UCLouvain

École polytechnique de Louvain

LINFO1115

Reasoning about a highly connected world graph theory, game theory, and networks

Peter Van Roy

Dylan Goffinet

2021-2022

Contents

1 | Graph Theory

1.1 Basic concepts

1.1.1 Graph

- A graph $G = (N, E)$ with N a set of nodes and E a set of edges, swhere an edge is a pair of two nodes.
- Example : $N = a, b, c, E = a, b, b, c$

1.1.2 Directed Graph

- Each edge has a direction, and is represented as a tuple with a *first* and a *second*
- Example : $E = (a, b), (b, c)$

1.1.3 Undirected Graph

- Edges have no direction, they are represented as a set
- Example : $E = a, b, b, c$

1.2 Paths and connectivity

1.2.1 Path

A path in a graph is a sequence of nodes such that each successive pair in the sequence is an edge of the graph.

1.2.2 Simple path

A simple path is a path where each node occurs at most once.

1.2.3 Cycle

A cycle is a path with 3 or more edges such that the first and last nodes are the same and otherwise all node are distinct.

1.2.4 Connected Graph

A graph is connected if there exists a path between every pair of nodes.

1.2.5 Component of a graph

A component is a subset of nodes that satisfies two properties :

1. The subset is *connected*.
2. The subset is *maximal* : there is no superset that is connected.

Strongly connected component

A component is strongly connected if it satisfies two properties :

1. Each node in the subset has a path to each other node in the subset.
2. The subset is *maximal* : it is not part of a larger subset with the property that each node can reach the other.

1.2.6 Analyzing a graph

- A graph can be divided into its components.
- For each component, we can study its internal structure. For example, removing a node (with its edges) may split it in more components.

1.2.7 Giant components

Large social networks often have a giant component, which is a component that contains a significant fraction of the graph's nodes.

1.2.8 Bipartite graph

A bipartite graph is a graph where the nodes can be separated into two sets such that each edge connects a node from one set to the other.

1.2.9 Clique or complete graph

There exist an edge joining every pair of nodes.

Labeled complete graph

A complete graph where every edge is labeled.

1.3 Distance in a graph

1.3.1 Length

The length of a path is the number of edges it contains.

1.3.2 Distance between two nodes

Length of the shortest path between these two nodes.

1.3.3 Breadth-first-search

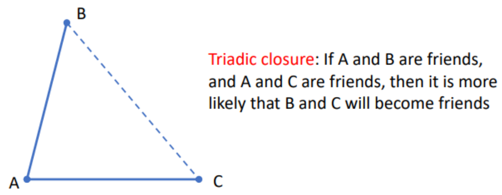
For finding distances between nodes.

2 | How to combine graph theory and sociology

2.1 Social networks

A graph where the nodes denote human beings and the edges denote a connection between two human beings.

Triadic closures



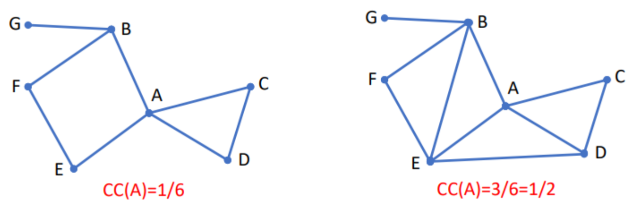
if two people in a social network have a common friend, then the likelihood that they will become friends in the future is increased :

- B and C have more opportunities to meet.
- B and C both thrust A so they have a basis for trusting each other.
- B and C have common interests with A so they may have common interests with each other.
- A has an incentive to bring B and C together since if they are not friends, it stresses A.

Strong triadic closure

If A and B are close friends as well as A and C, then it is especially likely that B and C will become connected.

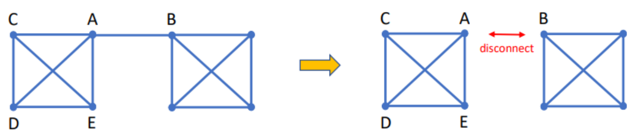
2.1.1 Clustering coefficient



The *clustering coefficient* $CC(A)$ of node A in a graph is :

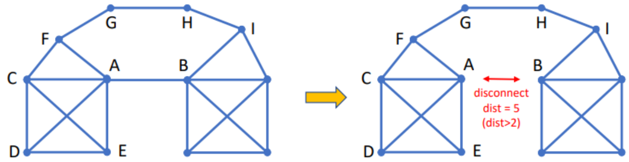
- $CC(A)$ = probability that two randomly selected friends of A are friends with each other.
- $CC(A) = \frac{X}{Y}$
  - X = the number of pairs of A's friends that are connected.
  - Y = the maximum number of pairs of A's friends that are connected.
- Triadic closures causes the clustering coefficient to increase

2.1.2 Bridges



The edge joining A and B is a *bridge* if removing this edge causes A and B to be in different components.

Local bridges



- The edge joining A and B is a *local bridge* if A and B have no friends in common, so deleting the edge will increase the distance between A and B to a value *strictly greater than 2*.
- Local bridges play roughly the same role as bridges, but in a less extreme way : they connect a node to another node that would otherwise be far away.

2.1.3 Neighborhood overlap

The neighborhood overlap generalizes local bridges : $\frac{X}{Y}$ where

- X = number of nodes who are neighbors of *both* A and B.
- Y = number of nodes who are neighbors of *at least one* of A and B.

3 | Networks in their surrounding context

3.1 Longitudinal studies

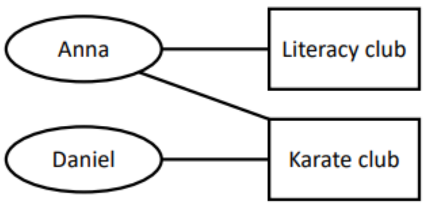
The network must be followed over time.

3.2 Homophily

They are two basic mechanisms that cause friends to resemble each other : selection and social influence (both are sociological concepts)

- *Selection* : people select friends with similar characteristics (an internal mechanism).
  - individuals drive the formation of new links.
- *Social influence* : people modify their behaviors to be closer to their friends (an external mechanism), also called "peer pressure".
  - Existing links drive the formation of new links

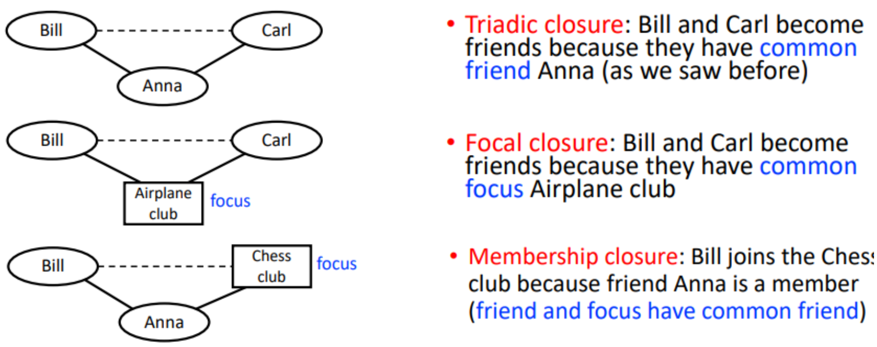
3.2.1 Affiliation network



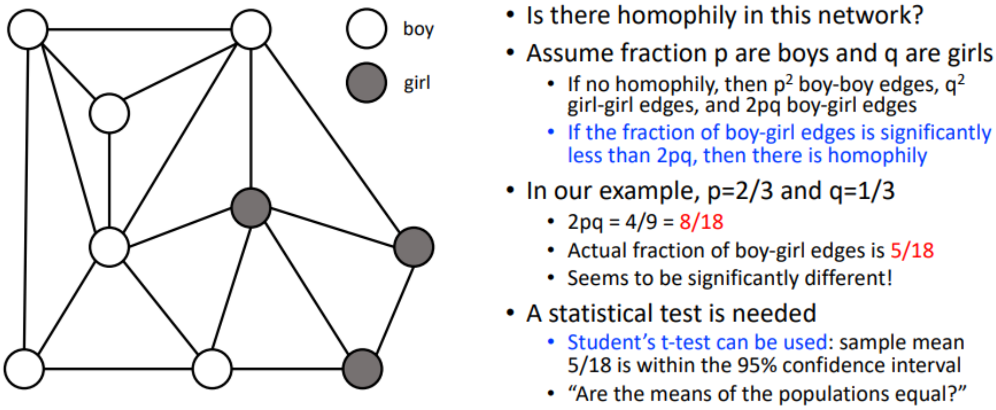
An affiliation network is a bipartite graph that shows which individuals are affiliated with which activities.

- The first set is the individuals.
- The second set is the foci.

Three forms of closures in a social-affiliation network



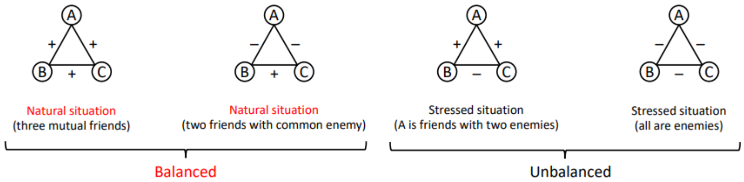
3.2.2 Measuring homophily



4 | Positive and negative relationships

4.1 Structural balance

4.1.1 Balanced and unbalanced triangles



4.1.2 Balance for networks

Local definition

A network is balanced if all the triangles in it are balanced

Global definition

If a labeled complete graph is balanced, then either all pairs of nodes are friends, or else the nodes can be divided into two groups such that nodes within a group are friends and nodes between groups are enemies.

4.1.3 Weak structural balance

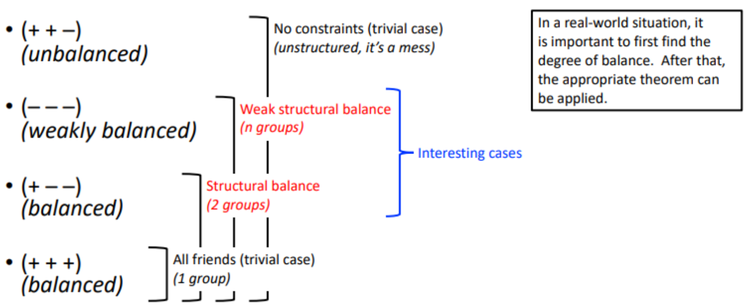
Local definition

A network is balanced if all the triangles in it are balanced or they are $(- - -)$.

Global definition

If a labeled complete graph is weakly balanced, then its nodes can be divided into groups such that any two nodes belonging to the same group are friends and any two nodes belonging to different groups are enemies.

Degrees of structural balance



5 | Introduction to game theory

Game theory studies the interaction of participants.

5.1 Game with two rational player

5.1.1 Game

- There is a set of participants, called the players.
- Each player has a set of options for how to behave : we will refer to these options as the player's strategies.
- For each choice of strategies (one by each player), each player receives a payoff. The payoff depends on the strategies selected by all players. Payoffs are generally numbers, and each player prefers larger payoffs to smaller ones.

Assumptions

1. All that a player care about is summarized in the payoff.
2. Each player knows everything about the game.
3. Each player chooses a strategy to maximize their own payoff, given a belief about the strategy used by the other player.

On shot game

Two players that play only once and where the players simultaneously and independently choose their actions.

Dynamic games

Games that are played sequentially over time.

Forms

Normal form

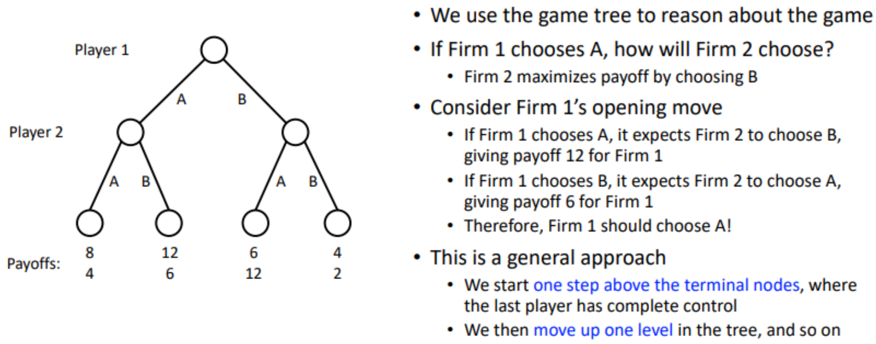
In normal form, each player commits ahead of time to a complete plan for playing the whole game.

		Firm 2			
		AA,AB	*AA,BB*	*BA,AB*	*BA,BB*
Firm 1	*A*	8, 4	8, 4	12, 6	12, 6
	B	6, 12	4, 2	6, 12	4, 2

- Another approach to analysis is to use the normal form
  - Each player makes a plan for how to play the entire game, covering all choices
  - This play corresponds to the player's strategy
- Firm 1 has two possible strategies, A or B
- Firm 2 has four possible strategies, which depend on what Firm 1 does
  - (A if A, A if B), (A if A, B if B), (B if A, A if B), (B if A, B if B) written as (AA,AB), …
  - We write it this way because Firm 2 has to enumerate all possibilities in advance!

Extensive form

In extensive form, each player makes an optimal decision at each intermediate step.



5.1.2 Strategies

Strictly dominant strategy

When a player has a strategy strictly better than all others, regardless of what the other player does (not always possible).

		Player 2	
		T'	T
Player 1	S'	0.48, 0.12	0.60, 0.40
	S	0.40, 0.60	0.32, 0.08

Mixed strategies

A random choice between two pure strategies is called a mixed strategy.

		Player 2	
		Head	*Tail*
Player 1	*Head*	-1, +1	+1, -1
	Tail	+1, -1	-1, +1

Best response

		Player 2	
		T'	T
Player 1	S'		$(P'_1, \_)$
	S		$(P_1, \_)$

- S for player 1 is best response to T for Player 2 if :
  - For all other strategies, S' of Player 1 : $P1(S,T) \geq P1(S',T)$
- S for player 1 is strictly best response to T for Player 2 if :
  - For all other strategies, S' of Player 1 : $P1(S,T) > P1(S',T)$

Nash equilibrium

- Each player's strategy is a best response to the other.
- Corresponds to free market.

		Player 2	
		C1	C2
Player 1	C1	a, b	e, f
	C2	c, d	g, h

$P(C2, C1)$ is a Nash equilibrium if $a < c$ and $h < d$

Coordination / anti-coordination games

Games with multiple Nash equilibria.

Principle of indifference

If Player 1 believes that Player 2 will play H more than $\frac{1}{2}$ of the time, then Player should play T (and conversely).

Choice of $q = \frac{1}{2}$ by Player 2 makes Player 1 indifferent to H or T.

Each player should randomize to make the other player indifferent.

Pareto optimality

- A choice of strategies (one by each player) is Pareto optimal if there is no other choice in which all players receive payoffs at least as high, and at least one player receives a strictly higher payoff.
- Corresponds to regulation.

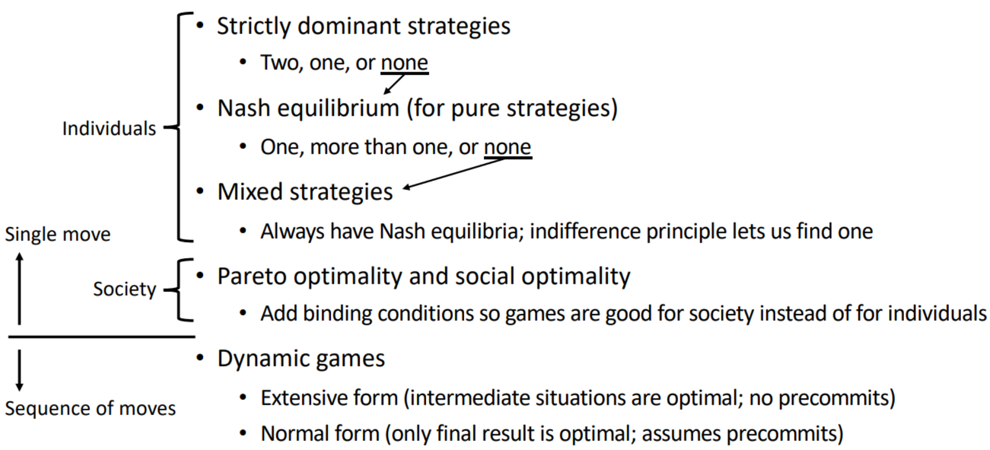
		Player 2	
		C1	C2
Player 1	C1	90, 90	86, 92
	C2	92, 86	88, 88

$(88, 88)$ is a Nash equilibrium, whereas the 3 others options are Pareto optimal.

Social optimality

- A choice of strategies (one by each player) is socially optimal if it maximizes the sum of the players' payoff.
- It is also Pareto optimal.

5.2 Summary

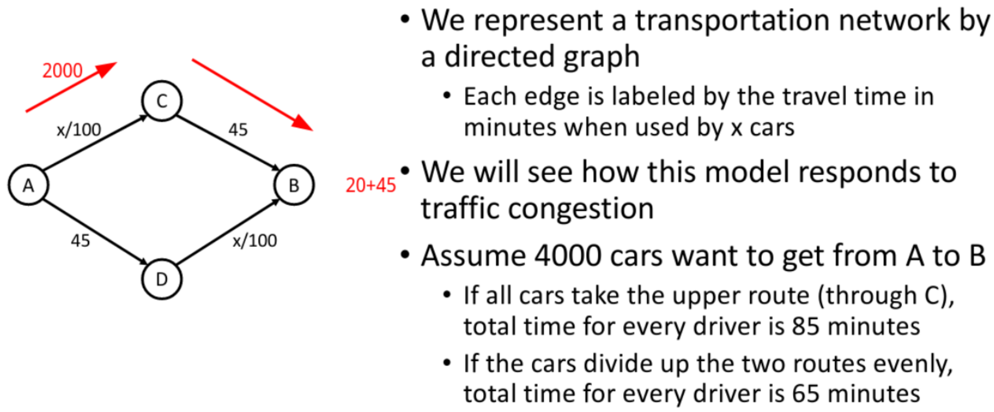


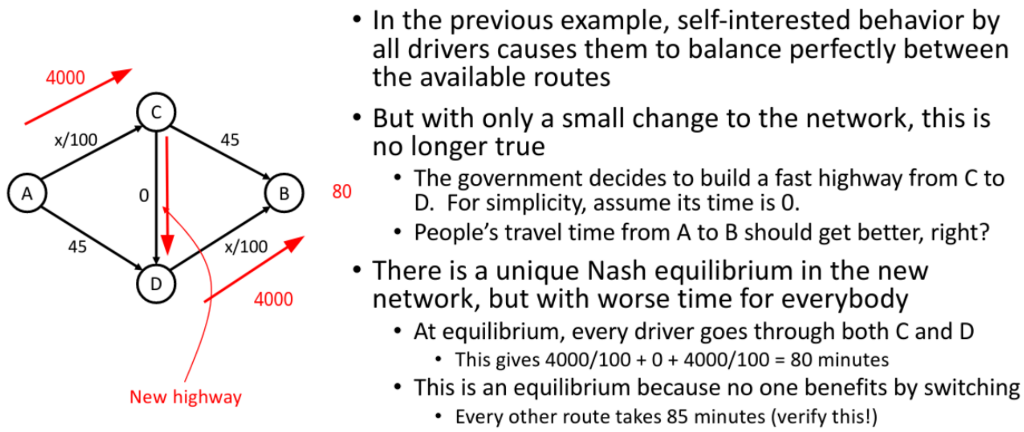
6 | Car traffic networks

6.1 Braess' Paradox

In a network representing a certain traffic (here, a car traffic network), adding a new edge (a new strategy) can result in a strictly worse Nash equilibria than the previous one. In short, there are many situations where adding more choices leads to worse situations.

Example :





7 | Auctions

7.1 When are auctions appropriate ?

When the value of the objet is :

- known : an auction is inappropriate.
- unknown : an auction is appropriate.

7.1.1 Two kind of unknown values

Independent, private value

- Each buyer is interested for its own personal use
- The values can be different because the buyers have different tastes

Common value

- Assume each buyer plans to resell the item if bought, then the item has an unknown but common value regardless of who buys it.
- Winner's course : if there are many bidders, the winning bidder will very likely overestimate the common value and lose money on the resell of the object.

  → To counter it, we should bid lower that our private value of the good.

7.2 Type of auctions

7.2.1 Ascending-bid auctions

- English auctions.
- Realtime.
- Seller gradually raises the price until only one bidder remains.
- Equivalent to second-price auctions.

7.2.2 Descending-bid auctions

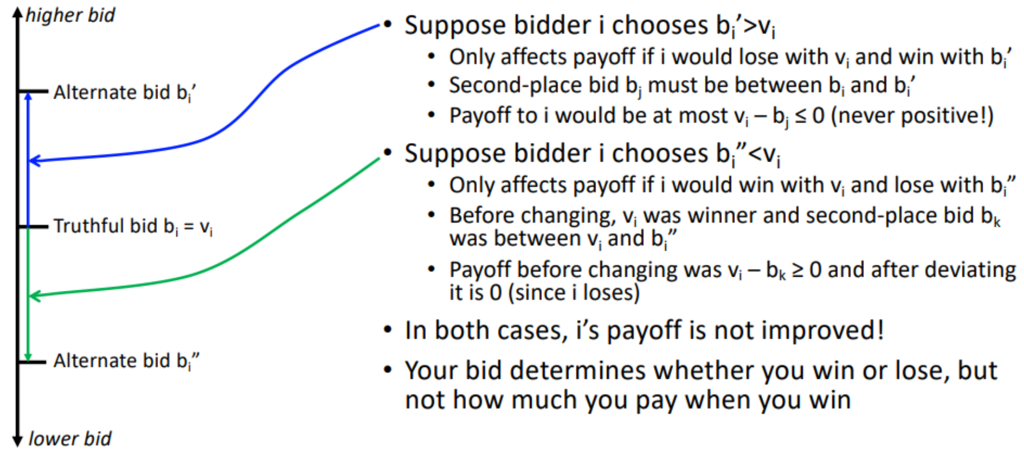
- Dutch auctions.
- Realtime.
- Seller gradually lowers the price until the first bidder accepts.
- Equivalent to first-price auctions.

7.2.3 First-price sealed-bid auctions

- Simultaneous bids.
- The highest bidder wins and pays the value of their id.
- Payoff to bidder i with value v_i and bid b_i is defined as follows :
  - if b_i is not the winning bid, then the payoff to i is 0.
  - if b_i is the winning bid, then the payoff to i is $(v_i - b_i)$.

7.2.4 Second-price sealed-bid auctions

- Vickrey auctions.
- Simultaneous bids.
- The highest bidder wins and pays the vale of the second-highest bid.
- Bidding your true value is a dominant strategy.
- Payoff to bidder i with value v_i and bid b_i is defined as follows :
  - if b_i is not the winning bid, then the payoff to i is 0.
  - if b_i is the winning bid, and some other b_j is the second-place bid, then the payoff to i is $(v_i - b_j)$.



8 | Markets

A market is :

- any place where two or more parties can meet to exchange goods or services, usually for money but sometimes directly through barter.
- a coordinating mechanism that uses prices to convey information among the parties to regulate production and distribution.

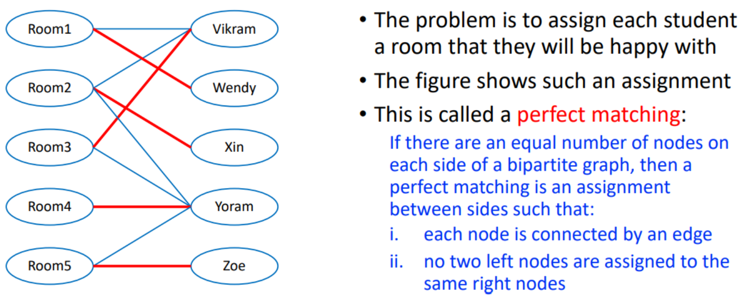
8.1 Matching markets

A matching market is a market that matches each buyer to a desired item.

8.1.1 Perfect matching

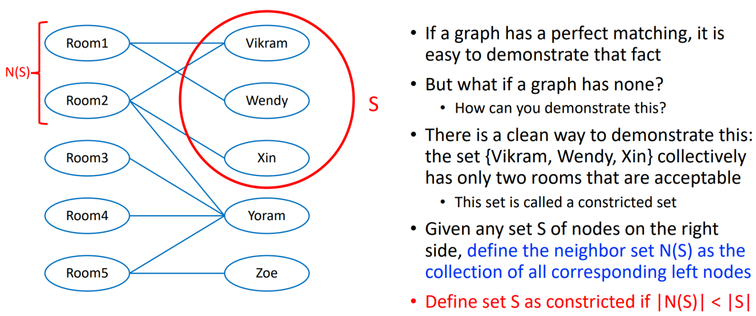
If there are an equal number of nodes on each side of a bipartite graph, then a perfect matching is an assignment between sides such that :

1. Each node is connected by an edge.
2. No two left nodes are assigned to the same right nodes.



8.1.2 Constricted sets

Given any set S of nodes on the right side, we define the neighbor set $N(S)$ as the collection of all corresponding left nodes. Set S is constricted if $|N(S)| < |S|$.



8.1.3 Matching theorem

> If a bipartite graph with equal numbers of left and right nodes has no perfect matching, then it must contain a constricted set.

8.1.4 Matching markets with prices

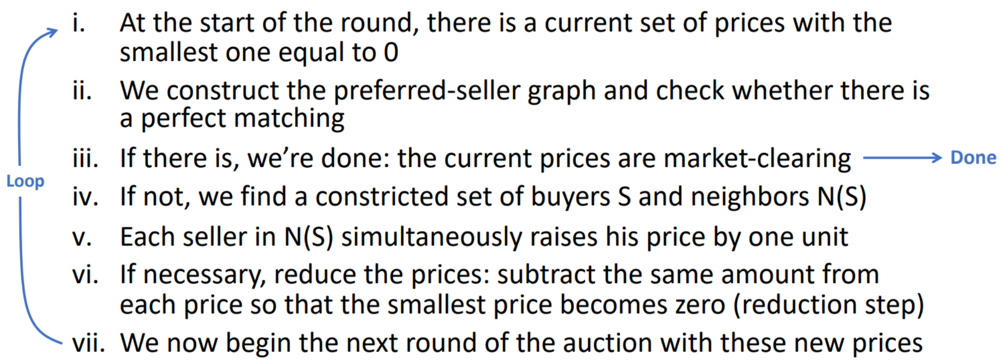
Prices and market clearing

Market clearing

- Market clearing means that the supply of an item is equal to the demand, so that there is no leftover of either.
- We say that a set of prices is market clearing if the resulting preferred-seller graph has a perfect matching.
- In a bipartite matching market, there exists a set of market-clearing prices for any set of buyer valuations.
- Market-clearing prices are always socially optimal :
  - For any set of market-clearing prices in a bipartite matching market, a perfect matching in the preferred-seller graph has the maximum total payoff for sellers and buyers, for any assignment of sellers to buyers.

Constructing market clearing prices

We construct a procedure that gradually increase prices and we show that this will always lead to market clearing.



8.1.5 Auctions and markets are related

The bipartite matching market generalize an ascending-bid auction :

- The ascending-bid auction has n buyers and 1 seller. The seller keeps increasing prices until one buyer wins the auction.
- The bipartite matching market has n buyers and n sellers. The sellers keep increasing the prices until each buyer has found one seller.

Valuations

Each person gives a numerical value to each item.

Quality

The sum of all persons' valuations.

Optimal assignment

It is an assignment of highest possible quality.

8.1.6 Summary

- Markets are a general form of interaction between participants
  - Markets use prices as a decentralized coordination mechanism
- We introduce an idealized market and study its behavior
  - A bipartite matching market with same numbers of buyers and sellers
  - Market clearing happens when all buyers are connected to different sellers
    - This corresponds to a perfect matching in our bipartite matching market
  - We show that prices can always be found to give market clearing
    - This gives an idealized view of how markets regulate prices
    - It is actually a kind of auction, generalizing the ascending-bid (English) auction
- In real-world markets, it is often assumed that "supply equals demand"
  - Market clearing makes precise the idea of "supply equals demand"
  - Prices are converging to values that cause "supply equals demand"
  - The idealized market can be adapted to real-world situations

8.2 Markets with intermediaries

8.2.1 Order book

Is a list of orders that buyers and sellers have submitted.

- Ask : lowest sell offer.
- Bid : highest buy offer.

8.2.2 Buyers, Sellers and Traders

Three fundamental principles

1. Individual buyers and sellers trade through intermediaries.
2. Buyers and sellers may have access to different intermediaries.
3. Buyers and sellers may trade a different prices.

Prices depend on the positions in the network

If a buyer can :

- only access one trader, then the trader increases his payoff.
- access several traders, then the buyer increases his payoff.

Prices and the flow of goods

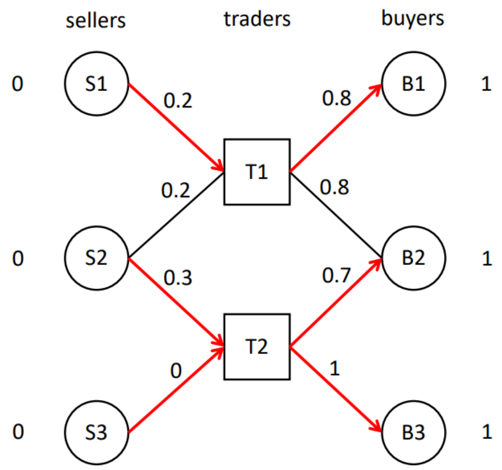
1. First step :
   - Each trader t offers a bid price b_{ti} to each seller i he his connected to.

     → This is where the strategic reasoning is done.
   - Each trader t offers an ask price a_{tj} to each buyer j he is connected to.
2. Second step :
   - Each seller and buyer chooses at most one trader to deal with.

     → The one with the best offer.
   - Each seller sells a copy (or not).
   - Each buyer buys a copy (or not).

At most one copy of the good moves along any edge in the network.

- Many goods can pass through one trader.
- A trader sells exactly what he receives.



Sometimes prices are pushed to the limit so there is no payoff.

Indifference

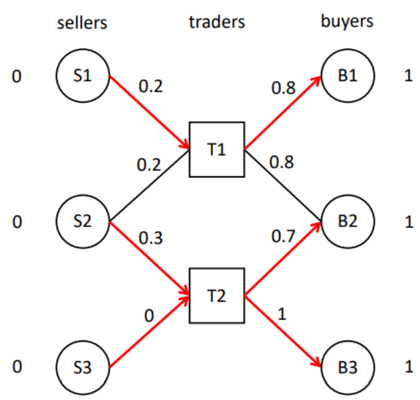
- Prices are pushed to the limit of an individual's willingness to trade.
- We allow trades with zero payoff with ties broken as needed.

Strategies

- All traders simultaneously choose ask and bid prices.
- A seller or buyer's strategy is a choice of neighboring trader (or not).

Payoff

- A trader's payoff is his total profit : sum of all ask prices of accepted offers to buyers minus sum f bid prices of accepted offers to seller.
- For seller i, payoff from selecting trader t is b_{ti}, while payoff from selecting no trader is 0 (note we only consider when $v_i = 0$).
- For buyer j, payoff from selecting trader t is $v_j - a_{tj}$ while payoff from selecting no trader is 0. in the former case, buyer receives the goods but pays a_{yj}.



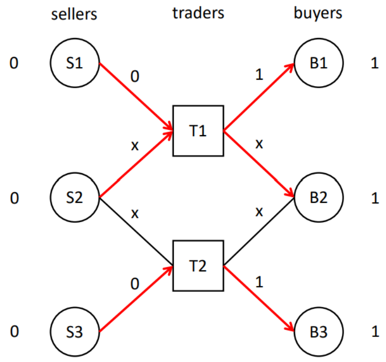
Payoffs for this example:

- T1's payoff is (0.8 – 0.2) = 0.6
- T2's payoff is (0.7 + 1) – (0.3 + 0) = 1.4

- S1's payoff is 0.2
- S2's payoff is 0.3
- S3's payoff is 0

- B1's payoff is (1 – 0.8) = 0.2
- B2's payoff is (1 – 0.7) = 0.3
- B3's payoff is (1 – 1) = 0

8.2.3 Equilibrium in trading networks



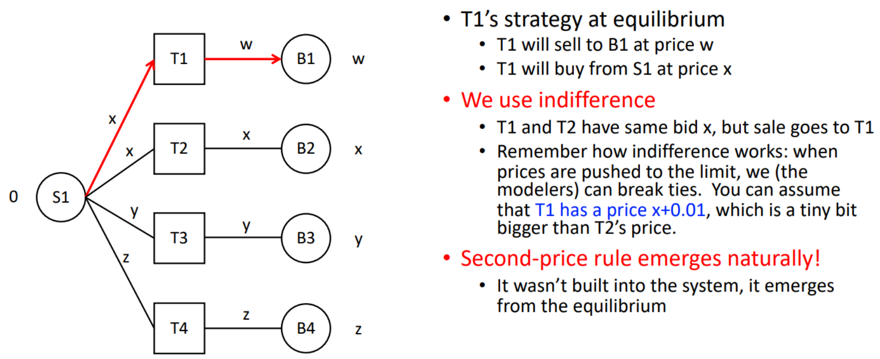
Monopoly

- Buyers and sellers are subject to monopoly when they access only a single trader.
- The only equilibrium is for the trader to set bid to 0 to the seller and ask 1 to the buyer.

Perfect competition

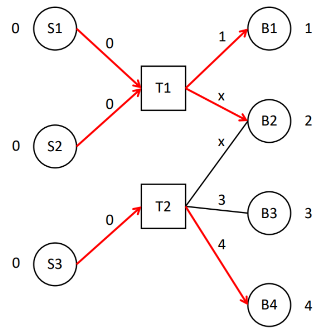
- Buyers and sellers are subject to perfect competition when they access more than one trader.
- Equilibrium for $T2$ occurs at common bid and ask of x.

Second-price auction

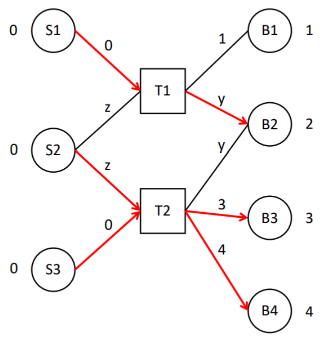


Ripple effect

The network is partially under the control of the participants.



- We first determine the equilibria for the original network
- All sellers and buyers are monopolized except for B2
  - Their payoffs are all zero
- B3 will not buy the good
  - Because of indifference
- B2 buys from T1
  - Because of indifference
  - x can range from 0 to 2



- We determine the equilibria for the changed network
- The two bids to S2 must be the same, also the two asks to B2 must be the same
- S2 will sell to T2 rather than T1
  - To T1 price would be maximum 2 and T2 could outbid that
- The ask y must be at least 1
  - Also, y cannot be greater than 2
- T2 buys two copies, T1 one copy
- The bid z must be at least 1
  - Also, z cannot be greater than 3

8.2.4 Social welfare in trading networks

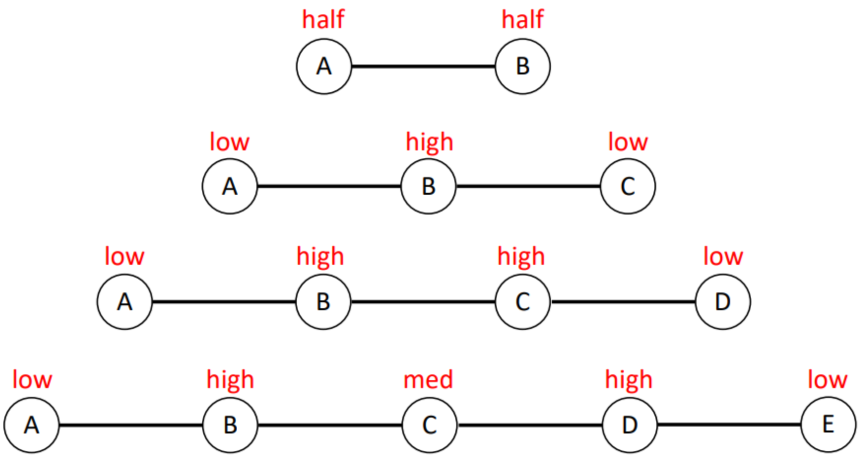
- The social welfare is simply the sum of $(v_j - v_i)$ over all moved goods.
- There exist values for the variable that maximize social welfare and picks an equilibrium.
- In every trading network, there is always at least one equilibrium.
- Social welfare increases as connectivity increases.

8.2.5 Summary

- There are often intermediaries in markets
  - Intermediaries (traders) simplify management for buyers and sellers
  - We call the resulting graph a trading network
- We have defined an idealized trading network
  - Sellers on one side, traders in the middle, buyers on the other side
  - We can compute the equilibria of trading networks
  - Two important patterns are monopoly and perfect competition
- Our idealized trading network has many realistic phenomena
  - Implicit competition, connection to second-price auction, ripple effects
- Social welfare increases as connectivity increases
  - Also, trader profits decrease as connectivity increases!
  - Trader profit depends on essential edges (deleting them changes social optimum)

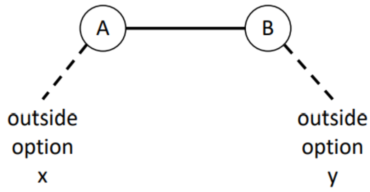
8.3 Bargaining and power in networks

8.3.1 Power in paths



8.3.2 A mathematical framework for negotiation

Nash bargaining solution



- A gets $x + (s/2)$ and B gets $y + (s/2)$ where $s = 1 - x - y$.
- Nash bargaining solution :
  1. A gets $(x + 1 - y)/2$
  2. B gets $(y + 1 - x)/2$

Ultimatum game

1. A is given a dollar and is told to propose a division with B, that is, A proposes how much to keep from himself and how much to give to B.
2. B is given the option to approve or reject the division.
3. if B approves, both get the proposed amount. If B rejects, both get zero.

Game theory

Game theory says that B should accept if the amount > 0.

Human way of playing

- If the amount offered is to unbalanced, B will feel cheated and will refuse A's offer.
- We say there is an Emotional payoff.

Stable outcomes

- An outcome of a network is stable if and only if it contains no instabilities.
- Stability is too weak in networks with small power difference

Outcome of a network exchange on a graph

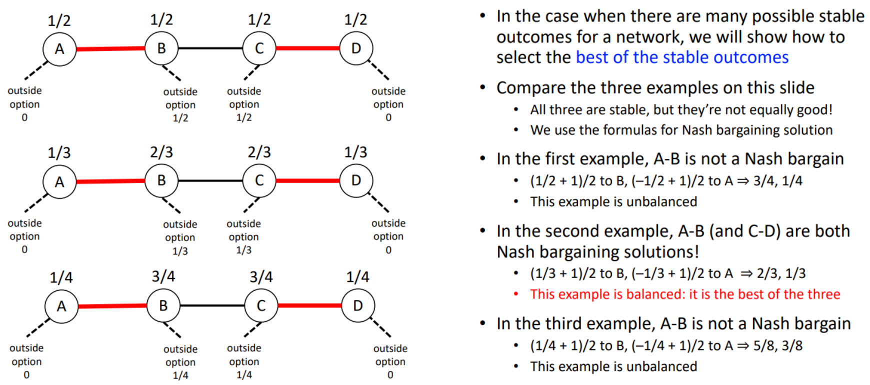
1. A matching on the set of nodes. This corresponds to the one-exchange rule, and some nodes may be left out.
2. A number associated with each node, its value, indicating how much the node gets from the exchange. If two nodes are matched, the sum of values should be 1. If a node is not in a matching, its value should be 0.

Instability

Given an outcome consisting of a matching and set of values, an instability in this outcome is an edge not in the matching, joining two nodes X and Y, such that $(X\text{'s value}) + (Y\text{'s value}) < 1$.

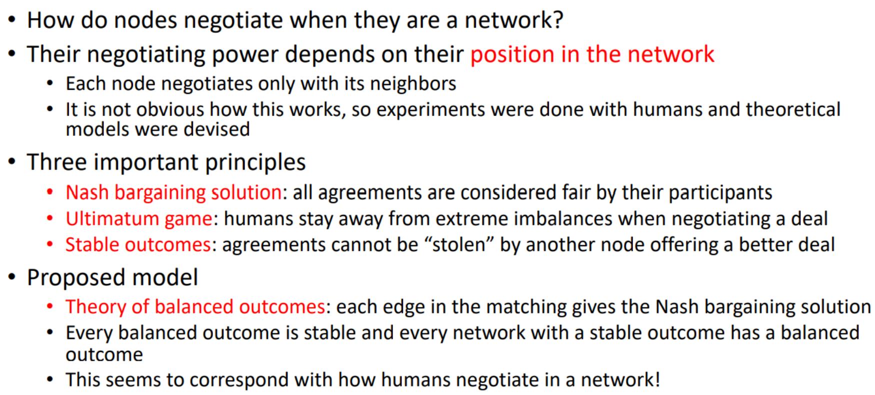
Theory of balanced outcomes

> An outcome (consisting of a matching plus node values) is balanced, if for each edge in the matching, the money split represents the Nash bargaining solution for the two nodes involved, given the best outside options for each node provided by the values in the rest of the network.



Outcome is balanced when all agreements are Nash bargaining solutions; balanced outcomes exist whenever stable outcomes exist.

8.3.3 Summary



9 | World Wide Web

9.1 Structure of the web

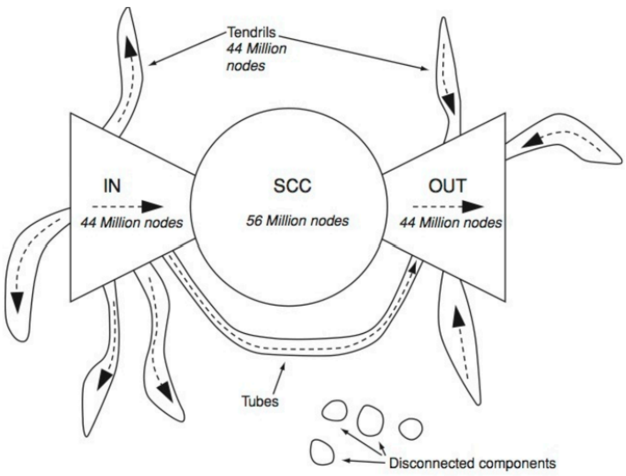
9.1.1 Hypertext

Each node is a page that contains links to other pages.

9.1.2 Network distribution

Pages are geographically distributed as internet nodes.

9.1.3 Bow-tie structure



SCC: Strongly connected component

9.2 Web 2.0

9.2.1 Increased abstraction level

1. Collaborative creation and maintenance of shared content, instead of individuals creating pages (Wikipedia).
2. Creation of services that manage data, instead of sets of pages (Gmail).
3. Connections between people instead of between pages (Facebook).

9.2.2 Social phenomena

Software that gets better the more people use it

Online services are more useful and more appealing as more people use them. A marketplace is more useful as more people put objects for sale and more buy objects. Recommendation systems and trust systems are more useful as more people use them.

The wisdom of crowds

Collaborative authorship on Wikipedia, group evaluation of news content on Digg, breaking news on Twitter. This process can also fail, because of herding and instability.

The "Long Tail"

Combining a small amount of very popular content with a huge amount of less popular content with "niche appeal". The less popular content is often bigger than the popular content.

9.3 Link analysis and Web search

9.3.1 The problem of search

- Inexpressiveness : subtle ideas cannot be expressed using keyword combinations.
- Synonymy : multiple words for the same concept, "scallions" vs "green onions".
- Polysemy : the same word has multiple meanings, "jaguar" gives cars instead of animals.
- Abundance problem : too much relevant information.

9.3.2 Hubs and authorities

- Authority : page we are looking for, a high-quality answer to the query.
- Hub : page that has a high-quality list to the pages we are looking for.

Algorithm to compute scores

- Start with all hub and authority scores equal to 1.
- Repeat k times :
  - For each page p, update auth(p) to be the sum of all hubs pointing to it.
  - For each page p, update hub(p) to be the sum of all authority pages it points to.
  - Normalize the authority scores by dividing by sum o all authorities (and also for hubs).

Limiting values

- Limiting values are independent of the initial values of hub ad authority scores.
- The normalized values converge to limits.

9.3.3 PageRank

- Uses only one value per page : its importance.
- Intuitively the PageRank value is a kind of "fluid" that circulates through the network, where nodes are containers and links are pages.

  → The amount of fluid is constant, so normalization is automatic.

Basic algorithm

Initialization

We assume that the network has n nodes.

→ $\forall p : pr(p) := \frac{1}{n}$

Iteration

1. Compute the amount of fluid on outgoing links of all pages p.

   → $\forall p : pr_{\text{link}}(p, p') := \frac{pr(p)}{m}$ (page p has m outgoing links, each to p')
2. Sum all fluid for incoming links of page p.

   → $\forall p : pr(p) := \sum_{p' \text{ incoming links}} pr_{\text{link}}(p, p')$

Problem with the basic definition

Some nodes can collect all the PageRank values.

Fixing the problem

1. Apply the basic update rule.
2. Do the following scaling : $\forall p : pr(p) := s * pr(p) + \frac{1-s}{n}$
    - This uses scaling factor s strictly between 0 and 1 (often 0.8 or 0.9). The PageRank fraction $(1-s)$ "evaporates" from each node and "rains" uniformly on all nodes.
    - This completes the cycle between all nodes.

Random walk definition

- Consider a person randomly browsing the Web :
    - They start by picking a page at random (each page equal probability).
    - They follow links for k steps and in each step they follow a random link.

      $\rightarrow$ This is called a random walk on the network.
    - We can prove that the probability of being on a page X after k steps is precisely the PageRank of X after k iterations of the basic update rule.
- The scaled update rule is formulated as follow :
    - At each step, the walker follows a random link with probability s, and "jumps" to a random page with probability $(1-s)$.
    - The walker mostly follows the link and sometimes jumps.

9.4 Summary

- The Web was invented around 1990 and was instrumental in the Internet becoming a widely used information network
    - The Web is fully decentralized, with individual users both creating and accessing content
    - The Web has an emergent "bow-tie" structure with a central strongly connected component
    - The original Web has evolved to Web 2.0, with content, services, and people becoming central, instead of individual pages (increasing level of abstraction)
- Web search solves the abundance problem for information retrieval
    - No group of humans can manually classify the Web's enormous and rapidly evolving content
    - Scalable algorithms determine relevant information using the structure of the graph itself (hubs and authorities, PageRank)
- Web search has evolved into an ecosystem that goes beyond the original PageRank
    - Search engines and companies are in a continuing struggle for ranking: search engines want to rank according to intrinsic importance, whereas companies want to rank highly
    - Search engines have added targeted advertising to the basic search functionality, turning search into a sustainable and useful business model

9.5 Random Walks, Matrices, and Page Rank

9.5.1 Random Walks on Graphs

Random Walks

Let $G(V, E)$ be a connected graph. Consider random walk on G :

- We start at node v_0 and at the t-th step, we are at node v_t
- We move to a neighbor of v_t with a probability of $\frac{1}{d(v_t)}$.
- The sequence of random walks is called the Markov chain.

Convergence

For any connected non-bipartite bidirectional graph, and any starting point, the random walk converges to a unique stationary distribution.

9.5.2 Stationary Distribution

Random walk converges to the stationary distribution : $\pi(v) = \frac{d(v)}{2m}$, where :

- $d(v)$ = degree of v, i.e. the number of neighbors of v.
- $m = |E|$, i.e. the number of edges is G.

9.6 Matrix Representations

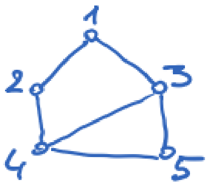
9.6.1 Definitions

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Adjacency Matrix

$$D = \begin{pmatrix} 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \end{pmatrix}$$

Diagonal matrix with $D_{i,i} = 1/d(i)$



$$M = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

Transition (random walk) Matrix
M=DA

9.6.2 Adjacency Matrix

A is $n * n$ adjacency matrix of $G = (V, E)$ where $A_{ij} = 1$ if there is a link between node i and node j (0 otherwise).

Matrix manipulations

- A give us all 1-hop paths.
- A^2 gives us the number of 2-hops paths.
  - A^k gives us the umber of k-hops paths.
- Let $v = (1\ 0\ 0\ 0\ 0)$ represent a message from the first node
  - vA^k indicates how many walks of length k from node 1 end up in node i.

9.7 Google PageRank

9.7.1 PageRank "Voting" formulation

- Each page has a budget of "votes" and distributes them evenly to all the outgoing links.
- The importance of node j is the sum of the votes on its in-links.

9.7.2 Google PageRank

- Principal eigenvector on the transition matrix of the web graph.
- Computed with Power iteration.
- Issues in directional graph : Dead end (nodes with no out-degree) and spider traps (loop inside a component).

How do we fix PageRank

- Make the graph strongly connected and aperiodic
- Make "tiny" links from each node to every other node, and keep the core of the initial graph.
- Google random walker will :
  - follow the "real" link with prob β
  - jump to a random connected page with prob $1 - \beta$

    → Random walk will escape any spider traps after 5-10 steps (β is in the range of 0.8 to 0.9), and will always teleport out of a dead end after the first step ($1 - \beta = 1$).

Problems with PageRank ?

- Measures "generic" popularity of a page, might miss topic-specific authorities.

  → Solved with Topic Specific PageRank.
- Susceptible to link spam (artificial link topologies create to boost page rank).

  → Solved with TrustRank.
- Uses a single measure of importance.

9.8 Beyond PageRank

9.8.1 Topic Specific PageRank

- Bias the random walk by giving more influence to the webpages that are close to a particular topic
- Teleport to "relevant pages" (teleport set, obtained from the "previous generation" search engines) instead of random ones.

9.8.2 SimRank

- Also known as PageRank with restarts
- Random walks from a fixed node on k-partite graphs with k types of nodes.

9.8.3 PageRank summary

- **“Normal” PageRank:**
  - Teleports uniformly at random to any node
  - All nodes have the same probability of surfer landing there
- **Topic-Specific PageRank also known as Personalized PageRank:**
  - Teleports to a topic specific set of pages
  - Nodes can have different probabilities of surfer landing there
- **Random Walk with Restarts (SimRank):**
  - Topic-Specific PageRank where teleport is always to the same node.

9.9 Web spamming and how to combat it

9.9.1 Web spamming

Deliberate action to boost a webpage position in ranking. Used by those with commercial interest to try to exploit search engines to bring people to their own site (regardless of the relevance).
Spam : techniques for achieving high relevance/importance for a web page.

9.9.2 Term-spam

Putting a specific word 1000 times on your page and/or copy a highly relevant page to yours and make it "invisible"

Google's solution to Term-spam

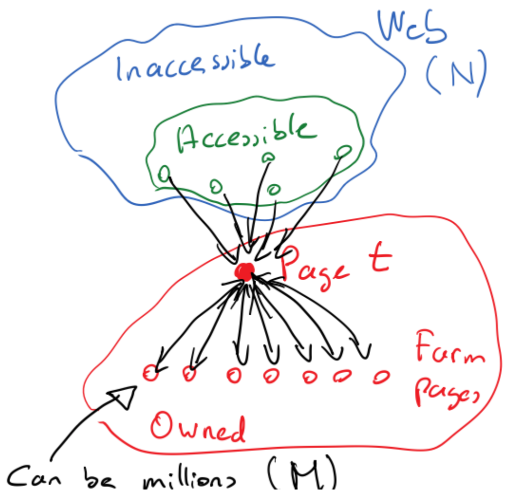
Believe what people say about you rather than what you say about yourself, and use PageRank as a tool to measure the relevance of Web pages.

9.9.3 Spam Farming

Concentrating PageRank on a single page by creating link structures (link farms) that boost Page rank of a particular page.

Webpages from spammers' point of view :

- Inaccessible pages.
- Accessible pages (blogs, comments, reviews where spammers can post).
- Owned pages (pages/domains controlled by spammers).



- x : PageRank contributed by accessible pages.
- y : PageRank of target page t.

  $\rightarrow y = \frac{x}{1-\beta^2} + c\frac{M}{N}$ where $c = \frac{\beta}{1+\beta}$

How to combat LinkSpam ?

- Identifying and Blacklisting of spam farms $\rightarrow$ Not good, arms race of hiding and detecting spam farms.
- TrustRank : teleport to trusted pages only.
- Teleport set = trusted pages, similar to Topic-Specific PageRank, each page gets a trust value between 0 and 1.

9.9.4 Spam Mass

Goal : estimate the fraction of a page's PageRank that comes from spam pages.

Solutions :

1. Mark as spam all pages bellow the trust threshold.

   → Legitimate pages but with naturally low page rank (e.g. new pages) will be ignored as we look at the absolute value of the page rank.

2. With r_p the PageRank of page p and r_p^+ the PageRank of p with teleport into trusted pages only :

   → The fraction of a page's PageRank that comes from spam pages : $r_p^- = r_p - r_p^+$

   ⇒ Spam mass of $p = \frac{r_p^-}{r_p}$

10 | Information cascades

Following the crowd

People connected in a network will influence each other, this results in herding or information cascade.

Definition of an information cascade

An information cascade has the potential to occur when people make decisions sequentially, with later people watching the actions of earlier people and inferring what the earlier people know (occurs when people abandon their own information in favor of inferences based on other people's actions, it is not a mindless imitation).

There are two kinds of cascades :

1. Information cascade

   → It is inference that makes the cascade.

2. direct-benefit cascade

   → It is benefit to you that makes the cascade.

10.1 A simple herding experiment

Simple experiment with :

- A decision to be made.
- People make the decision sequentially and each observes choices made earlier.
- Each person has some private information to help them decide.
- Each person makes inferences about other's private information from their actions.

We use Bayes' Rule to justify the reasoning used in the herding experiment.

10.2 Bayes' Rule

10.2.1 General conclusions

1. It is very easy for a cascade to occur.
2. A cascade can lead to suboptimal outcomes.
3. Cascades can be fragile.

10.2.2 Conditional probability and Bayes' Rule

"What is the probability of X, given that I know Y and the previous person asserted Z ?"

- We compute the probabilities of events
  - An event is something that may occur, but only once. Given event A, it has a probability of occurring that we denote $Pr[A]$.
  - Whether or not an event occurs depends on many random previous occurrences. We consider a sample space where each point in the space consists of a particular realization for each of these outcomes.

  .

- From there, we can compute the probability of $Pr[A|B]$ in terms of $Pr[B|A]$:

$$Pr[A|B] = Pr[A] * \frac{Pr[B|A]}{Pr[B]} \tag{10.1}$$

10.3 Summary

- We study decision making in groups of people
  - We study sequential decision making, where each person (1) has some private information, and (2) sees the decisions made by earlier people (but not their private information)
  - Using Bayes' Rule on conditional probabilities, we define a formal model for sequential decision making
- We show that even if people decide rationally, this can go wrong
  - It leads to information cascades, where a small amount of information will force everybody's decision to be the same (and it is often incorrect)
  - Another approach, independent decision making, does not lead to information cascades, so introducing independence is one way to avoid cascades
- We apply this to realistic situations
  - Cascades often occur even when situations differ from the formal model
  - However, cascades are fragile and it is usually possible to break the cascade if additional information is introduced into the decision making
  - In later lectures we will refine our understanding of cascades by adding direct benefits and by studying the effect of network structure (neighbors)

11 | Direct-Benefit Cascades (Network Effects)

Sometimes there is an explicit benefit to aligning your behavior with that of over, in economics they are called positive externalities.
An externality is a situation where the welfare of an individual depends on the actions of other individuals, without mutually agreed-upon compensations.

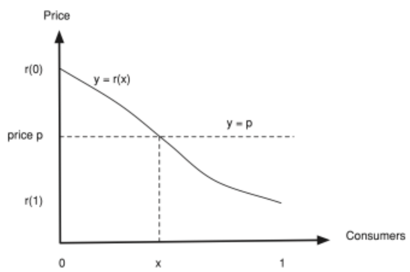
11.1 The Economy without Network Effects

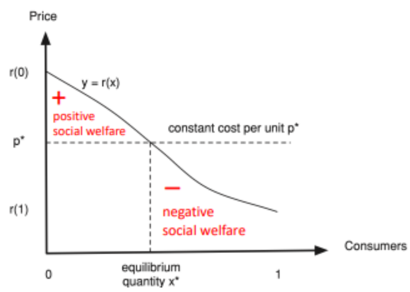
11.1.1 Studying network effects

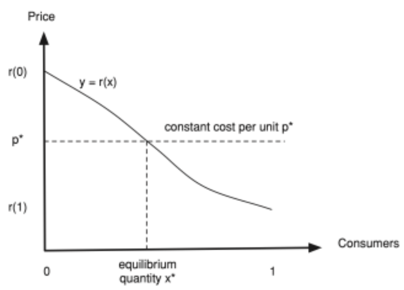
- All consumers are given names as real numbers from 0 to 1.
- Each consumer has a reservation price, which is the maximum amount he is willing to pay for one unit of the product.
  - Let $r(x)$ denote the reservation price of consumer x.
  - Reservation prices are arranged in decreasing order : if $r(x') > r(x)$, then $x' < x$.
- Suppose that the market price for a unit of the product is p (graphic 1).
  - If the price were r(0) or more, nobody would buy, conversely if the price were r(1) or less, everyone would buy

  $\rightarrow$ $r(1) < p < r(0)$.
- The equilibrium between price and demand is socially optimum (graphic 2).
  - With this, we can compute the social welfare (which for all consumers from 0 to x is the area between the curve $y = r(x)$ and the line $y = p*$.
- When the production cost is $p*$ per unit, the market price is $p*$ as it cannot be below $p*$ because of the production cost, nor above $p*$ because of competition (graphic 3).

  $\rightarrow$ The equilibrium of the product is $x*$, such as $r(x*) = p*$.







11.2 The Economy with Network Effects

Withe network effects, a potential consumer takes into account their own reservation price as well as the total number of users. This gives :

- $r(x)$: the intrinsic reservation price for consumer x.
- $f(z)$: the benefit to each consumer when fraction z has bought the product.

  $\rightarrow$ The actual reservation price of consumer x is $r(x)f(z)$, he buys if $r(x)f(z) \geq p*$.

11.2.1 Comparing the two models

Equilibrium price $p* =$

- $r(x*)$ in the model without network effects.
- $r(z)f(z)$ in the model with network effects.

11.2.2 Equilibria with network effects

We assume that all consumers expect that the fraction of the population buying the product is z, and they know what is the fraction actually buying.

→ This is a self-fulfilling expectations equilibrium as the fraction who actually purchase the product will be z.

- If everybody expects $z = 0$ to buy, then $r(x)f(0) = 0$ (and nobody will buy).

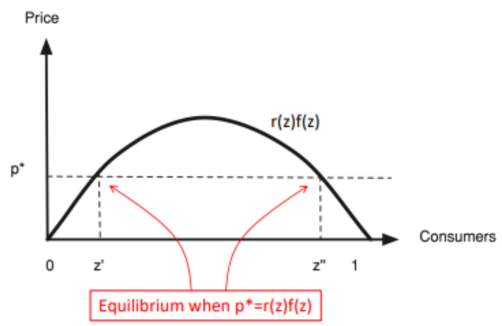
For $0 < z < 1$:

- The set of purchasers is exactly the consumers from 0 to z, and the price is the price of consumer z whose reservation price is $r(z)f(z)$.
- The equilibrium price is $p* = r(z)f(z)$.

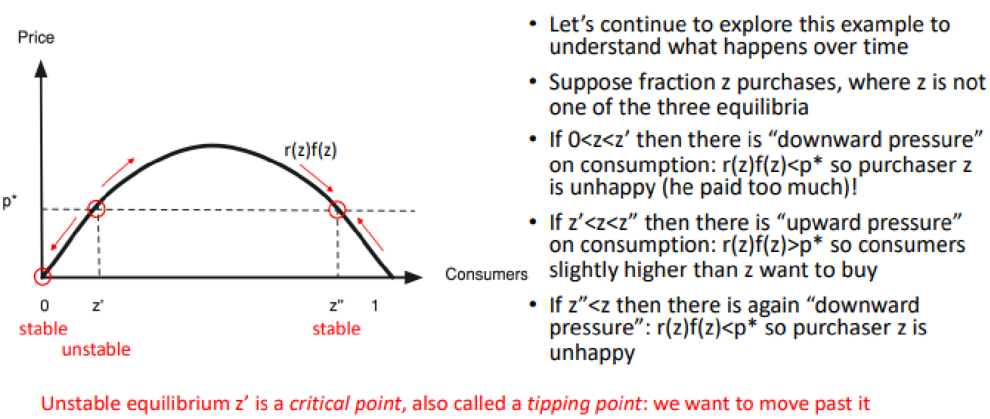
Example

For $r(x) = 1 - x$ and $f(z) = z$, $r(z)f(z) = z(1 - z)$. This means :

- If $p* > 1/4$, there are no equilibria other then for $z = 0$.
- If $0 < p* < 1/4$, there are two other equilibria : z' and z''



11.2.3 Stability, instability, and tipping points



11.2.4 Summary

- We introduced four models in steps to understand network effects
  - Economy without network effects (base case: "supply equals demand")
    - Purchasers do not take into account whether other users purchase or not
  - Economy with network effects (static view, self-fulfilling equilibrium)
    - The value of an item increases when there are more purchasers
  - Economy with network effects (dynamic view, with both stable and unstable equilibria)
    - Purchasers may have an incorrect belief of the total number of purchasers
    - This gives dynamic behavior with stable equilibria and unstable equilibria (tipping points)
  - Economy with network effects (individual effects plus network effects)
    - The product has value even for the first purchaser
    - This adds a new stable equilibrium close to zero, and sometimes it can "jump" to a high one
- We studied how the intuitions of these models apply in the real world
  - How companies succeed when their product requires a big user base
  - Remember this when you make your own company!

12 | Power Laws and Rich-Get-Richer Phenomena

- Web page popularity follows a power law
  - This is because Web page creation is correlated and not independent
  - The preferential attachment model explains the power law
- This is a special case of a rich-get-richer phenomenon
  - Popularity grows exponentially, so small differences get bigger
  - Initially, it is very unpredictable, subject to random fluctuations
  - When popularity is large it is more robust
- The long tail
  - Power laws mean that the popularity decreases very slowly
  - So a huge number of unpopular pages can have large aggregate popularity
  - Success of many Web sites depends on having a huge inventory

13 | Cascading Behavior in Networks

13.1 Diffusion in networks

The innovation diffuses in two waves :

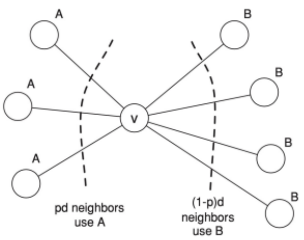
1. The first wave of person-to-person influence is informational.
2. The second wave of person-to-person influence is direct-benefit.

Homophily can act as a barrier to diffusion (homogeneous communities) as innovations tend to arrive from "outside" the system.

13.2 Modeling Diffusion through a Network

13.2.1 Defining a formal model

- A person's decisions are based on the decisions of their neighbors.
- A person has social network neighbors
  → The benefits of adopting a new behavior increase as more neighbors adopt it.
  → Simple self-interest dictates to adopt a new behavior as soon as a sufficiently high proportion of your neighbors adopted it.



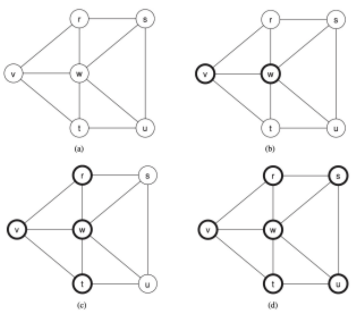
- Some of v's neighbors adopt A and some adopt B
  - How can v maximize its payoff?
- Assume v has d neighbors and fraction p have behavior A and fraction (1–p) have behavior B
  - If v chooses A, payoff is pda
  - If v chooses B, payoff is (1–p)db
- A is better choice if:

$$p \geq \frac{b}{a+b}$$

13.2.2 Cascading model

- All nodes initially use B as their behavior.
- A small set of initial adopters decide to use A.
- The remaining nodes evaluate their payoffs according to the coordination game.
- Some neighbors might decide to switch to A, and so forth.

Example



- Suppose the coordination game has a=3 and b=2 (A payoff is 1.5 times B payoff)
  - The threshold value q = b/(a+b) = 2/5
- Nodes v and w are the initial adopters
  - After one step, r and t will adopt A because 2/3 of their neighbors use A, and 2/3 > 2/5
  - After two steps, s and u will adopt A because 2/3 of their neighbors use A, and 2/3 > 2/5
  - It's a chain reaction: nodes converting will allow more nodes to convert, and so on

13.2.3 Complete cascade at threshold q

They are two possibilities :

1. The cascades runs for a while but then stops without flipping all nodes.
2. There is a complete cascade, where all nodes flip.

Consider a set of initial adopters with behavior A and all remaining nodes have behavior B. Nodes repeatedly evaluate the decision whether to switch from B to A using a threshold of q. If the resulting cascade of adoptions of A eventually causes all nodes to switch from B to A, then we say that the set of initial adopters causes a complete cascade at threshold q.

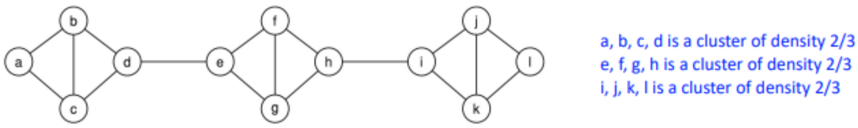
13.2.4 "Viral marketing"

- Tightly-knit communities can hinder the spread of an innovation.
- Techniques exists to push adoption beyond the barriers (raise quality to drop q, convince a small number of key people to switch).

13.3 Cascades and Clusters

13.3.1 Defining a cluster

A cluster of density p is a set of nodes such that each node in the set has at least a fraction p of its neighbors in the set.



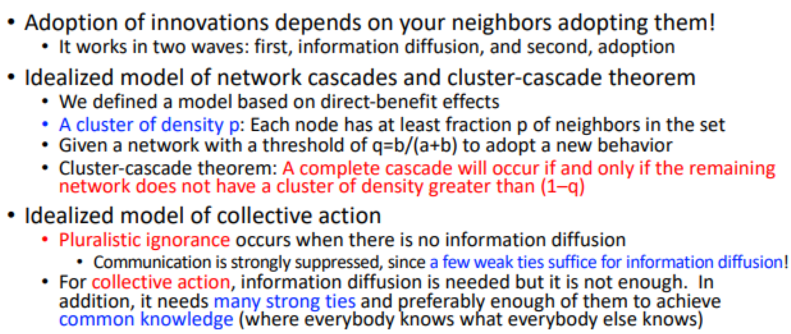
13.3.2 Cluster-cascade theorem

Consider a set of initial adopters of behavior A with a threshold of q for nodes in the remaining network to adopt behavior A. Then,

1. If the remaining network contains a cluster of density greater than $(1-q)$, the set of initial adopters will not cause a complete cascade.
2. Whenever a set of initial adopters does not cause a complete cascade with threshold q, then the remaining network must contain a cluster of density greater than $(1-q)$.

A complete cascade is blocked *if and only if* the remaining network contains a cluster of density greater than $(1-q)$.

13.3.3 Summary



14 | Convergence Toolkit

- **Social networks, graphs where nodes are humans (chapter 1-5)**
  - Closure: a social-affiliation network (which has both human nodes and center of interest nodes) tends to add links ("friends of friends become friends")
  - Structural balance: a friend/enemy network evolves toward balance: 2 enemy groups (if strongly balanced) or n enemy groups (if weakly balanced)
- **Games, where players make decisions to get payoff (chapter 6, 8)**
  - Games tend to converge somewhere between two extremes:
    * Nash equilibrium, which models a pure free market
    (which is based on reciprocal best response and indifference principle), and
    * Pareto optimality, which models government regulation
- **Auctions, where a group of players compete for an item (chapter 9)**
  - Traditional ascending-bid (like eBay) are second-price auctions, which tend to converge toward bidding your true value (which is a dominant strategy)
- **Markets, where a group of buyers and sellers interact (chapter 10)**
  - Market prices tend to converge toward market clearing ("supply equals demand"), and prices are a decentralized conflict resolution mechanism
- **Trading networks, adds traders as intermediate nodes (chapter 11)**
  - Economic networks with traders converge in two ways: monopoly maximizes profits to traders and competition maximizes profits to sellers and buyers
  - Increasing the network connectivity increases the overall benefit to society and reduces individual trader profits

- **Negotiation on networks, when nodes bargain (chapter 12)**
  - Negotiation power of a node depends on its position in the network; and negotiation converges according to the following three rules:
  - Stability: convergence tends to remove instabilities
  (instability = ability for a neighbor to sabotage an agreement)
  - Balance: convergence tends to a Nash bargaining solution for all nodes (which is a solution that is considered fair by its participants)
  - Avoiding extremes: humans avoid "all-or-nothing" division of benefits
- **PageRank algorithm, to determine quality of a Web page (chapter 13, 14)**
  - Page importance is modeled as fluid flow on the Web graph; always converging to equilibrium of flow and evaporation (equivalent to random walk with jumps)
- **Cascades, when each decision is based on others (chapter 16, 17, 19)**
  - Information cascades form easily: they form when decisions made sequentially tilt too strongly one way or another, which locks in all future decisions; on the other hand they are easily broken by bringing in new information
  - Direct-benefit cascades ("network effects") get benefits from a large community:
    * Tipping point (unstable equilibrium): below, fraction of users converges to zero, above, it converges to a high value. The goal is to get beyond it.
    * Lowering price: lowering the price can sometimes convert a tipping point into a narrow passageway that converges to a high value
    * Cluster: a tightly-knit community in a network can block network cascades; this can sometimes be avoided by convincing influenceable targets to switch
- **Power law of popularity, when network growth is "organic" (chapter 18)**
  - Preferential attachment: copying earlier decisions (such as links on older Web pages) converges to a power law and gives a significant long tail