Social and Information Networks

CSCC46H, Fall 2019
Lecture 12

Prof. Ashton Anderson
ashton@cs.toronto.edu
Today

Entire class!
Emphasis on final help
Final info:
Tuesday, Dec 17 2-5pm
You can have one handwritten, double-sided crib sheet (8.5x11)
Lecture 1
A Network!
Components of a Network

**Objects:** nodes, vertices

**Interactions:** links, edges

**System:** network, graph

\[ G(N,E) \]
Why study networks?

Networks are a **universal language for describing complex data**

Networks from science, nature, and technology are more similar than you might expect

**Shared vocabulary** between fields

CS, finance, tech, social sciences, physics, economics, statistics, biology

**Data availability** (and computational challenges)

Web/mobile, bio, health, medical

**Impact!**

Social networking, social media, drug design
The Internet in 1970

A first example

The Internet in 1970
Undirected and Directed Networks

Undirected

- **Links:** undirected (symmetrical, reciprocal)

![Undirected Network Diagram]

- **Examples:**
  - Collaborations
  - Friendship on Facebook

Directed

- **Links:** directed (arcs)

![Directed Network Diagram]

- **Examples:**
  - Phone calls
  - Following on Twitter
Connectivity of Graphs

**Connected component (undirected):**

- Any two vertices can be joined by a path
- No superset with the same property

A disconnected graph is made up of two or more connected components

**Bridge edge:** If we erase it, the graph becomes disconnected.

![Diagram of graphs with connected components and bridge edge](image-url)
Connectivity of Directed Graphs

Strongly connected directed graph
has a path from each node to every other node and vice versa (e.g., A-B path and B-A path)

Weakly connected directed graph
is connected if we disregard the edge directions

It is connected but not strongly connected (e.g., there is no way to get from F to G by following the edge directions)
Strongly Connected Component

Strongly connected component (SCC) is a set of nodes $S$ so that:

- Every pair of nodes in $S$ can reach each other
- There is no larger set containing $S$ with this property

Strongly connected components of the graph: $\{A,B,C,G\}, \{D\}, \{E\}, \{F\}$
**Fact:** Every directed graph is a DAG on its SCCs

- (1) SCCs partitions the nodes of $G$
  - That is, each node is in exactly one SCC
- (2) If we build a graph $G'$ whose nodes are SCCs, and with an edge between nodes of $G'$ if there is an edge between corresponding SCCs in $G$, then $G'$ is a DAG
Bow-tie Structure of the Web

203 million pages, 1.5 billion links [Broder et al. 2000]
Lecture 2
Adjacency Matrix

\[
A_{ij} = 1 \quad \text{if there is a link from node } i \text{ to node } j
\]

\[
A_{ij} = 0 \quad \text{otherwise}
\]

\[
A = \begin{pmatrix}
0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
1 & 1 & 1 & 0
\end{pmatrix}
\]

\[
A = \begin{pmatrix}
0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0
\end{pmatrix}
\]

Note that for a directed graph (right) the matrix is not symmetric.
**Bipartite Graph**

**Bipartite graph** is a graph whose nodes can be divided into two disjoint sets $U$ and $V$ such that every link connects a node in $U$ to one in $V$; that is, $U$ and $V$ are **independent sets**

**Examples:**

- Authors-to-papers (they authored)
- Actors-to-Movies (they appeared in)
- Users-to-Movies (they rated)

**“Folded” networks:**

- Author collaboration networks
- Movie co-rating networks
Node degree, $k_i$: the number of edges adjacent to node $i$
eq \text{e.g. } k_A = 4$

Avg. degree: $\bar{k} = \langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2E}{N}$

In directed networks we define an in-degree and out-degree. The (total) degree of a node is the sum of in- and out-degrees.

$k_C^{in} = 2$ \hspace{0.5cm} $k_C^{out} = 1$ \hspace{0.5cm} $k_C = 3$

Source: Node with $k^{in} = 0$
Sink: Node with $k^{out} = 0$

$\bar{k}^{in} = \bar{k}^{out}$
Connectivity: Degree Distribution

Degree distribution $P(k)$: Probability that a randomly chosen node has degree $k$

$N_k = \#\text{ nodes with degree } k$

Normalized histogram:
\[ P(k) = \frac{N_k}{N} \implies \text{plot} \]
Connectivity: Clustering Coefficient

What’s the probability that a random pair of your friends are connected?

$$C_i \in [0, 1]$$

$$C_i = \frac{e_i}{k_i^2} = \frac{e_i}{k_i(k_i - 1)/2} = \frac{2e_i}{k_i(k_i - 1)}$$

where $e_i$ is the number of edges between the neighbors of node $i$ and $k_i$ is the degree of node $i$

Average clustering coefficient:

$$C = \frac{1}{N} \sum_{i}^{N} C_i$$
Distance: definition

**Distance (shortest path, geodesic)** between a pair of nodes is defined as the number of edges along the shortest path connecting the nodes.

*If the two nodes are disconnected, the distance is usually defined as infinite.*

In **directed graphs** paths need to follow the direction of the arrows.

Consequence: Distance is **not symmetric**: $h_{A,C} \neq h_{C,A}$

$h_{B,D} = 2$

$h_{B,C} = 1$, $h_{C,B} = 2$
Distance: Graph-level measures

- **Diameter**: the maximum (shortest path) distance between any pair of nodes in a graph

- **Average path length** for a connected graph (component) or a strongly connected (component of a) directed graph

\[
\bar{h} = \frac{1}{2E_{\text{max}}} \sum_{i,j \neq i} h_{ij}
\]

where \( h_{ij} \) is the distance from node \( i \) to node \( j \), and \( E_{\text{max}} \) is the maximum number of edges (=n*(n-1)/2)

- Many times we compute the average only over the connected pairs of nodes (that is, we ignore “infinite” length paths)
Simplest Model of Graphs

Erdős-Renyi Random Graphs [Erdős-Renyi, ‘60]

$G_{n,p}$: undirected graph on $n$ nodes and each edge $(u,v)$ appears i.i.d. with probability $p$

Simplest random model you can think of
Random Graph Model

$n$ and $p$ do not uniquely determine the graph!

The graph is a result of a random process

We can have many different realizations given the same $n$ and $p$

$n = 10$
$p = 1/6$
Degree Distribution

Fact: Degree distribution of $G_{np}$ is Binomial.

Let $P(k)$ denote a fraction of nodes with degree $k$:

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

Select $k$ nodes out of $n-1$ \\
Probability of having $k$ edges \\
Probability of missing the rest of the $n-1-k$ edges

$$\bar{k} = p(n - 1)$$
Lecture 3
Networks & Communities

We often think of networks “looking” like this:

What can lead to such a conceptual picture?
Granovetter’s Answer

Two perspectives on friendships:

**Structural:** Friendships span different parts of the network

The two highlighted edges are structurally different: one spans two different “communities” and the other is inside a community

**Interpersonal:** Friendship between two people vary in strength, you can be close or not so close to someone
Triadic closure

Informally: If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future.
Triadic Closure

Triadic closure == High clustering coefficient

Reasons for triadic closure:
If B and C have a friend A in common:

- B is more likely to meet C
  (both spend time with A)
- B and C trust each other more
  (they have a friend in common)
- A has an incentive to bring B and C together
  (easier for A to maintain two disjoint relationships)
Granovetter’s Explanation

Granovetter makes a connection between the social and structural roles of an edge

- **First point: Structure**
  - Structurally embedded edges are also socially strong
  - Long-range edges spanning different parts of the network are socially weak

- **Second point: Information**
  - Long-range edges allow you to gather information from different parts of the network and get a job
  - Structurally embedded edges are heavily redundant in terms of information access
Network Vocabulary: Span and Bridges

**Define: **Span

The *Span* of an edge is the distance of the edge endpoints if the edge is deleted.

**Define: **Bridge edge

If removed, it disconnects the graph

Span of a bridge edge = $\infty$

**Define: **Local bridge

Edge of *Span* $> 2$

(any edge that doesn’t close a triangle)

Idea: Local bridges with long span are like real bridges
Granovetter’s Explanation

Model: Two types of edges:
- Strong (friend), Weak (acquaintance)

Model: Strong Triadic Closure property:
- Two strong ties imply a third edge
  If node A has strong ties to both nodes B and C, then there must be an edge (strong or weak) between B and C

Fact: If strong triadic closure is satisfied then local bridges are weak ties!
Granovetter’s theory leads to the following conceptual picture of networks.
NCAA Football Network

- **Nodes:** Teams
- **Edges:** Games played

NCAA conferences:
- Mid American
- Big East
- Atlantic Coast
- SEC
- Conference USA
- Big 12
- Western Athletic
- Pacific 10
- Mountain West
- Big 10
- Sun Belt
- Independents
Graph Partitioning

Two general approaches:

1. Start with every node in the same cluster and break apart at “weak links” (“divisive clustering”)

2. Start with every node in its own “community” and join communities that are close together (“agglomerative clustering”)
Graph Partitioning

**Definition:** the *betweenness* of an edge is how many (fractional) shortest paths travel through it

- For every pair of nodes A,B say there is one unit of “flow” along the edges from A to B
- Flow between A to B divides evenly among all shortest paths from A to B
- If k shortest paths, 1/k flow on each path
Girvan-Newman algorithm

Divisive hierarchical clustering based on the notion of edge **betweenness** (Number of shortest paths passing through an edge)

**Girvan-Newman Algorithm** (on undirected unweighted networks):

**Repeat until no edges are left:**

- (Re)calculate betweenness of every edge
- Remove edges with highest betweenness (if ties, remove all edges tied for highest)
- Connected components are communities

**Gives a hierarchical decomposition of the network**
How to Compute Betweenness?

Want to compute betweenness of paths starting at node A

Recall BFS goes layer-by-layer

BFS starting from A:
How to Compute Betweenness?

Count the number of shortest paths from A to all other nodes in the graph:

Work downwards:

1. Count the number of shortest paths from A to all other nodes in the graph.
2. For example, the number of shortest A-J paths is equal to the sum of the shortest A-G paths and the shortest A-H paths.
3. The number of shortest A-K paths is equal to the sum of the shortest A-I paths and the shortest A-J paths.

Diagram:

- A is connected to B, C, D, E, F, G, H, I, and J.
- The weights of the edges are shown.
- The counts for shortest paths are shown at each node.
How to Compute Betweenness?

How much flow goes from A to other nodes?

**Compute betweenness by working up the tree:** If there are multiple paths count them fractionally.

The algorithm:
- Add edge flows:
  - node flow = $1 + \sum$ child edges
  - split the flow up based on the parent value
- Repeat the BFS procedure for each starting node $U$

![Flow Diagram](image)
Lecture 4
Signed Networks

Networks with positive and negative relationships

Consider an undirected complete graph
Label each edge as either:

- **Positive**: friendship, trust, positive sentiment, ...
- **Negative**: enemy, distrust, negative sentiment, …
Theory of Structural Balance

Start with the intuition [Heider ’46]:

Friend of my friend is my friend
Enemy of enemy is my friend
Enemy of friend is my enemy

Look at connected triples of nodes:

Balanced
Consistent with “friend of a friend” or “enemy of the enemy” intuition

Unbalanced
Inconsistent with the “friend of a friend” or “enemy of the enemy” intuition
Define: A complete graph is *balanced* if every connected triple of nodes has:

- All 3 edges labeled +  
- or  
- Exactly 1 edge labeled +

![Diagram of balanced and unbalanced networks](image)
Local Balance $\rightarrow$ Global Factions

The Balance Theorem: Balance implies global coalitions
[Cartwright-Harary]

If all triangles are balanced, then either:

A) The network contains only positive edges, or
B) The network can be split into two factions: Nodes can be split into 2 sets where negative edges only point between the sets.

![Diagram showing two factions and a network with positive and negative edges.](image)
Structural Balance

A, B, and C are mutual friends: balanced.

A is friends with B and C, but they don't get along with each other: not balanced.

A and B are friends with C as a mutual enemy: balanced.

A, B, and C are mutual enemies: not balanced.

Figure 5.1: Structural balance: Each labeled triangle must have 1 or 3 positive edges.

Similarly, there are sources of instability in a configuration where each of A, B, and C are mutual enemies (as in Figure 5.1(d)). In this case, there would be forces motivating two of the three people to “team up” against the third (turning one of the three edge labels to a +).

Based on this reasoning, we will refer to triangles with one or three +'s as **balanced**, since they are free of these sources of instability, and we will refer to triangles with zero or two +'s as **unbalanced**. The argument of structural balance theorists is that because unbalanced triangles are sources of stress or psychological dissonance, people strive to minimize them in their personal relationships, and hence they will be less abundant in real social settings than expected.

What if we allow three mutual enemies?
Weak Structural Balance \rightarrow Many Global Factions

**Define:** A complete network is *weakly balanced* if there is no triangle with exactly 2 positive edges and 1 negative edge.

**Characterization of Weakly Balanced Networks:**
If a labeled complete graph is weakly balanced, then its nodes can be **partitioned**
(divided into groups such that two nodes belonging to the same group are friends, and every two nodes belonging to different groups are enemies)

**Global picture:** same thing as before, but with many factions, not necessarily two
Balance in General Networks

So far we talked about complete graphs

**Def 1: Local view**
Fill in the missing edges to achieve balance

**Def 2: Global view**
Divide the graph into two coalitions

**The 2 definitions are equivalent!**
Is a Signed Network Balanced?

Theorem: Graph is \textit{balanced} if and only if it contains no cycle with an odd number of negative edges [Harary 1953, 1956]

Proof by algorithm: We proved this by actually constructing an algorithm that either \textit{outputs} a division into coalitions or a cycle with odd number of negative edges.

Because these are the only two outcomes, this proves the claim.
**Signed graph algorithm:**

**Step 1:** Find connected components on + edges and for each component create a super-node

- Since nodes connected by a + edge must be in the same coalition
- If any – edge in the super node, done (cycle with 1 negative edge)

**Step 2:** Connect components A and B if there is a negative edge between the members

- Note there are only negative edges pointing out of a super-node (otherwise should’ve connected the two super-nodes that have a positive edge)
Homophily

Connections don’t form uniformly at random

**Null model**: what if they were forming at random?

**Measuring homophily**: are there fewer connections between nodes across traits than you’d expect at random?

**Homophily test**: If the fraction of cross-gender edges is **significantly** less than at random, then there is evidence of homophily.
CHAPTER 4. NETWORKS IN THEIR SURROUNDING CONTEXTS

Figure 4.2: Using a numerical measure, one can determine whether small networks such as this one (with nodes divided into two types) exhibit homophily.

And ultimately, one expects most links to in fact arise from a combination of several factors—partly due to the effect of other nodes in the network, and partly due to the surrounding contexts.

Measuring Homophily.

When we see striking divisions within a network like the one in Figure 4.1, it is important to ask whether they are "genuinely" present in the network itself, and not simply an artifact of how it is drawn. To make this question concrete, we need to formulate it more precisely: given a particular characteristic of interest (like race, or age), is there a simple test we can apply to a network in order to estimate whether it exhibits homophily according to this characteristic?

Since the example in Figure 4.1 is too large to inspect by hand, let's consider this question on a smaller example where we can develop some intuition. Let's suppose in particular that we have the friendship network of an elementary-school classroom, and we suspect that it exhibits homophily by gender: boys tend to be friends with boys, and girls tend to be friends with girls. For example, the graph in Figure 4.2 shows the friendship network of a (small) hypothetical classroom in which the three shaded nodes are girls and the six unshaded nodes are boys. If there were no cross-gender edges at all, then the question of homophily would be easy to resolve: it would be present in an extreme sense. But we expect that homophily should be a more subtle effect that is visible mainly in aggregate—as it is, for example, in the real data from Figure 4.1. Is the picture in Figure 4.2 consistent with homophily?

There is a natural numerical measure of homophily that we can use to address questions:

\[ p = \text{Probability that a node is white} \]
\[ q = \text{Probability that a node is red} \]

Prob an edge is between two white nodes? \[ p^2 \]
Prob an edge is between two red nodes? \[ q^2 \]
Prob an edge is between 1 red, 1 white? \[ 2pq \]

Homophily test: \[ 2pq = \frac{4}{9} = \frac{8}{18} \]

Observed: \[ \frac{5}{18} \]
How long is the typical shortest path?

Milgram devised a clever experiment

– Picked ~300 people in Omaha, Nebraska and Wichita, Kansas
– Asked each person to try get a letter to a particular person in Boston (a stockbroker), but they could only send it to someone they know on a first-name basis
– The friends then send to their friends, etc.

64 chains completed, 6.2 steps on average
6 Degrees: Should We Be Surprised?

Assume each human is connected to 100 other people

Then:

Step 1: reach 100 people
Step 2: reach $100 \times 100 = 10,000$ people
Step 3: reach $100 \times 100 \times 100 = 1,000,000$ people
Step 4: reach $100 \times 100 \times 100 \times 100 = 100M$ people

**In 5 steps we can reach 10 billion people**

What’s wrong here?

**Triadic closure:** 92% of new FB friendships are to a friend-of-a-friend [Backstom-Leskovec ‘11]
The Small-World Model

Rewiring allows us to “interpolate” between a regular lattice and a random graph
How to Navigate a Network?

“The geographic movement of the [message] from Nebraska to Massachusetts is striking. There is a progressive closing in on the target area as each new person is added to the chain”

Decentralized Search

The setting:
– Nodes live in a regular lattice, just as in Watts-Strogatz
– Each node has an “address”/location in the grid
– Node \( s \) is trying to route a message to \( t \)
– \( s \) only knows locations of its friends and location of the target \( t \)
– \( s \) does not know random links of anyone else but itself

Geographic Navigation: nodes will act greedily with respect to geography: always pass the message to their neighbour who is geographically closest to \( t \) (what else can they do?)

Search time \( T \): Number of steps it takes to reach \( t \)
What is success?

We know these graphs have diameter $O(\log n)$, so paths are logarithmic in shortest-path length.

We will say a graph is **searchable** if the decentralised search time $T$ is polynomial in the path lengths.

But it’s **not searchable** if $T$ is exponential in the path lengths.

<table>
<thead>
<tr>
<th>Searchable</th>
<th>Not searchable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search time $T$:</td>
<td>Search time $T$:</td>
</tr>
<tr>
<td>$O((\log n)^\beta)$</td>
<td>$O(n^\alpha)$</td>
</tr>
</tbody>
</table>
Kleinberg’s Model [Kleinberg, Nature ‘01]

Nodes still live in a grid, and know their neighbours

Each node has one random “long-range” link

**Key difference:** the link isn’t uniformly at random anymore, it follows geography

Prob. of long link to node \(v\):

\[
P(u \rightarrow v) \sim d(u, v)^{-\alpha}
\]

\(d(u, v)\) … grid distance between \(u\) and \(v\) (address distance, not shortest path)

\(\alpha\) … tunable parameter \(\geq 0\)
Kleinberg’s Model in 1-Dimension

Myopic search in general doesn’t find the shortest path!
Kleinberg’s Model in 1-Dimension

We analyze 1-dimensional case:

Claim: For $\alpha = 1$ we can get from $s$ to $t$ in $O(\log(n)^2)$ steps in expectation

Proof strategy:
Argue it takes $O(\log n)$ to halve the distance
$O(\log n)$ halving steps to get to target

The chains progress from the starting position (Omaha) to the target area (Boston) with each remove. Diagram shows the number of miles from the target area, with the distance of each remove averaged over completed and uncompleted chains.
Lecture 6
How is popularity distributed?

A deeper look at one of our central questions: how connected are people? **How many people do people tend to know?**

Most know some, and some know a ton

How is popularity *distributed* in the population?
A guess

Heights of males in the Italian army
Most values are clustered around a typical value

From "Height and the Normal Distribution: Evidence from Italian Military Data"
Node Degrees in Networks

Take a network, plot a histogram of $P(k)$ vs. $k$

**Plot:** fraction of nodes with degree $k$:

$$p(k) = \frac{|\{u|d_u = k\}|}{N}$$

Flickr social network
$n=584,207$, $m=3,555,115$
Plot the same data on log-log scale:

Flickr social network

\[ n = 584,207, \quad m = 3,555,115 \]

\[ P(k) \propto k^{-1.75} \]

Slope = $-\alpha = 1.75$
The main heavy-tailed distribution we will consider is the power law:

\[ p(x) \propto x^{-\alpha} \]

For example, Newton’s law of universal gravitation follows an “inverse-square law”, e.g. a power law:

\[ F(r) = G \frac{m_1 m_2}{r^2} \]

Where the distance \( r \) is the quantity that is changing.

To make it an actual distribution, include a normalizing constant \( c \)

\[ p(x) = cx^{-\alpha} \]
Height as a Power Law

Why is the mean of the power law so far out?
Power laws are everywhere
Network Resilience

Real networks are resilient to random failures

$G_{np}$ has better resilience to targeted attacks

Need to remove all pages of degree $>5$ to disconnect the Web

But this is a very small fraction of all web pages
Success is inherently unpredictable from quality.
MusicLab:

Who ends up here is pretty **random**!
Rich Get Richer

Example in networks: new nodes are more likely to link to nodes that already have high degree

Herbert Simon’s result:
Power-laws arise from “Rich get richer” (cumulative advantage)

Examples [Price ‘65]
Citations: New citations to a paper are proportional to the number it already has
Herding: If a lot of people cite a paper, then it must be good, and therefore I should cite it too
The Model Gives Power-Laws

Claim: The described model generates networks where the fraction of nodes with in-degree $k$ scales as:

$$P(d_i = k) \propto k^{-(1 + \frac{1}{q})}$$

where $q = 1 - p$

So we get power-law degree distribution with exponent:

$$\alpha = 1 + \frac{1}{q} = 1 + \frac{1}{1 - p}$$
Lecture 7
How to Organize the Web?

How do you organize the Web?

First try: Human curation
Web directories
  Yahoo, DMOZ, LookSmart

Second try: Web Search
Information Retrieval attempts to find relevant docs in a small and trusted set
  Newspaper articles, Patents, etc.

But: The Web is huge, full of untrusted documents, random things, web spam, etc.

So we need a good way to rank webpages!
Idea: links as votes!

If I link to you, that’s usually a good thing

1. Model the Web as a directed graph

2. Use the link structure to compute importance values of webpages

3. Use these importance values for ranking
Hubs and Authorities

Each page has a hub score \( h_i \) and an authority score \( a_i \)

HITS algorithm:
1. **Initialize** all scores to 1
2. Perform a sequence of hub-authority updates:
   — First apply Authority Update Rule
   — Then apply Hub Update Rule
3. **Normalize** (divide authority scores by sum over \( a_i \)'s and same for hubs)

(We normalize since the numbers get very big, and we only care about the relative sizes)
Hubs and Authorities: Example

Apply 2 rounds of hub and authority update steps on the graph below:

Note: in this example, values are very close to convergence after only 2 steps
PageRank: The “Flow” Model

A “vote” from an important page is worth more:

Each link’s vote is proportional to the importance of its source page.

If page $i$ with importance $r_i$ has $d_i$ out-links, each link gets $r_i / d_i$ votes.

Page $j$’s own importance $r_j$ is the sum of the votes on its in-links.

$$r_j = r_i / 3 + r_k / 4$$
Think of PageRank as a “fluid” that circulates around the network, passing from node to node and pooling at the most important ones.

PageRank Algorithm:
1. Initialize all nodes with $1/n$ PageRank
2. Perform $k$ PageRank updates:

**Basic PageRank Update Rule:** Each page divides its current PageRank equally across its outgoing links. New PageRank is the sum of PR you receive.

Page j’s PageRank Update equation: $r_j = \sum \frac{r_i}{d_i}$

Where $d_i = \text{out-degree of node } i$
PageRank: A Problem

In real graph structures, PageRank can pool in the wrong places.

Consider a slightly different graph:

What happens?

All the PageRank ends up here!
PageRank: A Solution

**Scaled PageRank:** only divide a fraction $s$ of the PageRank among outgoing links
The rest gets spread evenly over all nodes

In effect we create a complete graph

**Scaled PageRank Update Rule:** First apply Basic PageRank Update Rule, scale down the values by $s$, then divide the residual $(1-s)/n$ units of PageRank equally: $(1-s)/n$ to each.
PageRank: Random Surfer

Claim: The probability of being at page $X$ after $k$ steps of this random walk is equal to the PageRank of $X$ after $k$ applications of the Basic PageRank Update rule.

The Random Walk: Walker chooses a starting node at random, then at each step picks one of the out-links of its current node uniformly at random.
**Personalized PageRank**

**Goal**: Evaluate pages not just by popularity or global importance, but by how close they are to a given topic

**Solution**: change teleportation vector!

Teleporting can go to:
- Any page with equal prob. (normal PageRank)
- A topic-specific set of “relevant” pages
- A single page/node (random walk with restarts)
Update Rules as Matrix-Vector Multiplication

Recall Hub Update Rule:

\[ h_i \leftarrow M_{i1} a_1 + M_{i2} a_2 + \ldots + M_{in} a_n \]

This corresponds exactly to the simple matrix-vector multiplication \( h \leftarrow M a \)
Authority update rule is similar

\[ a_i \leftarrow M_{1i} h_1 + M_{2i} h_2 + \ldots + M_{ni} h_n \]

This corresponds exactly to the simple matrix-vector multiplication \( a \leftarrow M^T h \)

Transpose the matrix!
Convergence

Recall your eigenvectors and eigenvalues:

\[ Av = \lambda v \]

\( v \) is an eigenvector of \( A \), with corresponding eigenvalue \( \lambda \)

At convergence, performing additional hub-authority steps won’t change anything

Thus Hubs and Authorities converges to the leading eigenvector of \( MM^T \) and \( M^TM \! \)

\[ (MM^T)h^{(*)} = c \cdot h^{(*)} \]

(Full details in the reading)
PageRank Spectral Analysis

Recall the Basic PageRank Update Rule:

$$r_i^{(k+1)} = \sum_{i \rightarrow j} \frac{r_j^{(k)}}{d_i}$$

Define a new matrix $N$:

$$N_{ij} = \frac{1}{d_i}$$ for edges $i \rightarrow j$, 0 otherwise

where page $i$ has $d_i$ out-links

$$r^{(k+1)} = N_1 i r_1^{(k)} + N_2 i r_2^{(k)} + \cdots + N_n i r_n^{(k)}$$

$$r^{(k+1)} = N^T r^{(k)}$$

Similarly, PageRank converges to the leading eigenvector of $N^T$
Lecture 8
What is "rational" play?

Repeat!

44.4 is the new 66.6, and so on

The only "rational" move is guessing 0!

(of course, in real life not everyone is rational)
Exam-Presentation Game

What should you do?
If you knew your partner would study for the exam, what should you do? You’d choose exam (88 > 86)

If you knew your partner would work on the presentation, what should you do? You’d choose exam (92 > 90)

No matter what, you should choose exam!

<table>
<thead>
<tr>
<th></th>
<th>Your Partner</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Presentation</td>
</tr>
<tr>
<td>You</td>
<td></td>
</tr>
<tr>
<td></td>
<td>90, 90</td>
</tr>
<tr>
<td></td>
<td>92, 86</td>
</tr>
</tbody>
</table>
Basic Definitions

A game $G$ is a tuple $(P, S, O)$:

$P$ = set of Players
$S$ = a set of strategies for every player
$O$ = for every outcome (where every player is picking one strategy), a payoff for each player

Payoff matrix summarizes all of these (each dimension is a player, every row/column/etc is a strategy for one player, every cell expresses payoffs for each player)
Underlying Assumptions

Payoffs summarize **everything** a player cares about.

Every player knows everything about the structure of the game: who the **players** are, the **strategies** available to everyone, **payoffs** for each player and strategy.

Every player is **rational**: wants to maximize payoff and succeeds in doing so.

<table>
<thead>
<tr>
<th></th>
<th><strong>Your Partner</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Presentation</strong></td>
<td><strong>Exam</strong></td>
</tr>
<tr>
<td><strong>You</strong></td>
<td><strong>Presentation</strong></td>
</tr>
<tr>
<td></td>
<td>90, 90</td>
</tr>
<tr>
<td><strong>Exam</strong></td>
<td>92, 86</td>
</tr>
</tbody>
</table>
Fundamental Concepts: Strict Dominant Strategy

A strategy that is strictly better than all other options, regardless of what other players do

Exam is a strictly dominant strategy for both players
Sadly, (90,90) is not achievable with rational play
Even if you could commit to preparing for the presentation, your partner would still be better off studying for the final

<table>
<thead>
<tr>
<th></th>
<th>Your Partner</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Presentation</td>
<td>Exam</td>
</tr>
<tr>
<td>You</td>
<td>Presentation</td>
<td>90,90</td>
</tr>
<tr>
<td></td>
<td>Exam</td>
<td>92,86</td>
</tr>
</tbody>
</table>
Fundamental Concepts: Best Response

Let’s define some more of the fundamental concepts we just used. Strategy \( S \) by \( P_1 \) is a **best response** to strategy \( T \) by \( P_2 \) if the payoff from \( S \) is at least as good as anyone other strategy against \( T \)

\[
P_1(S,T) \geq P_1(S',T) \quad \text{for all other } S' \text{ by } P_1
\]

It’s a **strict best response** if:

\[
P_1(S,T) > P_1(S',T) \quad \text{for all other } S' \text{ by } P_1
\]

<table>
<thead>
<tr>
<th>Suspect 2</th>
<th>( NC )</th>
<th>( C )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( NC )</td>
<td>(-1, -1)</td>
<td>(-10, 0)</td>
</tr>
<tr>
<td>( C )</td>
<td>(0, -10)</td>
<td>(-4, -4)</td>
</tr>
</tbody>
</table>

S1’s best response to \( NC \) is: \( C \)
S1’s best response to \( C \) is: \( C \)
A **dominant strategy** for $P_1$ is a strategy that is a **best response** every strategy by $P_2$

A **strict dominant strategy** for $P_1$ is a strategy that is a **strict best response** every strategy by $P_2$

$$
\begin{array}{c|cc}
\text{Suspect 1} & NC & C \\
\hline
NC & -1, -1 & -10, 0 \\
C & 0, -10 & -4, -4 \\
\end{array}
$$

(Note: In Prisoner’s Dilemma, $P_1$ has a strict dominant strategy, so we expect $P_1$ to play it. There can be several dominant strategies, and it’d be unclear which one to expect)
Nash Equilibrium

In 1950, John Nash proposed a **simple** and **powerful** principle for reasoning about **behaviour in general games** (and won the Nobel Prize for it in 1994)

Even when there are no dominant strategies, **we should expect players to use strategies that are best responses to each other**

A pair of strategies \((S,T)\) is a **Nash equilibrium** if \(S\) is a best response to \(T\) and \(T\) is a best response to \(S\)
Mixed Strategies Example: Football

Players: Offense, Defense

Strategies: Run, Pass and Defend Run, Defend Pass

Payoff matrix:

<table>
<thead>
<tr>
<th></th>
<th>Defend Pass</th>
<th>Defend Run</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pass</strong></td>
<td>0, 0</td>
<td>10, −10</td>
</tr>
<tr>
<td><strong>Run</strong></td>
<td>5, −5</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

Mixed Nash:
\[ q = \frac{2}{3} \]
\[ p = \frac{1}{3} \]

No Nash equilibria in this game

O’s expected payoff for **Pass** when D plays p:
\[ 0(q) + 10(1-q) = 10-10q \]

O’s expected payoff for **Run** when D plays q:
\[ 5(q) + 0(1-q) = 5q \]

Defense makes Offense indifferent when q=2/3

\[-75x542]102\]
Lecture 9
Traffic modeled as a game

**Players:** Drivers 1, 2, 3…, 4000

**Strategies:** A-C-B, A-D-B

**Payoffs:** Negative drive time

- A-C-B time: \(-(x/100 + 45)\)
- A-D-B time: \(-(45 + y/100)\)

You want to lower your drive time, so we take the negative drive time as the “payoff”

Notice that this actually describes many equilibria: any set of strategies “2000 choose top, 2000 choose bottom” is an equilibrium (players are interchangeable, so any set of 2000 can be using ACB and any set of 2000 can be using ADB)

For any other set of strategies, deviation benefits someone (therefore isn’t an equilibrium)
6.2. REASONING ABOUT BEHAVIOR IN A GAME

A Related Story: The Prisoner’s Dilemma.

The outcome of the Exam-or-Presentation Game is closely related to one of the most famous examples in the development of game theory, the Prisoner’s Dilemma. Here is how this example works.

Suppose that two suspects have been apprehended by the police and are being interrogated in separate rooms. The police strongly suspect that these two individuals are responsible for a robbery, but there is not enough evidence to convict either of them of the robbery. However, they both resisted arrest and can be charged with that lesser crime, which would carry a one-year sentence. Each of the suspects is told the following story. “If you confess, and your partner doesn’t confess, then you will be released and your partner will be charged with the crime. Your confession will be sufficient to convict him of the robbery and he will be sent to prison for 10 years. If you both confess, then we don’t need either of you to testify against the other, and you will both be convicted of the robbery. (Although in this case your sentence will be less — 4 years only — because of your guilty plea.) Finally, if neither of you confesses, then we can’t convict either of you of the robbery, so we will charge each of you with resisting arrest. Your partner is being offered the same deal. Do you want to confess?”

To formalize this story as a game we need to identify the players, the possible strategies, and the payoffs. The two suspects are the players, and each has to choose between two possible strategies — Confess (C) or Not-Confess (NC). Finally, the payoffs can be summarized from the story above as in Figure 6.2. (Note that the payoffs are all 0 or less, since there are no good outcomes for the suspects, only different gradations of bad outcomes.)

So confessing is a strictly dominant strategy — it is the best choice regardless of what the other player chooses. As a result, we should expect both suspects to confess, each getting a 4-year sentence.

Prisoner’s Dilemma:

<table>
<thead>
<tr>
<th></th>
<th>Suspect 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>-1, -1</td>
</tr>
<tr>
<td></td>
<td>-10, 0</td>
</tr>
<tr>
<td>C</td>
<td>0, -10</td>
</tr>
<tr>
<td></td>
<td>-4, -4</td>
</tr>
</tbody>
</table>

Routing:

![Routing Diagram]

Braess’s Paradox
How bad can it get?

Routing:

Ratio between socially optimal and selfish routing (called the “Price of Anarchy”)?

This example: $80/65 = 1.23\times$ worse

Worst case: How bad can it get?

For selfish routing, “Price of Anarchy” = $4/3$
Game Theoretic Model of Cascades

Game Theory + Social Networks can help us think about this question!

Model every friendship edge as a 2 player coordination game

2 players – each chooses technology A or B

Each person can only adopt one “behavior”, A or B

You gain more payoff if your friend has adopted the same behavior as you

Local view of the network of node $v$
Calculation of Node $v$

Let $v$ have $d$ neighbours — some adopt $A$ and some adopt $B$

Say fraction $p$ of $v$’s neighbours adopt $A$ and $1-p$ adopt $B$

\[
\text{Threshold: } v \text{ chooses } A \text{ if } p > \frac{b}{a + b} = q
\]

Thus: $v$ chooses $A$ if:

\[
ap \cdot p \cdot d > b \cdot (1-p) \cdot d
\]
What are the impediments to spread?

Densely connected communities
• 1–3 are well-connected with each other but poorly connected to the rest of the network
• Similar story for 11–17
• Homophily impedes diffusion

A cluster of density $p$ is a set of nodes such that every node in the set has at least a $p$ fraction of its neighbours in the set

Nodes $\{1,2,3\}$ are a cluster of density $p = \frac{2}{3}$

Nodes $\{11,12,13,14,15,16,17\}$ are a cluster of density $p = \frac{2}{3}$
Simple Herding Model

Decision to be made (resto choice, adopt a new technology, support political position, etc)
People decide sequentially, and see all choices of those who acted earlier
Each person has some **private information** that can help guide their decision
People **can’t** directly observe what others **know**, but **can** observe what they **do**
Simple Herding Model

Model: \( n \) students in a classroom, urn in front
Two urns with marbles:

- “Majority-blue” urn has 2/3 blue, 1/3 red
- “Majority-red” urn has 2/3 red, 1/3 blue

50%/50% chance that the urn is majority blue/red

One by one, each student privately gets to look at 1 marble, put it back without showing anyone else, and guess if the urn is Majority-blue or Majority-red
Simple Herding Model

**Student 1:** Just guess the colour she sees

**Student 2:**
If same as first person, guess that colour.
But if different from first, then since he knows first guess was what first person saw, then he’s indifferent between the two. Guess what he saw

**Student 3:**
If first 2 are opposite colours, guess what she sees (tiebreaker)
If previous 2 are the same colour (blue) and S3 draws red, then it’s like he has drawn three times and gotten two blue, so she should guess majority-blue, *despite her own private information!*
Which is it?

"Broadcast"

Big media (CNN, BBC, NYT, Fox)
Celebrities (Biebs, Taylor Swift)

or

"Viral"

Organically spreading content
Chain letters
How to measure virality?

Solution: **average path length between nodes**

\[
\nu(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} \quad \text{Simple average!}
\]

Originally studied in mathematical chemistry [Wiener 1947] => “Wiener index”
Lecture 10
How Things Spread

Networks define how behaviours, ideas, beliefs, diseases, etc. spread

Last class: behaviour (adoption of an innovation or technology) and information

Today:

Epidemics
Epidemics

Which disease is more dangerous to the population?

vs.

= infected
= susceptible
Modeling Epidemic Diffusion

Biggest difference: model transmission as random

No decision-making, but also the processes by which diseases spread from one person to another are so complex and unobservable at the individual level that it’s most useful to think of them as random

Use randomness to abstract away difficult biological questions about the mechanics of spread
Branching Process

Model as a random process on a tree:

**Wave 1:** First person infected, infects each of $k$ neighbors with independent probability $p$

**Wave 2:** For each infected person, they infect each of $k$ neighbors with independent probability $p$

**Wave 3+:** repeat for each infected person

Here $k=3$
Branching Process: $R_0$

Only two possibilities in the long run: **blow up** or **die out**

How does it die out?
- Dies out if and only if none of the nodes on a given level are infected

Define **Basic reproductive number $R_0$:**
the number of expected new cases caused by an individual

$$R_0 = pk$$
Claim: Epidemic spread in the branching process model is entirely controlled by the reproductive number $R_0$:

- If $R_0 < 1$ then with probability 1 the disease dies out after a finite number of steps.
- If $R_0 > 1$ then with probability $> 0$ the disease persists by infecting at least one person in each wave.

“Go big or go home.”

$R_0 = pk$
SIR Epidemic Models

**S**  =  Susceptible

**I**  =  Infectious: node is infected and infects with prob **p**

**R**  =  Removed: after \( t_I \) time, no longer infected or infectious

Initially some nodes in **I** state, rest in **S** state.

Each node in **I** state remains infected for \( t_I \) time steps

During each step, each node has probability **p** of infecting each susceptible neighbour

After \( t_I \) time steps, no longer **S** nor **I**; removed to **R**
Now: SIS Epidemic Model

\[ S = \text{Susceptible} \]
\[ I = \text{Infectious: node is infected and infects with prob } p \]

Initially some nodes in \( I \) state, rest in \( S \) state.
Each node in \( I \) state remains infected for \( t_I \) time steps
During each step, each node has probability \( p \) of infecting all neighbors
After \( t_I \) time steps, node returns to \( S \)
Transient Contacts & Concurrency

A less random model: it matters in what order contact is made in the contact network.

Concurrency: having two or more contacts at once.
Epidemics vs. Behaviour

Simple vs. complex diffusion
Epidemics vs. behaviour

What's the difference?

Recall the small-world model
Simple Diffusion

Large world:

Small world:
Complex Diffusion

Large world:

Small world:
Lecture 11
Voting

We want to aggregate many individuals’ preferences

What are individual preferences?

Setup: a group of $k$ people are evaluating a finite set of possible alternatives
Individual preferences

Every person has a preference relation over the alternatives, denoted $>_{i}$ for player $i$

Must satisfy two properties:

**Complete**: all pairs of distinct alternatives $X$ and $Y$, either $X >_{i} Y$ or $Y >_{i} X$

**Transitive**: if $X >_{i} Y$ and $Y >_{i} Z$ then $X >_{i} Z$
Individual preferences

Another way of expressing preferences: ranked list

For example:

Ranked list $\rightarrow$ preference relation

Obviously complete and transitive

Preference relation $\rightarrow$ ranked list

Less obvious but still true
Voting Systems

**Voting system**: a method that takes a set of complete and transitive individual preference relations (or ranked lists) and outputs a group ranking.

When there's only two alternatives, what should we do?

**Majority Rule**: whoever is preferred by a majority of the voters wins, other one is second.

(let k be odd to avoid ties)
Majority rule with at least three alternatives can produce a *non-transitive* group ranking

**Cycle on preferences => non transitive => bad!**
Other systems?

Majority rule led to some **bad outcomes**

What about other strategies?

**Positional voting:** produce a group ranking directly from the individual rankings

  - Forget **pairwise comparisons**
  - Each alternative receives a certain **weight** based on its positions in all the individual rankings
Borda count

Heisman trophy in college football (and NBA MVP, etc.) all use the following method: get weight 0 for being picked last, 1 for being second last, ..., k-1 for being picked first

**Repeat for each voter, tally up the scores, and rank**

**Example: two voters, four alternatives**

Voter 1: A > B > C > D
Voter 2: B > C > A > D

A: 3 + 1 = 4
B: 2 + 3 = 5
C: 1 + 2 = 3
D: 0 + 0 = 0

Group ranking: B > A > C > D

**Called the “Borda Count”**
Borda count

Borda Count is susceptible to “irrelevant alternatives”
What voters think of Frozen should be irrelevant to how they feel about relative ranking of TG and CK

But it isn’t

This gives rise to another problem: voters can strategically misreport their preferences

For example, say voters 4 and 5 actually had the true ranking TG > CK > F

1,2,3: CK >i TG >i F
4,5: TG >i CK >i F

Borda: CK >i TG >i F

By lying and reporting TG >i F >i CK, they get TG to win
Good Voting Systems

What satisfies Unanimity and IIA and non-dictatorship?
With two alternatives, majority rule clearly satisfies all

Arrow’s Theorem [Arrow 1953]: With at least three alternatives, no voting system satisfies Unanimity, IIA, and Non-dictatorship

In general, there is no good voting system!
In practice, this means that there will always be inherent tradeoffs we have to choose from
Ideal Points

Assume the preferences lie on a one-dimensional spectrum, and each voter has an “ideal point” on the spectrum.

They evaluate alternatives by proximity to this ideal point.

Actually we can assume something weaker: each voter’s preferences “fall away” consistently on both sides of their favourite alternative.
Majority rule with single-peaked preferences

Recall majority rule: compare every pair of alternatives X and Y, and decide X > Y or Y > X by the majority of voters

Claim: If all individual rankings are single-peaked, then majority rule applied to all pairs of alternatives produces a group preference relation that is complete and transitive.

In other words, majority rule works!
Insincere Voting

We just assumed sincere voting

But there are very natural situations where a voter should actually lie, even though her goal is to maximize the probability that the group chooses the right alternative!

Example, information cascades-style:

Experimenter has two urns, 10 marbles each
One urn has 10 white marbles ("pure") and the other has 9 green and one white ("mixed")
Three people privately draw one marble and guess what urn it is, and all win money if the majority of them are right
Insincere Voting

Suppose you draw a white marble

→ Way more likely that urn is pure than mixed

If you draw a green marble

→ Know for sure it’s mixed

But what should you guess?

First, when will your guess actually matter?

If the two others agree, then your guess doesn’t change anything!

Only case where it matters is if they’re split

If they’re split, someone said mixed, so they know it’s mixed!

Then you should guess mixed to break the tie the right way!

Assuming others vote sincerely, you have an incentive to vote insincerely =>

everyone voting sincerely is not a Nash equilibrium