do not deviate). For those intersted, I have posted (in my CSC303 webpage) a number of other papers on elites in a social network and preferential attachment.

The Small World Phenomena

I already mentioned the small worlds phenomena. A mathematical explanation of this phenomiena (expecially how one hones in on a target recipient) was given by J. Kleinberg in a network formation model that explicitly forces a power law property.

The small world phenomena suggests that in a connected social network any two individuals are likely to be connected (i.e. know each other indirectly) by a short path. Moreover, such a path can be found in a decentralized manner

In Milgram's 1967 small world experiment, he asked random people in Omaha Nebraska to forward a letter to a specified individual in a suburb of Boston which became the origin of the idea of six degrees of separation.