14.15 Final Exam Cheat Sheet

Definitions, Graphs (L1-L2)

There are at most $\binom{n}{2}$ connections in a graph with n nodes and thus $2^{\binom{n}{2}}$ possible graphs.
Degree of seperation of a graph $\mathcal{G}$ with n nodes denoted d, is defined

$$d = \frac{\ln n}{\ln \lambda} \tag{1}$$

where $\lambda = \max_{g \in \mathcal{G}} |\mathcal{N}(\mathcal{G})|$.
Branching Tree Approximation: Assumption that neighbors aren't neighbors with each other (Imagines graph is a tree with no cycles)
Strength of Weak Ties: Idea that weak ties allow distant clusters of people to meet. Formalized through the **Strong-tie Tiadic Closure**:

$$(i,j),(j,k) \in E' \implies (i,k) \in E$$

where E' are the strong ties and E is the weak ties.

Graph Theory Terminology

A network/graph can also be represented by an **adjacency matrix G** where $g_{ij} \in \{0,1\}$ represents linkage.
Adjacency Matrices
From the adjacency matrix $\mathbf{G}$ we have that $\mathbf{G}_{ij}^k$ is the number of walks of length k between i and j (while for weighted graphs this is the sum of the values of all length k paths).
An undirected graph is **connected** if there is a path between all nodes. A **component** is a maximally connected subgraph of a graph. A graph can there be **partitioned** into its components, with the adjacency matrix (upon reordering) being written in block diagonal form.

Directed Graphs

For directed graphs, the graph is **strongly connected** if each node can reach any other through a *directed* path.
For a set $S \subset N$ the set of nodes T that can be reached from S are called the **out-component** of S while the set of nodes that can reach S are called the **in-component**. The strongly connected component of a node i is the intersction of the in and out component of i.

Summary Statistics (L1-L2)

Harder to analyze larger networks visually so we keep summary statistics, mainly:
a) Degree distribution/Density (eg. Poisson)
b) Diameter, Average Path length
c) Clustering
d) Centrality (importance of nodes)
e) Homophily (How are nodes of the same "type" linked)

Mean Degree, Density, Sparseness

The **mean degree** of a graph is the mean of the degrees of each node $d = \frac{1}{n}\sum_i d_i = \frac{2m}{n}$ (where $m = |\mathcal{E}|$).
The **density**, ρ is the fraction of all links that exist, ie.

$$\rho = \frac{m}{\binom{n}{2}} = \frac{\bar{d}}{n-1} \approx \frac{\bar{d}}{n}$$

A network is **sparse** if ρ is small, ie. $\lim_{n\to\infty} \rho_n = 0$.

Degree Distribution

The **degree distribution** is the distribution $p(\cdot)$ representing the fraction of nodes with degree d (or equivalently in Erdos-Renyi random graphs the distribution).

A graph is **d-regular** if all nodes have the same distribution $d = \frac{2m}{n}$ (so $p(\cdot)$ is the Dirac delta).

Many types of probability distributions:

Thin-Tailed: $p(d) \leq ce^{-\alpha d}$ for $\alpha, c > 0$. Tails fall off fast, so large degrees are unlikely.

Fat-Tailed: $p(d) = cd^{-\gamma}$ for constants $c, \gamma > 0$ (called a **power-law distribution**). Tails are *fat* so large degrees more likely than thin-tailed.

Diameter and Average Path Length

Average path length is given by $\sum_{i \neq j} l_{ij}/n(n-1)$.

The **diameter** is $d = \max_{(i,j)} d_{ij}$

We measure *clustering* in a graph $\mathcal{G}$ by the **overall clustering coefficient** (or **network transitivity**)

$$\mathrm{Cl}(\mathcal{G}) = \frac{\sum_{i \neq j \neq k} g_{ij} g_{ik} g_{jk}}{\sum_{i \neq j \neq k} g_{ij} g_{ik}} \tag{2}$$

The **individual cluster coefficient** of node i is given by

$$\mathrm{Cl}_i(\mathcal{G}) = \frac{\sum_{i \neq j \neq k} g_{ij} g_{ik} g_{jk}}{\sum_{i \neq j \neq k} g_{ij} g_{ik}} \tag{3}$$

where now the sum is over just j and k (abuse of notation).

The **average clustering coefficient** is given by

$$\mathrm{Cl}^{avg}(\mathcal{G}) = \frac{1}{n}\sum_i \mathrm{Cl}_i(\mathcal{G}) \tag{4}$$

Centrality/Closness

The **degree centrality** is simply $d/(n-1)$.

The **betweeness centrality** of a node k is given by

$$B_k = \sum_{(i,j) i \neq j, k \neq i,j} \frac{P_k(i,j)/P(i,j)}{\binom{n-1}{2}} \tag{5}$$

Closeness can be measured by either inverse average distance $(n-1)/\sum_{j \neq i} l_{ij}$ or **decay centrality**: $\sum_{j \neq i} \delta^{l_{ij}}$

Measures of centrality of nodes based on eigenvectors are called **eigenvector-based centrality**.

Homophily and Segregation

Homophily (or **Associative Matching**): extent to which nodes of same type are likely to be connected. Simple measure: fraction of links between individuals of di§erent types relative to uniform.

Eigenvector-based Centrality (L3)

Vanilla Eigenvector Centrality

We define the **eigenvector centrality** $\mathbf{c} = (c_i)$ of a graph G (written with adjacency matrix $\mathbf{g}$) to be a nonzero vector such that $\exists \lambda > 0$ such that

$$\lambda c_i = \sum_{j \neq i} g_{ji} c_j \quad \forall i \in N \Leftrightarrow \lambda \mathbf{c} = \mathbf{g}^T \mathbf{c} \tag{6}$$

(Note: g_{ji} is the links that *come to you*). Usually, we normalize by $\sum c_i = 1$.

A (directed) network is said to be **strongly connected** if there is a directed path between any two nodes. That is, $\exists l > 0$ such that $(g^l_{ij}) > 0$. The corresponding adjacency matrix is then said to be **irreducible**.

Theorem 1 (Perron-Frobenius)**.** *For any irreducible non-negative matrix* $\mathbf{g}$ *, its largest eigenvalue* λ_1 *is a positive real number, the components of the corresponding eigenvector* v_1 *are also all positive, and* v_1 *is the only non-negative eigenvector.*

Rate of convergence dictated by how fast $(\lambda_2/\lambda_1)^t \to 0$ as $t \to \infty$.

Katz-Bonacich Centrality and Leonteif Inverses

Unfortunately, if graph is not strongly connected, many nodes will have eigenvector centrality 0. This is partially resolved with the following fix: Given $\beta > 0$, the vector of **Katz-Bonacich Centralities with parameter** $1/\lambda$ is defined as the non-negative vector $\mathbf{c}$ such that

$$c_i = \frac{1}{\lambda} \sum_{j \neq i} g_{ji} c_j + \beta \quad \forall i \in N \Leftrightarrow \mathbf{c} = \frac{1}{\lambda} \mathbf{g}^T \mathbf{c} + \beta \mathbf{1} \tag{7}$$

The resulting solution is

$$\mathbf{c} = \beta \left(\mathbf{I} - \frac{1}{\lambda} \mathbf{g}^T \right)^{-1} \mathbf{1} = \beta \left(\mathbf{I} - \alpha \mathbf{g}^T \right)^{-1} \mathbf{1} = \beta \Lambda^T \mathbf{1} \tag{8}$$

where often we write $\alpha := 1/\lambda$ which we call the **decay parameter**. This is well defined iff $\alpha < 1/\lambda_1$.

We define the **Leonteif inverse of** $\mathbf{g}$ **with parameter** α to be

$$\Lambda = (\mathbf{I} - \alpha \mathbf{g})^{-1}$$

This implies the Katz-Bonacich centralities with $\beta = 1$ is simply $\Lambda \mathbf{1}$ where Λ is the Leontief inverse of $\mathbf{g}^T$. Intuitively, we have

$$\Lambda = \mathbf{I} + \alpha \mathbf{g} + \alpha^2 \mathbf{g}^2 + \cdots$$

which is the sum of the value of all length l weights discounted by α^l.

PageRank

Problem with Katz-Bonacich: weighs too highly the *incoming* influence from other pages (eg. Amazon sellers would be ranked too high!).

Pagerank modifies Katz-Bonacich by normalizing by out-degrees of each incoming node:

$$c_i = \alpha \sum_{j \neq i} \frac{g_{ji}}{d_j^{out}} c_j + 1 \Leftrightarrow \mathbf{c} = \alpha \mathbf{g}^T D^{-1} c + \mathbf{1} \tag{9}$$

where D is the diagonal matrix with entries $D_{ii} = \max\{d_i^{out}, 1\}$ which gives:

$$\mathbf{c} = \left(\mathbf{I} - \alpha \mathbf{g}^T D^{-1} \right)^{-1} \mathbf{1} \tag{10}$$

This is usually normalized as $\hat{\mathbf{c}} = \frac{1-\alpha}{n} \mathbf{c}$.

Intuitive interpretation of Pagerank: fraction of periods at which web surfer who randomly moves to a neighbor with probability α and jumps with probability $1 - \alpha$.

Iterative algorithm for Pagerank:
1. Initialize $c_i(0) = 1/n$ for all i.
2. Compute

$$c_i(t+1) = \frac{1-\alpha}{n} + \alpha \sum_{j \neq i} \frac{g_{ji}}{d_j^{out}} c_j(t) \tag{11}$$

Katz's First Measure

Dividing by out-degree but without "free parameter" β, we get **Katz's First Measure**:

$$\mathbf{c} = \mathbf{g}^T D^{-1} \mathbf{c}$$

Production Networks (L4)

Before: Robert Lucas (1977) argues that idiosyncratic shock effects dissipate due to law of large numbers. This, however, ignores *network* structure.

Leonteif Approach to Fluctuations

To study transmission structure: model production economy with n sectors in a **production network**.

The economy is described by a set of weights W and an efficiency parameter $\alpha < 1$. Sector i requires αw_{ij} units of every good j to produce one unit of i. The production of x_i is given by

$$x_i = \frac{1}{\alpha} \min \left\{ \frac{x_{i1}}{w_{i1}}, \cdots, \frac{x_{in}}{w_{in}} \right\} \tag{12}$$

This assumption that inputs are required in fixed quantities is called the **Leonteif production technology**.

To model the consumer demand, suppose the production levels of each sector i is given by $\mathbf{x} = [x_1, \cdots, x_n]^T$ and the consumer demand is given by $\mathbf{y} = [y_1, \cdots, y_n]^T$. We have the remaining (not used for other industries) supply of produce is given by

$$\mathbf{y} = \mathbf{x} - \alpha W^T \mathbf{x} = (\mathbf{I} - \alpha W^T) \mathbf{x} \tag{13}$$

For invertible $\mathbf{I} - \alpha W$ (ie. $\alpha < 1/\lambda_1$) the solution is given by the Leonteif inverse:

$$\mathbf{x} = \Lambda \mathbf{y} \tag{14}$$

Equal to Katz-Bonacich centrality when demand is $\mathbf{1}$! Thus, goods with high *in-centrality* produce a lot. Higher $\alpha \implies$ more demand, sum converges slower.

The Leontiff inverse (written as a geometric sequence) gives us the intuition: the effect of a demand shock on sector i on the production of sector j is the sum of the value of all walks (where each edge is discounted suitably) with the net effect on the total production

$$\Delta \text{GDP} = \sum_j \Delta x_j = \mathbf{1}^T (\mathbf{I} - \alpha W^T)^{-1} \mathbf{e}_i = (\Lambda \mathbf{1})^T \mathbf{e}_i$$

That is, the impact of sector i on GDP is equal precisely to i's Katz-Bonacich centrality in W.

Ideas: demand shocks with high "out-centrality" have biggest effect on GDP as shocks travel "upstream" and doesn't depend on where we start in terms of demand (initial $\mathbf{y}$).

Conclusions from Leonteif Input/Output Model

Some conclusions to draw:

- Most influential demand shocks are from those firms that supply many others as demand shocks travel "upstream".
- Supply shocks (ie. changes in $\mathbf{x}$) travel "downstream".
- Aggregate volatility from idiosyncratic shock effects does not neccesarily dissipate as $1/\sqrt{n}$, depending on network structure.

Moral of the story: indirect linkages are also important.

Degroot Learning Model (L5-L6)

Degroot Learning as a Model of Social Learning

Degroot Learning Model: Rule of thumb social learning model of how people in network update their opinions over time and reach consensus.

(a) Given a finite set of nodes N and discrete timesteps $t = 1, 2, \cdots$

(b) Nonnegative **Updating matrix** W, that is **row-stochastic**, ie. $\sum_{j=1}^{n} W_{ij} = 1, \quad \forall i \in \{1, \cdots, n\}$

(c) Each agent has **initial belief** $x_i(0) \in [0, 1]$

(d) **Degroot Update/Averaging Procedure**:

$$x_i(t+1) = \sum_{j=1}^{n} W_{ij} x_j(t) \tag{15}$$

or in matrix form, where $x(t) = [x_1(t), \cdots, x_n(t)]^T$

$$x(t) = W x(t-1) \tag{16}$$

That is, the beliefs are updated according to a Markov Chain with transition matrix W.

We say a matrix is **aperiodic** if the gcd of all directed cycles is 1. A sufficient condition is that of **self-loops**, ie. $\exists i$ such that $W_{ii} > 0$.

Recall all **ergodic** Markov Chains converge to a stationary distribution. For finite directed graphs, we have all *strongly connected*, *aperiodic* markov chains are ergodic, and will thus converge to a stationary distribution.

Lemma 2. *For a strongly connected network, convergence implies consensus.*

The **consensus** vector is thus defined $\lim_{t\to\infty} W^t x(0)$. These ideas are summarized by the following theorem:

Theorem 3. *If the directed network W is strongly connected and aperiodic, then $\lim_{t\to\infty} W^t = W^*$ exists. Moreover, the rows of the matrix W^* are all the same: ie.*

$$(W^* x)_i = (W^* x)_j \forall x \in \mathbb{R}^n, \quad \forall i, j \in [n]$$

A matrix W is said to be **primitive** if $\exists l > 0$ st W^l has all positive entries. It can be shown W is primitive iff it is strongly connected and aperiodic.

We have the **long run influence** s_i of the agent i is the ith element of any of the rows of W^*. Formally, we call the vector $\mathbf{s} = [s_1, \cdots, s_n]^T$ a **long-run influence vector** if $\sum s_i = 1$ and

$$x_1^* = \mathbf{s}^T x(0)$$

where x_1^* is the consensus.

Due to convergence, we get the following:

Theorem 4 (Long Run Influences are Right Eigenvectors)**.** *We have if* $\mathbf{s}$ *is the long-run influence vector,*

$$\mathbf{s}^T W = \mathbf{s}^T \Leftrightarrow W^T \mathbf{s} = \mathbf{s}$$

That is, an agent's long run influence is precisely her eigenvector centrality.

We can thus simply calculate the long run influence and beliefs by calculating the right eigenvectors of W and then calculating $\mathbf{s}^T x(0)$

Wisdom of the Crowd

Main idea: each initial estimate $x_i(0)$ is an iid, unbiased estimate of some true state of the world θ.

Suppose we have a growing network of size n. The long-run consensus is then $\mathbf{s}^T(n)\mathbf{x}(0)$. By law of large numbers, this sum converges in probability to θ as $n \to \infty$ iff

$$\lim_{n\to\infty} \max_{i\le n} s_i(n) = 0$$

Thus, a large network is wise iff each node's long run influence is small.

Equal Long-Run Influence

A nonnegative matrix W is called **doubly-stochastic** if both all rows *and columns* sum to 1.

Theorem 5. *In Degroot Learning, everyone has equal long-run influence (ie. each $s_i = 1/n$) iff the matrix W is doubly-stochastic.*

Generalizations of Degroot

Some generalizations of Degroot Learning:

Time-varying weight on Belief (DeMarzo, 2000): If agents realize they are learning over time, they can put more weight on own belief more:

$$x_i(t) = (1 - \lambda_t) x_i(t-1) + \lambda_t \sum_{j=1}^{n} W_{ij} x_j(t-1) \tag{17}$$

where λ_t decreases over time. Still converges if $\sum \lambda_t \to \infty$.

Ignoring People with Distant Beliefs (Krause, 2000): Update based only on neighbors with "similar" opinions:

$$x_i(t) = \frac{\sum_{j\in N_i : |x_i(t-1) - x_j(t-1)| < d} x_j(t-1)}{|\{j \in N_i : |x_i(t-1) - x_j(t-1)| < d\}|} \tag{18}$$

No Initial Opinions (Banerjee, 2021): Only subset $S \subset N$ starts with beliefs and use Voronoi sets to generalize.

Aggregation vs. Diffusion

Degroot learning is an example of **aggregation**: everyone starts with opinions and we study how they combine or get *aggregated*. In **diffusion** models of learning only a few people start with information and we ask how this spreads.

Static Random Graphs (L6-7)

For larger networks, it is usually not tractable to analyze the particular structure of individual networks. It can instead, be more useful to understand the statistical properties of how these networks are formed.

In L6-L7 we study **static networks** where all the network nodes are formed "at once". In L8 we study **dynamic networks** where nodes are formed "over time".

The Erdos-Renyi Model

Erdos-Renyi Random Graph: Either link formages as:

(a) Fixed number of links m and they are chosen randomly, *or*
(b) (Gilbert) Each link forms independently with some probability $p \in (0, 1)$

We focus on the second case. Based on p, we will often get very different properties of the graphs.

Node Degrees

We have since the links form independently, the expected number of links

$$\mathbb{E}[\# \text{ of links}] = \frac{n(n-1)}{2}p$$

and by the Weak Law of Large numbers, it concentrates around the mean.

Each node has **mean degree** $\lambda = (n-1)p \approx np$ and so the **density** is simply $\lambda/(n-1) = p$.

An ER graph is called **sparse** if the link probability $p \to 0$ as $n \to \infty$.

The degree distribution of each node follows a binomial distribution:

$$p(D = d) = \binom{n-1}{d} p^d (1-p)^{n-1-d} \tag{19}$$

For fixed λ (as $n \to \infty$), ie. also get that the degree distribution of each node is approximated roughly by the **Poisson Distribution**:

$$p(d) \approx \frac{e^{-\lambda}\lambda^d}{d!} \tag{20}$$

so the ER model is often also called the **Poisson Random Graph Model**. As the degree distribution falls off exponentially, the ER model generates a **thin-tailed** degree distribution (so almost all nodes will have similar degrees, with roughly no "hubs").
The probability a node is isolated is then simply

$$\mathbb{P}(D = 0) = e^{-\lambda}$$

Friendship Paradox: for large n, your friend will have degree approximately $\lambda + 1$.

Clustering

The probability that potential triangle becomes triangle is p so for sparse graph, expected clustering $\to 0$ as $n \to \infty$ so both overall and individual clustering $\to 0$.

Diameter/Average Path Length

We use the branching tree approximation (friends of friends aren't my friends). In a component of size k, the average path length and diameter are both $d = \frac{\log k}{\log \lambda}$. This is rather small and so the phenomenom is called **small worlds**.

Threshold Functions

For any property A, we say the function $t(n)$ is a **threshold function** if

$$P(A) \to 0 \quad \text{if} \quad \lim_{n\to\infty} \frac{p(n)}{t(n)} = 0$$

and

$$P(A) \to 1 \quad \text{if} \quad \lim_{n\to\infty} \frac{p(n)}{t(n)} = \infty$$

A property is **monotone** if (N, E) satisfies it, then so does any (N, E') where $E' \subset E$. If a threshold function exists, we say that a **phase transition** occurs at the threshold.

Some examples:

(a) $t(n) = 1/n^2$ is a phase transition for the property that the graph has an edge
(b) More generally, a threshold function for the emergence of a k-node tree is $1/n^{k/(k-1)}$.
(c) Threshold function for emergence of a cycle of length k is $t(n) = 1/n$ (for *any* k!)
(d) A **giant component** is a component that contains a positive fraction of all nodes in the network.

Component Structure

Depending on the growth of $p(n)$ we get different component structure:

(a) For $p(n) << 1/n$, all components will be small $O(\log(n))$
(b) For $p(n) >> 1/n$, a giant component (size $O(n)$) will emerge. That is, $t(n) = 1/n$ is a threshold for a giant component.
(c) (**Erdos-Renyi Theorem**) As $p(n)$ increases further, the component will increase in size until $p(n) \approx \log n/n$ at which point the graph is connected. That is, $t(n) = \log n/n$ is a threshold for connectivity.

Giant Components

As discussed before, the threshold for a giant component is given by $t(n) = 1/n$.

For $p(n) >> 1/n$ what is the size of the giant component? For fixed λ, as $n \to \infty$, we have the fraction of points in the giant component q satisfies:

$$q \approx 1 - e^{-\lambda q} \Leftrightarrow \lambda = -\frac{\log(1-q)}{q} \tag{21}$$

This gives us that:

1. $q = 0$ until $\lambda = 1$
2. q increases as a concave function of λ
3. $q \to 1$ as $\lambda \to \infty$

.

Configuration Models

The Erdos Renyi leads to a degree distribution following $\approx$ the Poisson Distribution (which has thin tails). This is not always realistic so we make some adjustements:

Configuration Model: specify the desired degree distribution in advance, then generate a random network with approximately this degree distribution. More specifically, start with a **degree sequence** $(d_1, \cdots, d_n)$ which specifies the desired degree for each node $i \in N$ (with $\sum d_i$ even). Then, enumerate a list consisting of each node i d_i times. Then, randomly pick two elements and form a "link" between them and delete those entries.

Degree and Excess Degree Distribution

As in ER, the clustering $\mathbb{E}[d]/n \to 0$ as $n \to \infty$.
Now recall, although a randomly chosen node will have degree $\langle d \rangle = \frac{\sum_{i \in N} d_i}{n}$, the degree of the node of a randomly chosen end of a randomly chosen link is

$$\sum_{i \in N} \left(\frac{d_i}{2m}\right) d_i = \frac{\sum_{i \in N} d_i^2}{2m} = \frac{n\mathbb{E}[d^2]}{n\mathbb{E}[d]} = \frac{\langle d^2 \rangle}{\langle d \rangle} = \langle d \rangle + \frac{\mathrm{Var}(d)}{\langle d \rangle} \quad (22)$$

where $\langle d \rangle = \frac{2m}{n}$.
In the configuration model, however, each stub connects to each of the other $2m-1$ other stubs with equal probability. This probability distribution is called the **excess degree distribution**.
Thus, we get the excess degree distribution is the same as the degree distribution for the configuration model. Namely, the excess degree distribution is given by

$$\frac{d}{2m} nP(d) = \frac{dP(d)}{\langle d \rangle}$$

Giant Components

For the configuration Model, under the branching tree assumption, we thus get from the excess degree distribution, the expected number of distance-k neighbors is

$$\langle d \rangle \left(\frac{\langle d^2 \rangle}{\langle d \rangle} - 1\right)^{k-1} \quad (23)$$

The value $\left(\frac{\langle d^2 \rangle}{\langle d \rangle} - 1\right)$ is called the **reproduction number**.

Molloy-Reed Criterion: In general, a giant component exists iff the reproduction number ≥ 1 implying divergence of the number of distance-k neighbors. This is only the case if

$$\frac{\langle d^2 \rangle}{\langle d \rangle} \geq 2 \quad (24)$$

The Morris-Contagion Model

If you immunize a fraction π of the population (they can no longer be in the giant component) then

$$R = \left(\frac{\langle d^2 \rangle}{d} - 1\right)(1 - \pi)$$

so a giant component emerges iff

$$\pi < \pi^* = 1 - \frac{\langle d \rangle}{\langle d^2 \rangle - \langle d \rangle} \quad (25)$$

π^* is called the **contagion threshold** because to prevent infection of a positive fraction of nodes, must remove π^* of them.

Small-World Models

ER has an *unrealistically* concentrated degree distribution with low clustering. Configuration model has unrealistically low clustering.
This motivates the **small world model** which has

- Realistic degree distribution
- Small diameter and average path length
- High clustering

More specifically, the **Watz-Strogatz Small-World Model** starts with a ring network of n nodes where each node is connected to its $2k$ closest neighbors (this has high clustering/diameter/patch length).
Then, randomly *rewire* a small fraction p of the nodes (as in ER). More specifically, the average path length $\approx \log n$, interval length $\approx 1/kp$ and diameter $\leq\approx \log n$. In this sense, it is a "linear interpolation" between a regular ring network and an ER graph.

Power Laws, Dynamic Network Formation and Preferential Attachment (L8)

Power Laws

We now focus on "dynamic" graph formation. Indeed, many application networks appear to have degree distribution that scales as a **power law distribution**, ie. $p(k) = k^{-\alpha}$ for α. This is controversial, because power laws have "fat tails".
More formally, a nonnegative RV X has a **power law distribution** if its cdf tail falls polynomially with power α, ie.

$$\mathbb{P}(X \geq x) \sim cx^{-\alpha}$$

For example, in a **Pareto Distribution** the cdf

$$\mathbb{P}(X \geq x) = \left(\frac{x}{t}\right)^{-\alpha}$$

satisfies a power law distribution. The density function is given by $p(x) = \alpha t^\alpha x^{-\alpha-1}$, with the density having infinite variance if $\alpha \leq 2$ and infinite mean if $\alpha \leq 1$.

Practically, how to test if it's a power law distribution? *Idea*: visualize on a log-log scale and check if it decays roughly linearly.

Some other power law terms:

- **Power Law**: $\mathbb{P}(X \geq x) \approx cx^{-\alpha}$
- **Scale-Free**: $P(cX) = f(c)\mathbb{P}(X)$ where f satisifes some power law. Scale-free everywhere usually satisfy a power law so we use the terms interchangeably.
- **Heavy-Tailed**: $\lim_{x\to\infty} e^{\lambda x}\mathbb{P}(X \geq x) \to \infty$ for all $\lambda > 0$, ie. doesn't die slower than exponential.
- **Fat Tailed**: tails die $\approx$ slower than exponential.

In summary, Power Law $\approx$ Scale Free $\subset$ heavy tailed $\subset$ fat tailed.

Uniform Attatchment Model

In an **uniform attatchment model**, nodes are born over time and form m links to existing nodes when born, uniformly at random. Older nodes will have higher degrees, but since random, only very old nodes will have substantially higher average degrees.

Fraction of nodes with degree $\geq d$ equals $e^{-(d-m)/m}$ so it has thin tails and does *not* generate a power law distribution.
Because the generation process is stochastic, we keep track of only the *expected* degrees rather than the realized properties which is an example of the **mean-field approximation**.

The rate of the degree growth over time is given by:

$$\frac{d}{dt}d_i(t) = \frac{m}{t}, \quad d_t(t) = m \implies d_i(t) = m + m\log\frac{t}{i} \quad (26)$$

To see that older nodes have higher expected degree, remark for any degree d and time t, if $i(d)$ is the node such that $d_{i(t)} = d$, the fraction of nodes have expected degree $\leq d$ is $F_t(d) = 1 - \frac{i(d)}{t}$

which gives us the fraction of nodes with expected degree less than d is

$$F_t(d) = 1 - e^{-(d-m)/m} \tag{27}$$

which is "thin-tailed" as in ER. In fact, it can be shown as $t \to \infty$, F_t is also the realized degree.

Preferential Attatchment

With dynamic graph formation there is the idea of **"rich gets richer"**: degree distribution depends heavily on initial conditions. This idea of the growth of the quantity depending on how much they already have is called **cumulative advantage** or **preferential attatchment** (Barabasi and Alber 1999).

Treatment is similar to uniform attatchment, but rather than linking randomly upon generation, new nodes link to pre-existing nodes with *probability proportional to their degrees.* Instead, existing node i forms a new link from a new node at time t is

$$m\frac{d_i(t)}{\sum_{j=1}^{t} d_j(t)} \tag{28}$$

Older nodes will then have much higher degree distribution, leading to a power law degree distribution. The differential equation for expected degree growth is given by:

$$\frac{d}{dt}d_i(t) = \frac{d_i(t)}{2t}, \quad d_t(t) = m \implies d_i(t) = m\left(\frac{t}{i}\right)^{1/2} \tag{29}$$

Now, the growth of older nodes grows as $\sqrt{t}$ rather than $\log t$, so older nodes have degree distribution that grows much quicker than in uniform attatchment. The cdf (fraction of nodes with degree greater than d) is given by:

$$\mathbb{P}_t(D \geq d) = \left(\frac{m}{d}\right)^2 \implies \mathbb{P}_t(d) = 2m^2d^{-3} \tag{30}$$

Pareto, Log Normal, Zipf, and Girat

Another heavy-tailed distribution is the **log-normal distribution**, ie. where the log of the random variable is normally distributed. Log-normal distributions are heavy-tailed but tails are thinner than polynomial.

By the CLT, the *geometric* mean converges to log normal so the growth rate converges to log-normal, which is called the **law of proportional effect** or **Gibrat's Law**.

On the other hand, **Zipf's Law** showed that the population of cities falls as $1/t$, so the population is better approximated by a Pareto Distribution.

How do we differentiate? The idea is if there are:

- A fixed number of growing cities with iid growth rates $\implies$ log-normal (Gibrat's Law)
- Growing number of growing cities with iid growth rates $\implies$ Pareto distribution, in a process called a **Kesten process**

Diffusion Through Networks and Society (L9-10)

We start with a class of diffusion models called **Compartmental Models** because at each point in time each individual is in one of several states or compartments, namely:

- **Susceptible (S)**: the individual has not yet been affected and is susceptible to infection
- **Infectious (I)**: Individual is currenly affected, and can pass the infection to others
- **Removed (R)**: Individual is no longer affected and is now immune.

We will cover three main types of Compartmental Models:

(a) **SI Model**: Once you're infected, you stay infectious forever. Thus, everyone gets infected eventually. The **Bass Model** generalizes, that allows "innovation" of disease in addition to infection.

(b) **SIR Model**: Once infected, infectious for a while, but then recover forever. Now, typically the reproduction number R_t drops below 1 so not everyone can be infected. The point at which this happens is called the **herd immmunity threshold**.

(c) **SIS Model**: Like SIR, but you can recover and then get infected again. Instead, a **steady state** is reached where $R_t = 1$.

The simplest type of diffusion models assume a **homogenous population** where everyone is equally likely to meet everybody else. More realistically, we instead assume **heterogenous contact rates** where some people are more likely to meet others.

Diffusion of Innovation

Suppose there is some innovation that spreads through a diffusion process. The **Bass Model** simply tracks the spread by spliting adoption based on the result of **innovation** or **imitation**. If $F(t)$ is the fraction of the population that has adopted the product, p is the **innovation rate** and q is the **imitation rate**, we get

$$F(t+1) - F(t) = (1 - F(t))(p + qF(t)) \tag{31}$$

We assume $p + q \leq 1$ to avoid double counting. The SI model is the case in which $p = 0$.
It is easier to analyze the continuous version:

$$\frac{dF}{dt} = (1 - F(t))(p + qF(t)) \tag{32}$$

which gives

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}} \tag{33}$$

The resulting F is called the **adoption curve** of the product. If $p > q$ the adoption curve is **concave**; if $p < q$ the adoption curve is**S-shaped.**

Concave or S-Shaped Adoption?

At high adoption levels, the curve will always be concave due to saturation, but whether the curve at low adoption levels will be concave or s-shaped depend on the simple relation between p and q. Namely,

- If q is higher, ie. the adoption rate is higher, this will speed up adoption so the curve will be S-shaped.
- If p is higher, higher adoption slows down adoption and there are fewer non-adopters to innovate.

SIR Model

For SIR Models, we have a susceptiblity fraction $S(t)$, infectious fraction $I(t)$, and a recovered fraction $R(t)$ where $S(t) + I(t) + R(t) = 1$. Two additional parameters:

- **Transmission Rate** β
- **Recovery Rate** γ

The dynamics of the SIR Model is modeled by three equations:

First, the **susceptible equation**, ie. the susceptible fraction $S(t)$ meets an infected person with probability $I(t)$ and gets infected with probability β:

$$\frac{dS}{dt} = -\beta S(t)I(t) = -\gamma R_0 S(t)I(t) \tag{34}$$

S(t) decreases over time (monotonically).

Second, the **infection equation**:

$$\frac{dI}{dt} = \beta S(t)I(t) - \gamma I(t) = \gamma R_0 S(t)I(t) - \gamma I(t) \tag{35}$$

where γ of the $I(t)$ recover.

Finally the recovery equation,

$$\frac{dR}{dt} = \gamma I(t) \tag{36}$$

On average, an infected person infects $R_0 = \beta/\gamma$ others before they recover so we call it the **basic reproduction number**.

The SIR model has no closed form solution but:

(a) The infectious share of the population is maximized when $\dot{I}(t) = 0$, ie. $R_0 S(t) = 1$ (which is called the **herd-immunity threshold**).
(b) R(t) keeps increasing with $\lim_{t\to\infty} R(t) = 1 - \lim_{t\to\infty} S(t)$ (since $\lim_{t\to\infty} I(t) = 0$ since we have a SIR model).
(c) The difference $R(\infty) - R(t)$ where $R_0 S(t) = 1$ is called the degree of **overshooting**.

What Percentage of the Population Get Sick?

What percentage of the population get sick? Through some algebra we can get:

$$R(\infty) = 1 - e^{-R_0 R(\infty)} \tag{37}$$

This is precisely the expected number of elements in the giant component of an Erdos-Renyi graph! Thus, by Molloy-Reed, if $R_0 \le 1$ then $R(\infty) = 0$, while it increases rapidly as a concave function of R_0.

Identically, the SIR model can be seen as drawing a link from i to j if "when i gets sick, i will infect j before i recovers (if j isn't sick already)" with the link probability being R_0/n. This viewpoint in terms of random graphs is called **percolation**.

Key Lessons From the SIR Model

We can learn some main takeaways from the SIR Model:

- I(t) grows exponentially when $S(t) \approx 1$ and falls exponentially when $S(t) \approx 0$ so (roughly) normal
- What is the value of lockdowns (ie. setting $\beta = 0$)? It is impossible to avoid reaching the herd immunity threshold, but overshooting can be reduced by imposing lockdown as soon as herd immunity is reached.
- What about vaccinations? Vaccinating a fraction π of the population reduces $S(0)$ to $1 - \pi$ which can greatly reduce $R(\infty)$.

Heterogeneous-Agent SIR and Homophily

The SIR equations can be extended to the heteogeneous case, where each each person can have a different "degree" d. Each meeting is then with a degree d individual with probability $dP(d)/\langle d\rangle$. This is precisely equivalent to the setup of the configuration model and $R(\infty)$ is the size of the GC of the configuration model.

Let $\tilde{P}(d) = \frac{P(d)d}{\langle d\rangle}$ be the excess degree distribution. Let q be the probability the branching process does not die out starting from an arbitrary node. We then get the recursion:

$$1 - q = \sum_{d=0}^{\infty} P(d)(1-q)^d \tag{38}$$

The fraction of nodes in the giant component of the configuration model then converges to q. Although this does not have a closed form, we can determine when a giant component exists (doesn't die out), for which the disease takes over a fraction of the population: this holds iff the Molloy-Reed criterion $\frac{\langle d^2\rangle}{\langle d\rangle} > 2$ which means more variacne in diesease makes it easier to invade. In the SIR Model (ER Network) this occurs iff $R_0 > 1$ while for a regular network this occurs if $R_0 > 2$. For scale free $(0 \le \gamma \le 3)$, a GC exists even if R_0 is arbitrarily small.

Can introduce **homophily** by making agents more likely to meet agents of their own type, ie. p of the meetings are with those of same degree. You can show as you increase p, the threshold value of R_0 for disease to invade decreases in p, so with more homophily the disease invades easier.

SIS Model

In the SIS model, the individual becomes susceptible again after recovering. The equations are given by:

$$\dot{S}(t) = \gamma I(t) - \beta S(t)I(t) \tag{39}$$
$$\dot{I}(t) = \beta S(t)I(t) - \gamma I(t) \tag{40}$$

where β is the transmission rate, γ is the recovery rate (and $S(t) + I(t) = 1$). With $R_0 := \beta/\gamma$ being the reproduction rate, we can rewrite these as:

$$\dot{S}(t) = \gamma I(t) - \gamma R_0 S(t)I(t) \tag{41}$$
$$\dot{I}(t) = \gamma R_0 S(t)I(t) - \gamma I(t) \tag{42}$$

For an SIS model, the **steady-state infection level** I is given by the stationary point of S:

$$\dot{S} = \gamma I - \gamma R_0(1-I)I = 0 \implies I = \begin{cases} 1 - \frac{1}{R_0} & \text{if } R_0 \ge 1 \\ 0 & \text{if } R_0 < 1 \end{cases} \tag{43}$$

Heterogeneous-Agent SIS

For the heterogeneous case, let $I_d(t)$ be the share of degree d nodes infected at time t. There is a positive study-state infection level iff

$$\frac{\langle d^2\rangle}{\langle d\rangle} > 1 \tag{44}$$

To see this, note if $I_d(t)$ is the fraction of degree d nodes that is infected, the share of **infected people** is given by $\sum_d P(d)I_d(t)$. Each **meeting** is with an infected invididual with probability:

$$\theta(t) = \frac{\sum_d dP(d)I_d(t)}{\langle d\rangle} \tag{45}$$

which is a good measure of the **infection level**. At the steady state, we will have

$$I_d = \frac{d\theta}{d\theta + 1} \tag{46}$$

with θ satisfying

$$\theta = \sum_d \frac{d^2 \theta P(d)}{\langle d \rangle (d\theta + 1)} \tag{47}$$

That is, there is a positive steady-state infection level iff

$$H(\theta) = \sum_d \frac{d^2 \theta P(d)}{\langle d \rangle (d\theta + 1)}$$

has a non-zero fixed point. One can find through differentiation that this holds iff $\langle d^2 \rangle / \langle d \rangle > 1$.

Behaviorial SIR Models

In practice, people take actions to prevent and/or reduce infection so there is a behaviorial dynamic to SIR models. For high $I(t)$, for example, people may pay a "cost" (eg. going out less) to lower the transmission rate β.

More specifcially, suppose at each point an individual can pay a cost **c** (called the action **vigilance**) to eliminate their risk of getting infected. Suppose individuals percieve the **harm** they suffer if they get infected as h. It is worth it to pay the vigillance cost if

$$c < \beta I(t) h \leftrightarrow I(t) > \frac{c}{\beta h} = I^* \tag{48}$$

If $I(t) < I^*$, no one is vigilant, while if $I(t) > I^*$, everyone is vigillant. If $I(t) = I^*$ people are indifferent so we get what we called a **mixed-strategy equillibrium**.

Thus, in practice, the epidemic will follow three phases:

1. **Rising Phase**: Initially, $I(t) \approx 0$ so no one is vigilant. Then, $I(t)$ starts rising so the epidemic proceeds as in SIR.
2. **Plateu Phase**: Once $I(t)$ hits I^*, it remains exactly so as otherwise $I(t)$ can't rise above I^* (so we would eventually get herd immunity) or everyone becomes vigillant.
3. **Declining Phase**: Once herd immunity is reached, everyone stops being vigillant and $I(t)$ falls back to 0 in the standard SIR model.

Seeding and Contagion (L11)

Whereas the simplest models of diffusion through networks are "mechanical" (works without any outside intervention), but in practice interventions and strategic decisions play key roles.

Optimal Seeding

For some types of technologies, eg. spread of beneficial new technologies, we want to help spread diffusion. This problem is called **optimal seeding**, ie. which nodes to "seed" first to maximize spread of diffusion.

The diffusion model takes the following form:

- Each node is either **informed** on **uninformed**, and also either **participates** or **does not participate**.
- Each informed node that participates tells each neighbor about the program with probability α_P. Each informed node that does not participate tells its neighbors with probability α_N.
- Newly informed node participates with probability $p + \lambda f$. p is called the **baseline participation probability** and λ measures the **endorsement effect**.
- The process repeats T times.

The **BSS** experiment finds $\alpha_P \approx 0.35$, $\alpha_N \approx 0.05$, with $\lambda \approx 0$.

Diffusion Centrality

How do you measure which nodes are important/central (which leads to high importance in optimal seeding)? **Diffusion Centrality** is a relevant centrality measure (Banerjee), defined by

$$\left[\sum_{t=1}^{T} (\alpha g)^T \right] \cdot \mathbf{1} \tag{49}$$

where here, $\alpha = \alpha_P = \alpha_N$ for simplicity. The diffusion centrality represents the expected total number of times that someone hears the information passing from node i.

For $T = 1$, this is just the degree centrality (times α). As $T \to \infty$, either it converges to a scaled Katz-Bonacich centrality (if $\alpha < 1/\lambda_1$) or eigenvector centrality (if $\alpha > 1/\lambda_1$). Banerjee find this is a better measure of centrality in the diffusion case. Indeed, they find that by asking villagers who the "gossips" (highly influential, central people) are, they find they are more important in eigenvector and diffusion centrality measures.

Complex Contagions: Do Seeds Really Matter?

If central nodes get infected anyway, does seeding matter? Akbarpour, Malladi, and Saberi (2023) show that for a simple diffusion model, for a large enough graph, optimal seeding does not really matter. That is, random seeding works quite well in standard diffusion model. Some caveats, however:

(a) This only holds for *simple* diffusion models, where each infected node infects susceptible neighbors with equal probability α.

(b) Does matter under **complex contagion**, where must be infected by *several* neighbors to become infected, ie. number of infected neighbors must reach a **threshold**.

Intuitively, we find that in fact some idea of complex contagion goes on in things like more advanced technologies. Indeed, in many social situations, especially those with a "coordination" aspect, an individual wants to change their behavior iff **sufficiently many** of their neighbors do to.

Threshold Models

A good model for complex contagion is thus the **threshold model** or **threshold contagion**. Centola and Macy (2007) study it in the context of the Wattz-Strogatz small world model and find that rewiring helps simple contagion but only helps in complex contagion up to a point, as they need close-knit groups to get complex contagion started.

One such example is the **Morris Contagion Model** with a graph $G = (N, E)$, with each agent(node) taking an action $a \in \{0, 1\}$ iff more than a fraction $q \in (0, 1)$ of the neighbors take action 1 (with tiebreaks, eg. exactly fraction q being broken arbitrarily). Intuitively, q measures the "inherent quality" of action 0 relative to action 1.

What are the **equilbria**/study states of the model? To analyze this, we start by defining **cohesiveness**.

For any $p \in [0, 1]$ we say a set $S \subset N$ is **p-cohesive** if, for each $i \in S$, at least a fraction p of i's neighbors are also in S. That is,

$$\min_{i \in S} \frac{|N_i(G) \cap S|}{|N_i(G)|} \geq p \tag{50}$$

Intuitively, there can be equlbria if society contains highly cohesive groups that can "coordinate" on different norms. Indeed, we get the following theorem:

Theorem 6 (Cohesiveness Equality)**.** *For any $S \neq \emptyset$, N, it is an equlibrium for everyone in S to choose action 1 and everyone else to choose action 0 iff S is q-cohesive and $N \backslash S$ is $(1-q)$-cohesive.*

The interpretation is that the more cohesive sets there are in a network, there are is more scope for diverse patterns in the behavior. This tells us when an equilibrium *occurs* but does not specify when exactly this may happen. This is the question of **contagion**.

Contagion

Suppose we start by seeding some subset of nodes $S \subset N$ and letting all those nodes have $a = 1$ and all other nodes have $a = 0$. We now repeat the diffusion process, with the condition that the diffusion is **progressive** (ie. we don't allow switching back from $a = 1$ to $a = 0$). We say there is a **contagion from** S if this process results in $a = 1$ taking over.

Theorem 7 (Contagion Theorem)**.** *There is a contagion from S iff, for any superset $S \subset S'$ the set $N \backslash S'$ is not $(1-q)$-cohesive.*

This theorem tells us that even for superior innovations (low q), they can not be spread if the network is highly cohesive. Conversely, inferior innovactions can spread if the network is seeded well. This has some limits as we will see soon:

Consider the extension to an infinite graph, where we define **contagion from** S means that given a finite initially infected set S, adoption grows without bound. A striking result says that an inferior innovation ($q < 1/2$) can *never* spread.

Theorem 8. *For any infiite graph where each node has finite degree, if $q > 1/2$ then, for any finite seed set S, contagion does not spread (infinitely far) from S. That is, it can spread for a while but not forever.*

The proof follows from the fact that the **interface**, ie the number of edges connected to one infected and one non-infected node decreases over time for inferior q.

Strategic Network Formation (L12)

Some networks are better described as being formed *strategically*, rather than randomly. There can be both strategic **1-sided linking** and **2-sided linking**, where linking decisions are either unilateral or bilateral, resulting in different structures.

The framework is given as follows: there are n agents in the network corresponding to the n nodes with each node i recieving a **payoff** as a function of the final network

$$u_i = \mathcal{G} \to \mathbb{R}$$

We take these as a given and study how a network forms in anticipation.

Efficiency and Utility Functions

A network G is **Pareto efficient** if there is no network G' such that $u_i(G') \geq u_i(G)$ for all i with strict inequality for *some* i.

A network G is **utalitarian efficient** if there is no network G' such that $\sum_{i \in N} u_i(G') \geq \sum_{i \in N} u_i(G)$.
A utilitarian efficient network is Pareto efficient but not the other way around.

One such distance-based utility function is given by:

$$u_i(G) = \sum_{j \neq i} b(l_{ij}(G)) - d_i(G)c \tag{51}$$

where l_{ij} is the distance, d_i is the out-degree, c is the cost of maintaining a link, and $b : \mathbb{N} \to \mathbb{R}$ is a strictly decreasing benefit function depending on the distance with $b(\infty) = 0$. This models someone who benefits from being connected (esp. at short distances) but dislikes maintaining the links themselves. For this, the utalitarian efficent network is simply a star network leading to the **law of the few**.

2-sided Linking

An undirected network G is **pairwise stable** if

(a) For any $(i, j) \in G$ we have $u_i(G) \geq u_i(G \backslash \{i, j\})$
(b) For any $(i, j) \notin G$, if $u_i(G \cup \{i, j\}) > u_i(G)$ than $u_j(G \cup \{i, j\}) < u_j(G)$

That is, no agent can gain unilaterally by cutting a link and no pair of agents can gain strictly by adding a link.

The star network is optimal for the distance-based utility function for two sided linking:

Theorem 9. *In the distance-based utility model with 2-sided linking, the unique utilitarian efficient network is*

1. *The complete network if $c < b(1) - b(2)$*
2. *An undirected star, if $b(1) - b(2) < c < b(1) + \frac{n-2}{2}$*
3. *The empty network, if $c > b(1) + \frac{n-2}{2} b(2)$*

1-sided Linking

A directed network G is a **Nash Equilibrium** if $u_i(G) \geq u_i(G')$ for any network G' where $g_{jk} = g'_{jk}$. That is, each agent chooses her set of out-links optimally taking all the other out-links as a given.

For one-sided linking, if paying the cost c results in two one-sided links, we get theorem 9 but with $c/2$. Otherwise, we can also get a *cycle* is optimal.

Pairwise Stable Networks

Although the star is efficient, it is not always stable:

Theorem 10. *In the distance-based utility model with two-sided linking, the following hold:*

(a) If $c < b(1) - b(2)$, the unique pairwise stable network is the complete network. (Recall that for these parameters, it is also efficient.)
(b) If $b(1) - b(2) < c < b(1)$, the star is pairwise stable. (For some parameters, there are also other pairwise stable networks.)
(c) If $c > b(1)$, then in any pairwise stable network every player has either no links or at least two links. (In particular, the star is not pairwise stable.)

Couauthor Model

Distance-based utility model captures settings where link-formation comes with externalities. As before, if there is positive gain from linking, we say there are **positive externalities**. In other settings, more links may cause **negative externalities**.

For example, consider the **coauthor model** where

$$u_i(G) = \sum_{j \in N_i} \Big(\frac{1}{d_i(G)} + \frac{1}{d_j(G)} + \frac{1}{d_i(G)\, d_j(G)} \Big) \tag{52}$$

Intuitively, it is best for each author to only have one coauthor:

Theorem 11. *In the coauthor model, if n is even, the unique utilitarian efficient network consists of $n/2$ disjoint pairs.*

The efficient network however is not stable. Stable networks instead consist of disjoint cliques of very di§erent sizes:

Theorem 12. *In the coauthor model, if $n \geq 4$, any pairwise stable network is inefficient and can be partitioned into fully connected components, each with a different number of members. Furthermore, if m is the size of one such component and m' is the size of the next-largest component, then $m' \geq m^2 - 1$.*

Game Theory (L13)

We call multi-person decision problems, **games**. The formal analysis of these sorts of problems is called **game theory**.

More specifically, a **game** consists of

- a set of **players** $N = \{1, \cdots, n\}$
- a **strategy set** S_i for each $i \in N$
- a **payoff function** $u_i : S_1 \times \cdots \times S_n \to \mathbb{R}$ for each player $i \in N$.
- A vector $(s_1, \cdots, s_n) \in S := S_1 \times \cdots \times S_n$ a **strategy profile**.

We write s_{-i} to denote a vector of strategies for all players except i (ie. strategies of all of **i's opponents**). Goal is for each player to maximize their *own* payoff $u(s_i, s_{-i})$.

Static Games of Complete Information

We consider **static games of complete information**, where the game is played all at once, not over time, and there is no uncertainty in the game.

In 2 player games we can represent the game as a **payoff matrix**. In the **prisoner's dilemna**, the two players get 2 for both cooperating, 1 for both defecting, and 3 if one defects and 0 for the other when they cooperate.

How should a player choose their strategy? This is called a **solution concept**. A strategy $s_i \in S_i$ for player i is called **strictly dominant** if for every alternative strategy $s_i' \in S_i$, we have $\forall s_{-i} \in S_{-i}$, $u_i(s_i, s_{-i}) > u(s_i', s_{-i})$. In prisoner's dilemma, both defecting is optimal.

Pure Strategy Nash Equilibrium

In most games, players will have rather than a strictly dominant strategy, a **Nash Equilibrium**; each player plays optimally, *taken as given* what everyone else is doing.

Given opponent's strategy s_{-i}, a **best response** for player i is an optimal action against s_{-i}: $u_i(s_i, s_{-i}) \geq u(s_i', s_{-i}), \quad \forall s_i' \in S_i$.

A **pure strategy Nash Equilibrium (PSNE)** is a strategy profile $s^* \in S$ such that

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*) \quad \forall s_i \in S_i, i \in N \tag{53}$$

Intuitively, each player is playing a best response. If every player has a strictly dominant strategy, then the unique PSNE will be playing this strictly dominant strategy.

Mixed Strategy Nash Equilibrium

A **mixed-strategy** σ_i for each player i is a probability distribution over S_i. Denote by Σ_i the set of mixed-strategies of player i and let $\Sigma = \Sigma_1 \times \cdots \Sigma_n$. A player i's **payoff** for a mixed strategy profile $\sigma = (\sigma_1, \cdots, \sigma_n)$ is her expected payoff under independent randomization:

$$u_i(\sigma) = \sum_{s \in S} \mathbb{P}_\sigma(s) u_i(s) \tag{54}$$

$$= \sum_{(s_1, \cdots, s_n) \in S} \left(\prod_{j=1}^{n} \sigma_j(s_j) \right) u_i(s_1, \cdots, s_n) \tag{55}$$

A **mixed-strategy Nash Equilibrium (MSNE)** is an optimal mixed-strategy profile, ie. a $\sigma^* \in \Sigma$ such that

$$u_i(\sigma_i^*, \sigma_{-i}^*) \geq u_i(\sigma_i^*, \sigma_{-i}^*) \quad \forall \sigma_i \in \Sigma_i, i \in N \tag{56}$$

There can be infinitely many mixed strategies. The following theorem allows us to check which of these are Nash Equilibria by looking only at the pure strategies within the support:

Theorem 13. *σ^* is a NE iff*

$$u_i(\sigma_i^*, \sigma_{-i}^*) \geq u_i(s_i, \sigma_{-i}^*) \quad \forall s_i \in S_i, i \in N \tag{57}$$

Existance of Nash Equilibria

A game is **finite** if every player's strategy set S_i is finite. We have the following key theorem:

Theorem 14. *Every finite game has a Nash Equilibrium (including mixed-strategy NE)*

Network Traffic/Routing Games (L14-15)

Network Traffic is a type of game on a network where multiple individuals need to get from some point A to some other point B on a network while minimizing their own travel times. This will lead to Nash Equilibria which can be **socially efficient** or inefficient (maximizing total utility/lowering average delay).

We say that an **equilibrium routing** is a routing pattern that results from each agent choosing the route that minimizes her own delay. Equilibrium routing does not always correspond to social efficiency since agents don't take into account **negative extranality** imposed on other agents by their actions.

We define this concept conceptually as follows: in a game with negative payoffs, the **price of anarchy** (POA) is the ratio of the total cost borne by all agents at the worst equilibrium to the total cost of the social optimum. The corresponding cost at the best equilibium is called the **price of stability** (POS) and

$$1 \leq POS \leq POA \tag{58}$$

General Traffic Model

Consider a simple routing game where mass is normalized to 1 and there is one starting point or origin point and one destination point. Let P denote the set of paths from the origin to destination and let x_p be the **flow** on the path $p \in P$ (the share of mass that takes path P). Each $i \in E$ has a **latency function** $l_i(x_i)$, where x_i is the total flow on link i given by

$$x_i = \sum_{p \in P : i \in P} x_p \tag{59}$$

A **routing pattern** x is a probability distribution on P. The **total delay** of a routing pattern x is

$$C(x) = \sum_{i \in E} x_i l_i(x_i) = \sum_{p \in P} x_p \sum_{i \in P} l_i(x_i) \tag{60}$$

Socially Optimal and Equilibrium Routing

A routing pattern x is **socially optimal** if it is a solution to the problem

$$\min_x \sum_{i \in E} x_i l_i(x_i) \tag{61}$$

subject to

$$\begin{cases} x_i = \sum_{p \in P : i \in p} x_p & \forall i \in E \\ \sum_{p \in P} x_p = 1, & x_p \geq 0 \quad \forall p \in P \end{cases}$$

Alternatively, a routing pattern x is an **equilibrium** if for any path $p \in P$ with $x_p > 0$ there does not exist a path $p' \in P$ such that

$$\sum_{i \in p'} l_i(x_i) < \sum_{i \in p} l_i(x_i) \tag{62}$$

That is, for any path with nonzero probability mass, there is no other path that is unilaterally better.

In the context of non-atomic routing (continuum of players) this Nash Equilibrium is called **Wardrop Equilibrium**.

Inefficiency of Equilibrium Routing

One can show that equilirbium routingg can be arbitrarily inefficient, in fact with POA $\approx \infty$ (eg. take a route with latency x^k and another with 1).

Improving Efficiency in Routing Networks

How can one reduce traffic in a routing network? Two contenders are to build new links and increase capacity, or to add congestion pricing (eg. tolls) to incentivize offsetting externalities. We focus on the second case.

Can reducing a latency function l_i of a link ever increase socially optimal delay? No, we can retain the same pattern. Adding a link always decreases optimal delay as well. Adding a new link *can* however increase *equilibrium* delay. This is called **Braess's Paradox**, which shows that closing routes can reduce commuting time, even if the number of commuters does not fall.

Congestion Pricing

How can we tackle this? We can add **congestion pricing**, ie. a *toll* to some routes. In general, we can set the toll on a link i equal to the externality of using link i, evaluated at the social optimum x^*:

$$t_i = x_i^* l_i'(x_i^*) \tag{63}$$

which is called a **Pigouvian tax** (or **Piguouvian toll**). This gives a socially optimal pattern as given by the following theorem:

Theorem 15. *With Pigouvian tolls, the socially optimal routing pattern x is also an equilibrium routing pattern.*

This is the idea behind setting taxes to equal externalities to increase efficiencies.

Routing Games as Potential Games

A **potential game** is one in which there exists a function $\phi : S \to \mathbb{R}$ called a **potential function** such that, for any player i and two strategies s_i, s_i' switching from s_i to s_i' has the same effect on player i's payoff as it has on the potential. The **maxima** of the potential function corresponds to the **equilibria** of a game.

Formally, a **potential function** is a function $\phi : S \to \mathbb{R}$ such that $\forall i \in N$, $s_i, s_i' \in S_i$ and $s_{-i} \in S_{-i}$:

$$u_i(s_i', s_{-i}) - u_i(s_i, s_{-i}) = \phi(s_i', s_{-i}) - \phi(s_i, s_{-i}) \tag{64}$$

Intuitively, this potential function reflects *all* player's incentives simultaneously. Some examples of a potential game is a **common interest game**, where all players have the same payoff $u : S \to \mathbb{R}$.

Potential games have some key properties, one of which is the existence of PSNEs:

Theorem 16. *Every finite potential game has a Pure Strategy Nash Equilibrium (PSNE). Conversely, every PSNE is a local maximum or saddle point of ϕ.*

The key result is that all routing games are in fact potential games which implies the existence of PSNEs:

Theorem 17. *Every routing game is a (convex) potential game and therefore has a unique PSNE. In particular, the potential function is given by*

$$\phi(x) = \sum_{j \in E} \sum_{i=1}^{x_j} l_j(i) \tag{65}$$

Potential Games and the Price of Stability

We showed for routing games with arbitrary latencies the state of anarchy can be arbitrarily large. For affine functions, however, the situation changes. A latency function $l_j(x_j)$ is **affine** if it can be written as $l_j(x_j) = a_j x_j + b_j$ for constants $a_j, b_j \geq 0$.

Theorem 18. *In any routing game with affine latency functions, $POS \leq 2$*

Network Effects (L15-16)

A social or economic environment exhibits **network effects** when each individual's optimal choice depends on some average of other choices. There are **local network effects** if this depends only on the neighbors of the node.

Strategic Complements

A simple example of a market without network effects is when there are a large number of individuals i who choose between products $s_i \in \{0, 1\}$ with a taste parameter $x_i \in \mathbb{R}^+$ which is the preference of 1 over 0 which has pdf $F(x_i)$.

Theorem 19. *In the unique equilibrium, fraction $S = 1 - F(c)$ of the population chooses $s_i = 1$ (ie. only people who have preference c or higher choose 1).*

Now suppose we consider network effects. More specifically, suppose the utility of using a new product is better if more people use it. Such a property is called **positive consumption externality**. That is, now the preference is given by $h(S)$ where

$h : [0,1] \to \mathbb{R}$ is a continuous, increasing function. Equilibrium is reached when

$$S = 1 - F\left(\frac{c}{h(S)}\right) := D(S) \tag{66}$$

for which a solution always exists by the IVT. There can be, however, multiple equilibria, which is given by the following:

Theorem 20. *Assume that $c > 0$ and $h(0) = 0$ (so that $S = 0$ is always an equilibrium). There are multiple equilibria if and only there exists $S > 0$ such that $S \leq D(S)$*

Welfare Comparisons for Multiple Equilibria

If there are multiple equilibria in a game with strategic complements and positive externalities, the equilibria can be **Pareto ranked**:

Theorem 21. *In our model of a market with network effects, if S and S' are both equilibrium levels of consumption of product 1 and $S > S'$, then S Pareto-dominates S'*

Comparitive Statics

How do changes in parameters affect behavior in the model? This is the problem of **comparative statics** which is difficult when there are multiple equilibria. For strategic complements (network effects) we focus only on greatest and smallest equilibria.

Theorem 22 (Topkis Theorem)**.** *In our model of a market with network effects, suppose that x_i increases for a positive fraction of agents and remains the same for everyone else. Then the largest and smallest equilibrium values of S both weakly increase (i.e. either increase or stay the same).*

The intuition is that increasing x_i has both a **direct effect** in increasing S and correspondingly since $h(S)$ increases more people adopt (**indirect effects**).

Stability and Tipping

What equilibria are likely to emerge when there are multiple equilibria? It is useful to think about the *dynamics* of the market to reach equilibrium. When $D(S) < S$, the share of people taking action 1 will decrease while the opposite occurs when $D(S) > S$. This process is called **best-response dynamics**.

From this, we see that equilbria such that $D(S)$ crosses S from *below* is **stable** and the ones that cross from below are unstable, called **tipping points**.

Theorem 23 (Samuelson Correspondence Principle)**.** *In our model of a market with network effects, the largest and smallest equilibrium values of S are both stable. Intermediate equilibrium values may be unstable.*

Residential Choice

We now study *local* network effects thorugh a specific example: residential choice (where to live). We first focus on residential choice with "neighborhood effects".

Residential Choice with Neighborhood Effects (**Becker-Murphy model**): Suppose there are two types of people $i \in \{0, 1\}$ (low skill and high skill) and two neighborhoods $j \in \{0, 1\}$ (low amenities, high amenities) with 50% of each type.

- Suppose the utility of a type-i person living in neighborhood j with share x is given by $u_i(x, j)$.
- Everyone benefits from high-skill neighbors, ie. $\frac{\partial u_i}{\partial x} > 0$ (called **neighborhood effect**)
- high amenity utility is always better: $u_i(x, 1) > u_i(x, 0), \forall i, x$.

Unlike in strategic complements, there is now a "fixed" supply of each good. Also, the price difference between goods are fixed in advance. The externality also now depends on *which* people consume the same good as you rather than the *quantity* of people. A type-i person's **willingness-to-pay** to live in neighborhood 1 when the share of high-skill people in neighborhood 1 is

$$WTP_i(x) := u_i(x, 1) - u_i(1 - x, 0) \tag{67}$$

The **difference** in willingness-to-pay between high and low skill people is given by:

$$\Delta WTP(x) := u_1(x, 1) - u_1(1 - x, 0) - u_0(x, 1) + u_0(1 - x, 0) \tag{68}$$

If:

- $\Delta WTP(x) > 0$: more high-skill people move to neighborhood 1, more low-skill people move to neighborhood 0
- $\Delta WTP(x) < 0$: more high-skill people move to neighborhood 0, more low-skill people move to neighborhood 1
- $\Delta WTP(x) = 0$: no one moves.

Now, an **equilibrium** is a price difference p between the two neighborhoods and a share x of high-skill people in neighborhood 1 such that either:

- **Segregation with positive sorting**: $x = 1$ and
  $$WTP_1(1) \geq p \geq WTP_0(1) \tag{69}$$
  This exists iff $\Delta WTP(1) \geq 0$ which happens intuitively when high-skill people value peer effects and/or amenities more than low-skill people.
- **Segregation with negative sorting**: $x = 0$ and
  $$WTP_1(0) \leq p \leq WTP_0(0) \tag{70}$$
  This exists iff $\Delta WTP(0) \leq 0$ which happens intuitively when high-skill people value peer effects more than low-skill people and/or value amenities less than low-skill people.
- **Integration**: $x \in (0, 1)$ and
  $$WTP_1(x) = p = WTP_0(x) \tag{71}$$
  This occurs iff segregation with positive and negative sorting are both strict equilbira (ie. $\Delta WTP(1) > 0$ and $\Delta WTP(0) < 0$) so by continuity it crosses 0.

Such an equilibrium is called **stable** if $\frac{\partial \Delta WTP(x)}{\partial x} < 0$.

Skill and peer effects are said to be **complements** if $u_1'(x, j) > u_0'(x, j)$ for all x, j and said to be **substitutes** otherwise. The former intuitively means that high-skill individuals value a marginal increase in neighbor's skill more than low skill people do, while the latter is the opposite.

Theorem 24. *If skill and peer effects are complements, any integrated equilibrium is unstable. If skill and peer effects are substitutes, any integrated equilibrium is stable.*

Additively Seperable Peer Effects

Consider a neighborhood effects model where the utility is **additively seperable** where everyone cares equally about peer effects, ie. there exists v, w_i such that

$$u_i(x, j) = v(x) + w_i(j) \tag{72}$$

In this case, skills and amenities are **complements** if $w_1(1) - w_1(0) > w_0(1) - w_0(0)$ and the opposite is true for substitutes.

Theorem 25. *For additively seperable utilities, if skill and amenities are complements, the unique equilibrium is segregation with positive sorting. If skill and amenities are tsubstitutes, the unique equilibrium is segregation with negative sorting.*

Efficiency in Equilibria

If share x of high-ability people live in neighborhood 1, the **utalitarian welfare** is defined as:

$$\frac{x}{2}u_1(x,1) + \frac{1-x}{2}u_1(1-x,0) + \frac{1-x}{2}u_0(x,1) + \frac{x}{2}u_0(1-x,0)$$

A utalitarian efficient assignment will maximize this total welfare. Now remark, if the peer effects v is sufficiently concave, the value will be higher in the middle of $x \in (0,1)$ rather than at the endpoints which can push integration to be utalitarian efficient although the equilibria is segregation (with positive sorting). Intuitively, this is because the high-skill people don't take into account externalities of leaving neighborhood 0.

Residential Choice with Homophilly (Schelling Segregation Model)

In the **Schelling Segregation Model**, the local effects of homophilly are more pronounced although there are no prices on the notion of efficiency. The setup is as follows:

- There are agents of two types randomly arranged on a grid with a few empty spaces
- A neighbor is **satisfied** if a percentage p of her (up to 8) neighbors are also of the same type and **unsatisfied** otherwise.
- Each timestep, an unsatisfied agent is chosen and moved to the closest empty space so that she's satisfied.
- Process continues until everyone is satisied.

For $p = 0.5$ found around 80% end up satisfied in equilibrium.

Tipping Point Theory

Card et. al analyzed "tipping point" behavior where when the minority population is above some threshold or *tipping point*, the share of the minority population increases while the opposite is true below a certain threshold.

Suppose a fraction x of a neighborhood are minorities and $D(x)$ of those who move in are minorities. Just like strategic complements, we get equilibria when $x = D(x)$. Increase the demand, $D(x)$ by enough removes integrated equilibria. The points at which the integrated equilibria disappears is called a **tipping point** and the minority share jumps discontinuously in response to small changes in the share.

Game Theory of Local Networks (L17)

How do we think about game theory of models with local strategic complements (coordination game) or substitutes (anti-coordination game)?

Suppose we have n players on a graph with adjacency matrix g. Consider the static game where each player i takes a static action $x_i \geq 0$. Denote a strategy profile $x := (x_1, \cdots, x_n)$. Consider payoffs of the form

$$u_i(x) = b_i\left(x_i + \delta \sum_{j \neq i} g_{ij} x_j\right) - c_i x_i \tag{73}$$

where $b_i(\cdot)$ is an increasing, concave, differentiable **benefit function**, δ is a coordination parameter, $c_i > 0$ is a **cost parameter**, measuring the cost or effort.

We have $\delta > 0$ is **strategic substitutes** (anti- coordination game: work less hard when others pick up the slack) while $\delta < 0$ is **strategic complements** (use an app like facebook: more they use the more you do) due to concavity of $u_i(\cdot)$.

Best Responses

For a given player i and opponent strategy x_{-i}, the optimal response is given by taking the derivative

$$b_i'\left(x_i + \delta \sum_{j \neq i} g_{ij} x_j\right) = c_i \tag{74}$$

so

$$x_i^* = \begin{cases} \bar{x}_i - \delta \sum_{j \neq i} g_{ij} x_j & \text{if } \delta \sum_{j \neq i} g_{ij} x_j \leq \bar{x}_i \\ 0 \quad \text{otherwise} \end{cases} \tag{75}$$

where $\bar{x}_i$ is the solution to $b'(\bar{x}_i) = c_i$.

Local Public Goods

An example of this is given by a scenario where each player shovels snow. An agent can get out when the sum of the effort x_i of her and her neighbors exceeds 1. Then, $\delta = 1$ and $\bar{x}_i = 1$ with an equilibrium being achieved when:

$$\forall i \text{ such that } x_i > 0, x_i + \delta \sum_{j \neq i} g_{ji} x_j = 1 \tag{76}$$

$$\forall i \text{ such that } x_i = 0, \delta \sum_{j} g_{ij} x_j \geq 1 \tag{77}$$

An equilibrium such that each agent x_i takes either 0 or 1 is said to be **specialized**. This is one special case of the above problem. Which agents work in a specialized equilibrium?

A set of nodes $S \subset N$ in a network is a **maximally independent set** if no two nodes in S are linked and every node not in S is linked to at least one node in S.

Theorem 26. *In every network, a maximal independent set exists.*

The proof is trivial by construction. From this, we can then generate specialized equilibria:

Theorem 27. *For any $S \subset N$, there is a specialized equilibrium where everyone in S takes $x_i = 1$ and everyone else takes $x_i = 0$ iff S is a maximal independent set. In particular, a specialized equilibrium always exists.*

What if $0 < \delta < 1$ (strategic substitutes)? This leads to a system of linear equations given by **??** based on whether $x_C = 0$ or $x_P = 0$ or neither.

Linear Quadratic Payoffs

A simialr type of best-response game is when the payoff function is linear quadratic:

$$u_i(x) := \bar{x}_i x_i - \delta \sum_{j \neq i} g_{ij} x_i x_j - \frac{1}{2} x_i^2 \tag{78}$$

for some arbitrary constant $\bar{x}_i$. The equilibria of this game are identical to the previous (purely linear) one.

For a symmetric adjacency matrix G, the linear quadratic game is a potential game with potential function

$$\phi(x) = \sum_i \left(\bar{x}_i x_i - \frac{1}{2} x_i^2 \right) - \frac{1}{2} \sum_{i,j} g_{ij} x_i x_j \tag{79}$$

Linear Quaratic Games with Strategic Complements

Now suppose the payoffs are (where $\delta < 0$)

$$u_i(x) = x_i + \alpha \sum_{j \neq i} g_{ij} x_i x_j - \frac{1}{2} x_i^2 \tag{80}$$

where here $\alpha := -\delta$. The first order conditions are given by

$$x_i = 1 + \alpha \sum_{j \neq i} g_{ij} x_j \Leftrightarrow \mathbf{x} = \mathbf{1} + \alpha g \mathbf{x} \tag{81}$$

which gives optimal solution:

$$\mathbf{x} = (I - \alpha g)^{-1} \mathbf{1} = \Lambda \mathbf{1} \tag{82}$$

which is precisely the Katz-Bonacich Centrality (in the transpose).

Network Markets (L18-L19)

How does network phenomena affect economic interactions? We now take the perspective of **networked markets**. Sometimes, we can have a free, Laissez-Faire market (traditional economic markets) but other times there are restrictions on which agents can trade with each other. Other times, there can be heterogenous valuations for sellers and a small number of goods, etc. which are more like *networks*.

The key features of "networked markets" are the following:

- What are the **competitive** or **market-clearing** prices of the network buyers and sellers? What are the properties of the resulting **competitive equilibria**?
- How does the bnetwork structure determine **bargaining power**?
- How are these properties determined in **intermediated markets**?

Matching Markets

In a **matching market** (or **assignment game**), we have a bipartite network with buyers I and sellers J. For simplicity, assume $|I| = |J|$ and each buyer only wants **one good** and each seller has only **one good** to sell. Next, suppose the value of goods are *heterogeneous* and given by a matrix $\mathbf{v}$ with $v_{ij} \geq 0$ representing the value buyer i values seller j's good. The outcomes are **matchings** and **prices**.

The payoffs for buyer i buying j's good at price p_j is given by:

$$i\text{'s payoff} = v_{ij} - p_j \tag{83}$$

$$j\text{'s payoff} = p_j \tag{84}$$

Competitive Equilibria

A **competitive equilibrium** is a bijection $M : I \to J$ and a nonnegativve price vector $p = (p_j)_{j \in J}$ such that for each each buyer i, if $j = M(i)$, then

$$v_{ij} - p_j \geq v_{ij'} - p_{j'} \quad \forall j' \in J \tag{85}$$

$$v_{ij} - p_j \geq 0 \tag{86}$$

M is called the **equilibrium assignment** (or **allocation**) with the condition that M being a bijection being called the **market-clearing condition**, p is the **market-clearing price vector**, and the condition that each buyer prefers their assigned good to any other for p is the **individual optimization condition**.

An equivalent definition of a competitive equilibrium is a price vector $\mathbf{p} = (p_j)_{j \in J}$ together with a perfect matching in the graph where there is a link (i, j) iff j is a **preferred-seller** for i (ie. it follows equation 85)

Efficiency

Are any competitive equilibria efficient? Define the **total value** generated by a matching $M : I \to J$ to be

$$V(M) = \sum_{i \in I} v_{iM(i)} \tag{87}$$

There exists (since $|I|, |J| < \infty$) an **optimal total value**

$$V^* = \max_M V(M) \tag{88}$$

The following theorem shows that competitive equilibria are indeed efficient:

Theorem 28 (First Welfare Theorem)**.** *If (M, p) is a competitive equilibrium, then $V(M) = V^*$*

Perfect Matchings

Competitive equilibriums are efficient but do they always exist? We first define some graph-theoretic terminology in terms of bipartite graphs. A **perfect matching** in a graph G is a set of edges with no common vertices. In a bipartite graph, it is a bijection $\varphi : I \to J$ where I and J are the two parts.

Theorem 29. *There is always at least one competitive equilibrium.*

The idea is to consider the set $N(S)$ which is the set of all neighbors of nodes in S and then use the following key graph theoretic result:

Theorem 30 (Hall's Marriage Theorem)**.** *A bipartite graph between nodes $i \in I$ and $j \in J$ with $|I| = |J|$, has a perfect matching iff for all $S \subset I$, we have $|S| \leq |N(S)|$*

The key idea is to use alternating paths between nodes in the matching and nodes not in the matching.

The idea is then to make the algorithm take the form of an "auction", raising the prices of over-demanded goods. The key idea is to start with zero prices and then raise the prices of things in constricted sets before normalizing by the minimum price. In economics, this process of adjusting prices to balance supply and demand is called **tatonnement**.

Bargaining Power

Bargaining power refers to the share of value that each individual gets in economic transactions. To study this and isolate the structure of the network, we today return to the matching markets setting but with $v_{ij} \in \{0, 1\}$ ie. binary. The payoffs are then also binary.

Stability of a Network Market

We can determine the bargaining powers of players in the network by considering the "stable outcomes" of the bargaining procedures. Formally, for a subset of agents $M \subseteq N$, let $v(M)$ be the maximum value that they can create by trading on their own. For our setting, this corresponds to a maximum matching in the subnetwork M (where (i, j) is a link iff $v_{ij} = 1$). A payoff vector $(u_i)_{i \in N}$ is stable, if for all $M \subseteq N$, we have

$$\sum_{i \in M} u_i \geq v(M) \tag{89}$$

with the set of all stable payoffs being called the **core** of the game.

Some facts about stable outcomes:

Theorem 31. *We have that*

(a). *If player i is unmatched in some maximal matching (that is, she is **under-demanded**) then she gets a payoff of 0 in every stable outcome.*

(b). *If player j is linked to some under-demanded player i, he gets a payoff of 1 in every stable outcome. Such a player is said to be **over-demanded**.*

Thus, for cases where our nodes are either under or over demanded, the stable payoffs are settled.

Dulmage-Mendelsohn Decomposition

We call a node that is neither over or under demanded, **perfectly-matched**. This comes from the following:

Theorem 32 (Dulmage-Mendelsohn Decomposition)**.** *For a bipartie graph G with under-demanded, over-demanded, and perfectly matched sets U, O, P, in every maximum matching:*

1. *Every node in O is matched to a node in U.*
2. *Every node in P is matched to another node in P.*

For perfectly-matched nodes, any price p is possible in stable outcomes. This is called the **indeterminacy of contract**.

Bilateral Bargaining

If we make some assumptions about the details of the bargaining game, then we can find equilibria. That is, rather than thinking about general markets, we take a specific game-theoretic viewpoint. We take two choices:

- **Ultimatum Bargaining**: each party makes a take-it-or-leave-it offer to the other.
- **Alternating-offers bargaining**: the parties take turns making offers until they agree on a price.

Ultimatum Bargaining

A **subgame perfect equilibrium (SPE)** is a strategy profile that remains a Nash Equilibrium conditional on reaching any point in the game.

Theorem 33. *The ultimatum bargaining game has a unique SPE. In it, Seller's strategy is to offer $p = 1$, and Buyer's strategy is to accept any price.*

Alternating-offers Bargaining

The game ends if B rejects in an ultimatum game. This is unrealistic in general scenarios: instead, in alternatin-offers bargaining:

- In even periods ($t = 0, 2, 4, \cdots$, S offers a price p. Then B Accepts or Rejects. If B Accepts, they trade at price p and the game ends. If B Rejects, go on to the next period.
- In odd periods, the order flips.
- If the game ends at price p at time t, the payoffs are $\delta_S^t p$ and $\delta_p^t(1-p)$ for B where $\delta_S, \delta_B \in (0, 1)$ are **player discounts.**

Theorem 34. *The alternating-offers bargaining game has a unique SPE: S always offers price $p_S = \frac{1-\delta_B}{1-\delta_S\delta_B}$ and accepts all prices greater or equal to $p_B = \delta_S p S$; symmetrically, B always offers price p_B and accepts all prices less than or equal to p_S*

Intermediation on Networks

In general, the buyers and sellers might not be able to sell directly and goods are re-sold through many intermediaries.

We consider a directed acyclic network G with initial nodes $s \in N$ with only out-links, a final node $b \in N$ with only in-links and all other nodes have both in and out links. In each period t, some node is the current **owner** of the good (with s being th einitial owner). Goods are sold to downstream neighbors and when good reaches b, utility of 1 is begotten and game ends. In a simple line network, the price of the good *doubles* at each step along the supply chain, with downstream intermediaries getting higher payoffs then upstream.

Trust and Cooperation (L20)

Before, we thought about the network effects of bargaining. What strategies can groups use to support trust or cooperation? We model this by **repeated games**: the long run relationship of a set of players playing the same game over and over.

Repeated Games: Model

A **repeated game** $G^T(\delta)$ is a finite static game G with action sets $A_1, \cdots, A_n$ and payoff functions $g_i : \prod A_i \to \mathbb{R}$ (called the **statge game**) and a **time horizon** T, discount factor $\delta \in [0, 1]$.

In each period $t = 1, \cdots, T$, the player simutaneously chooses actions $(a_1^t, \cdots, a_n^t)$ after observing all previous actions. The payoffs are

$$u_i = \sum_{t=1}^{T} \delta^{t-1} g_i(a_1^t, \cdots, a_n^t) \quad \forall i \in N \tag{90}$$

For infinite T, we need $\delta < 1$. Some definitions:

- A **history** in a repeated game is the action profiles $(a_1^\tau, \cdots, a_n^\tau)_{\tau=1}^{t-1}$ played up to the current period t.
- A **strategy** for a player i in a repeated game is a function from histories to current period actions.
- A **Nash equilibrium** is a strategy profile where each strategy is a best response to others.
- A **subgame perfect equilibrium (SPE)** is a strategy profile where each strategy is a best response to the others, conditional on reaching any possible history.

Repeated Prisoner's Dilemma

Suppose we consider the repeated prisoner's dilemna. We split into two cases:

i. If T is finite, by backward induction, (D, D) is the unique SPE.

ii. If T is infinite, **grim-trigger** is a unique SPE.

The result is summarized by the following:

Theorem 35 (Finite Repeated Games SPE v)**.** *If a static game G has a unique NE then, for any $T < \infty$ and $\delta \in [0, 1]$, the T-times repeated game $G^T(\delta)$ has a unique SPE. In the SPE, the unique NE of G is played in every period.*

We define the **grim-trigger** stratgey to be the strategy of starting with C and switching to D only if anyone ever players D after. The corresponding result for grim trigger is the following:

Theorem 36 (Infinite RG Grim-Trigger SPE). *In the infinitely repeated PD, grim trigger strategies are a SPE i§* $\delta \geq \frac{1}{2}$

This will not tell us that the players *will* cooperate, however, as there can be many equilibria. This idea is rigorized in the following result:

Theorem 37 (Folk Theorem). *Consider any action profile* a *such that there exists a SPE where* $u_i(a^{NE}) < u_i(a)$ *for all* $i \in I$. *There exists some* $\bar{\delta} < 1$ *such that, for any* $\delta > \bar{\delta}$ *there is a SPE where* a *is played in every period.*

Social Norms and Decentralized Repeated Games

Define cooperative equilibria is being those that incur cost to benefit someone else. We define cooperative equilibria in the game theoretic view as SPE in repeated games. We usually view social norms as arising in decentralized repeated games, where not everyone interacts directly with each other in every period.

One such simple model of decentralized society is a **repeated game with random matching**: here there are N players, with at each player at time $t = 1, 2, \cdots$ splitting randomly into pairs to play a symmetric, two-player stage game with some $j(i,t)$. Player i's total per-period payoff is then:

$$\sum_{t=1}^{\infty} \delta^{t-1} u(a_i^t, a_{j(i,t)}^t) \tag{91}$$

Note that players only observe actions in their own matches.

Can cooperation occur in such a model? One can show this is impossible if $N \to \infty$ as they can "disappear into the crowd" but this is not the case for finite N (and correspodning large δ).

Contagion Strategies

The anonymous matching version of trigger strategies (where if anyone of your partners defected before, you defect) are called **contagion strategies** as intuitively the defecting *action* "spreads" like a contagion in society.

Theorem 38. *In the prisoner's dilemma with anonymous random matching, contagion strategies are a Nash Equilibrium when players are sufficiently patient.*

Patience and Small Groups

In general, we learn where it is easier to cooperate when players are more patient, ie. δ is higher. Having more information also encourages cooperation. But who is most important for cooperation? This can not be done for general symmetric games, but we can do this for games on networks.

In general, we find cooperation to be easier in smaller groups. This is formalized by remarking that the critical discount factor $\bar{\delta}$ needed for contagion to support cooperation is increasing in N. Alternatively, suppose N is fixed but allow each player to play with K other players by telling them the outcomes of all her past matches. Then, the higher K leads to faster contagion spread and a lower critical discount factor. Some examples of this include the Maghribi in the Western Mediterranean, Jews in the NY diamond industry, etc.

Cooperation on a Network

Now suppose that we play a cooperation game on a network, with the players arranged in a network with edges E. In each period $t = 1, 2, \cdots$, each player i chooses a **level of cooperation** (or **effort**) $x_i \in \mathbb{R}_+$. The effort of a given player benefits *everyone* in the network, with player i's payoff being:

$$u_i(x) = \sum_{j \neq i} f(x_j) - x_i \tag{92}$$

for some increasing and concave function $f(\cdot)$. The network matters only in which *information* it provides, not in terms of any payoffs.

What is the maximum level of cooperation in any NE? Suppose we use contagion stratgies and there is a vector of "correct" effort levels $\mathbf{x} = (x_1, \cdots, x_n)$ such that each player i exerts x_i if every neighbor exerts x_j. If they observe they take some $x_j' \neq x_j$, they take effort 0 forever.

Note that if a player i deviates today, then player j stops cooperating at $d(i,j)$ from now. Thus, a vector of cooperation levels $\mathbf{x} = (x_1, \cdots, x_n)$ can be supported in equilibrium iff for each player i, we have the equilibrium payoff:

$$\frac{1}{1-\delta}\left(\sum_{j \neq i} f(x_j) - x_i\right) \geq \sum_{t=0}^{\infty} \delta^t \sum_{j : d(i,j) > t} f(x_j) \tag{93}$$

which gives

$$x_i \leq \sum_{j \neq i} \delta^{d(i,j)} f(x_j) \tag{94}$$

which corresponds to the maximum payoff (when solved for the system of equations), which is different for each player depending on the system of equations. Such a system always has a solution using iterated fixed point style convergence methods.

When is it possible to support more cooperation? When players are more patient (ie. δ is higher, the maximum x_is are larger). Alternatively, there is more cooperation when the network is "denser" with $d(i,j)$ smaller.

Recursive Centrality

Which individuals cooperate more? Intuitively, it is those that either have *more* distance t neighbors or those who themselves have more distance t neighbors. This idea is similar to Katz-Bonacich centrality, but because our functions $f(\cdot)$ are nonlinear, they are more general. We formalize this through the notion of **recursive centrality** (which is only a partial ordering).

More specifcally, i is **1-more central** than j if i has more distance-t neighbors than j for every t, ie.

$$|N_i(t)| \geq |N_j(t)| \quad \forall t \in \mathbb{N} \tag{95}$$

Recursively, we say i is **s-more central** than j if for every t there is an injection $\psi : N_j(t) \to N_i(t)$ such that, for each $k \in N_j(t)$, $\psi(k)$ is $(s-1)$-more central than k. Intuitively, i's distance-t neighbors are more central than j's distance t neighbors. We say i is **recursively more central** than j if it is s-more central for all s.

Theorem 39. *If i is recursively more central than j, then* $x_i \geq x_j$.

Information Aggregation (L21)

How do crowds make choices? Many decisions in society are made by groups of individuals who have differing amounts of information. Are crowds smarter or dumber than expert individuals?

Condorcet Jury Theorem

How do we formalize this notion of wisdom of crowds? Suppose we have a jury that all have the same preference of convicting a defendant iff they are guilty. Each juror/voter has different information, an independent noisy signal of the true state (guilty or innocent).

Theorem 40 (Condocert Jury Theorem)**.** *If all jurors vote according to their information vote convict if your signal indicates guilt, vote acquit if your signal indicates innocence then for a large jury, with high probability the majority will vote to convict if the defendant is guilty and vote to acquit if the defendant is innocent.*

This is an easy consequence of the law of large numbers. However, this assumes sincere voting. We now take the "game theory" version of the jury theorem, assuming strategic voting.

A defendant is either guilty or innocent, $\theta \in \{G, I\}$. Prior probability the defendant is guiliy is given by $p \in (0,1)$. There are N jurors, who must jointly decide whether to convict ($x = G$) or acquit ($x = I$). Assume they get a payoff of 0 if the correct decision is made, $-z$ if $x = G$ but $\theta = I$, or $-(1-z)$ if $x = I$ and $\theta = G$.

If there is only one juror, she would convict if and only if $\beta \geq z$, ie. the probability of guilt exceeds z.

Each juror, however, has different information, ie. a conditionally iid signal $s \in \{G, I\}$ with distribution

$$\mathbb{P}(s = G|\theta = G) = \mathbb{P}(s = I|\theta = I) = q > 0.5 \tag{96}$$

The decision is $x = G$ if at least a threshold k^* voters vote G. Some leading examples:

- Majority rule: N is odd, $k^* = \frac{N+1}{2}$
- Unanimity rule: $k^* = N$

The classical Condocert Jury theorem is more rigorously phrased as follows:

Theorem 41. *With majoiry rule and sincere voting, $\mathbb{P}[x = \theta]$ is increasing in N and converges to 1 as $N \to \infty$.*

Game-Theoretic CJT

Suppose we now take a Game-Theoretic view, where we assume voters vote to maximize payoff, ie. $\mathbb{P}(x = \theta)$. Now suppose we have a single player and suppose that $p > q$ (ie. the prior is larger than the signal). But then, by Bayes' Rule, we get that

$$\mathbb{P}(\theta = G|s = I) = \frac{\mathbb{P}(\theta = G \cap s = I)}{\mathbb{P}(\theta = G \cap s = I) + \mathbb{P}(\theta = I \cap s = I)} \tag{97}$$

$$= \frac{p(1-q)}{p(1-q) + (1-p)q} > \frac{1}{2} \tag{98}$$

ie. the signal is not enough to turn over the prior. Thus, sincere voting is not an equilibrium. This extends to the multi-player case, ie.

$$\mathbb{P}\big(\theta = G \,\big|\, s = I \,\cap\, V_{N-1} = 0\big) \;=\; \frac{1}{1 + \dfrac{1-p}{p}\dfrac{q}{1-q}} \tag{99}$$

so sincere voting is *not* an equilibrium in game-theoretic CJT. Then what are the equilibria? We claim these are the **symmetric, responsive equilibria** where all voters use the same signal to vote function.

Theorem 42. *Suppose $k^*(N) = \alpha N$ for some $\alpha \in (0,1)$ (require at least fraction α guilty votes to convict). As $N \to \infty$ for any sequence of symmetric responsive equilibria, $\mathbb{P}(x = \theta) \to 1$*

Thus, except in unanimity, the CJT survives in "Bayesian" Nash equilibrium. Thus, information aggregation still occurs, except in unanimity.