Graph Theory with Applications

University of Houston, Spring 2023

MATH 4315 (Dr. Josic)

Contents

TEXT MOSTLY COPIED FROM THE BOOKS. I'll get to typing real notes eventually.

1 Week 3 reading

1.1 van Steen

1.1.1 Chapter 3.1

directed graph examples:

- streets, with one-way streets
- who knows whom, social networks
- wireless networks, messages can be sent from A to B but not necessarily B to A

Definition 1.1 (directed graph)**.**
A collection of vertices and arcs. Arcs go from u (tail) to v (head).

Definition 1.2 (underlying graph)**.**
Replace all the arcs with undirected edges.

Can transform an undirected graph into directed one by associating an orientation to each edge.

Definition 1.3 (neighbours)**.**

- in-neighbour: adjacent vertices having an arc with v as head.
- out-neighbour: adjacent vertices having an arc with v as tail.
- neighbours: union of in and out

Directed graph is **strict** if no loops and no two arcs with same endpoints have same orientation.

- out-degree
- in-degree

Example: limit the rate of incoming messages to ensure nodes are not overloaded.

Theorem 1.1.
sum of in-degree or sum of out-degree = number of arcs.

ADJACENCY MATRIX

- strict if all entries are 1 or 0
- in-degree is sum of column
- out-degree is sum of row
- not necessarily symmetric

CONNECTIVITY

- directed walk: vert, arc, vert arc
- directed trail: direct walk in which all arcs are distinct

- directed path: direct trail in which all vertices are distinct
- directed cycle: directed trail in which all vertices distinct except start and end.
- strongly connected: if exists directed path between every pair of distinct vertices.
- weakly connected: underlying graph is connected.

examples: need communication networks from any node to another, similar to transportation. need strongly connected network

breadth-first algorithm. find **reachable** (if exists directed path).

Theorem 1.2.
There exists an orientation D(G) for a connected undirected graph G that is strongly connected iff $\lambda(G) \geq 2$.

1.2 Easley, Kleinberg

1.2.1 Chapter 3

TRIADIC CLOSURE

- triadic closure: if two people have a friend in common, then they are likely to become friends as well.
- clustering coefficient: fraction of pairs of A's friends that are connected to each other by edges divided by the total number of nodes - 1

reasons for triadic closure:

- opportunity for B and C to meet
- exists a basis for trust
- incentive that A wants B and C to be gucci

Bearman and Moody: teenage girls with low clustering coefficient are more like to unalive themselves

Definition 1.4 (bridge)**.**
An edge that joins two nodes that if deleting the edge would cause the nodes to be in different components. (The only route between endpoints).

bridges are rare in reality.

Definition 1.5 (local bridge)**.**
Deleting a bridge that causes the distance to be > 2. The **span** of a local bridge is the distance its endpoints would be if the bridge was deleted.

an edge is a local bridge when it does not form the side of any triangle in the graph. Granovetter: people get their jobs from weak ties, maybe from local bridges because local bridges provide access to new information. Weak connections are valuable.

- classify edges as strong or weak
- strong tie: close friendships

- weak tie: not close friendships, acquaintances

STRONG TRIADIC CLOSURE PROPERTY Granovetter: A violates the strong triadic closure property if it has strong ties to B and C, but there is no edge at all between B and C.

If node A satisfies the **strong triadic closure property** and is involved in at least two strong ties, then and local bridge it is involved in must be a weak tie.

Onnela et al.:

- who talks to whom network through cell phone provider.
- Nodes: correspond to cell phone users, edge if they made phone calls to each other.
- Data exhibits giant component, a single connected component contained 84 percent of users.
- Strength of an edge: number of minutes in phone call. Sort all edges by tie strength, rank in percentiles.
- delete edges by strongest tie, giant component shrinks gradually
- delete edges by weakest tie, giant component shrinks rapidly, then broke

Definition 1.6 (neighborhood overlap)**.**
Edge connecting A to B. Ratio of (number of nodes who are neighbors of BOTH A and B)/(number of nodes who are neighbors of AT LEAST ONE of A or B). Do not count A or B. Local bridge has ratio 0.

Neighborhood overlap increase as tie strength increases.

Marlow et al.:

- link 1: mutual connection if they sent messages to each other
- link 2: one-way user sent, whether or not got one back
- link 3: maintained relationship if followed info from friend, whether or not communication took places
- passive engagement results in a highly connected network whereas if active, probably smaller

Example: high school friends vs currnet friends. mutual connection much sparser than maintained relationship or ALL

HUBERMAN ROMERO WU:

- following user: passive engagement, weak ties
- DM's, replies: active engagement, strong ties
- strong ties (friends with 2+ dms) vs followees
- strong ties stabilize at 50, while can have many followees
- each strong tie requires time and investment, so is small given limit by time. weak ties are easier created, not needed to be maintained

Definition 1.7 (embeddedness)**.**
embeddedness of an edge is the number of common neighbors shared by two endpoints. the numerator of neighborhood overlap.

Embeddedness: if two individuas are connected by an embedded edge, easier for them to trust each other. Social deterrence in harming friends.

Examples: individual success in a company depends on access to local bridges. Burt: conection to local bridges has advantages.

- early access to information in disparate parts
- amplifier of creativity, synthesis of multiple ideas
- gatekeeping: regulates access between two groups, source of power for node,

Structural hole: informal definition, represents empty space in netweork that don't interact closely.

COLEMAN, BOURDIEAU, BORGATTI, JONES, EVERETT:

- social capital (physical, human, economic, cultural).
- property of a group, property of an individual
- emphasize benefits of triadic closure and embedded edges: enforcement of norms, reputation, protection
- Burt: social capital as closure vs. brokerage. Closure in groups vs ability to broker interactions between groups

GRAPH PARTITIONING

WAYNE ZACAHARY karate club: dispute between president and instructor split the club into two rival clubs.

- divisive methods: removing spanning links, divides up network as they go
- agglomerative methods: start from bottom up, assemble chunks

GIRVAN NEWMAN

traffic: for every pair of nodes A and B in the graph connected by a part, imagine 1 unit of fluid flow along the adges from A to B. Flow divides itself evenly along all the possible shortest paths from A to B.

betweeness of an edge. the total amount of flow it carries. LINTON FREEMAN

- find the edges with the highest betweennness and delete them. this may cause the graph to partition, which whould be the first level of partition
- recalculate all betweenness, and remove the highest ones. repeat until desired resolution

ZACHARY

minimum-cut

- include tie strength
- delete edges of minimum total strength to separate two specific nodes.

GIRVAN NEWMAN and ZACHARAY accurately split this however did not do node 9 correctly.

LESKOVEC et al. partition only works with smaller networks, otherwise nodes become too highly connected to work.

COMPUTING BETWEENNESS

- Perform breadth-first search of the graph, starting at A
- determine number of shortest paths from A to each other node
- determine flow from A to all other nodes that use each edge.

Shortest path: sum of the number of shortest paths to all nodes directly above it in the breadth-first search.

Flow values:

- start at bottom node. A single unit of flow arrives.
- go up. add all the flows coming in, N. Look at the flows going out, and sum the number of shortest path for each node, M. Do N + 1, and divide by M, then scale for each shortest path number.

Total: sum up all betweenness for each edge. then divide by two for the double counting.

1.2.2 Chapter 4

- homophily: principle that we tend to be similar to our friends
- homophily as reason for triadic closure.
- homophily test: m probability of male, f probability of female. if the fraction of cross-gender edges is significantly less than $2mf$, then there is evidence for homophily (use statistical measures)
- an edge is **heterogenous** if connects two different nodes.

Example:
Moody: middle school and high school: division based on race and by age and school

- selecting friends as active choice or unconscouis choice
- immutable characteristics (race)
- mutable characteristics (behavior, activity, beliefs, opinions)
- people can modify mutable characteristics: social influence / socialization
- social influence as reverse of selection: links shape people characteristics, selection: people create links

COHEN KANDEL

- social influence is moderate, selection is comparable to social influence
- drug use: if homophily is from social influence, then drug intervention can work on a group. if homophily is from selection, then drug intervention can only work on a individual.

CHRISTAKIS FOWLER

- tracked obesity status and social network structure over 32 years
- obese and non obese people clustered in homophily
- significant evidence for obesity as a social influence, can spread through social network

Affiliation networks:

- bipartite graphs
- person connected to activity
- Example: board of directors of various companies
- pattern of participation in structure actrivities

Social-affiliation networks:

- two kinds of edges
- 1 for social, 1 for affiliation
- triadic closure: social (triadic closure), B-C to a focus (social link when focus in common) (focal closure), A-B to focus (friend joins focus cause A is already doing it like drugssss) (membership closure)
-

KOSSINETS WATTS steps to methodology

- take two snapshots
- for each k, identify all pairs who have exactly k friends in common in snap 1
- $T(k)$ be the fraction of the ppairs that have formed in the snap 2. This is probability.
- plot

Used email communication of 22k uundergraduate and graduate students over 1 year period.

Results: probability increases as more common friends, 1 to 2 more friends have even greater impact. Also studied by LESKOVEC et al. on Linkedin flickr delicious wahoo answers

focal closure through shared classes. DIMINISHING RETURNS asfter 3 classes.

membership closure through LIVEJOURNAL and WIKIPedia DIMISNISHINNG RETURNS again as connection to people increases

WIKIPedia ratio: number of articles edited by both / number of articles edited by at least one selection: rapid increase in similarity before first contact. social influence continues slower increase in similarity after first contact. Fastest increase before contact. LImitations: aggregate effect, very acerage study of population, not indiciative of the individual.

SEGREGATION MOBIUS AND ROSENBLAT: African American per city block in Chicago, homophily creates chops restaurants and business.

SCHELLING MODEL.

- global patterns of spatial segregation can arise from the effect of homophily
- agents (population of individuals), type X or type 0
- types represent immutable characteristic
- arrange in cells, with neighbors all around. 8 neighbors except for boundary
- each angent wants to have at least some other agents of its own type as neighbors.

- assume threshold t. if neighbors are fewer than t, then the agent wants to move
- agents move in a sequence of rounds: move each unsatisfied ONCE to a place where it will be satisfied. if any agent unsatisfied, then next sequence of rounds starts.
- similarity in 150 by 150 with 10,000 agents each type. Model produced large homogeneous regions. segregation.
- no one intends segregation, but cause segregation to grow at the expense of integrating ones.
- as homophily draws people together along immutable characteristics, there is a natural tendency for mutable characteristics to change in accordance with the network structure
- analogy to mixing liquids VINKOVIC KIRMAN

2 Week 4 reading

2.1 van Steen

2.1.1 Chapter 3.2

Weighted graph is a graph for which each edge e has an associated real-valued number $w(e)$ called its weight. For any subgraph $H \subseteq G$, the weight of H is the sume of the weights of its edges.

if u, v not adjacent, then $w(u, v) = \infty$.

geodesic. Shortest path that has the smallest weight among all paths between two nodes in G.

Dijkstra's algorithm

- Let $S_t(u)$ be set of vertices that has been found.
- assign to each vertex (L_1, L_2) where L_1 is the vertex connecting it, L_2 the total weight of the path from v_0.
- get the neighbor set.
- check which neighbor has the shortest path weight.
- add it to the list
- get the neighbor set. assign L-values.

Dijsktra's algorithm creates a **tree** that is **rooted** at v_0.

2.1.2 Chapter 3.3

Edge coloring motivation: consider devices that need to migrate data, device cannot be involved in more than one migration at once. What is the most efficient way to migrate data?

Let G be connected loopless graph. G is k-**edge colorable** if there exists a partitioning of $E(G)$ into k disjoint sets such that no two edges from the same E_i are incident with the same vertex.

THe minimal k for a graph is the the **edge chromatic number**, denoted by $\chi'(G)$.

For any simple graph, either $\chi'(G) = \Delta(G) or \Delta(G) + 1$. Δ is the maximal degree of a vertex in G.

Let G be simple connected graph. G is k-**vertex colorable** if there exists a partitioning of $V(G)$ into k disjoint sets such that no two vertices from the same V_i are incident with the same vertex.

The minimal k is the **chromatic number** of G, $\chi(G)$

Class -scheduling problem.

Chromatic number of a graph is difficult. No efficient solution exists.

For a simple connected graph, $\chi(G) \leq \Delta(G) + 1$.

For any planar graph G, $\chi(G) \leq 4$.

For any planar graph G, there is a vertex with $\delta(v) \leq 5$.

For any planar graph G, $\chi(G) \leq 5$.

2.2 Easley, Kleinberg

2.2.1 Chapter 5

Label a triangle edges with + or -. Heider, social pyschology. Looking at triangle, 4 distinct ways up to symmetry to label the three edges with + or -.

- three +: all friends, balanced.
- 1 + , 2 - : two friends, one mutual enemy. balanced.
- 2 + , 1 - : implicit forces trying to get B-C to be friends, or side with one vs the other
- 3 - : all enemies. forces motivates teaming up against another person.

Structual Balance Property For every set of three nodes in a complete graph, if three edges connecting them have 1 + or 3+, then structural balance.

if a complete graph can be divided into two sets of mutual friends, with complete mutual antagonism between the two sets, then it is balanced. This is the only way for a complete graph to be balanced, other than all edges positive.

Balanced Theoerem. By Harary. 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 X and Y, such that the pair of people in X likes each other and same with Y and everyone in X hates everyone in Y.

local property implies strong global property.

Proof

Pick all the friends of A and all the enemies. Separate. If BC (friends of A) friends check. If BC enemies, violates strucutral balance and also violates balance theorem.

If DE (enemies of A) friends, check. If DE enemies, then violates structural balance, violates balanced network.

If BD (1 friend, 1 enemy). If enemies, check. If friends, violates structura balance.

ANTAL KRAPIVSKY REDNER:

Random labeling, then repeatedly look for a triangle that is not balance and flip one of the edges to make it balanced. Dynamical process, models situation where people continuall reassess their likes and dislikes of other, also physical systems that reconfigure to minimize energy.

MOORE: invoked US support of Pakistan. US improving relations with China, China's enemy is India, India's enemy is Pakistan, so US support Pakistan. Then, North Vietnam friends with India, Pakistan nopt friends with Bangladesh.

ANATAL KRAPIPVSKY REDNER used shifting allianced in WWI.

trust and distrust in social ratings, social web sites

Weak structural balance property There is no set of three nodes such that the edges among them consist of 2 + and 1 - .

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

Proof. similar to balanced network proof.

Incomplete graphs: can have a +, -, or empty between any pair of nodes.

- a graph is balanced if you can fill in missing label edges in such a way that the signed complete graph is balanced.
- a graph is balanced if it is possible to divide the nodes into two sets, and mutual enemies between sets.

The two definitions are equivalent.

If the graph contains a cycle with an odd number of negative edges, then the graph is not balanced.

A signed graph is balanced if and only if there are no cycles with an odd number of negative edges.

Procedure

- identify supernodes: blobs are connected internally with positive edges and the only edges going between two supernodes are negative.
- perform breadth first search, check for odd cycles.
- if all the edges in the reduced graph connect to nodes in adjacent layers of the bread-first search, then there is a way to label the nodes in the reduced graph which provides a balanced graph. Otherwise, there is an edge connected two nodes in the same layer of the breadth first search, which there is an odd cycle in the original graph, so unbalanced

Let $0 \leq \varepsilon < \frac{1}{8}$ and $\delta = \varepsilon^{1/3}$. If at least $1 - \varepsilon$ of all triangles in a labeled complete graph are balanced, then either

- there is a set consisting of at least $1 - \delta$ of the nodes jin which at least $1 - \delta$ of all pairs are friends or else
- the nodes can be divided into two groups such that $1 - \delta$ satisfy the balanced theorem.

PROOF number of triangles in a complete graph: $\frac{1}{6}n(n-1)(n-2)$.

Assume at most an ε of all triangles are unbalanced. Define the weight of a node to be the number of unbalanced triangles it belongs to. We seek a node of low weight (node that is in relatively few unbalanced triangles.

Total weight of all nodes is at most $\frac{1}{2}\varepsilon(n(n-1)(n-2))$. Pick a GOOD NODE whose weight is at most $\frac{1}{2}\varepsilon(n-1)(n-2)) < \frac{\varepsilon}{2}n^2$. Split the graph according to the good node. all the friends of A in X and all the enemies in Y.

3 Week 5 reading

3.1 van Steen

3.1.1 Chapter 4.1

Definition 3.1.
A **tour** of a graph is a closed walk that traverses each edge in G. An **Euler tour** is a tour in which all edges are traversed exactly once.

Theorem 3.1.
A connected graph has an Euler tour if and only if it has no vertices of odd degree.

Theorem 3.2.
A connected graph has an Euler trail ($u \neq v$ if and only if it has exactly two vertices of odd degree. The trail originates and ends in the odd-degree vertices.

Definition 3.2.
Fleury's algorithm:

- choose an arbitrary v_0
- construct a trail. Choose an edge incident to v_k, but not in the trail. Make sure that the edge is not a cut edge of the induced subgraph $G - E(W_k)$, unless there is no other option.
- repeat step two until done.

Fleury's algorithm constructs an Euler tour if it exists.

Chinese postman problem: find a closed walk that covers all edges of G but with minimal weight. Applications:

- garbage collection
- mail delivery
- maintain links on a Web page

Solving the problem: transform a non-Eulerian graph to an Eulerian one by duplicating edges, minimally. Then apply Fleury's algorithm.

Special case: two vertices of odd degree: Use Dijkstra's algorithm to find a path with minimal weight, then duplicate edges in that path.

Postman problem

- for each pair of distinct odd-degree vertices, find a minimum weight path.
- construct a weighted complete graph on $2k$ vertices in which odd degree vertices are joind by an edge hjaving weight of the Path.
- fin the set of edges such that the sum of weights in minimal and no two edges are incident with the same vertex.
- for each edge connecting the odd-degree vertices, duplicate the edges in graph.

The reulting graph is Eulerian with minimal weight, then apply Fleury's algorithm to find a minimuim weight Euler tour.

4 Week 6 reading

4.1 Easley, Kleinberg

4.2 Chapter 6

interconnectedness at the level of the behavior, game theory. game theory is desinged to address outcomes that choices but dependent on choice made by people they interact with as well.

examples:

- soccer: penalty kicks

- competitive pricing
- bidding
- choosing a route
- international relations
- PEDs in sports
- evolutionary biology (success and failure of mutations)

game theory: decision-makers interact with one another, and satisfaction is dependent on everyone.
example: college exam vs presentation with partner.
game:

- set of players
- strategies on how to behave
- payoff from game: players prefer larger payoffs. use payoff matrix.

assumptions:

- everything the player cares about is summarized in payoff
- each player knows everything about the structure of the game
- rationality: each player wants to maximize payoff. each player also selects the most optimal strategy.

strictly dominant strategy: when a player has a strategy that is strictly better than all other options regardless of what the other player does. example: studying for exam is stricly dominant, so both study and each gets average of 88.

Prisoner's dilemma

- two suspects
- strategy: confess or not confess
- payoffs: if C/N-C, 0/-10. if C/C, -4/-4. if NC/NC, -1/-1

Confessing is a strictly dominant strategy. it is the best choice regardless of what the other player chooses.

Application to PEDs in sports, arms races. Best outcome is to use drugs, but harms yourself as well. Using drugs is the strictly dominant strategy but there is a better outcome for both of them if they don't play rationally. Similar to arms races in military.

Prisoner's Dilemma only shows up when the conditions are right. If the test is easier, then changes strategies and better outcomes.

best response: the best choice of one player given a belief about what the other player will do. S is strategy by player 1, T is strategy by player 2.

$$P_1(S,T) \geq P_1(S',T)$$

S is the optimal strategy, S' any other strategy. **strict best response**: strict inequality. **dominant strategy**: a strategy that is the best response to every strategy of the other **strictly dominant strategy**: strict best repsponse to every strategy of player 2.

Nash equilibrium When there is no dominant strategy. Suppose that player 1 chooses S and player 2 chooses T. (S, T) is a **Nash equilibrium** if S is a best response to T and vice versa.

Computing Nash equilibrium: check all the pairs or compute each players best response to each strategy of the other player and then find strategies that a mutual best response.

coordination game: two player's shared goal is to coordinate on the same strategy. There can be more than one Nash equilibria, so you need extra stuff to figure out how the players will go. For example, approaching cars in the US will choose the right side to avoice collision, but elsewhere, they might choose left instead.

Battle of the sexes: an unbalanced coordination game where one Nash equilibria favors one player, but being in equilibrium is better than not. Husband and wife need to choose one move, but one prefers rom-com and the other action movies. choosing one will favor the other.

Stag Hunt game: if two people work together, they can catch the stag/. if they work by themselves, they can catch only a hare. (Writings of Rousseau). If one tries stag and the other hare, difference in payoff! Same as unbalanced coordination, but if they miscoordinate, one who tried to get high-payoff gets penalized more.

Hawk Dove game goal is anticoordination. 1 player must go agressive, the other must go passive, otherwise if they go the same, then mutual destruction and harm. Game has two Nash equilibria, but without more information, can't tell who is going to do what.

No Nash equilibria: enlarge the set of strategies to include random actions.

attack defense game: matching pennies: show H or T. lose penny if match, wins if they don't match.

zero-sum games: games with the property that the payoffs of the players sum to zero in every outcome. games where players' interest are in direct conflict: connection to D-Day Normandy and corresponding German response. There are no pair of strategies that are best responses to each other.

Introduce probability: p, q probability of H, $1 - [p, 1 - q$ probability of T. **mixed strategies**: mixes the options of H and T.

Suppose Player 2 chooses H with q. and suppose Player one chooses pure strategies $(p = 1 or 0)$. Then if pure strategy H, then expected payoff to player 1 is $1 - 2q$. If pure strategy T, then expected payoff to player 2 is $2q - 1$.

Nash equilibrium: pair of strategies that is a best response to the other. in matching pennies, $1 - 2q = 2q - 1 \implies q = 1/2$

Result: Choice of $q = 1/2$ makes Player 1 indifferent to choosing H or T. This makes Player 1 nonexploitable.

mixed strategy: each player should randomize so as to make the opther player iindifferent between their two alternatives.

Nash's result: Every game with finite number of players and any finite number of strategies has at least one mixed-strategy equilibrium.

examples:

- active randomizxation: tennis (where to serve the ball), card player (call a bnluff?), rock paper scissors
- two species of animals engage in 1-1 attack defense games for food. Suppose each animal is hard wired to choose a pure strategy, but for the population as a whole, half of it chooses one the other half choose the other

Run pass game

- if def match off, then off gets 0.

- off runs when def pass, off gets 5
- if off pass when def run, then off gets 10

p probability offense pass, q probnability defense defends pass.

DEFENSE POV

- payoff when OFFENSE passes: $0q + 10(1 - q) = 10 - 10q$
- payoff when OFFENSE runs: $5q + 0(1 - q) = 5q$

set equal $\implies q = 2/3$

OFFENSE POV

- payoff when DEFENSE passes: $0p - 5(1 - p) = 5p - 5$
- payoff when DEFENSEE runs: $-10p + 0(1 - p) = -10p$

set equal $\implies p = 1/3$

pPENALTY -KCICK game

kicker can aim left or right, goalie can dive left or right. CAVEATS, most kickers are right footed, kickers still have good chance of getting in even if the goalie dives correctly. SAME premise as matching pennies.

data matches analysis of goalie and kicker indifference expected payoff calculations. wowee surprising

FINDING ALL NASH EQUILIBRIA

- check all 4 pure outcomes.
- look for mixing equilibria.

Pareto optimality. if there is no other choice of strategies in which all players receive payoffs at least as high, and at least one playerr receives a strictly higher payoff.

examples: a superior choice IF the players agree to a binding agreement. instead of thinking individually.

social optimality. if a choice of strategies mmaximized the sum of the player's payoffs.

Outcomes that are SOCIALLY OPTIMAL must ALSO be PARETO OPTIMAL

multiplayer game. Suppose a game has n players. Each player has a set of possible strategies. An outcome of the game is a choice of a strategy for each player. Each player has a payoff function, $P_i(S_1, \ldots, S_n)$. A strategy S_i is the **best response** by player i to a choice of strategies of the OTHER players if $P_i(S_1, S_2, \ldots, S_i, \ldots S_n) \geq P_i(S_1, S_2, \ldots, S_i', \ldots S_n)$ for ALL strategies S_i' available to player i. A $textbf{Nash equilibrium}$ if each strategy of an outcome it contains is a best response to all the others.

strictly dominated. if there is some other strategy available to the SAME player that produces a STRICTLY higher payoff in response to EVERY CHOICE of strategies by the OTHER players.

FACILITY LOCATIOn game

ITERATED DELETION OF STRICTLYT DOMINATED STRATEGIES: keep deleting the strategies that are strictly dominated to reduce the game to simpler things. Effective method of seaching for Nash equilibria.

PROOF THAT THE SET OF NASH EQUILIBRIA REMAINS THE SAME THROUGH ONE ROUND OF DELETION.

Weakly dominated strategies. replace the STRICT inequality with a greater than OR EQUAL to. But deletions of weakly dominated strategies can eliminate Nash equilibria.

DYNAMIC GAMES: games over time, where responses can change over time. use a GAME-TREE to represent extensive form. matrix is normal form, can translate but lose some properties of the game.

5 Week 7 reading

5.1 Easley, Kleinberg

5.1.1 Chapter 7

evolutiuonary game theory: no individual is overtly reasoning or even making explicit decisions. behaviors involve interaction of multiple organisms, success depends on its behavior with others. fitness of an individual evaluated in context of whole population, fittest genes get passed on

analogy:

- strategy is geneticially determind characteristics and behaviors
- fitness is payoff

EXAMPLE: BEETLES

- large beettle vs small beetle: large gets more ofood
- same size: equal portions of food
- in all cases large needs more food to be sustained

difference to game theory: BEETLE CANT CHOOSE SIZE

beetle's overall fitness is average fitness it experiences from its pairwise interactions. Overall fitness determines reproductive success.

evolutionarily stable if when the whole population is using this strategy, and small group of invaders using a different strategy will eventually die off over multiple generations.

fitness of an organism in a population is the expected payoff it recieves from an interaction with a random member of the population

strategy T **invades** strategy S at level x if for small $x > 0$, x fraction uses T and $1 - x$ uses S.

evolutionarily stable if there is a small positive number y such that when any other strategy T invades S at ANY level $x < y$, fitness of S is strictly greater than the fitness of an organism using T

Large beetle is evolutionarily stable in the game. Small beetles cannot drive out Large beetles, but Large beetles can beat out small any time.

Game is same as Prisoner's Dilemma, also arms race. Evolution by natural selection causes fitness to decrease, but doesn't violate because by decreasing fitness, the environment changes (natural selection assumes fixed environment) to a hostile one.

EXAMPLES

- height of neighboring tress. size game
- roots of plants. They will get equal, but devote more resources to win over the other. If a wall is placed, less competition so they are more reproductive success. If two different root development, then then size game.
- TURNER, CHAO. Virus Phage 6. Phage 2 is less effective by itself, but in presence of Phage 6, it is more fit. Prisoner's Dilemma, Phage 2 is evolutionarily stable.

General description: Symmetric game Let $x > 0$ small such that x uses T and $1 - x$ uses S

$$\begin{bmatrix} (a,a) & (b,c) \\ (c,b) & (d,d) \end{bmatrix}$$

S is evolutionarily stable if for all sufficiently small values of $x > 0$ if,

$$a(1-x)+bx > c(1-x)+dx \implies a > c \quad \text{or} \quad a = c \text{ and } b > d$$

$(S, S$ is NASH EQUILIBRIUM if $a \geq c$. Thus, if S is evolutionarily stable, then S, S is a Nash equilibrium.

MIXED STRATEGIES

- each individual hard-wired to play pure strategy
- each individual hard-wired to play a mixed strategy

Suppose Oragnism 1 uses mixed strated S with p and T with $1-p$, and Organism uses mixed strategy S with q.

Expected payoff for player 1:

$$V(p,q) = pqa + p(1-q)b + (1-p)qc + (1-p)(1-q)d$$

evolutionarily stable mixed strategy for a symmetric game p if exists small $y > 0$ such that when any other mixed strategy q invades p at any level $x < y$, fitness of p is strictly greater than fitness of q

For some y, for all $0 < x < y$, $q \neq p$,

$$(1-x)V(p,p) + xV(p,q) > (1-x)V(q,p) + x(q,q)$$

If p evolutionarily stable mixed strategy, the (p,p) is a mixed Nash equilibrium.

scenario: there is an undesirable, fitness lowering role in a population, but if some people don't do it, then the whole collective suffers

Difference between Prisoner's Dilemma and Hawk-Dove game:

- player can choose to be "helpful" or "selfish" in both games
- Prisoner's: payoff penalty from selfishness is not that bad, so equilibrium.
- Hawk-Dove: payoff penalty from selfishness is so bad that one has to avoid it.

6 Exam 1 reveiw

simple graph, adjacency matrix, connected components (equivalence relations), path, cycle, directed graph, graph diameter, breadth-first-search algorithm, degree of a node, graphicality, connected simple graph theorem, weighted graph, geodesic on weighted graph, dijkstra's algorithm, edge coloring, 4 colour theorem, Euler tours, Hamiltonian cycle, Euler tour proof, Fleury's algorithm, sum of in degree is the sum of out degrees is the number of arcs: directed graphs, clustering coefficient, bridges vs local bridge, strong triadic closure property, local bridge implies weak tie, betweenness, betweenness algorithm, homophily, structural balance, shelling model, game, nash equilibrium, mixed strategies, evolutionary stable strategies, strict nash inplies ess implies nash

7 Week 9 reading

7.1 Easley, Kleinberg

7.1.1 Chapter 10

bipartite matching problem EXAMPLE: assigning rooms to students, 1 student to a room, students get multiple choices of which room they want. graph is bipartite, two categories: list of rooms and list of students, can't connect between same type.

perfect matching Given an equal number of nodes on each side of a bipartite graph, a **perfect matching** is an assignment such that each node is connected by an edge to the node it is assigned to, and no two nodes are assigned to the same node. It is a choice of edges in a bipartite graph such that each node is the only one endpoint.

SHOWING NO PERFECT MATCHING EXISTS find a constricted set: S is constricted if $|S| \geq |N(S)|$

Matching theorem. If a bipartite graph with equal numbers of nodes on the left and right has no perfect matching, then it must contain a constricted set. (König, 1931. Hall, 1935)

Having a constricted set is the only obstacle to having a perfect matching.

with valuations: number that indicates how much they prefer. collection of individuals evaluating a collection of objects. The **quality** of an assignment is the sum (or maximum) of the individual's valuations.

optimal assignment: choice of assignemnt such that maximum quality, maximizes total happiness of everyone. Generalization of the perfect matching problem: if edge exists, valuation 1. if not, valuation 0.

EXAMPLE collection of sellers, collection of buyers. Each buyer has a valuation for each house. Valuation of buyer j for house i is v_{ij}. Suppose each seller i sells house at $p_i \geq 0$. Buyer's **payoff** is $v_{ij} - p_i$, buyer wants to maximize payoff. Rules: if quantity is maximized in tie, then buyer maximizes payoff by picking anything. If payoff is negative for every choice of seller, then no purchase and payoff is 0. Preferred seller of buyer is the seller that maximizes payoff. **market clearing**: set of prices is market clearing if preferred seller graph has perfect matching.

For any set of buyer valuations, there exists a set of market-clearing prices. For any set of market-clearing prices, a perfect matching in the resulting preferred-seller graph has the maximum total valuation of any assignment of sellers to buyers. A set of market-clearing prices and a perfect matcing in the preferred seller graph produces the maximum possible sum of payoffs to all sellers and buyers.

AUCTION PROCEDURE: Set of curent prices, smallest set at 0. Construct the preferred-seller graph and check for perfect matching. If no, find constricted set of buyers S and $N(S)$. Each seller in $N(S)$ raises price by 1. Reduce the prices so that the smallest price is 0. Next round of auction.

Proof: Potential of a buyer: maximum payoff. Potential of a seller: current price charged. Potential energy of a auction. sum of the potential of all participants. at each round, potential energy decreases by at least one since $|S| > |N(S)|$. POtential starts at P_0 and the auction must come to an end within P_0 steps.

Relation to single-item auction: introduce fake sellers, set price at 0. Do the usual auction procedure, but maintain fake sellers at 0. As long as buyers have real seller as the matching, then constricted set. Last buyer remaining gets the item, paying the second highest valuation.

PROOF OF MATCHING THEOREM: Outline:

- take assumed bipartite graph
- consider a matching that includes as many nodes as possible, maximum matching
- try to enlarge it by looking for a way to include one more node in the matching, but fail.
- when this fails, show that it produces a constricted set.

process:

- start with a matching that is NOT maximal.
- start at unmatched node
- look for a matching that would include it while still including current matching
- undo a matching that allows the unmatched to match AND lets the others be matching.
- matching has been enlarged.
- (follow a zig-zag path through graph, adding unused edges to matching while removing currently used edges from the matching)

alternating path: a simple path (no repeat of any nodes) that alternates between nonmatching and matching edges.

CLAIM: In a bipartite graphing with a matching, if there is an alternating path whose endpoints are unmatched nodes, then the matching can be enlarged.

augmenting path: an alternating path with unmatched endpoints.

SEARCHING FOR AUGMENTING PATH. (adapts the breadth-first search). GIven a bipartite graph with a matching.

1. start at any unmatched node (on the right).
2. do breadth first search. DON'T INCLUDE edges that would NOT make an augmenting (alternating path)

layers alternate between matched edges and unmatched edges. If the search EVER produces a layer that contains an unmatched node FROM THE LEFT-HAND SIDE, then an augmented path has been found.

Show that if alternating bread-first search fails, then there exists a constricted set

observations. Given an alternating breadth-first search that fails.

- odd layer: left side, even layer: right-side
- each odd layer contains same number of nodes as next layer.
- there are strictly more nodes in the even layers than the odd layers. (unmatched node is layer 0)
- every node in an even layer has all itsneighbors in the graph present

conclusion: set of nodes in all even layers in a failed alternating breadth-first search forms a constricted set. The neighbors of the nodes in the even layer are the nodes in the odd layers of the search, and we noted that there are strictly more nodes in the even layers than the odd layers. So $|S| > |N(S)|$. In other words, for any bipartite graph with a matching and take a node W unmatched. Then either there is a augmenting path beginning at W or there is a constricted set containing W.

Official proof of matching theorem: Consider a bipartite graph with equal number of nodes on left and right, with no perfect matching. Take a maximal matching on it. Since the matching is not perfect, there exists unmatched node from the right. There must be NO augmenting path from that node because the matching is already maximal. Since there is NO augmenting path from that node, performing alternating breadth-first search FAILS, and so a constricted set exists.

HOW TO COMPUTE A PERFECT MATCHING: Start with trivial matching. Then perform alteternating breadth first search to to enlarge the matching. Keep enlarging one by one until there is no perfect matching or there is one. DOES NOT GUARANTEE A MAXIMUM MATCHING

MAXIMUM MATCHING: if there is NO augmenting path beginning at ANY node on the right hand side, then the current matching has maximum size.

KOZEN: make all the unmatched nodes on the right constitute layer 0, then GO. If an unmatched node from the left is ever reached in some layer, then follow the path and an augmenting path is created.

7.1.2 Chapter 11

8 Week 10 reading

8.1 Easley, Kleinberg

8.1.1 Chapter 15

cost-per-click model. each slot has a specific clickthrough rate associated with it, which is the number of clicks per hour that an ad placed in that slot will receive. First, we assume that advertisers know the clickthrough rates. Second, we assume that the clickthrough rate depends only on the slot itself and not on the ad that is placed there. Third, we assume that the clickthrough rate of a slot also doesn't depend on the ads that are in other slots. I Now, from the advertisers' side, we assume that each advertiser has a revenue per click: the expected amount of revenue it receives per user who clicks on the ad. Here too, we will assume that this value is intrinsic to the advertiser and does not depend on what was being shown on the page when the user clicked on the ad will also assume that if an advertiser wants to be shown in a given slot, it wants to be shown there for every single search on the given query In the basic setup of a search engine's market for advertising, there are a certain number of advertising slots to be sold to a population of potential advertisers. Each slot has a clickthrough rate: the number of clicks per hour it will receive, with higher slots generally getting higher clickthrough rates. Each advertiser has a revenue per click, the amount of money it expects to receive, on average, each time a user clicks on one of its ads and arrives at its site.

clickthrough rate: r_i revenue per click: v_j benefit of advertiser j: $r_i v_j$

If there are more advertisers than slots, we simply create additional "fictitious" slots of clickthrough rate 0 (i.e., of valuation 0 to all buyers) until the number of slots is equal to the number of advertisers. The advertisers who are matched with the slots of clickthrough rate 0 are then simply the ones who don't get assigned a (real) slot for advertising. Similarly, if there are more slots than advertisers, we just create additional "fictitious" advertisers who have a valuation of 0 for all slots

regular matching market method, obtain market-clearing prices

Truthfull bidding Recall from Chapter 9 that when bidders are charged the full value of their bids, they will generally underreport, and this is what happened here. Bids were shaded downward, below their true values. More problematically, because the auctions were running continuously over time, advertisers constantly adjusted their bids by small increments to experiment with the outcome and to try to slightly outbid competitors. This resulted in a highly turbulent market and a huge resource expenditure on the part of both the advertisers and the search engines, as the constant price experimentation led to prices for most queries being updated extremely frequently.

truthful bidding is a dominant strategy for second-price auctions – it is at least as good as any other strategy, regardless of what the other participants are doing. This dominant-strategy result means that second-price auctions avoid many of the pathologies associated with more complex auctions. But what is the analogue of the second-price auction for advertising markets with multiple slots?

To put it another way, each individual is charged a price equal to the total amount everyone would be better off if this individual weren't there. We will refer to this as the Vickrey-Clarke-Groves (VCG) principle, after the work of Clarke and Groves, who generalized the central idea behind Vickrey's second-price auction for single items [112, 199, 400]. For matching markets, we will describe an application of this principle due to Herman Leonard [270] and Gabrielle Demange [128]; it develops a pricing mechanism in this context that causes buyers to reveal their valuations truthfully.

First, let S denote the set of sellers and B denote the set of buyers. Let V_S^B denote the maximum total valuation over all possible perfect matchings of sellers and buyers – simply the value of the socially optimal outcome with all buyers and sellers present. Total harm is the price to pay:

$$p_{ij} = V_{B-j}^{S} - V_{B-j}^{S-i}$$

VCG MECHANISM

1. Ask buyers to announce valuations for the items. (These announcements need not be truthful.)
2. Choose a socially optimal assignment of items to buyers – that is, a perfect matching that maximizes the total valuation of each buyer for what they get. This assignment is based on the announced valuations (since that's all we have access to.)
3. Charge each buyer the appropriate VCG price;

Essentially, what the auctioneer has done is to define a game that the buyers play. They must choose a strategy (a set of valuations to announce), and they receive a payoff – their valuation for the item they get, minus the price they pay. What turns out to be true, though it is far from obvious, is that this game has been designed to make truth-telling – in which a buyer announces her true valuations – a dominant strategy. We will prove this in the next section; but before this, we make a few observations

- The market-clearing prices defined there were posted prices, in that the seller simply announced a price and was willing to charge it to any buyer who was interested. The VCG prices here, on the other hand, are personalized prices: they depend on both the item being sold and the buyer to whom it is being sold
- The market-clearing prices in Chapter 10 were defined by a significant generalization of the ascending (English) auction: prices were raised step-by-step until each buyer favored a different item
- l generalization of the sealed-bid second-price auction. At a qualitative level, we can see that the "harm-done-to-others" principle is behind both the second-price auction and the VCG prices, but in fact we can also see fairly directly that the second-price auction is a special case of the VCG mechanism.

Claim: If items are assigned and prices computed according to the VCG mechanism, then truthfully announcing valuations is a dominant strategy for each buyer, and the resulting assignment maximizes the total valuation of any perfect matching of items and buyers.

PROOF If buyer j decides to lie about her valuations, then one of two things can happen: either this lie affects the item she gets, or it doesn't. If buyer j lies but still gets the same item i, then her payoff remains exactly the same,

$$\begin{aligned} v_{ij} - p_{ij} &\geq v_{hj} - p_{hj} \\ v_{ij} + V_{B-j}^{S-i} &\geq v_{hj} + V_{B-j}^{S-h} \\ v_{hj} + V_{B-j}^{S-h} &\leq V_B^S \end{aligned}$$

GENERALIZED SECOND-PRICE AUCTION is a generalization of the second-price auction for a single item. However, as we will see, GSP is a generalization only in a superficial sense, since it doesn't retain the nice properties of the second-price auction and VCG.

In the GSP procedure, each advertiser j announces a bid consisting of a single number b_j – the price it is willing to pay per click. it is up to the advertiser whether or not its bid is equal to its true valuation per click, v_j . Then, after each advertiser submits a bid, the GSP procedure awards each slot i to the i-th highest bidder, at a price per click equal to the $(i + 1)$-st highest bid. paying a price per click equal to the bid of the advertiser just below them.

First, we'll see that GSP has a number of pathologies that VCG was designed to avoid: truth-telling may not constitute a Nash equilibrium, there can in fact be multiple possible equilibria, and some of these may produce assignments of advertisers to slots that do not maximize total advertiser valuation. On the positive side, we show in the next section that there is always at least one Nash equilibrium set of bids for GSP, and that among the (possibly multiple) equilibria, there is always one that does maximize total advertiser valuation. The analysis leading to these positive results about equilibria builds directly on the market-clearing prices for the matching market of advertisers and slots, thus establishing a connection between GSP and market-clearing prices. Hence, while GSP possesses Nash equilibria, it lacks some of the main nice properties of the VCG mechanism from Sections 15.3 and 15.4.

AD QUALITY: add a factor of q_{ij}

Claim: In any matching market, the VCG prices form the unique set of market-clearing prices of minimum total sum

8.1.2 Chapter 16

As a first example, suppose that you are choosing a restaurant in an unfamiliar town, and based on your own research about restaurants, you intend to go to restaurant A. However, when you arrive you see that no one is eating in restaurant A, whereas restaurant B next door is nearly full. If you believe that other diners have tastes similar to yours, and that they too have some information about where to eat, it may be rational to join the crowd at B rather than to follow your own information. To see how this is possible, suppose that each diner has obtained independent but imperfect information about which of the two restaurants is better. Then if there are already many diners in restaurant B, the information that you can infer from their choices may be more powerful than your own private information, in which case it would in fact make sense for you to join them regardless of your own private information. In this case, we say that herding, or an information cascade, has occurred. This terminology, as well as this example, comes from the work of Banerjee [40]; the concept was also developed in other work around the same time by Bikhchandani, Hirshleifer, and Welch [59, 412].

Consider, for example, the following experiment performed by Milgram, Bickman, and Berkowitz in the 1960s [298]. The experimenters had groups of people, ranging in size from just one person to as many as fifteen people, stand on a street corner and stare up into the sky. They then observed how many passersby stopped and also looked up at the sky. They found that with only one person looking up, very few passersby stopped. If five people were staring up into the sky, then more passersby stopped, but most still ignored them. Finally, with fifteen people looking up, they found that 45 of passersby stopped and also stared up into the sky

) There is a decision to be made – for example, whether to adopt a new technology, wear a new style of clothing, eat in a new restaurant, or support a particular political position. (b) People make the decision sequentially, and each person can observe the choices made by those who acted earlier. (c) Each person has some private information that helps guide their decision. (d) A person can't directly observe the private information that other people know, but he or she can make inferences about this private information from what they do

9 Week 11 reading

9.1 Easley, Kleinberg

9.1.1 Chapter 17

Why would someone imitate decision of someone else? Direct-effect benefits, network effects. Example: adoption of technology.

externality: any situation in which welfare of an individual is affected by the actions of other individuals without mutual agreed compensation. **positive** when welfare increases. Externalities have to be uncompensated. Example: market with huge number of potential consumers, each consumer is small enough to not make huge impact on market. Small effect on aggregate. Consumers are $x \in [0, 1]$. Each consumer wants at most 1 unit of good. Willingness to pay:

- intrinsic interest
- the number of other people using the good: larger the using population, the more willing to pay

NO NETWORK EFFECTS Reservation of consumer x, $r : [0, 1] \to \mathbb{R}$, is inverse demand function, monotone decreasing. $x < y \implies r(x) > r(y)$. Market price: $r(0) < p < r(1)$. Intermediate value theorem: there exists x_0 such that $r(x_0) = p$. Everyone whose reservation price is p or greater will buy ($\{x \mid r(x) \geq p\} \implies [0, x_0]$). x_0 fraction of population buys it.

EQUILIBRIUM: Suppose good can be produce at constant p_0 per unit. Assume there are many potential producers so that none are large enough to be able to influence the market price of good. Assume $r(0) > p_0 > r(1)$. Supply: there exists unique x_0 such that $r(x_0) = p_0$. x_0 is the equilibrium quantity. If less than x_0 population purchase, then there are consumers who have an incentive to purchase, upward pressure on consumption. If more, then consumers who regret purchase, downward pressure on consumption. Equilibrium is socially optimal.

NETWORK EFFECTS: Katz, shapiro, varian. z fraction of population is using good, reservation price is $r(z) \cdot f(z)$, intrinsic interest r times the benefit to each consumer from having a fraction of population us is. $f(z)$ is increasing, controls how much more valuable a product is when more people are using it. Assumptions:

- $f(0) = 0$
- f continuous
- $r(1) = 0$

PERFECT CONSUMER PREDICTIONS: self fulfilling expectations equilibrium: if consumers all think z, then z will happen. $p_0 = r(z)f(z)$ EXAMPLE: $r(z) = 1 - z, f(z) = z$. Max at $z = 1/2, p(z) = 1/4$. If $p_0 > 1/4$, no equilibrium since no one buy. so only equilibrium is $z = 0$. If $p_0 < 1/4$, two solutions. Either little people, little confidence, or many people, much confidence in product.

10 Week 13 Reading

10.1 Easley, Kleinberg

10.1.1 Chapter 21

Branching processes start with tree, with a root, then root spreads to k people with probability p. Small p vanishes, large p continues and spreads to many people. Only two possibilities are to infect no one (dying after finite steps), infect every wave. **basic reproductive number**, denoted $R_0 = pk$ for branching process. If $R_0 < 1$, finite, dies out. If $R_0 > 1$, disease persists with positive probability, but not sure depending on p value. Critical point $R_0 = 1$. Small changes in p or k can wildly affect the disease spreading. Two public health measures:

- quarantine: reduce k
- better hygience: reduce p

SIR Epidemic model A node can be:

- susceptible: has not caught disease, can be infected
- infectious: caught disease, can spread to neighbors
- removed (recovered): after finishing up the infection, node is removed and can't spread disease, can't get disease.

Use digraph. Progress is controlled by p, probability of contagion and t_1 the duration of infection.

- some nodes are in I state
- each node that enters I state remains infectious for fixed number of steps
- during t_1 steps, node has probability p to pass to S neighbors.
- after t_1 steps, it is R state, inert.

SIR model is most appropriate for disease that each individual only catches once in their lifetime. Branching process is special case $t_1 = 1$.

Extensions of the SIR:

- assign different probabilities between two nodes based on closeness of contact
- separate I into stages and having different probabilities

Networks that don't have tree structure means R_0 dichotomy doesn't exist.

SIR and $t_1 = 1$ as percolation. Before start, for each node flip coin with probability p is open. Once all edges flipped, every node that can be reached will be infected.

SIS model. Same as SIR but no removed, only back to susceptible. Possibility of reinfection. Can be extended to include different probabilities.

Life cycle of disease. SIR definitely ends on a finite graph, burns through supply of nodes. SIS epidemic also ends on finite graph with probability 1, there will be some point no node has disease all at once. For SIS, can have critical probabilites depending on contact network. Can represent SIS as SIR by making a copy of all nodes at time $t + 1$, then looking at eges between them.

Oscillations in diseases Syphilis as example. Key ingredients:

- temporary immunity
- long range links in contact network

Long range links produce coordination in timing of flare-ups in dispersed parts, subside, and create temporary immunity. Long range contact: peak. Immunity: trough.

SIRS model Same as blah blah, but Recovery state is only temporary, for t_R steps. To produce large fluctations, need lots of long range connections. Need small world properties, long-range linkes make it possible for things that happen in one part of network to quickly affect another elsewhere.