Lecture 1: Introduction to Social and Economic Networks

Alexander Wolitzky

MIT

6.207/14.15: Networks, Spring 2022

This Course

Introduction to study of networks, focusing on formal analysis of social and economic networks, at a relatively advanced undergraduate level.

- Exciting + active area at intersection of econ and CS.
- Touches on many topics of current interest: spread of pandemics, supply chain breakdowns, misinformation and polarization on social networks...

Prereqs: basic probability at the level of 6.041, 14.30, or 18.600

- Will also freely use basic calculus and linear algebra
- No knowledge of economics, game theory, or graph theory is assumed

A typical lecture in this course will consist of a relatively deep dive into one or more canonical models of some network phenomenon, along with focused discussion of real-world applications.

Today's lecture is different: whirlwind tour of topics and questions in networks, to try to give a sense of what this subject is about.

Course Logistics

Class attendance and active participation is expected and essential.

Readings:

- Lecture slides (required)
- Textbook readings from Easley and Kleinberg, *Networks, Crowds, and Markets*, and Jackson, *Social and Economic Networks* (recommended)

Assignments:

- Weekly PSets (40%)
- Midterm (30%) and Final (30%)
- For grad students enrolled in 14.150: additional literature review / research proposal.

See the syllabus for details.

Introduction

- What are networks? Why study networks? Which networks? What tools?
- Networks are a way of representing interactions among some kind of units.
  - In the case of social and economic networks, these units (**nodes**) are usually individuals or firms.
  - The connections between them (**links**) can represent any of a wide range of relationships: friendship, business relationship, communication channel, etc.
- The study of networks can encompass an extremely wide range of interactions.
  - Communication networks (e.g., Web, internet, gossip).
  - Transportation.
  - Organizational structure.
  - Trade and intermediation.
  - Credit and financial flows.
  - Friendship or trust.
  - Spread of epidemic diseases.
  - Diffusion of products, innovations, and ideas.

An Example: Understanding and Mitigating the Covid-19 Pandemic

The simplest epidemiology models ignore network structure and heterogeneities among different types of people, and analyze the spread of a disease in a uniform population where everyone is equally likely to infect everyone else.

- These standard models—"SIR" and "SIS" models—are very important, and we'll cover them in this course.

The Network Economics of Covid-19 (cntd.)

However, the standard models leave out two crucial ingredients, which are at the heart of this course.

1. **Physical network structure and heterogeneities ("graph theory"):** Some people are much more active than others, and people occupy different positions in the network of social interactions (which does evolve over time, but not as much as a model based on uniform random mixing would suggest).
2. **Individual behavioral responses ("game theory"):** In response to a major shock like a pandemic, people adjust their behavior "strategically." When the disease is raging, many people will reduce their activity levels even if there is little or no government intervention. When infection levels are lower, many people will increase activity, even if the government tries to maintain social distancing/lockdowns.

The Network Economics of Covid-19 (cntd.)

Accounting for network-type effects and behavioral responses can dramatically affect predictions and policy responses.

The **herd immunity threshold** is the fraction of the population that must be immune to the virus (through either past infection or vaccination) to cause infections to fall.

- Accounting for network structure and hetereogeneities can have a big effect on this threshold: if some people are much more active than others, they're more likely to get sick first, and when they become immune it has a disproportionate effect on the virus's ability to spread. Herd immunity threshold may be lower once account for networks/heterogeneities.
- On the other hand, once many people are immune, networks/heterogeneities may make it harder to prevent the virus from flaring up again, because if non-immune people cluster together (geographically, socially, etc.) this can give the virus a toe-hold that would not exist if non-immune people were evenly spread throughout the population.

The Network Economics of Covid-19 (cntd.)

The effectiveness of lockdowns/social distancing measures is dramatically affecting by behavioral responses.

- During the pandemic, individual activity/mobility (e.g., measured by cell phone activity tracking) has been only loosely correlated with government interventions.
- Accounting for behavioral responses crucially affects what policies are feasible and desirable. For example, early in the pandemic some argued for "targeted lockdowns" of high-risk groups like the elderly. But if people benefit more from being active when others are active, releasing some people from lockdown may make it more difficult to keep others locked down.
  (This is an example of **strategic complementarity**.)

The Study of Social and Economic Networks

The pandemic is just one example where sound analysis depends on understanding both social network structure and strategic behavior.

- Many human interactions are "networked" and also involve individual decisions.
- So, much network analysis has some focus on social and economic networks (even when the main interest may be on understanding physical or communication networks).
    - E.g., social network structures, such as Facebook, are superimposed on the Web.

Social and Economic Networks (cntd.)

- Social and economic networks are characterized by both the **pattern of linkages** and the **interactions and decisions that take place over the links.**
  - Will you lend your friend money? Follow their advice? Imitate their behavior? Share a news story with them?
- Most of these decisions are made in a goal-oriented, "strategic" way.
- Hence, we analyze them using game theory: the mathematical analysis of strategic interactions among multiple individuals.
- Studying social and economic networks fundamentally combines **graph theory** (physical modeling of network structure) and **game theory** (strategic modeling of interactive behavior).
- You will learn a substantial amount of graph theory and game theory in this course, albeit focused on the parts of these fields most relevant for studying social and economic networks.

How to Understand Networks?

Networks are complicated objects. This complexity affects how we study networks.

- In a network with N nodes, there are $N\left(N-1\right)/2$ possible links. (Why?)
- Each of these links can be present or absent, so there are $2^{N(N-1)/2}$ different networks with N nodes.
- $\implies$ there are more networks with 20 nodes than there are elementary particles in the universe!
- And most networks we'll discuss have many more than 20 nodes. E.g., there are 2 billion websites.

How to Understand Networks? (cntd.)

The vast number of networks means we can never hope to fully understand each of them.

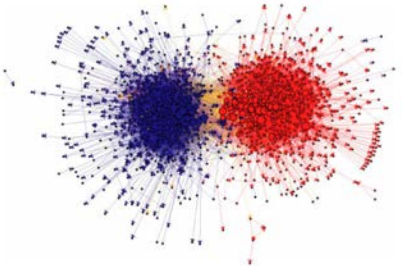
- Instead, develop theoretical frameworks for identifying key properties of networks and the processes that play out on them.
- To some extent these frameworks will be application-dependent, but there are also important commonalities.
- E.g., there are commonalities, but also important differences, in the factors that determine the spread of an epidemic disease vs. the spread of a new technology.

This course will introduce you to some of the main theoretical frameworks and tools for analyzing social and economic networks, and will explore how they are applied in different real-world settings.

Rest of this Lecture

- Mention some more motivating examples.
- Introduce basic language for modeling networks (more details next lecture)
- Discuss some canonical examples from sociology and economics
- Preview some topics from the rest of the class

Visual Examples: 1

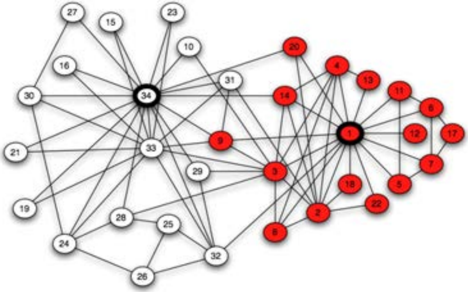


© ACM. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

Network structure of links between political blogs prior to 2004 US Presidential election reveals two well-separated clusters (Adamic-Glance, 2005).

- If we didn't already know US had 2 main political parties, we could tell from this.

Visual Examples: 2

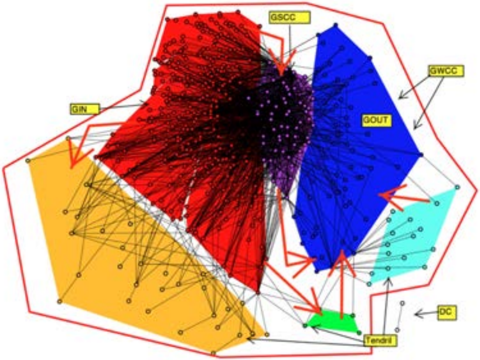


© Easley & Kleinberg. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

Social network of friendships within 34-person karate club (Zachary, 1977).

- Club eventually split apart into white and red groups.
- Zachary used the idea of a "minimum cut" to correctly predict the split based on the network (except node #9).

Visual Examples: 3



Courtesy of Elsevier, Inc., https://www.sciencedirect.com. Used with permission.

Network of loans among financial institutions (Bech-Atalay, 2008).

- The connectivity structure is informative of what roles different institutions play in the financial system.
- Who's "too inter-connected to fail"?

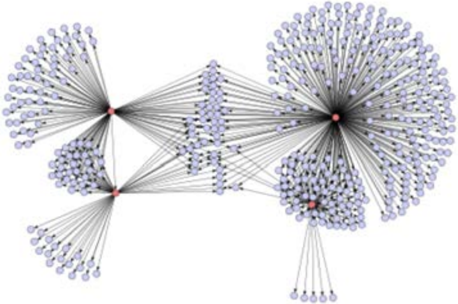
Visual Examples: 4

© Valdis Krebs. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

The spread of a tuberculosis outbreak (Andre et al. 2007).

- Black=TB patients, other nodes are patients' contacts who did not develop TB.
- The spread of epidemic diseases is a classic example of a **dynamic process** on a network.

Visual Examples: 5

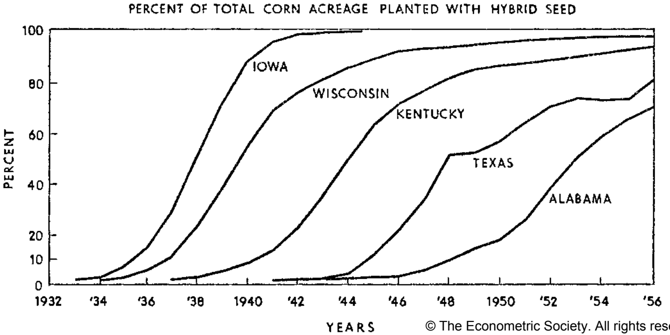


© ACM. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

The network of email recommendations for a Japanese graphic novel (Leskovec et al. 2007). Another diffusion process.

- But different from spread of disease, because recommending a new product (or adopting a new technology) is a **choice**.
- Again, processes on networks often involve both physical and social/economic components.

Visual Examples: 6



© The Econometric Society. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

Percentage of total corn acreage planted with hybrid seed (Ryan and Gross, 1943; Griliches, 1957).

- S-shaped adoption curves indicative of social learning / "word-of-mouth": early adopters discover innovation directly, then adoption accelerates as people learn from their neighbors.
- Similar curve describes infections by an epidemic disease.

Conceptual Example: Do We Live in a Small World?

- Early 20th century Hungarian poet and writer Frigyes Karinthy first came up with the idea that we live in “small world”. He suggested that any two people among the 1.5 billion people on Earth were linked through at most five acquaintances.
- The sociologist Stanley Milgram made this famous in his study “The Small World Problem” (1967).
- Milgram asked some residents of Wichita and Omaha to send a letter to a target person in Boston by sending it to a personal acquaintance, who would then do likewise, and so on until the target person was reached.
- This let Milgram measure how many “intermediate nodes” would be necessary to link the original sender and the target.
- 42 of the 160 letters supposedly made it to their target, with a median number of intermediaries equal to 5.5.

Do We Live in a Small World? (cntd.)

- Hence was born the idea of six degrees of separation.
- Can you think why Milgram's procedure could give misleading results? (Too low an estimate? Too high?)
- There are similar studies for other types of networks.
- For example, Albert, Jeong, and Barabasi ("Diameter of the World Wide Web," 1999) estimated that in 1998 it took on average 11 clicks to go from one random website to another (at the time there were 800 million websites).
- What do these "small world" results imply about other aspects of network structure?
  - Are they surprising?
  - How should we interpret these results?

Interpreting Small Worlds: A Simple Model

- Suppose that each node has λ **neighbors** (e.g., each website has links to λ other websites).
- Each of my λ neighbors will then have λ neighbors themselves.
- Suppose, unrealistically, that my neighbors don't have any neighbors in common (this can't be exactly correct because my neighbors have at least one common neighbor, namely me, but sometimes it's a good approximation).
- Then in two steps, I can reach λ^2 other nodes.
- Repeating the same reasoning (and maintaining the same simplifying assumption), in d steps I can reach λ^d other nodes.
- Now imagine that this network has $n \approx \lambda^d$ nodes.
- This implies that the "degree of separation" (average distance between two nodes) is approximately

$$d = \frac{\ln n}{\ln \lambda}.$$

Interpreting Small Worlds (cntd.)

- Suppose we do take the formula $d = \ln n / \ln \lambda$ as a plausible first approximation for the "degree of separation".
- Suppose that as in Karinthy's time the world population is about 2 billion. How many friends would everyone need to have for the degree of separation to equal 5?
- Solving our formula for λ, we obtain

$$\lambda = \exp\left(\frac{\ln n}{d}\right).$$

  Plugging in $n =$2 billion and $d = 5$ gives

$$\lambda = \exp\left(\frac{\ln 2,000,000,000}{5}\right) \approx 72.$$

- So, Karinthy's "five degrees of separation" was not a crazy guess (or at least it matches our simple model).

Interpreting Small Worlds (cntd.)

- When Milgram wrote in the mid-1960s, world population was about 3.5 billion. For the degree of separation to equal 6, the required number of friends of each individual would be about

$$\exp\left(\frac{\ln 3,500,000,000}{6}\right) \approx 39.$$

- Not too different a number.

Interpreting Small Worlds (cntd.)

- The simplifying assumption that my neighbors aren't neighbors with each other is called **branching process approximation**.
  It imagines the network is a "tree," when really it isn't.
- Sometimes this is very useful; other times it can be misleading.
- This assumption rules out **triangles** or **cycles** (and more generally **clustering**, a measure of the density of triangle), which are very common in social networks, web links, and other networks.
- Interestingly, however, in **Erdos-Renyi** random graphs, where links form uniformly at random, we will see that average distance can be approximated for large n by $d = \ln n / \ln \lambda$ (where λ is the expected degree of a node).
- This is because cycles are relatively rare in such graphs.
- We will study Erdos-Renyi random graphs in detail in Lecture 5.

Interpreting Small Worlds (cntd.)

- Lack-of-cycles is a way in which Erdos-Renyi random graphs, though mathematically convenient, are not good approximations to social networks.
- In reality, the shortest path between remote people usually passes through special "connectors", such as their most popular friend, cousin in a different city, or political representative.
- Models of **small world networks** modify the Erdos-Renyi random graph model to try to capture this type of pattern (albeit not always perfectly).
- We will study these models in Lecture 6.

Networks as Graphs

- We typically represent networks mathematically as **graphs**, which formalize the patterns of relationships (**links**) between different units (**nodes**).
- Graphs can be **directed** or **undirected**, depending on what kind of relationship they represent.
  - E.g. web links are directed, Facebook friend graph is undirected, actual friend graph is. . . ?
- Graphs can also be **weighted** or **unweighted**, depending on whether links differ their importance, capacity, likelihood of materializing, etc.
  - E.g. weak vs. strong ties
- At the simplest level, a (unweighted) graph is

$$G = (N, E)$$

N = set of n nodes in the graph (e.g. websites, individuals)
E = set of edges, linking nodes in the graph

Notation for Links and Directed/Undirected Graphs

- We write $i \in N$ if i is a node in this network, and $(i, j) \in E$ if there is a link from i to j.
  - Sometimes abbreviate to $ij \in E$.
- In an undirected graph, $(i, j) \in E$ if and only if $(j, i) \in E$.
- In a directed graph, $(i, j) \in E$ does not necessarily imply that $(j, i) \in E$.

Alternative Notation: The Adjacency Matrix

- We can also use the notation $g_{ij} = 1$ if $(i, j) \in E$ and $g_{ij} = 0$ otherwise.
- Then the $n \times n$ matrix $g = (g_{ij})_{(i,j) \in n \times n}$ contains all information about who is linked to whom.
- This matrix is called the **adjacency matrix** of the graph.
- As we will see starting next class, we can often do simple matrix algebra on g to derive properties of networks, for example determining the number of paths of a certain length between any two nodes.
- For a weighted graph, we can write $g_{ij} > 0$ if $(i, j) \in E$ and $g_{ij} = 0$ otherwise.
  - In this case, the magnitude of g_{ij} corresponds to the strength of the link.

Importance of Networks in Sociology?

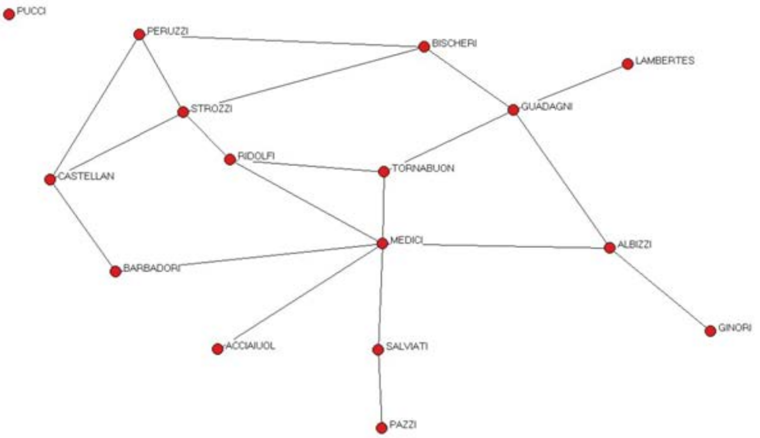
Let's use this notation to understand some famous case studies on network analysis in sociology and economics.

- Sociology is largely about group interactions: thus, network structure is naturally important.
- Notions such as **social capital**, **power**, or **leadership** may be best understood by studying the network of interactions within groups.
- Traditional sociology is largely descriptive and nonmathematical.
- In recent decades, researchers have asked whether the study of networks can bring more analytic power to sociology.
- For example, what network statistics is "social power" related to? What kind of relationships and linkages does a leader need to have in a community?
- Network analysis can also be informative about the dynamics of groups (as in the karate club example).

Example: Power in a Network... in Renaissance Florence

- Cosimo de Medici ultimately formed the most politically powerful and economically prosperous family in Renaissance Florence, dominating Mediterranean trade.
- The Medicis were initially less powerful than many other important families, both in terms of political dominance of Florentine institutions and economic wealth.
- How did they achieve prominence?
- It could just be luck. In social science, we have to be very careful to distinguish luck from systematic patterns, and correlation from causation. Trying to use network structure to explain everything is a road to pseduo-science.
- But in this case there is a very plausible network-based explanation, offered by Padgett and Ansell (1993) "Robust Action and the Rise of the Medici", who argue that the Medici became the most powerful family because of their position in the social network of Florence.

Power in a Network (cntd.)



Courtesy of Matthew O. Jackson. Used with permission.

Network of marriages in 15th century Florence.

- Think "marriage" = "political alliance", not "soul mate".
- How can we measure the power / "centrality" of the Medici?

Power in a Network (cntd.)

- The Medici are linked by marriage to more families that anyone else (6 links, vs. 4 for the Strozzi and Guadagni)
- But Padgett and Ansell argued that the power of the Medici was not just due to their number of connections, but to their "centrality" in the social network, and especially to their role as "brokers" that tie the network together.

Power in a Network (cntd.)

- One measure of centrality in a network that can capture this brokerage role is "betweenness centrality":
- Let $P(i,j)$ be the number of shortest paths connecting family i to family j.
- Let $P_k(i,j)$ be the number of shortest paths connecting these two families that include family k.
- The **betweenness centrality** of node k is then defined as

$$B_k \equiv \sum_{(i,j):i \neq j, k \neq i,j} \frac{P_k(i,j)/P(i,j)}{(n-1)(n-2)},$$

with the convention that $P_k(i,j)/P(i,j) = 0$ if there's no path from i to j.

- That is, for each pair of families (i,j) calculate the fraction of the shortest paths that go through family k, and then average this over all pairs (i,j) not including k.

Power in a Network (cntd.)

- It turns out that this betweenness measure B_k is very high for the Medicis, 0.522.
- The second-highest B_k is 0.255 (the Guadagni).
  - This would have been hard to "eye-ball".
- So, perhaps, the Medicis were uniquely well-positioned to coordinate different families' actions, and were essential in holding together the network of alliances in Florence.
- Their main rivals were the Albizzi, who were much less central.
- In a key showdown in September 1434, the Albizzis called on other families to send armed men to prevent the Medicis from taking over the government, but the other families did not respond in time, while the Medicis quickly got armed support from several families.
- Is this a good measure of "social power"? Of political power? What does it contribute to our understanding? What does it leave out?

Importance of Networks in Economics?

- Economics is about the allocation of scarce resources: exchange, cooperation and competition, learning and information aggregation, technology adoption, etc.
- In reality, much of this allocation takes place in networked settings, where participants with close, long-term relationships interact in the context of both other individuals' relationships and economic institutions like firms and marketplaces.
- But, much of economics studies one of two extremes:
  1. **Markets** with anonymous interactions among a large number of participants.
  2. **Games** with a small number of players.

  - E.g. competitive equilibrium at one end, bargaining and auctions at the other.
- Can we develop new insights by systematically analyzing the underlying network of relations?

An Example of "Network Effects": Finding a Job

How do people find jobs?

- Myers and Shultz (1951) *The Dynamics of a Labor Market* and Rees and Shultz (1970) *Workers in an Urban Labor Market* documented that most workers found their jobs through "a social contact".
- Granovetter (1973) "The Strength of Weak Ties": most people find jobs through acquaintances, *not* close friends.
- Is this suprising?
- Yes and no. No because people have many more acquaintances than friends, but also because of clustering: if 1 and 2 are close friends, and 2 and 3 are close friends, then 1 and 3 are very likely to know each other.
  So more likely to get referrals to a manager *whom you don't already know* through an acquaintance than a close friend.
  - This is known as the **strength of weak ties**.
- This form of "social capital" provided by weak ties is similar to the "brokerage" example just discussed.

Triadic Closure

Let's see a simple example of network analysis, which formalizes Granovetter's "stength of weak ties."

- Let's represent a weighted (undirected) graph as $G = (N, E, E')$, where $E' \subset E$ represents "strong ties".
- $(i, j) \in E$ means that i and j are acquaintances, while $(i, j) \in E'$ means that i and j are close friends.
- The **strong-tie triadic closure** property is the following:

$$\text{if } (i, j) \in E' \text{ and } (j, k) \in E', \text{ then } (i, k) \in E.$$

- This property is often violated, so we may consider a "probabilistic" version, where we say that the conditional probability that $(i, k) \in E$ given $(i, j) \in E'$ and $(j, k) \in E'$ is greater than the unconditional probability that $(i, k) \in E$: that is,

$$\begin{aligned} & \mathbb{P}\left((i, k) \in E \mid (i, j) \in E' \cap (j, k) \in E'\right) \\ > \; & \mathbb{P}\left((i, k) \in E \mid (i, j) \in E \setminus E' \cap (j, k) \in E'\right). \end{aligned}$$

Triadic Closure (cntd.)

- Now let's say that a worker i can get a job with a manager k **through a referral from** j if and only if j is close friends with k **but** i and k do not know each other.
- Assume the probabilistic version of strong-tie triadic closure, and assume also that

$$\mathbb{P}\left((j,k) \in E' \mid (i,j) \in E'\right) = \mathbb{P}\left((j,k) \in E' \mid (i,j) \in E \backslash E'\right)$$

- Then we can show that close friends are less useful for finding jobs than acquaintances.
- Formally, let R be the event that i gets a job with k through a referral by j, and $\mathbb{P}(R)$ denote the probability of this event.
- Then

$$\mathbb{P}\left(R \mid (i,j) \in E' \right) < \mathbb{P}\left(R \mid (i,j) \in E \backslash E'\right),$$

  potentially explaining Granovetter's findings.
- Intuition: if $(i,j) \in E'$ and $(j,k) \in E'$, then it's likely that $(i,k) \in E$, so i doesn't "need" j to get the job.

Proof

We have

$$
\begin{aligned}
& \mathbb{P}\left(R \mid (i,j) \in E'\ \right) \\
= \ & \mathbb{P}\left((j,k) \in E' \cap (i,k) \notin E \mid (i,j) \in E'\ \right) \\
= \ & \mathbb{P}\left((j,k) \in E' \mid (i,j) \in E'\ \right) \\
& \times \mathbb{P}\left((i,k) \notin E \mid (i,j) \in E' \cap (j,k) \in E'\ \right) \\
< \ & \mathbb{P}\left((j,k) \in E' \mid (i,j) \in E \backslash E'\ \right) \\
& \times \mathbb{P}\left((i,k) \notin E \mid (i,j) \notin E \backslash E' \cap (j,k) \in E'\ \right) \\
= \ & \mathbb{P}\left(R \mid (i,j) \in E \backslash E'\right),
\end{aligned}
$$

where the inequality uses equality of the first terms and probabilistic triadic closure.

More Economic Examples

- How do people learn about new products?
  - Japanese graphic novel example.
  - "Cult followings".
- How does a new technology spread?
  - Hybrid corn example.
  - Word-of-mouth from early adopters leads to a distinctive adoption pattern.
- How do people form their political, social and religious opinions?
  - Imitate family, friends, and neighbors?
  - More sophisticated information aggregation by talking and observing friends and news sources?
  - How do social networks matter?

The Impact of the Internet

- The rise of the internet and other advances in information and communication technology have changed the nature of social networks.
- For example, how do the internet and social media change what information we obtain and how we process it?
  - Recall the political blogs picture. Obviously, the internet does not guarantee that each individual hears a greater diversity of opinions, or ultimately learns the truth.
  - Can greater access to information cause "herding"—excessive copying of others behavior and information—instead of the "wisdom of crowds"?

Wisdom or Folly of Crowds?

- "Wisdom of the crowds": combining the information of many, particularly people with different perspectives and diverse experiences, leads to better decisions.
- Marquis de Condorcet and Francis Galton: "average of a group is wiser than its members"
- Galton: people at a fair guessing the weight of an ox: "The average competitor was probably as well fitted for making a just estimate of the dressed weight of the ox, as an average voter is of judging the merits of most political issues on which he votes."

  But Galton found the **average** estimate was very accurate.
- Condorcet jury theorem: apply the law of large numbers to opinions that are independent random draws from a distribution with mean equal to the "truth".
- These perspectives suggest that large groups or networks can reach better and more accurate decisions.

Wisdom or Folly of Crowds? (cntd.)

Easy to find optimistic quotes on the wisdom of crowds:

"No one in this world, so far as I know, has ever lost money by underestimating the intelligence of the great masses of the common people."
–H.L. Mencken

"A large group of diverse individuals will come up with better and more robust forecasts and make more intelligent decisions than even the most skilled decision maker."
–James Surowiecki

But what about herding, "groupthink", or "echo chambers" in large groups/networks?

"If the blind lead the blind, both shall fall into the ditch."
–Matthew 15:14.

What determines when crowds are wise vs. when they herd?

The Rest of the Course

- Graph theory and social networks.
- Random graphs and network formation.
- Spread of epidemics and other diffusion processes.
- Production networks and supply chains.
- Intro to game theory
  (static, dynamic, incomplete information).
- Traffic and congestion on networks.
- Network effects in markets.
- Pricing, bargaining, and intermediation in networks.
- Trust and cooperation in networks.
- Auctions, prediction markets, and information aggregation.
- Social learning: wisdom or folly of the crowd?

MIT OpenCourseWare
https://ocw.mit.edu

14.15 / 6.207 Networks
Spring 2022

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms.