Economics of Information Networks

Stephen Turnbull

Division of Policy and Planning Sciences

Lecture 3: November 22, 2018

Abstract

We consider technology adoption with network externalities. This lecture is still incomplete.

Note: There is no class on November 29 due to entrance examinations.

Adoption with network externalities

- Describe the situation with a simple game:

	New	Old
New	α, α	γ, δ
Old	δ, γ	β, β

Table 1: Static technology adoption game

where the network externalities are indicated by the parameters $\alpha > \gamma$, $\alpha > \delta$, $\beta > \gamma$, and $\beta > \delta$ (that is, matched technology is always better than mismatched technology for both players).

- Two inefficient equilibria can arise:

  Excess inertia $\alpha > \beta$ but (Old, Old) is observed.

  Excess momentum $\alpha < \beta$ but (New, New) is observed.

Concrete instances

- With the constraints on the parameters, it is approximately a game of pure coordination, with the equilibria described later.
- Without the constraints, it is a generalized game of coordination, where there is a social benefit if both players play the "right" strategy. (Normally in game theory we associate "right" with "equilibrium," or sometimes "efficient." Here the meaning is stronger: "efficient and fair.")
- This schema fits both the *game of pure coordination* (left) and the *prisoners' dilemma* (right):

	New	Old
New	$4, 4$	$0, 0$
Old	$0, 0$	$1, 1$

	New	Old
New	$4, 4$	$0, 5$
Old	$5, 0$	$1, 1$

Table 2: Two generalized coordination games

Solutions

- It is easy to solve these games.

Pure coordination
- By process of elimination, both mismatched cases are not equilibrium: both players want to deviate.
- It doesn't matter that if they both deviate at the same time, they end up mismatched in disequilibrium again. This is the "best response (or Cournot) dynamic," but it is a poor algorithm for finding solutions. Alternating Cournot works in this case, but not necessarily in others.
- The two matched cases are both strict equilibria.
- The equations for the mixed strategy equilibrium have a unique solution, also an equilibrium.
- The (New, New) equilibrium is "good," the others are "bad."

Prisoners' dilemma
- Old is a strictly dominant strategy for both players, so the (Old, Old) equilibrium is unique.
- It's "bad," in fact, the worst possible.

Stories

- More important than solving games and assessing their equilibria is "telling stories" about how those payoffs might come to be.
- Both example games are symmetric, so we tell the story from the point of view of one. The other is derived by exchanging the strategies and payoffs.

Stories

Pure coordination Here, we are probably looking at *consumers* of the technology. The basic benefits of the two variants are similar, so $\gamma = \delta = c$, and for convenience of computation we set $c = 0$. Of course we could just add c to all parameters, get the same basic schema, and it would have the same equilibria.

Then we suppose the New variant of the technology has better communication capabilities, or potentially attracts more users, giving higher network externality benefits.

Prisoners' dilemma How can we tell this story?

We need to look at the *other* side of the market. This is a case of the *network providers* who can adopt new technology.

But how is it that the Old variant can have higher payoffs in competition with the New? Note that New has lower payoffs in competition with Old than sticking with Old. So perhaps in competition Old is attracting all the customers away from New (New probably starts smaller), and ends up as a (near-) monopolist. Metcalfe's Law justifies the huge payoff.

Dynamic technology adoption model

- Technology evolves (improves) over time: value to users of best known technology at time t is T_t (not including network externalities).
- Value to users of technology actually in use at time t is V_t.
- Overlapping generations model to provide *exogenous inertia*. Users choose technology when young and are "locked in" when old. Note that this means that old users are *dummy players* who have no interesting choices to make, and we interpret T_t and V_t to be *lifetime payoffs*.
- Network externalities may cause users to keep old technology, so that

$$V_t = \begin{cases} T_t & \text{if} \quad \text{adopt new} \\ V_{t-1} & \text{if} \quad \text{keep old} \end{cases}$$

  and V_t can be equivalent to T_{t-s} for "large" s.

- However, the choice is always V_{t-1} *versus* T_t since $T_t > T_{t-s}$ for $s > 0$.

Network externalities in the dynamic model

- As usual network externalities depend on user population. N_t is the population of the generation which is young at time t.
  - This implies that the population of the *old* generation at time t is N_{t-1}.
- A somewhat general specification of utility is

$$U_t = \begin{cases} u(T_t, N_t) & \text{if} \quad t \text{ adopts } T_t \\ u(T_t, N_t + N_{t-1}) & \text{if} \quad t \text{ keeps } V_{t-1} \end{cases}$$

- Generation t user adopt T_t if $u(T_t, N_t) \geq u(T_t, N_t + N_{t-1})$.
- An important determinant of adoption is whether technology and network size are complements or substitutes in consumption: does the marginal utility of better technology increase or decrease with larger network size?
  - We look at the extreme cases of perfect complements and perfect substitutes to see how it affects the adoption decision.

Notes

- Generations need not be in terms of the human life cycle. It might also be in terms of the corporate investment cycle, or the life of a technology.
  - Take care with the "life of technology" interpretation: obsolescence is somewhat endogenous.
- Shy's analysis is inaccurate in general: lifetime payoffs should depend on the choice of the next generation due to network externalities. *(You are not responsible for this detail on the examination in 2017.)*

The case of complements

- We look at the extreme case of *perfect complements*, where $u(T, N) = \min\{T, N\}$.
- The technology choice doesn't actually depend on T_t!
  1. If $V_t \leq N_t$, the young at t will choose the new technology to get $u(T_t, N_t) = \min\{T_t, N_t\} \geq V_t = \min\{V_t, N_t + N_{t-1}\} = u(V_t, N_t + N_{t-1})$.
  2. If $V_t > N_t$, the young receive $u(V_t, N_t + N_{t-1}) = \min\{V_t, N_t + N_{t-1}\} > N_t = \min\{T_t, N_t\} = u(T_t, N_t)$.
- We call Case 2 the technology stagnation case. If population is stable, you surely reach that case, and stay there forever after.

The case of substitutes

- Preferences under *perfect substitutes* are additive, where $u(T, N) = T + N$.
- Comparing the two possibilities for the young generation, we have $u(T_t, N_t) - u(V_t, N_t + N_{t-1}) = (T_t + N_t) - (V_t + N_t + N_{t-1}) = (T_t - V_t) - N_{t-1}$, and the young generation will choose the new technology when the benefit to changing technology is greater than the external benefits from the old generation.
- Since technology is assumed to strictly improve over time, even if in one period it doesn't overcome the network externality, eventually it does. At that point, the technology may jump several generations.
  - Consider the widely deprecated Windows ME and Windows Vista editions of Microsoft Windows.

Technology duration under perfect substitutes

- The duration of a technology is the time from adoption to becoming obsolete (replaced by a new technology).
- For simplicity, assume population is constant: $N_t = N$, for all t.
- Assume technology improves linearly: $T_t = \lambda t$.
- The current technology (giving V_t) was adopted at some time s, and $V_t = \lambda s$.
- Thus, the adoption decision is made when $T_t - V_t = \lambda t - \lambda s \geq N$, and the duration is $\Delta = t - s = N/\lambda$. (Technically, it should be the ceiling of $Delta$, which is the smallest integer greater than or equal to Δ.
- The concept of *lock-in* may be defined in this context as follows: *If increasing* N *leads to an increase in technology duration,* lock-in *due to network externalitiesis present.*
- In overlapping generations models like this one, lock-in is always present.

Homework #8

Due: December 6, 2018 at 11:00. Submit to `turnbull@sk.tsukuba.ac.jp` with `Subject: Homework #8 01CN901`.

Recall Shy's *overlapping generations model* of technology adoption, where the whole generation decides when young which technology to use, and keeps it when old. In general, you would expect each generation to *anticipate* the next generation's choice, which would affect their valuations via network externalities when old.

Tell a "story" about how the network externalities of the old generation might be zero, so that Shy's analysis is correct in assuming that T_t and V_t are independent of next the generation's choice.

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 standard term for this case), or may have zero or more links (if necessary, 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. Because graph structure is defined by the links, 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.
- If A is the tail and B is the head of a directed path p, we say "B is *reachable* from A via p."
- 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."
- 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. This is equivalent to saying that all nodes in the path are unique, and the end is not the same as the beginning.

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," because there is no other kind of component.
- 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.

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.

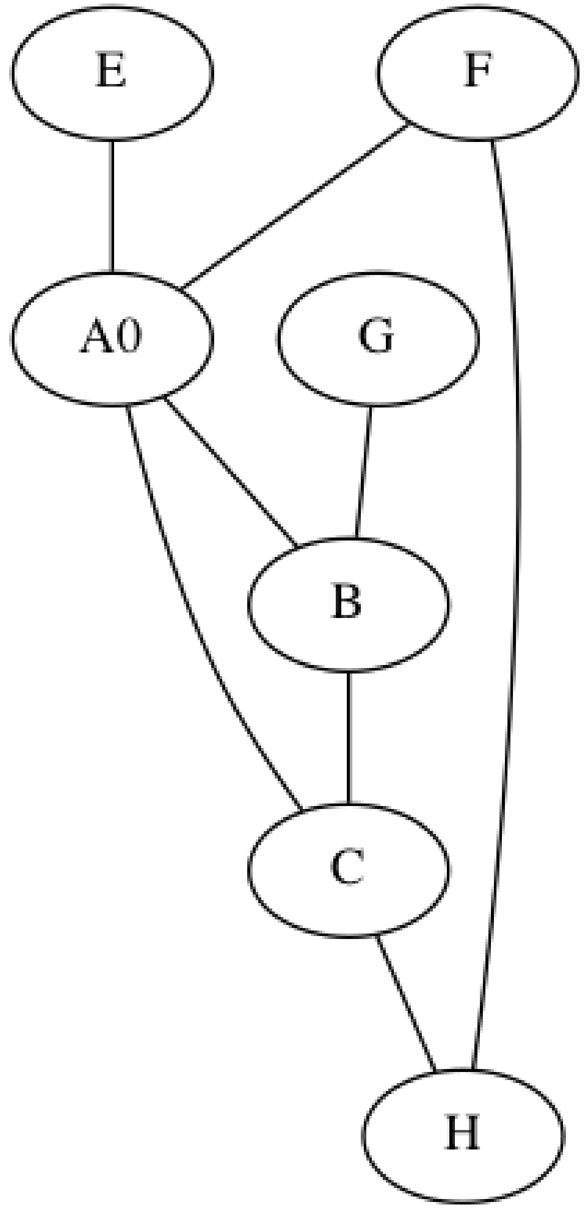
Bridges

- 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.

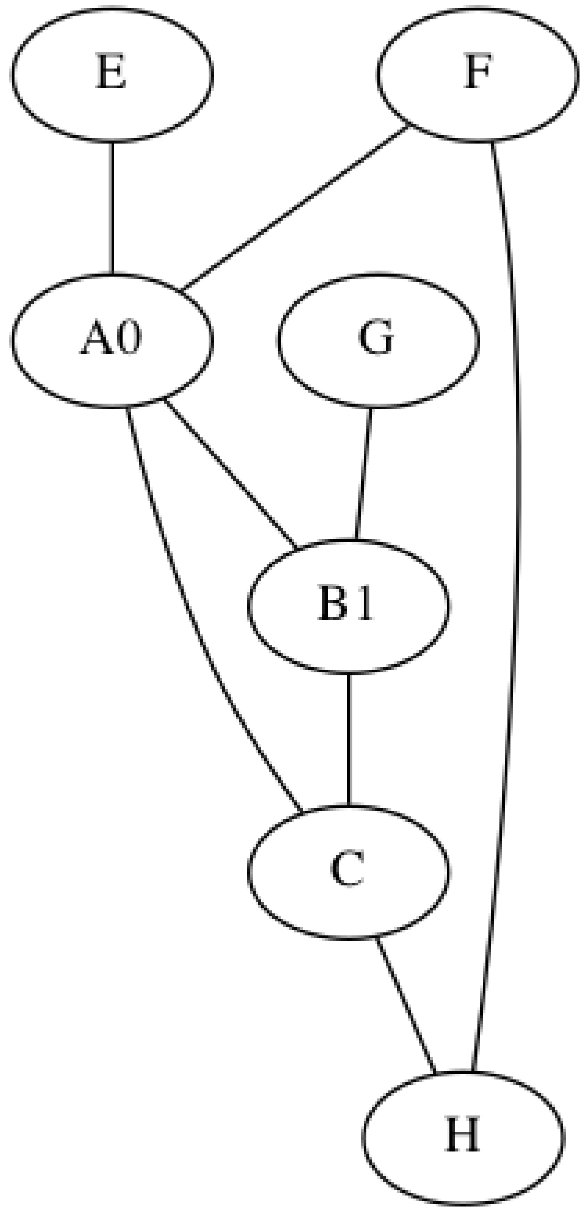
Distance

- 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.
  - We don't talk about the diameter of disconnected graphs because the distance between nodes in different components is undefined. (There is no "shortest path" because there is no path at all.)
  - We don't talk about the diameter of connected sets of nodes that aren't components because it's not clear what to do with paths that go outside the component.
- *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.

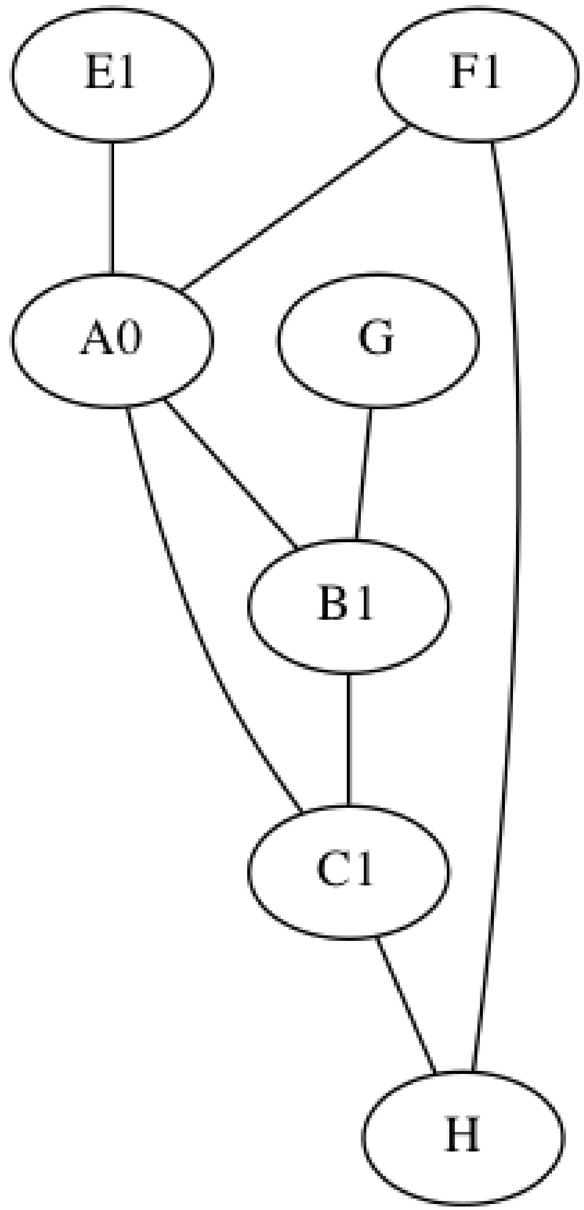
Breadth-First Search: Start



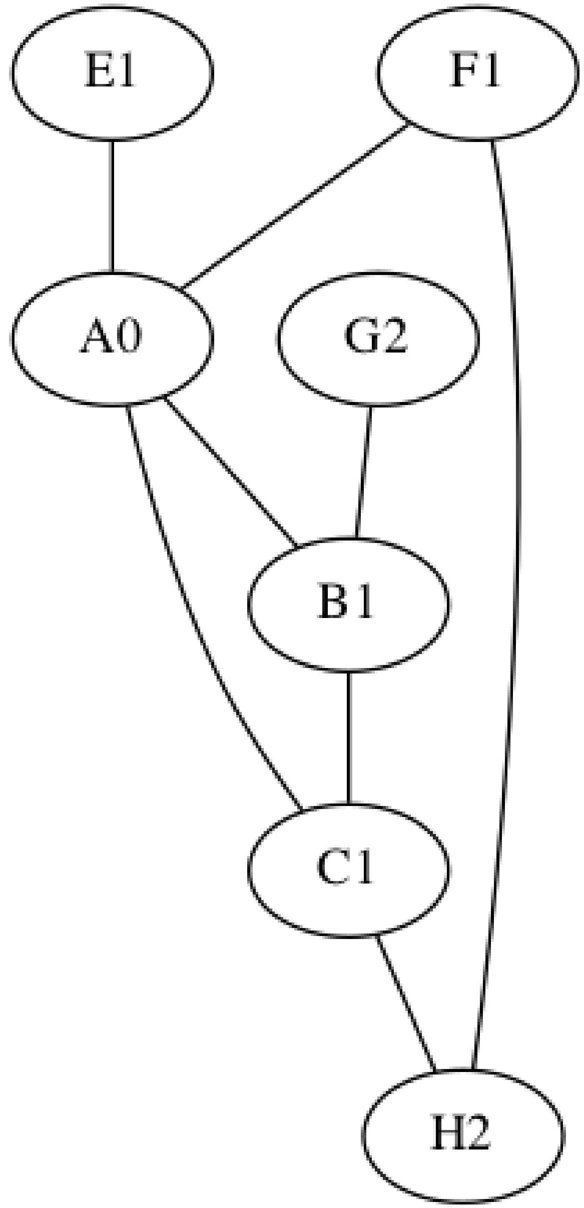
Breadth-First Search: First Neighbor



Breadth-First Search: All Neighbors



Breadth-First Search: All Distance=2 Nodes



Quantitive Analysis of Graphs

- We start by considering 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."
- Real information networks have many properties that depend on the exact configuration of links.

Connected Graphs

- A *path* through a graph is a sequence of nodes (where n is the length of the sequence) such that each pair of consecutive nodes is linked.
  - If the graph is directed, neighboring links must be connected head to tail.
- A *cycle* is a path such that node n is connected to node 1.
- A set of nodes in a graph is *connected* if for all nodes x and y in the set there is a 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 (1) it is connected and (2) there is no link such that if that link is removed the graph is no longer connected.
    - Example: Every cycle is strongly connected. If one link is removed between two nodes, they are still connected "the long way around!"

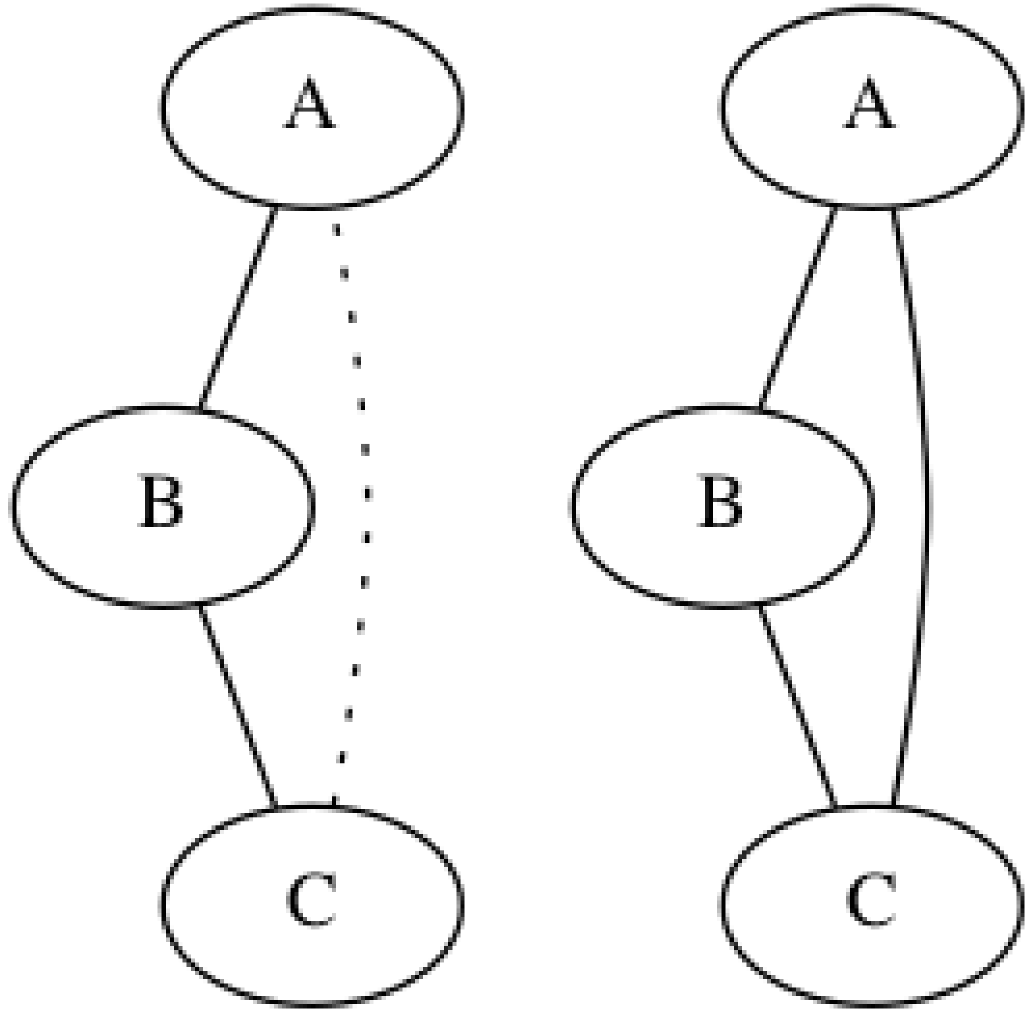
Triadic Closure

- 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.

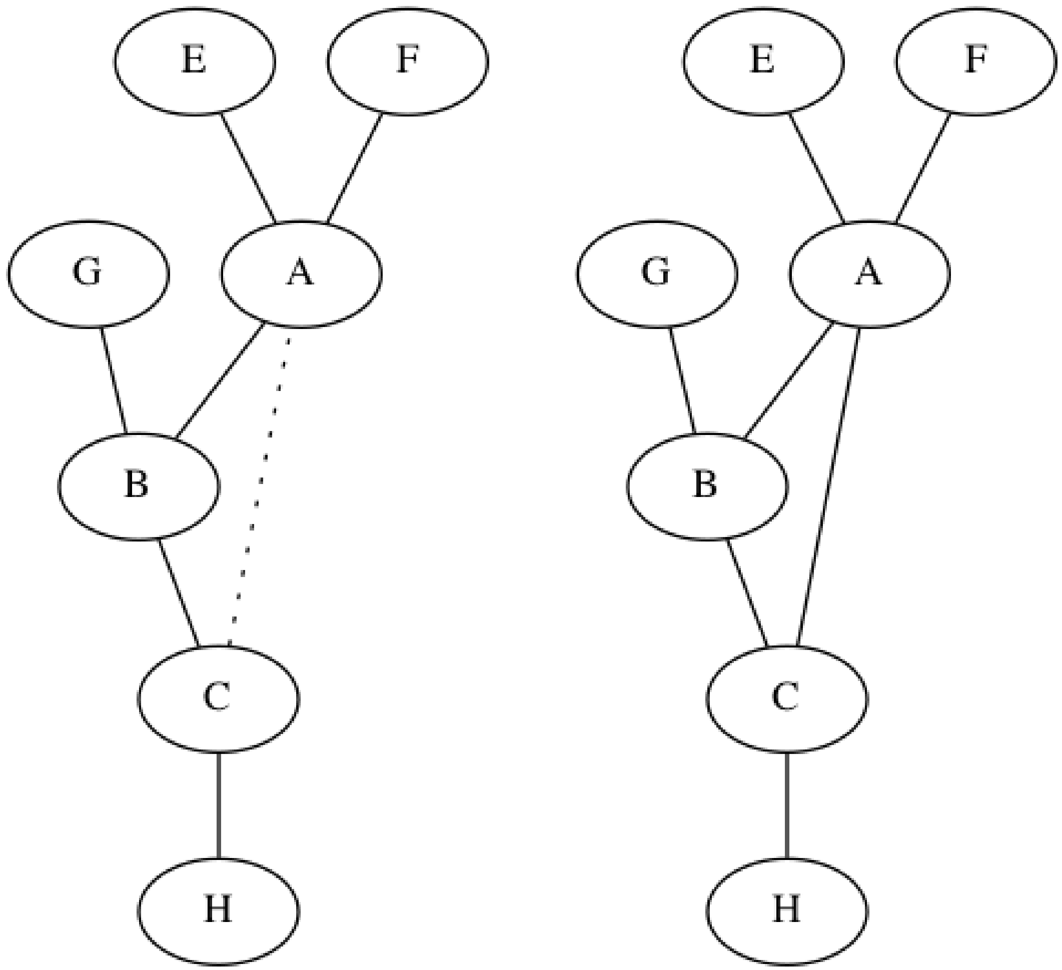
Triadic closure: definition

- Consider three nodes in a graph. If A is connected to B and B is connected to C, the question arises whether 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.
- The term *triadic closure* refers to both the process of creating a triangle (*e.g.*, via introduction), and to the condition 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.

Triadic closure: whole graph



Triadic closure: subgraph



The strength of weak ties

- Granovetter[] observed that people frequently learned of opportunities such as job opportunities through personal contacts who were *not* considered friends. That is, the "tie" to these acquaintances was "weak."
- Based on the process of triadic closure, we would expect that
  1. links to "acquaintances" are bridges
  2. much information is commonly known among friends
  3. there may still be some competition among friends

  concluding that information about such opportunities often would flow from acquaintances naturally.
- Why doesn't triadic closure operate with acquaintances? We propose a *weighted graph* model with strong and weak links. d *Strong links* correspond to friends, and *weak links* to acquaintances, as you would expect.

Strong triadic closure

- "Important opportunities" like mid-career job changes arise infrequently, so the triadic closure process should convert (local) bridges to non-bridges relatively quickly, and we would expect that information flow via bridges should be uncommon.
- The words "friend" and "acquaintance" suggest a solution: a weighted graph, where some links are *strong* and others *weak*.
- In such a graph we define *strong triadic closure* as the condition where if strong links exist between A and B and between A and C, there is a link (strong *or* weak) between B and C.
- **Theorem:** In a strong triadically closed graph, if A has any other strong links, and is the endpoint of a (local) bridge, the bridge is a *weak* link.
  - Note: the interpretation of a local bridge as an "acquaintance" makes a lot of sense: they are both not part of the (not necessarily strong) triadic closed group of friends and local bridge is a weak link.

Strong triadic closure

- Important opportunities arise infrequently
- Local bridges should convert quickly
- Information shouldn't flow by bridges
- Friends, not acquaintances
- Need to justify acquaintances (weak ties)
- This is why we generalize to a weak-strong graph, weak = acquaintance, strong = friend

A note on “strong”

- “Strong version of a condition” *vs.* usage of “strong” in “strong triadic closure.”
  - A *strong condition* usually is more restrictive, and applies to fewer cases.
  - *Strong triadic closure* is the opposite; the condition is less restrictive than in triadic closure (it only applies when the links are both strong, so more graphs are strong triadic closed than are triadic closed).

Graph structure corresponds to real phenomena

- Theorem: *if*
  1. graph is strong triadic closed
  2. node A has a bridge
  3. node A has another strong link

  Then the bridge is a weak link.

Interpretation

- Strong triadic closure implies group of friends is closed
- So bridges allow weak links to non-friends (acquaintences)
- The theorem is purely mathematical, but it implies something that we now know about graph structure from Granovetter, namely that there is reason to expect bridges to weak links.
- This is what mathematics is for: helping us understand *why* certain properties of behavior tend to "cluster."

Discussion of graph with local bridge

- Show that graph is strongly triadically closed, though not triadically closed.
- Point out gatekeepers (several).
- Interpretation as members of a college class, with connections across companies.

“Smoothing” strength

- We needed to distinguish strong and weak links to analyze Granovetter’s observation. An obvious generalization that is sometimes useful is to make strength a numerical quantity.
  - Practical measures of link strength in communication networks include number of calls, and cumulative length of calls, in a fixed period.
  - In surveys, we might ask whether a person is *unrelated*, an *acquaintance*, or a *friend*.

"Smoothing" bridges

- In many practical applications, the fraction of edges that are local bridges) is tiny.
- We can make this smoother by defining *neighborhood overlap* of an edge AB as
$$\frac{|N(A) \cap N(B)|}{|N(A) \cup N(B)|}$$
where $N(X)$ is the set of nodes that are neighbors of X.
- In data sets where we have a quantitive measure of strength, it often shows a correlation with neighborhood overlap.

Overlap *vs.* strength in cellphone calls

[graph here]

Giant component in cellphone data

- Giant components usually don't have many local bridges.
- People usually have many correspondents, and they will typically converse with several. Seems very likely that will be substantial overlap among neighbors (triadic closure).
- However, we expect to see weak ties that link "communities."
  - In the bridge figure we can see three or four such triadically closed "communities."
  - *E.g.* a company with several departments.

Decomposing the giant component

- Delete links one at a time, starting with the strongest.
  - The giant component gradually shrinks as individuals lose all their ties.
- Now try the same process, starting with the weakest.
  - The giant component shrinks more quickly as individuals lose all their ties, but
  - also because it fragments into smaller components when bridges are deleted, and some of these bridges link components of similar size.
  - The fragmentation doesn't happen until late in the strongest first process because these are much less likely to be bridges (or in general have low neighborhood overlap).
- This strongly suggests that weak links join "communities."

Research opportunities

- The theoretical work is not fully worked out.
  - Some advanced mathematics is used, but there remain important concepts (such as "community") that remain undefined.
  - Probably there are elementary theorems (like the one about strong triadic closure and the weakness of bridges) to be proved.
- Experiments like changing the order of link deletion according to a parameter like link strength often help clarify the nature of concepts, so are valuable contributions even if you are not a strong mathematician.
  - (Homework) What do you think would happen in case of deletion in order of decreasing neighborhood overlap? Increasing order?
- Many opportunities to study real networks, as well as to develop theorems that describe why they behave (show relationships) as they do. (Book points this out.)

Applications of tie strength

- Besides strong/weak, what can we say?
- Try different ways of measuring strength, and how that affects results.
- Social networks allow examination of relationships in ways that subjects define, so are not artifacts of the research design.
  - Consider office colleagues one of whom is a salesperson, they may call each other a lot. But each may rarely call their wives, because they don't do it at work, and otherwise life is very regular.
- Relationships like Facebook friend or Twitter follower are declared by users themselves, and have well-defined semantics according to the rules of the social network.
  - They correspond imperfectly to the usual usage in everyday language.

Friends on Facebook

- Facebook friends aren't necessarily real-life friends, but this is easier to deal with than the reification of "friend" as measured by number of phone calls.
- On the other hand, Facebook friendship is necessarily symmetric, which is not always true in real life.
- With Facebook, you can get an exact list of friends from the Facebook API. Not necessarily true of people trying to list real-life friends.
  - However, in case of karate club, there's a small universe known to the research so can be more accurate (*e.g.*, give a list).

Behavior on social networks

- Existence of flamewars on email lists and newsgroups starting from the 1980s.
  - Even virtual violence (deleting files, for example).
- Similar existence of flaming, trolling, doxxing on modern social networks.
- Social networks perhaps even more flammable than mailing lists, since it is (at least, can be) more public.
- Are they really different?
- Are behavioral differences due to changes in individuals' thinking and behavior, or is it emerging from network structure?
  - *E.g.*, text-based communication does not express subtleties of emotion (even with smilies and emoji).

A remark on Twitter feeds

- Compare personal experience:
  - People I follow don't seem to post much. Is that because they aren't posting much or because Twitter isn't putting them in my feed because they don't seem to post things I'm interested in?
  - I don't like Trump, and I don't like my Republican Senator. But I follow my Senators and Representative to be well-informed in future elections. However, I see only nasty opposition in the replies to the Republican Senator. Is that all there is, or has Twitter figured out I don't like him and so only show the nasty replies?

Interaction on Facebook

Graph references for this lecture are from
`http://www.cs.cornell.edu/home/kleinber/networks-book/networks-book-ch03.`

- Some researchers looked at intensity of interaction among Facebook users.
- When people have Facebook friends, do they actually interact with them?
- Some people use Facebook to coordinate a group. Interaction looks small, although the group make interact offline frequently and intensely.
- Others post comments each way (*e.g.*, a family spread out geographically.
- Researchers defined three degrees of communication between friends:
  - reciprocal communication (each member sent messages to the other),
  - one-way communication (only one member has sent a message to the other), and
  - maintained relationship (one way), clicking on content or multiple profile visits.

Apparent nesting of relationships by strength

This page refers to Figure 3.8.

- We look at the *network of neighbors of one member*. The graph shows the various graphs associated with a particular member.

- According to the definitions in this research, *every* reciprocal communication link is also a one-way communication link.

- For this member, it's easy to visually confirm that the one-way communication links are pretty much all maintained relationship links.

  – I will drop "communication" and "relationship" below.

  I did not expect this. In using Twitter, there are people whose profiles I check *because* I don't "at" them, get "at"-ed by them, *etc.*

- Thinning takes place by "area," not uniformly in the graph. Suggests multiple mechanisms for finding new relationships.

  – Can this be measured statistically? If so, look at distribution of number of mechanisms.

Friendship strength

This page refers to Figure 3.9.

- The graph shows the average count of each kind of link over the members with a particular number of friends.
- We see that on average the quantitative relationship is verified.
  - This does *not* mean that the nesting relationship is verified.
  - Reciprocal links are logically a subset of the one-way links, but
  - I would like to see some statistics on the percentage of maintained links that are also one-way and reciprocal links. (Those are different statistics.)
- The book points out that even people with 500 friends on Facebook on average limit themselves to about 40 maintained links, 20 one-way links, and 10 mutual links.

Increasing friendship strength

This page refers to Figures 3.9 and 3.10.

- The book points out that even people with 500 friends on Facebook on average limit themselves to about 40 maintained links, 20 one-way links, and 10 mutual links.
- I'm more interested that the asymptotes do not seem to be horizontal: people with more friends have more strong relationships of each type.
- This is different for Twitter: the asymptote is horizontal, there seems to be a cap on strong ties.
  - Not clear what this says: could be a limitation of Twitter clients.
  - Could be a limitation of the methodology (no measure of attention paid to tweets, which corresponds to maintained relationships).
  - It could be a different kind of social network, different style of communication: *needs more research*.
    - Note that Facebook's *News Feed* resembles Twitter to a certain degree.