Economics of Information Networks

Stephen Turnbull

Division of Policy and Planning Sciences

Lecture 4: December 8, 2022

Abstract

In Lecture 4, Part 1 we look at power laws, a kind of statistical distribution that seems to characterize many economic collections: sales of books, videos, and music recordings, number of followers in social networks, and other measures of "popularity". Lecture 4 Part 2 goes into more detail about some basic features of graphs, and useful measurements. Lecture 4 Part 3 follows up on Lecture 3's discussion of information cascades with a more general discussion of cascading behavior in networks.

Popularity and power laws

- Consider books, where popularity is measured by number of readers.
- It turns out that a readership of k is achieved by a fraction of publications proportional to $1/k^3$. Kleinberg and Easley report two different measures:
  - the number of *purchasers* of a book, and
  - the total number of *citations* of a scientific paper.
- So the probability of sales of k books is $\mathcal{P}(k) = Ak^{-3}$ where $\sum_{k=1}^{\infty} \mathcal{P}(k) = 1$. So $\mathcal{P}(k)$ satisfies the laws of probability: $\mathcal{P}(k) \leq 1$ and the total probability of all possible $k = 1$.
- So probability of total sales is proportional to a power of total sales, which is why it's called a "power law."
- A useful trick is to graph a distribution in the *log-log* form, that is, $\log \mathcal{P}(k)$ *vs.* $\log k$, because $\log \mathcal{P}(k) = \log A - a \log k$.

Why not a power law?

- Since the kinds of networks we're interested in are composed of large numbers of people, the obvious distribution is the *normal distribution* family.
- But popularity distributions are not normal. They're asymmetric, and typically monotonically decreasing: the most frequent class of books (or Twitter accounts) are the ones with very few sales (or followers).
- One alternative is *lognormal distribution*, where the logarithm of sales is normally distributed. The tail of books with few sales is short, and the support of the distribution is positive.
- Another possibility would be a negative exponential distribution, $\mathcal{P}(k) = \alpha e^{-\alpha k}$.
- But popularity distributions have *fat tails*: outliers are much more likely (equivalently outliers at a given percentile rank are much bigger) than these distributions.

A theory of power law distributions

- Under certain conditions the *Central Limit Theorem ensures* that an aggregate will be normally distributed, so we need a theory of why not.
- We look for violations of the conditions of the theorem: specifically, independence of the quantities aggregated.
- Often, the quantities are actually binary: did a person buy a particular book or not? does a page have a link to a particular page?
- How does one book purchase become (statistically) dependent on another? Or one web page's link targets become dependent on other pages? These are basic questions when investigating if lack of independence might generate non-normality.
- As with information cascades, we will arrange for feedback from other pages into a page's decision where to link.

A “rich-get-richer” model

1. We are given a parameter p, $0 < p < 1$.
2. The first page is $j = 1$. It links to itself.
3. Set $j = j + 1$.
4. Create page j, and choose a random Web page i from the existing pages ($i.e.$, one of 1, ..., $j - 1$ with equal probability).
   (a) With probability p, page j includes a link *to* this page i.
   (b) With probability $1{-}p$, page j *copies* the link from page i.
5. Repeat from Step 3.

The "rich-get-richer" model, *cont.*

- As the number of pages gets large, the distribution of in-link counts (the number of pages that link to this page) comes to approximate a power law. The limit exponent a on the number of links k in the distribution depends on p turns out to be $a = 1 + \frac{1}{1-p}$.
- This model allows the creation of a single link from page j. A fancier model could repeat Step 4 to create multiple, independently generated links from page j. This also generates a power law.

Preferential attachment

- How does this model produce *preferential attachment* (the technical term for "rich get richer")? Consider a page i with ℓ in-links. In Step 4b, page i has ℓ chances out of $j-1$ to be linked again: its chance of being linked in this case is ℓ times that of a page with only one in-link. At this step, each page's chance of being linked is proportional to its number of in-links.
- The probability of being linked in in Step 4b in this model is approximately the expected rate of growth of the number of links of the model. When the rate of growth of a variable is proportional to its current level, it's *exponential growth*.
- The pages with many in-links will accrue more in-links faster than those with a few, so those pages will experience explosive popularity.

Homework #17

Due: December 29, 2022 at 11:00. Submit to `turnbull@sk.tsukuba.ac.jp` with `Subject: Homework #17 0AL0200`.

1. Write a program to simulate the "rich-get-richer" model. Any language is fine (I prefer Python, though). The most important part is the *stopping* rule. *E.g.*, stop when the page with the most links reaches N links, say at $N = 100$. The program should output the distribution of link counts.
2. (Be careful!) What is the stopping rule in the "rich-get-richer" model as presented above?
3. Use a linear regression of the log-log form to determine the value of a that corresponds to various settings of p.
4. (Optional) Have the program output a graph of the distribution as an image.
5. (Optional) Have the program output a graph of the distribution in log-log form as an image.
6. (Optional) Enhance the program to allow more than one link per page. Are the distributions produced power laws?

Homework #18

Due: December 29, 2022 at 11:00. Submit to `turnbull@sk.tsukuba.ac.jp` with `Subject: Homework #18 0AL0200`.

1. In an information cascade, we observe preferential attachment. Explain.
2. In the "rich-get-richer" model of web page linking, the preferential attachment is just a mechanical rule. Is preferential attachment in our information cascade examples (the restaurant line and the red/blue urns) "mechanical"? Explain how preferential attachment arises in an information cascade.
3. Describe at least two models of preferential attachment in web page linking based on human psychology. One should explain why pages *copy* links as in the "rich-get-richer" model. Another should give a reason that doesn't involve copying. (*Good* models are an area of active research. An undergraduate student can understand and contribute here, but a *really good* model—and a bit of luck—could get you a Nobel Prize in Economics. I'm not kidding.)

Path dependence

- Information cascades are useful: we get useful information by watching others' actions. But very bad outcomes may occur if the cascade is *all* the information we get. Which happens depends on the order of arrival.
- In the red/blue urn example, dramatically different outcomes depend on the first three draws. This is an extreme example of *path dependence*.
- A poor technology with a positive network externality can dominate its industry, as long as it starts with a clear lead in users over superior rivals. More path dependence.
- An optimization algorithm in a non-convex problem may converge to a poor local optimum depending on the initialization of the algorithm.
- This is very different from the convergence to a long-run limit that many models (and statistical methods!) depend on.

Is path dependence real?

- What if the most popular “idols” and “talents” had to start over? Would they again rise to the top or was their success a path-dependent accident of preferential attachment? Such an experiment is impossible.
- The web makes related experiments possible. Salgankik, Dodds, and Watts constructed a web site that offered specific music downloads by unfamiliar artists, of varying quality. Visitors were shown the list of tunes and their download count.
- They duplicated the site 9 times, with independent download counts. (One site had no counts.)
- The result was wide variation in the distributions, although the best songs always ended up in the top half, and the worst in the bottom half.
- The version without counts displayed much lower variation in download counts across the songs.

The long tail

- In looking at power laws so far, we have emphasized the upper tail of the distribution of sales (links, followers): the blockbusters, the outliers with huge numbers, the influencers.
  - This reflects economies of scale and specialization. Produce one of these, and you're rich.
- But it's not necessarily true that they have a dominant fraction of all sales, especially if there's a preference for variety (or wide dispersion of tastes). An alternative approach is like a supermarket: handle everything. Then the question is what sales volume you can achieve on less popular items.

Supermarkets *vs.* Internet services

- Note that for a supermarket there's a real problem of shelf life. You have to sell meat and fish and fresh-baked goods on the same day, for example. But not so for books and DVDs!
- There's also a question of storage space for books and DVDs. But this doesn't apply to Internet distribution of e-books and streaming videos! (Or, at least, not in the same way.)

It's power laws all the way down

- To examine the *long tail* of unpopular items, change viewpoint from the fraction of items selling *exactly* k units to the number selling *at least* k units.
- Changing *fraction* to *number* is just rescales by the number of items. It changes the labels on the vertical axis, but not the curve in the graph.
- The change from *exactly* to *at least* is a real change, although just of "viewpoint" on the underlying data. This distribution is also a power law: $F(k) = \int_k^\infty Ax^{-\alpha}dx = -\frac{A}{1-\alpha}k^{1-\alpha}$!
- Next, how many units of the j-th most popular item are sold? But this is exactly the inverse of the F defined above: if $j = F(k)$ is the number of books that have sold k units or more, then the j-th most popular book is exactly the one that sold k units. Draw the curve, then flip the axes.
- The right end of the "flipped" curve is the "long tail": niche items selling few units. A power law!

Mathematics and meaning

- We've seen three different power laws while examining the "long tail":
  - the original power law distribution of item popularity derived from preferential attachment
  - the power law distribution of "at least this popular" items
  - the inverse power law of percentile rank (the popularity of the j-th item).
- They're quantitatively quite different—it is a big mistake to use one of them to answer a question appropriate for the others. *E.g.*, don't try to use the *item popularity* distribution to answer questions about *sales volume of the items* in the tail.
- Each is useful in its own way. For example the *inverse power law* can be used to answer "how much revenue comes from the items with sales rank less than 1000?" (by integrating).
- Mathematical transformations allow us to ask various questions from the same simple data.

Beyond the math

- While we won't discuss flamewars and social media bullying in detail, clearly these phenomena gain much of their power from preferential attachment. Disinformation is related to information cascades, people believing things because other people believe them.
- Of course there are important psychological aspects, such as *motivated reasoning* (picking facts and relationships that support what you want to believe, and ignoring facts and relationships that contradict your prejudices). But informational effects amplify problems.
- We can also point to potentially opposing aspects of search engines and recommender engines.

“The Algorithms”

- Search engines and social media *curate* content: they choose content for you, because all available content is an overwhelming (almost unimaginable) quantity. The selection rules are often called “the Algorithm”.
- On the one hand, search engines tend to rank candidate links by popularity, not by value of content. This can lead to information cascades.
  - Social media “algorithms” have been shown to encourage “information bubbles” which reinforce prejudice and political division, and to contribute to radicalization.
- To exploit their stocks of “long tail,” unpopular items little-known to consumers, vendors like Amazon and Netflix offer *recommender* engines which help consumers to navigate the plethora of little-known items, and enjoy products they would otherwise never hear of.
  - These engines may encourage, rather than undermine, diversity.

Basic notions of networks

- Network structure is described using graphs.
  - Networks often have additional properties attached to them.
- Graphs consist of
  - *nodes*, also called *objects*, which may be interpreted in many ways, *e.g.*, people or places, and
  - *links*, usually called *arrows* in directed graphs, which represent relationships of various kinds. Links are also often called *edges*.
  - Additional properties of networks may be attached to either nodes or links.

Basic types of graphs

- Graphs may be restricted to zero or one link between each node (the usual case, and there is no special term for this case), or may have zero or more links. To distinguish from the usual case, this is called a *multigraph*.
- Links may be automatically symmetric, which is call an *undirected graph*, or they may be in principle asymmetric, called a *directed graph*.
  - In a directed graph, even if not a multigraph, there may be two links between a pair of nodes, going in opposite directions.
- A graph with numerical property attached to each link is called a *weighted graph*, and one with a qualitative property attached to each link is called a *colored graph*.
  - There is no particular name for graphs with properties attached to nodes. Graph structure is defined by the links, so attaching properties to nodes is not very useful in graph theory.

Basic structures in graphs

- A *path* in a graph is a sequence of links, each of which shares an endpoint with its next link.
- In a directed graph, *path* usually means a *directed path*, in which each pair of links is connected head to tail. But in discussions of "connectedness," *path* means an *undirected path*.
- When A and B are the endpoints of a path p, we say "A and B are *connected* by p." If A is the tail and B is the head of a directed path p, we say "B is *reachable* from A via p."
- The *length* of a path is the number of links in the path.
- A *cycle* is a path where each node can be considered to be both the beginning and the end of the path. A cycle of length 1 is called a *loop*.
- A *simple path* is a path that contains no cycles.

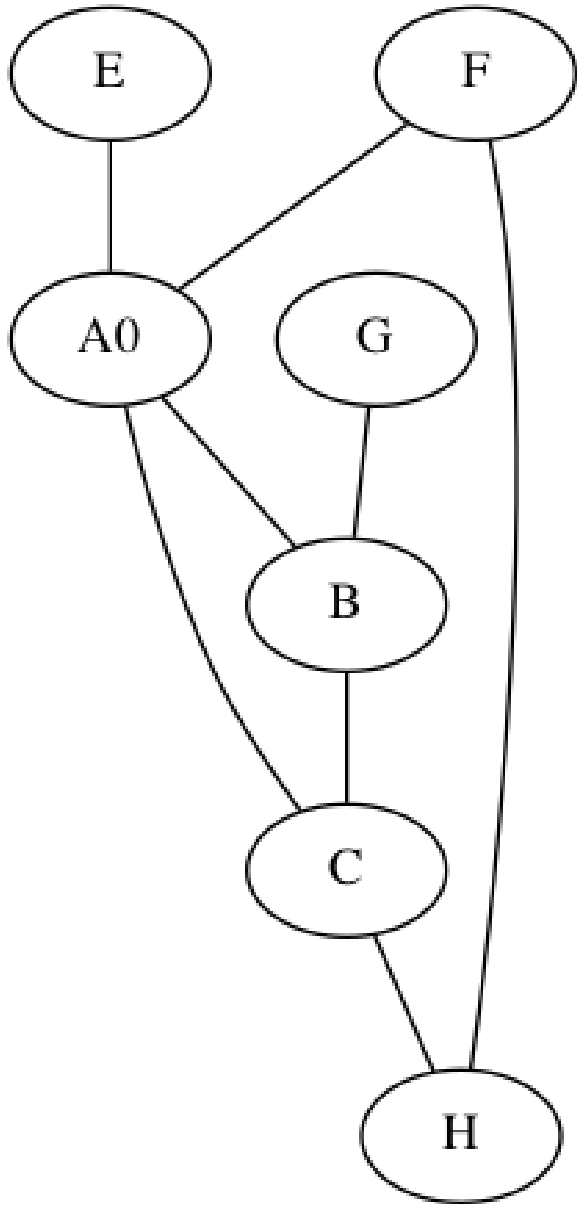
Connectedness

- Two nodes are *connected* if there is an undirected path from one to the other.
- A set of nodes is *connected* if for each node in the set there is a path to every other node in the set.
- A set of nodes is a *connected component* if (1) the set is connected, and (2) no node in the set is connected to a node not in the set. More concisely, it is a *maximal connected set*.
  - "Connected component" is often shortened to just "component." There is no other kind.
- A *connected graph* consists of a single connected component. A *disconnected graph* contains more than one component.
- In social and economic networks, we often observe the phenomenon of a single *giant component*, even if the graph is technically disconnected because of some tiny components or solitary nodes.

Distance

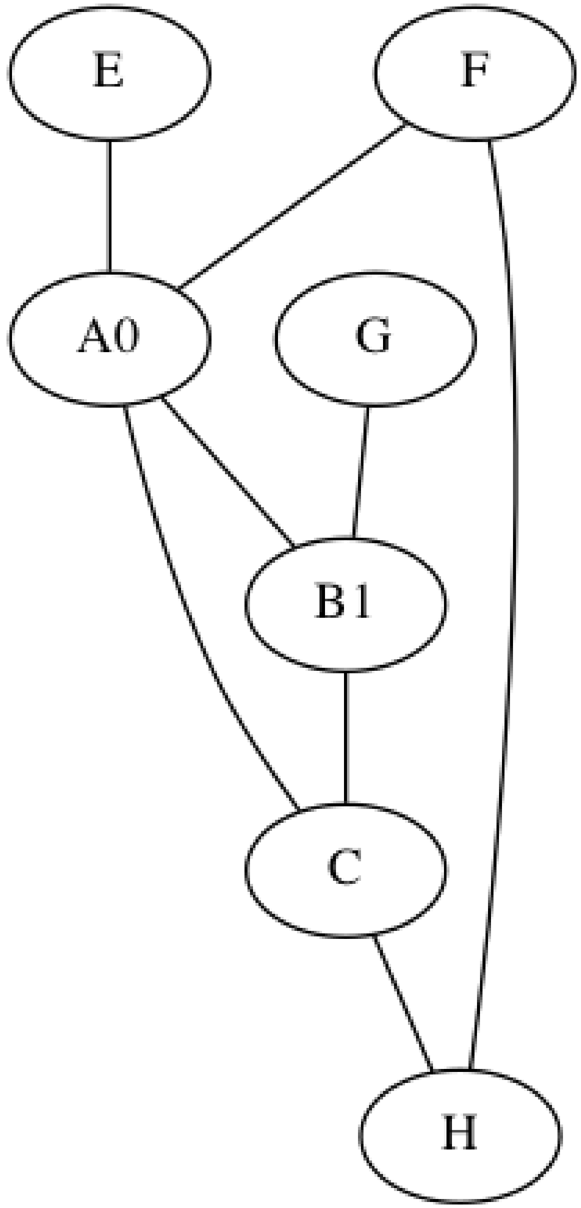
- The *distance* between two nodes is the length of the shortest path connecting them. (This path is necessarily a simple path, and need not be unique.)
- By convention, the distance from a node to itself is defined by the *empty path* which starts and ends at the node and contains no links. The distance therefore is 0. (Note: there is a different empty path for each node.)
- The *diameter* of a component is the maximal distance taken over all pairs of nodes in the component.
  - The diameter of a disconnected graph doesn't exist: the distance between nodes in different components is undefined. (There is no "shortest path" if there is no path at all.)
  - The diameter of connected sets of nodes that aren't components isn't useful. It's unclear what to do with paths that go outside the component.

Breadth-First Search: Start

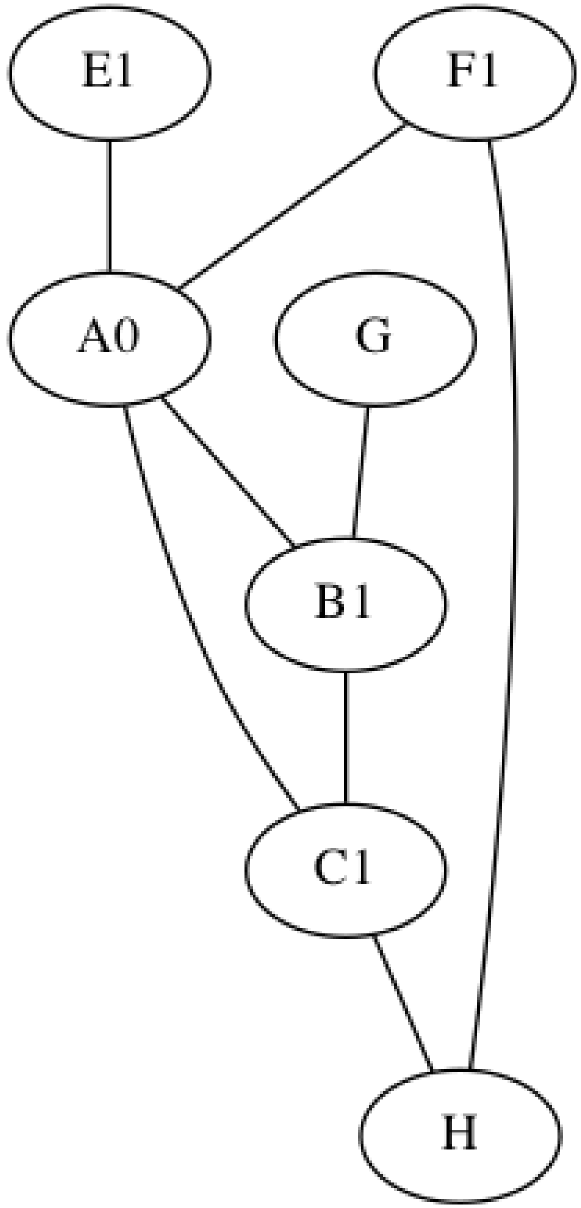


Breadth-first search is an efficient algorithm to (1) determine the distance from one node to another, and (2) determine all distances between all pairs of nodes.

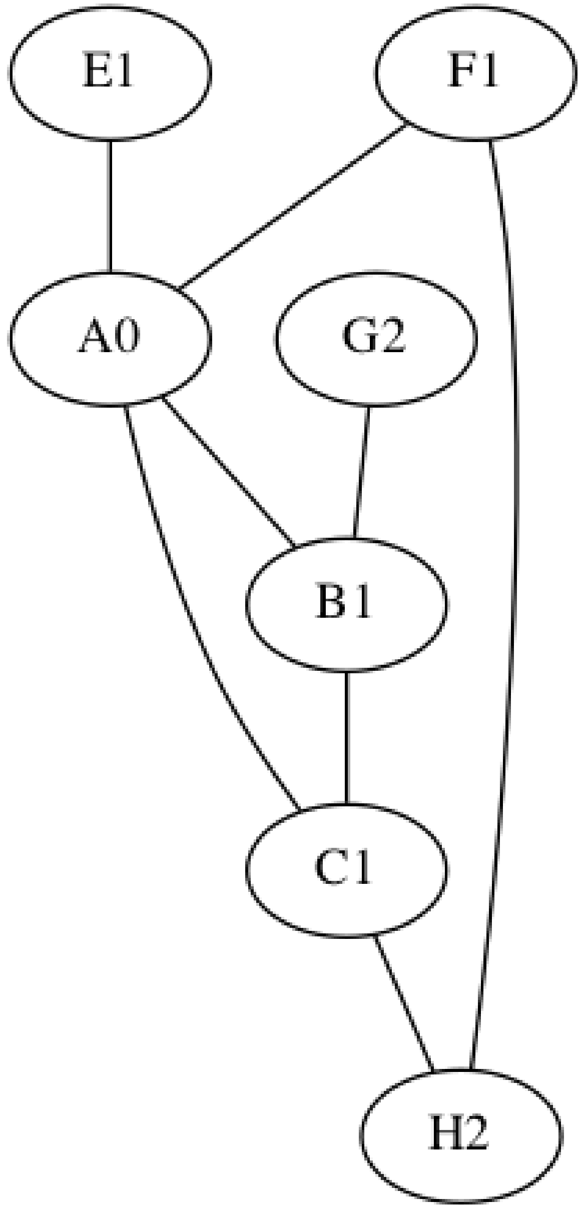
Breadth-First Search: First Neighbor



Breadth-First Search: All Neighbors



Breadth-First Search: All Distance=2 Nodes



At distance = 2, all nodes have been found. This implies that if a target node exists, it has been found.

Betweenness

- A node A is *pivotal* for B and C if it lies on every shortest path between B and C.
- A node A is a *gatekeeper* for B and C if it lies on every path between B and C.
- A node A is a *local gatekeeper* for neighbors B and C if there is no link between B and C.

Bridges

- A *bridge* is a link if deleting it would cause its component to become disconnected.
- A *local bridge* is a link such that deleting it increases the distance between its endpoints to more than two, or is a bridge.
- The *span* of a local bridge is the length of the shortest alternative route between its endpoints.

Quantitive Analysis of Graphs

- To do economics, we need to compare values. Generally we want to make *marginal* comparisons, trading a little of one value for some of another, until we achieve balance. So we need to have quantitive measures of values. What can we measure in graphs?
- We start with *undirected* graphs with at most one link between objects.
- We have considered the *star graph* with a special central node that can be used to model a market, and the *complete graph* that underlies Metcalfe's quantitative model of "network externalities."
  - In these special graphs, we *count* the *number of links* or the *number of nodes* to determine value. In economics we often transform these values by a *cost* or *utility* function.
- Real information networks have many properties that depend on the exact configuration of links.

Connected Graphs

- A set of nodes in a graph is *connected* if for all nodes x and y in the set there is an undirected path from x to y.
  - A graph is *connected* if the set of all nodes is connected.
    * Note: in a directed graph, the nodes need not be linked head to tail. The graph may be connected even though there are pairs of nodes with no path between them.
    * Example: every path, considered as a subgraph, is connected.
  - A set of nodes is *strongly connected* if it (1) is connected and (2) has no link which if removed the graph is no longer connected.
    * Example: Every cycle is strongly connected. Remove any link, but its nodes are still connected "the long way around!"

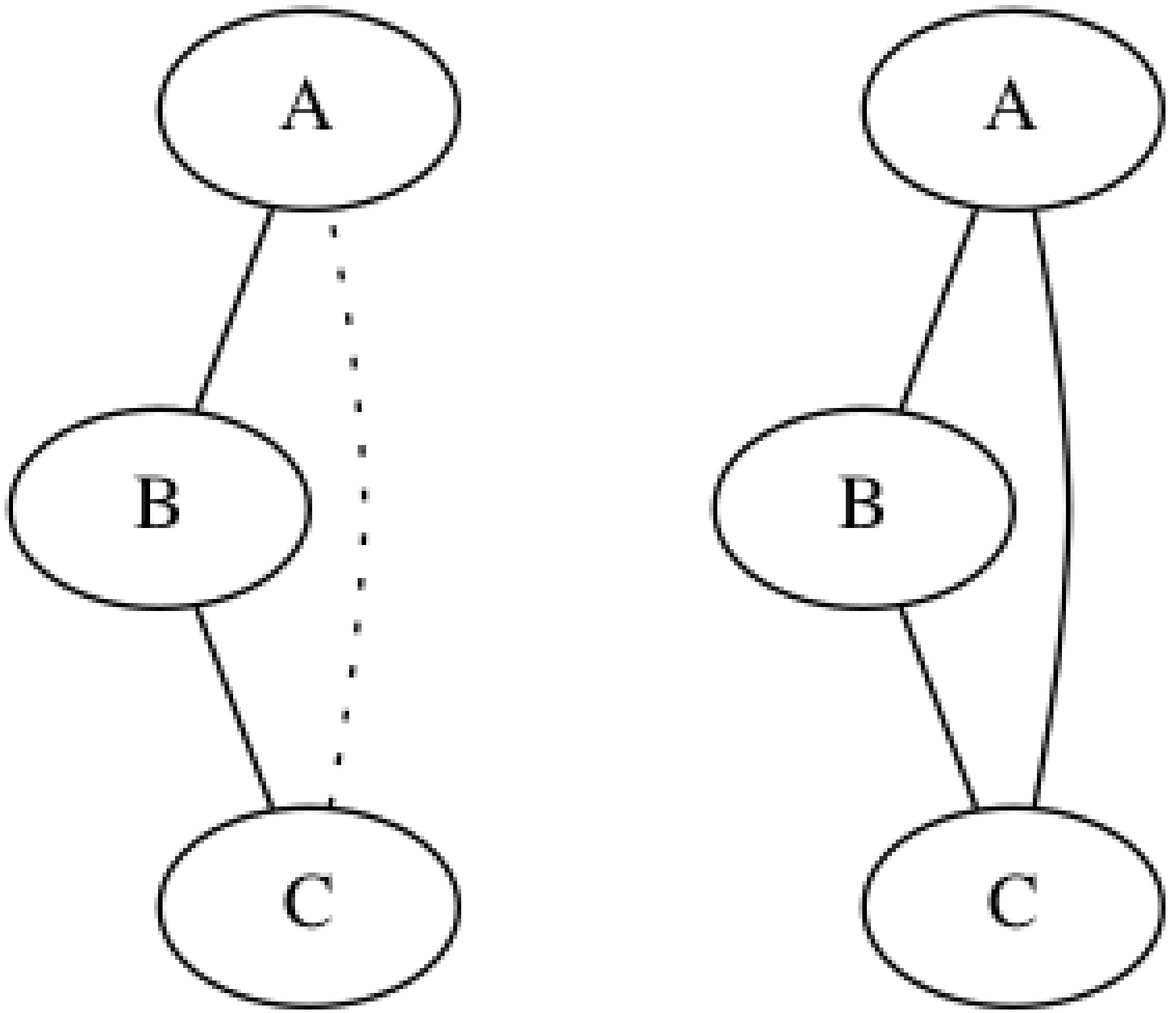
Triadic Closure

- We can start considering "completeness" in a non-trivial way when we have at least three nodes.
  - We typically ignore loops in information networks—you don't need to tell yourself what you already know.
  - With only two nodes, you can *define* completeness, but it's not very interesting. If the graph is incomplete, it has no connections!
- In fact, it's useful to look at all the connected subsets of exactly three nodes in a graph, which gives the idea of triadic closure.

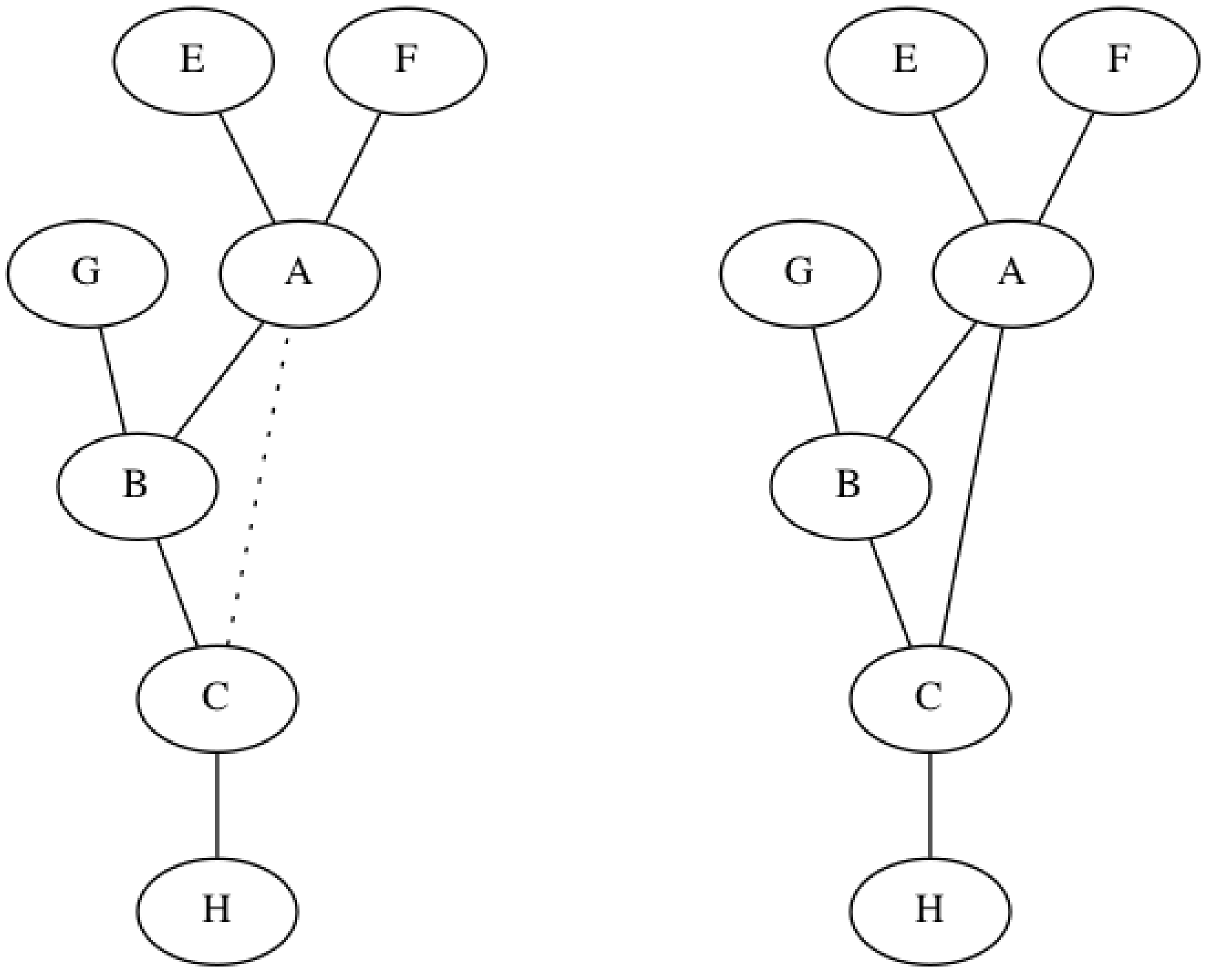
Triadic closure: definition

- Consider three nodes in a graph. If A is connected to B and B is connected to C, is ABC is the shortest path between A and C?
- In social situations, there is the phenomenon of *introduction*: when A knows B and B knows C, it is common that B introduces A to C, creating a new link from A to C.
- *Triadic closure* refers to both the process of creating a triangle (*e.g.*, via introduction), and to a state where three nodes and their links form a triangle.
- The *clustering coefficient* of a node is the fraction of pairs of neighbors of the node which are neighbors of each other. It measures the amount of triadic closure around the node.
- We can also consider the fraction of all triads that are closed in the whole graph (or in any connected component), but this is not so interesting.

Triadic closure: whole graph



Triadic closure: subgraph



Diffusion in Networks

- **Diffusion** is a general term that means that some substance gradually spreads out through a fluid.
- It is a *dynamic* process that takes place over time.
- We generalize that term to mean that something spreads out by moving node to node in a network.
- Until now we have mostly considered the *static* structure of networks, *i.e.*, the structure at a point in time. Even *triadic closure* was considered not as an ongoing process but comparing one instantaneous state of the network to another.
- We then considered how the structure of the network affected aggregates, or how immediate neighbors relate.
- Now we want to consider an intermediate level of structure.

Act Locally, Affect Globally

- Perhaps you know the slogan, "act locally, think globally".
- This is the level we want to consider.
- Think about the COVID-19 pandemic, and how many governments have policies restricting *international* travel that they would never apply to *domestic* travel. (Or Japan's contradictory policy!)
- We may adapt our political views to those of friends and neighbors, or even choose residence to ensure that those around us have similar views.
- Many years after Microsoft Word had clearly won the wordprocessor competition, WordPerfect remained the choice of American lawyers. Partly this was due to comprehensive suites of WordPerfect macros for creating legal documents, but part of it was due to the need to collaborate with others in the legal community.
- Thus network *structure* interacts with individual *behavior*. The process of *triadic closure* is a step in this direction. Now we go much farther.

Diffusion of Innovations

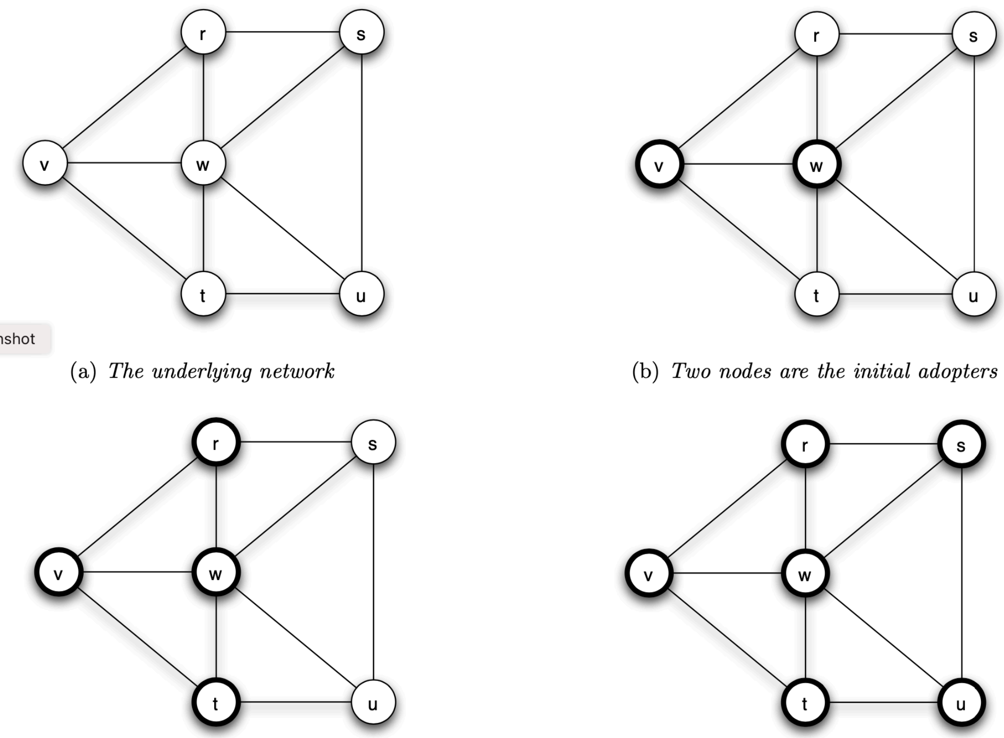
- In economics, *invention* is not so important. Communication of that invention to others is far more important. That is the process of **innovation**.
- Sociologists studied this in the 20th century. Studies of *hybrid seed corn* (Ryan and Gross [1943]) and *tetracycline* (Coleman, *et al.* [1966]) showed that although these products were introduced to users by salesmen, it was recommendations by colleagues that drove adoption.
- Direct benefits that derive from using tools compatible with others in your group, called *network externalities*, are important influences on adoption. In *Metcalfe's Law* we considered the case where everybody wants to communicate with everyone else, but what if you only care about your network neighbors? Telephone companies try to exploit this effect with "friends and family" discounts and *non-portability* of phone numbers and email addresses.

Adoption Coordination on Networks

- Suppose there are two brands, A and B.
- For each neighbor using A you get benefit a if you use A but 0 if you use B. Similarly B is worth b for matching brand else 0.
- You should use A if $p \geq \frac{b}{a+b}$, else B, where p is the proportion of neighbors using A. p is the *adoption threshold*.
- Without loss of generality, suppose $a > b$. In particular for the example, $a = 3$ and $b = 2$.
- Note that "all A" and "all B" are both equilibria!

Switching from B to A

Consider this network. The general theorem that all B is an equilibrium holds. But if two nodes switch to A together, A will diffuse through the network and B disappears.



(a) *The underlying network*

(b) *Two nodes are the initial adopters*

(c) *After one step, two more nodes have adopted*

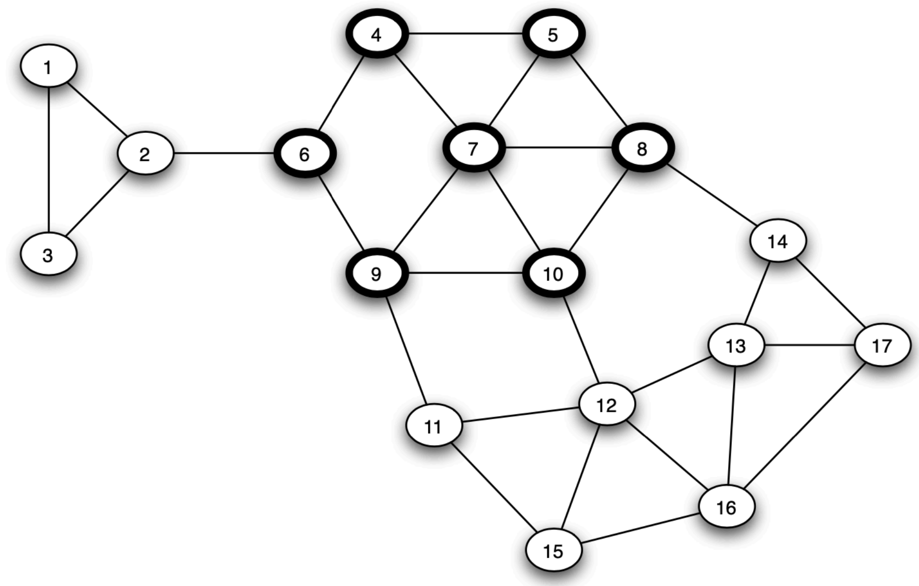
(d) *After a second step, everyone has adopted*

Figure 19.3: Starting with v and w as the initial adopters, and payoffs $a = 3$ and $b = 2$, the new behavior A spreads to all nodes in two steps. Nodes adopting A in a given step are drawn with dark borders; nodes adopting B are drawn with light borders.

A Partial Adoption Cascade

Here we show only the end state, where the equilibrium has only some of the nodes adopting the more profitable brand. Note that the adoption cascade crosses the local bridge at 4–6 (or 9–6) but can't cross a global bridge (6–2) or the local bridges 9–11, 10–12, and 8-14. Local condition is not enough!

(a) *Two nodes are the initial adopters*



(b) *The process ends after three steps*

Figure 19.5: Starting with nodes 7 and 8 as the initial adopters, the new behavior A spreads to some but not all of the remaining nodes.

Homework #19

Due: December 29, 2022 at 11:00. Submit to `turnbull@sk.tsukuba.ac.jp` with `Subject: Homework #19 0AL0200`.

1. In the smaller network, use the formula "use A if $p \geq \frac{b}{a+b}$, else B" to show that all B is an equilibrium when $a > 0$ and $b > 0$.

2. Show that starting from "all B," if just one node adopts A, diffusion fails and the network reverts to all B in equilibrium.

3. In the larger network, verify that the state in the image is the equilibrium, stopping with a partial cascade.

Examples of Partial Cascades

- Note how this differs from the Metcalfe's Law world where eventually all users end up with the same brand. That's because in that world everyone is connected to everyone else.
- People generally will maintain political attitudes despite being on the "boundary" of the neighborhood where their position is possible.
- Some industries (video editing, graphics) still favor Macintosh computers despite the general dominance of Windows. This is partly because of better performance for those applications, but also due to the local network effects creating an "island" of Mac users.
- Kleinberg and Easley point out that even in the larger network, if you raise the value of A to $a = 4$, eventually you get a complete cascade.
- Another strategy is to find *key users* in the "island" of non-users. For example, if A can convince 12 or 13 to switch, the cascade will restart and all nodes 11–17 will eventually switch. But 11 or 14 doesn't help.

Homework #20

Due: December 29, 2022 at 11:00. Submit to `turnbull@sk.tsukuba.ac.jp` with `Subject: Homework #20 0AL0200`.

1. Starting from the situation of the partial cascade equilibrium, show that with $a = 4$, $b = 2$ the adoption cascade becomes complete.

Clustering in Networks

- How do we figure out when a complete cascade will occur?
- This depends on clustering of nodes in the network. In the large network, the set $\{1, 2, 3\}$ is a **cluster** in the sense that each member has more neighbors in the set than outside of it.
- We define a **cluster of density** q in a network to be a set of nodes such that each member has a proportion of its neighbors in the cluster greater than or equal to q.
- The set of all nodes is a cluster of density 1.
- The union of two clusters of density q is also a cluster of density q, even if they are not connected!
- **Note:** the definition of *cluster of density* q is purely a property of the network, and does not have any connection to the adoption threshold.

Clusters *vs.* Cascades

Theorem Consider a set of initial adopters of brand A, with a threshold of p for nodes in the remaining network to adopt brand A.

1. If the remaining network contains a cluster of density $q > 1-p$, then the set of initial adopters will not cause a complete cascade.
2. Moreover, whenever a set of initial adopters does not cause a complete cascade with threshold p, the remaining network must contain a cluster of density $q > 1-p$.

In this sense, clusters and (complete) cascades are incompatible.