Collaborative insurance, unfairness and discrimination

Arthur Charpentier

Workshop on decentralized insurance and risk sharing, Chicago, 2024

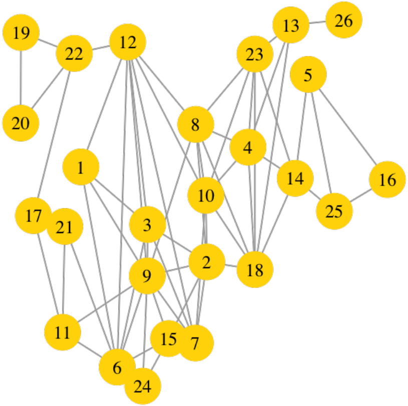
Disclaimer

This is an ongoing project, and, despite my efforts, it might still contain errors.
Some concepts and results presented in those slides are probably either extremely vague, or wrong.
All apologies.

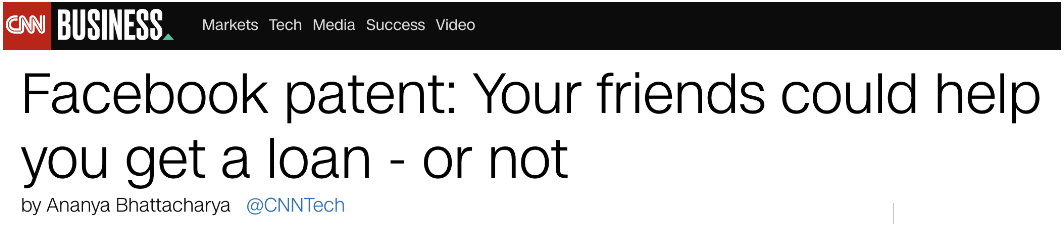
Agenda

Peer-to-peer insurance

Peer-to-peer insurance is a reciprocity insurance contract through the Collaborative consumption concept (...) (It is sometimes paired with the concept of the "sharing economy") (...) A group can be set up by the policyholders, forming a social network somewhat like Facebook. W



Agenda



CNN BUSINESS Markets Tech Media Success Video

Facebook patent: Your friends could help you get a loan - or not

by Ananya Bhattacharya @CNNTech

"*you apply for a loan and your would-be lender somehow examines the credit ratings of your Facebook friends. If the average credit rating of these members is at least a minimum credit score, the lender continues to process the loan application. Otherwise, the loan application is rejected,*" Bhattacharya (2015), see also Meyer (2015)

Agenda

Unfair (algorithmic) decision

Let us define three main criteria to evaluate if a given classifier is fair, that is if its predictions are not influenced by some of these sensitive variables. For example gender, ethnicity, sexual orientation or disability. W

Classical concepts related to (group) fairness are related to (conditional) independence

$$\begin{cases} m(\boldsymbol{X},S) \perp\!\!\!\perp S, \text{ independence} \\ m(\boldsymbol{X},S) \perp\!\!\!\perp S \mid Y, \text{ separation} \\ Y \perp\!\!\!\perp S \mid m(\boldsymbol{X},S), \text{ sufficiency} \end{cases}$$

or conditional probabilities

$$\begin{cases} \mathbb{P}[m(\boldsymbol{X},S) > t | S = A] = \mathbb{P}[m(\boldsymbol{X},S) > t | S = B], \text{ demographic parity} \\ \mathbb{P}[m(\boldsymbol{X},S) > t | S = A, Y = y] = \mathbb{P}[m(\boldsymbol{X},S) > t | S = B, Y = y], \forall y, \text{ equalized odds} \\ \mathbb{P}[Y = 1 | m(\boldsymbol{X},S) = t, S = A] = \mathbb{P}[Y = 1 | m(\boldsymbol{X},S) = t, S = B], \forall t, \text{ calibration} \end{cases}$$

or related expected values (weak conditions) for regression problems.

This course in one slide...

> **Fairness (machine learning)**
> An important distinction among fairness definitions is the one between group and individual notions. Roughly speaking, while group fairness criteria compare quantities at a group level, typically identified by sensitive attributes (e.g. gender, ethnicity, age, etc.), individual criteria compare individuals. In words, individual fairness follow the principle that ”similar individuals should receive similar treatments”, W

Individual fairness could be related to a “*perceptions of unfair treatment,*” Brown et al. (2006), Van Houtven et al. (2005) or Gonzalez et al. (2021)

Can be related to “*individual counterfactual explanation,*” Rudin (2019), see also Rudin (2019), Burkart and Huber (2021),Karimi et al. (2020) or Karimi et al. (2021).

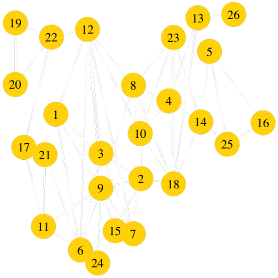
This course in one slide...

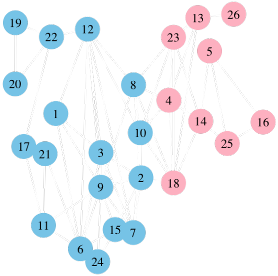
> **Graph (discrete mathematics)**
> a graph is a structure amounting to a set of objects in which some pairs of the objects are in some sense "related" W

But when observations are individuals in a connected network, those concepts are harder to quantify (difference local topology, Wu et al. (2017))...

Demographic information (overall) should be studied from a local perspective, with a network perspective

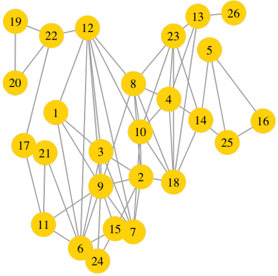
As a person in group either *A* or *B*, I feel discriminated, because among my "neighbors" (or connections), I have the feeling that I have a higher premium





This course in one slide...

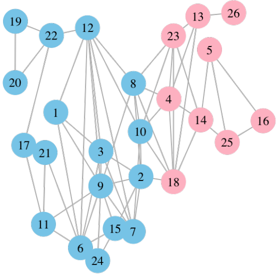
> **Graph (discrete mathematics)**
> a graph is a structure amounting to a set of objects in which some pairs of the objects are in some sense "related" W



But when observations are individuals in a connected network, those concepts are harder to quantify (difference local topology, Wu et al. (2017))...

Demographic information (overall) should be studied from a local perspective, with a network perspective

As a person in group either *A* or *B*, I feel discriminated, because among my "neighbors" (or connections), I have the feeling that I have a higher premium



Risk Sharing and Convex Order

References: Ohlin (1969); Kaas et al. (1994); Müller and Stoyan (2002); Denuit et al. (2006); Shaked and Shanthikumar (2007); Kaas et al. (2008); Denuit and Dhaene (2012a); Rüschendorf (2013) (etc.)

In those slides, all random variables are supposed to have a finite expected value.

Convex Order

"*The convex ordering approach to solve the optimal insurance decision problem was first adopted by* Ohlin (1969) *of minimizing a measure of the dispersion of the retained and ceded losses. The crucial mathematical tool employed by* Ohlin (1969) is *the 'Karlin–Novikoff once-crossing criterion' by* Karlin and Novikoff (1963) *for (increasing) convex ordering. Later,* Gollier and Schlesinger (1996) *used the same approach to extend the result of* Arrow (1974) *through maximizing an increasing convex order preserving objective functional of the terminal wealth. More recently, this approach was re-exploited to solve various optimal insurance decision problems*," Cheung et al. (2015)

> **Mean-preserving spread**
> a mean-preserving spread (MPS)* is a change from one probability distribution A to another probability distribution B, where B is formed by spreading out one or more portions of A's probability density function or probability mass function while leaving the mean (the expected value) unchanged. W
> * Rothschild and Stiglitz (1978)

Majorization and Convex Order

Following Hardy et al. (1929, 1934), and Marshall and Olkin (1979), for random vectors in $\mathbb{R}_+^n$,

Definition 2.1: Majorization (1)

Consider two sorted vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ ($x_1 \geq x_2 \geq \cdots \geq x_n$ and $y_1 \geq y_2 \geq \cdots \geq y_n$) such that $\displaystyle\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$, then $\boldsymbol{x} \preceq_M \boldsymbol{y}$ (majorization order) if $\displaystyle\sum_{i=1}^k x_i \leq \sum_{i=1}^k y_i$, $\forall k$.

$$\left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}, \frac{1}{n}\right) \prec_M \left(\frac{1}{n-1}, \frac{1}{n-1}, \cdots, \frac{1}{n-1}, 0\right) \prec_M \cdots \prec_M (1, 0, \cdots, 0, 0)$$

$\boldsymbol{x} \prec_M \boldsymbol{y} \leftrightarrow \boldsymbol{x}$ exhibits "less dispersion"
"less variability" than $\boldsymbol{y}$

Majorization and Convex Order

Definition 2.2: Majorization (2)

Consider two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ such that $\displaystyle\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$, then $\boldsymbol{x} \preceq_M \boldsymbol{y}$ if

$$\text{either} \sum_{i=1}^k x_{[i]} \leq \sum_{i=1}^k y_{[i]},\ \forall k \quad \text{or} \sum_{i=1}^k x_{(i)} \geq \sum_{i=1}^k y_{(i)},\ \forall k$$

where $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ (increasing) while $x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[n]}$ (decreasing).

If $\boldsymbol{x} \in \mathcal{S}_n$, and if $e_i = (0, \cdots, 0, 1, , 0 \cdots, 0)$, then $\dfrac{1}{n}\mathbf{1} \prec_M \boldsymbol{x} \prec_M e_i,\ \forall i$

Majorization and Convex Order

Proposition 2.1: Majorization

$\boldsymbol{x} \preceq_M \boldsymbol{y}$

$$\Longleftrightarrow \frac{1}{n}\sum_{i=1}^n h(x_i) \leq \frac{1}{n}\sum_{i=1}^n h(y_i) \text{ for any convex function } h$$

$$\Longleftrightarrow \frac{1}{n}\sum_{i=1}^n (x_i - d)_+ \leq \frac{1}{n}\sum_{i=1}^n (y_i - d)_+ \text{ for } d \in \mathbb{R}$$

$\Longleftrightarrow$ $\boldsymbol{x} = D\boldsymbol{y}$ for some doubly stochastic matrix D, i.e.

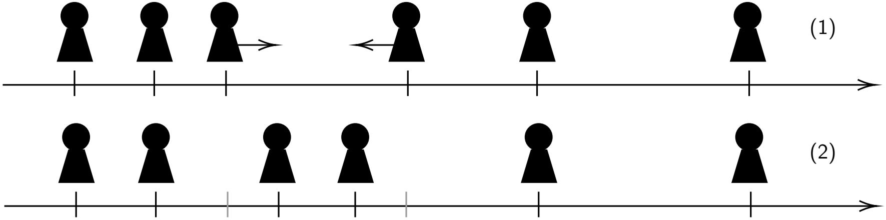
$$\sum_{k=1}^n D_{i,k} = \sum_{k=1}^n D_{k,j} = 1,\ \forall i, j$$

$\Longleftrightarrow$ $\boldsymbol{x} = P_1 \cdots P_k \boldsymbol{y}$ for finitely some Pigou-Dalton transfer matrices P_j

$(P_j = \alpha\mathbb{I} + (1-\alpha)T$ for some $\alpha \in (0,1)$ and $T = 0$ except $T_{i,j} = T_{j,i} = 1)$

Majorization and Convex Order

(Muirhead)-Pigou-Dalton transfers are classical notions when studying inequalities (Muirhead (1903); Dalton (1920), see also Atkinson (2015))



$$\boldsymbol{y}^{(2)} \preceq_M \boldsymbol{y}^{(1)} \longleftarrow \begin{cases} y_i^{(2)} = y_i^{(1)}, \ \forall i \neq j, k \text{ where } j < k \\ y_j^{(2)} = y_j^{(1)} + h, \\ y_k^{(2)} = y_k^{(1)} - h \end{cases}$$

Majorization and Convex Order

Definition 2.3: Convex order

Consider two random variables X and Y, $X \preceq_{CX} Y$ if $\mathbb{E}[h(X)] \leq \mathbb{E}[h(Y)]$ for any convex function h.

$\Longleftrightarrow$ Y is a mean-preserving spread of X, i.e. $Y \overset{\mathcal{L}}{=} X + Z$, where $\mathbb{E}[Z|X] = 0$.

$\Longleftrightarrow$ $\mathbb{E}[(X-d)_+] \leq \mathbb{E}[(Y-d)_+]$ for all $d \in \mathbb{R}$.

$\Longrightarrow$ $\mathbb{E}[X] = \mathbb{E}[Y]$ and $\text{Var}[X] \preceq \text{Var}[Y]$.

X exhibits "less dispersion"
"less variability" than Y

$$X \boxed{\preceq_{CX}} Y$$

"X is preferred over Y"

Majorization and Convex Order

Proposition 2.2: (Proposition 6 in Kaas et al. (2002))

Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ denote a collection of variables and $\boldsymbol{X}^+$ a comonotonic version, then $\boldsymbol{a}^\top\boldsymbol{X} \preceq_{CX} \boldsymbol{a}^\top\boldsymbol{X}^+$, $\boldsymbol{a} \in \mathbb{R}^n_+$.

Proposition 2.3: (Proposition 3.4.48 in Denuit et al. (2005))

Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ denote a collection of i.i.d. variables, if $\boldsymbol{a} \prec \boldsymbol{b}$ for the majorization order, $\boldsymbol{a}^\top\boldsymbol{X} \preceq_{CX} \boldsymbol{b}^\top\boldsymbol{X}$.

Proposition 2.4

Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ denote a collection of i.i.d. variables, and $\boldsymbol{p}$ some n-dimensional probability vector. Then $\boldsymbol{p}^\top\boldsymbol{X} \preceq_{CX} X_i$ for any i.

Majorization and Convex Order

Definition 2.4: Risk-sharing scheme

Consider two random vectors $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ and $\boldsymbol{X} = (X_1, \cdots, X_n)$ on $\mathbb{R}_+^n$. $\boldsymbol{\xi}$ is a risk-sharing scheme of $\boldsymbol{X}$ if $\boldsymbol{X}^\top \mathbf{1} = \boldsymbol{\xi}^\top \mathbf{1}$, i.e. $X_1 + \cdots + X_n = \xi_1 + \cdots + \xi_n$ almost surely.

Definition 2.5: Componentwise Convex, Denuit and Dhaene (2012b), Carlier et al. (2012)

Consider two random vectors $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ and $\boldsymbol{X} = (X_1, \cdots, X_n)$ on $\mathbb{R}_+^n$. $\boldsymbol{\xi} \preceq_{CCX} \boldsymbol{X}$ if $\xi_i \preceq_{CX} X_i$.

Definition 2.6: Desirable Risk Sharing Scheme

A risk sharing scheme $\boldsymbol{\xi}$ of $\boldsymbol{X}$ is desirable if $\boldsymbol{\xi} \preceq_{CCX} \boldsymbol{X}$.

Majorization and Convex Order

For example, $\xi_j = \overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i, \ \forall j$

- it is a risk sharing since $\xi_1 + \cdots + \xi_n = X_1 + \cdots + X_n$
- it is desirable (componentwise convex-order) : $\xi_j \preceq_{CX} X_j, \ \forall j$

Definition 2.7: Linear risk-sharing scheme

Consider two random vectors $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ and $\boldsymbol{X} = (X_1, \cdots, X_n)$ on $\mathbb{R}_+^n$. $\boldsymbol{\xi}$ is a linear risk-sharing scheme of $\boldsymbol{X}$ if $\boldsymbol{X}^\top \mathbf{1} = \boldsymbol{\xi}^\top \mathbf{1}$ almost surely, and $\boldsymbol{\xi} = \boldsymbol{R}\boldsymbol{X}$ for some matrix $\boldsymbol{R}$ $n \times n$ with positive entries.

Majorization and Convex Order

Definition 2.8: Row-stochastic matrix

A row-stochastic matrix (or left-stochastic matrix) $\boldsymbol{R}$ has positive entries, and $\boldsymbol{R}\mathbf{1} = \mathbf{1}$, i.e.,

$$\boldsymbol{R} = [R_{i,j}] \text{ where } R_{i,j} \geq 0, \text{ and } \sum_{i=1}^{n} R_{i,j} = 1 \ \forall j.$$

Definition 2.9: Column-stochastic matrix

A column-stochastic matrix (or left-stochastic matrix) $\boldsymbol{C}$ has positive entries, and $\mathbf{1}^\top \boldsymbol{C} = \mathbf{1}^\top$, i.e.,

$$\boldsymbol{C} = [C_{i,j}] \text{ where } C_{i,j} \geq 0, \text{ and } \sum_{j=1}^{n} C_{i,j} = 1 \ \forall i.$$

Majorization and Convex Order

Definition 2.10: (Doubly) Stochastic matrix

A double-stochastic matrix $\boldsymbol{D}$ has positive entries, and $\boldsymbol{D}\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^\top \boldsymbol{D} = \mathbf{1}^\top$, i.e.,

$$\boldsymbol{D} = [D_{i,j}] \text{ where } D_{i,j} \geq 0, \text{ and } \sum_{i=1}^{n} D_{i,j} = \sum_{j=1}^{n} D_{i,j} = 1 \ \forall i, j.$$

This corresponds to the "zero-balance conservation" property in Feng et al. (2023).

Proposition 2.5: Birkhoff–von Neumann theorem,

Any doubly stochastic matrix can be expressed as a barycenter of permutation matrices $\boldsymbol{P}_1, \cdots, \boldsymbol{P}_m$.

Majorization and Convex Order

If $\boldsymbol{D}$ be some $n \times n$ doubly-stochastic matrix, $\xi = \boldsymbol{D}\boldsymbol{X}$ is a linear risk sharing of $\boldsymbol{X}$ that satisfies the "zero-balance conservation" property.

Proposition 2.6: Comparing Linear Risk Sharing

Consider two linear risk sharing schemes ξ_1 and ξ_2 of $\boldsymbol{X}$, such that there is a doubly stochastic matrix $\boldsymbol{D}$, $n \times n$ such that $\xi_2 = \boldsymbol{D}\xi_1$. Then $\xi_2 \preceq_{CCX} \xi_1$.

Majorization and Convex Order

Proposition 2.7: Linear Risk Sharing

Let $\boldsymbol{R}$ be some $n \times n$ row-stochastic matrix, and given $\boldsymbol{X}$, a positive vector in $\mathbb{R}^+$, define $\xi = \boldsymbol{R}\boldsymbol{X}$. Then ξ is a linear risk sharing of $\boldsymbol{X}$.

Following Martínez Pería et al. (2005), we can define an ordering based those row-stochastic matrices

Definition 2.11

Consider two linear risk sharing schemes ξ_1 and ξ_2 of $\boldsymbol{X}$. ξ_1 weakly dominates ξ_2, denoted $\xi_2 \preceq_{wCX} \xi_1$ if and only if there is a column-stochastic matrix $\boldsymbol{R}$, $n \times n$ such that $\xi_2 = \boldsymbol{R}\xi_1$.

Networks

References: Gibbons (1985); Trudeau (1993); Barabási (2002); Diestel (2006); Harris (2008); Easley et al. (2010); Evans (2017); Hougaard and Moulin (2018); Pass (2019) (etc.)

Networks

Definition 3.1: Network

A (directed) network $\mathcal{G} = (V, E)$, where, as a convention, $V = \{1, \cdots, n\}$ denote either nodes, or vertices, and $E \in \{0, 1\}^{n \times n}$ represents the relationships.

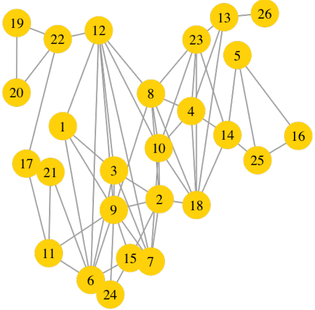
Definition 3.2: Adjacency Matrix

$A_{ij} \in \{0, 1\}$, and $A_{ij} = 1$ if and only if i and j are linked,

$$A_{ij} = \begin{cases} 1 \text{ if } (i,j) \in E \\ 0 \text{ otherwise} \end{cases}$$

There are no self-loops, i.e. $A_{i,i} = 0$. If the matrix is symmetric $(A_{ij} = A_{ji})$, the network is undirected.

Networks




```
> library(igraph)
> g
IGRAPH 8d07103 U--- 26 61 --
+ attr: gender (v/c)
+ edges from 99d7971:
  [1] 1--12 1-- 3 1-- 9 1-- 6 2-- 3
  [6] 2-- 8 2--10 2--18 2-- 7 2--15
 [11] 2-- 9 3--12 3-- 9 3-- 6 3-- 7
 [16] 4--18 4--10 4-- 8 4--23 4--13
 [21] 4--14 5--14 5--25 5--16 6--11
 [26] 6--21 6--12 6-- 9 6--15 6--24
 [31] 7-- 9 7--12 7--10 8--12 8--23
 [36] 8--10 8-- 9 8--18 9--12 9--15
+ ... omitted several edges
```

Networks

Definition 3.3: Neighbors

(immediate) neighbors of node i are

$$N_i = \{j \in V : (i,j) \in E\}.$$

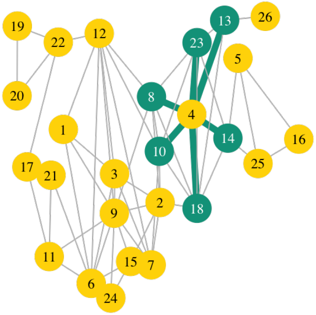
Proposition 3.1: Neighbors

$$N_i = \{j \in V : A_{i,j} > 0\}.$$

Definition 3.4: Extended Neighborhood

(immediate) extended neighbors of node i are

$$\overline{N}_i = N_i \cup \{i\}$$




```
> neighbors(g, 4)
+ 6/26 vertices
[1]  8 10 13 14 18 23
```

Networks

Definition 3.5: Neighbors of neighbors

Neighbors of neighbors of node i are

$$N_i^{(2)} = \{j \in V : (A^2)_{i,j} > 0\}.$$

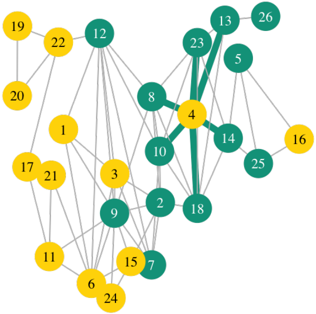
where classically, $(A^2)_{i,j} = \sum_{k=1}^{n} A_{i,k}A_{k,j}$

Definition 3.6: 2-Neighbors

Neighbors of order 2 of node i are

$$\overline{N}_2(i) = \{j \in V : \exists k \leq 2,\ (A^k)_{i,j} > 0\}.$$

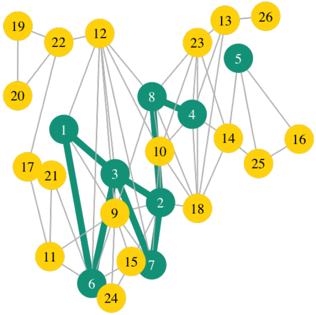
Note that $\overline{N}_2(i) = N_i \cup N_i^{(2)}$.



Networks

Definition 3.7: Subgraph of $\mathcal{G}$

Given two networks $\mathcal{G} = (E, V)$ and $\mathcal{G}' = (E', V')$, $\mathcal{G}'$ is a subgraph of $\mathcal{G}$ (denoted $\mathcal{G}' \subset \mathcal{G}$) if $E' \subset E$ and $V' \subset V$.



Induced subgraph

an induced subgraph of a graph is another graph, formed from a subset of the vertices of the graph and all of the edges, from the original graph, connecting pairs of vertices in that subset W

Definition 3.8: Induced subgraph of $\mathcal{G}$

Given a network $\mathcal{G} = (V, E)$ and a subset of vertices $V' \subset V$. The induced subgraph $\mathcal{G}_{V'} = (V', E')$ is the graph whose vertex set is V' and whose edge set consists of all of the edges in E that have both endpoints in V' (denoted E').

Networks

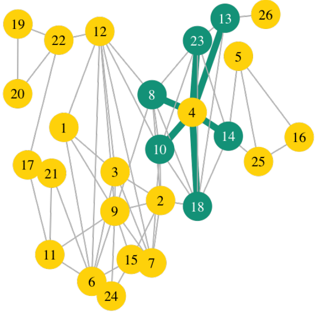
Set $E_i = \overline{N}_i$ and

$$V_i = \{(i,j) \in E, \text{ where } j \in N_i\}.$$

Definition 3.9: Induced subgraph of neighbors

Given a node i in a network (E, V), the induced subgraph of node i is $\mathcal{G}_{\overline{N}_i}$, also denoted $\mathcal{G}_i = (E_i, V_i)$.

E.g. $\mathcal{G}_4 = (E_4, V_4)$



Networks

Definition 3.10: Degrees

Row i contains list of vertices connected to vertex i,

$$d_i = \sum_{j=1}^{n} A_{i,j} = \boldsymbol{A}_{i,\cdot}^{\top}\mathbf{1} = \#N_i.$$

Let $\boldsymbol{d} = (d_i)$ denote the vector of degrees, and $\boldsymbol{D} = \text{diag}(\boldsymbol{d})$.

Definition 3.11: Normalized Adjacency Matrix

$\boldsymbol{A}_0 = \boldsymbol{D}^{-1}\boldsymbol{A} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ is the normalized adjacency matrix.

Networks

(for directed networks, this corresponds to "out degrees")

Definition 3.12: Walk

A walk from node i to node j is a sequence of edges, (i, v_1), (v_1, v_2), (v_2, v_3), $\cdots$, (v_{k-1}, v_k), (v_k, j)

Definition 3.13: Path

A walk where all the vertices are distinct is a path.

Definition 3.14: Connected graph

There exists a path that connects very pair of nodes in the network.

Networks

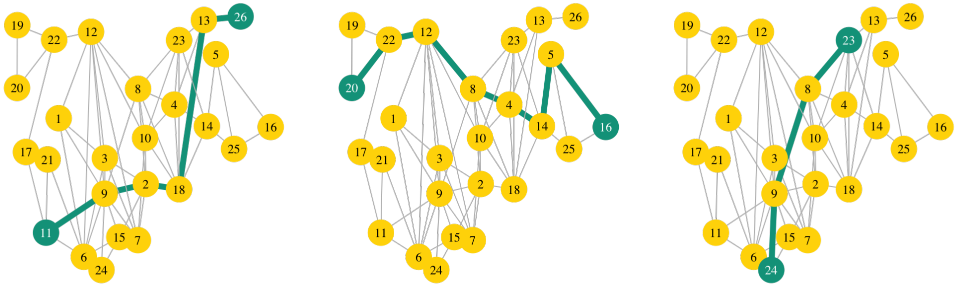
Definition 3.15: Shortest path

A geodesic between nodes i and j is a "shortest path" (i.e., with minimum number of edges) between these nodes. $d_{sp}(i,j)$ is the distance between nodes i and j.

Conveniently suppose that the set of vertices V is $\mathcal{I}_n = \{1, 2, \cdots, n\}$.

Networks

Exemples of (shortest) paths.



```
> shortest_paths(g,from=11,to=26)
$vpath
$vpath[[1]]
+ 6/26 vertices,
[1] 11  9  2 18 13 26
```

```
> shortest_paths(g,from=20,to=16)
$vpath
$vpath[[1]]
+ 8/26 vertices,
[1] 20 22 12  8  4 14  5 16
```

Networks

Definition 3.16: Random walk

Random walk with transition matrix $\boldsymbol{P} = \text{diag}(\boldsymbol{d})^{-1}\boldsymbol{A}$.

Let x_t denote the node reached at time t, and $\boldsymbol{p}(t) \in \mathcal{S}_n \subset \mathbb{R}_+^n$ the probability vector associated with $\{x_t = i\}$. Then

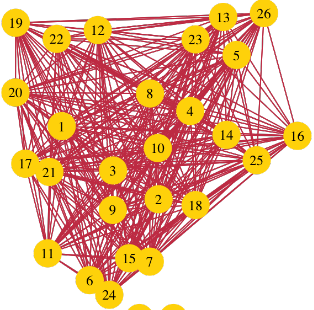
$$\boldsymbol{p}_{t+1} = \text{diag}(\boldsymbol{d})^{-1}\boldsymbol{A}\boldsymbol{p}_t.$$

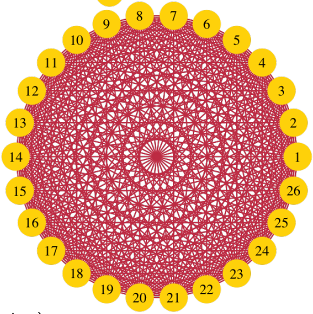
The stationary distribution is $\boldsymbol{\pi} = \lim_{t\to\infty} \boldsymbol{p}_t$.

Proposition 3.2: Unique Stationnary Distribution

$\boldsymbol{\pi}$ exists and is unique if the network is connected and aperiodic.

Random Graphs: Regular Graph (Dirac)



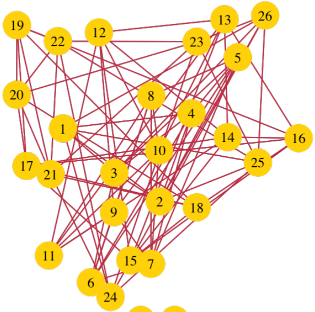


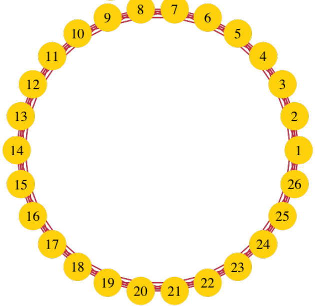
Definition 3.17: Complete graph

A complete graph is a simple undirected graph in which every pair of distinct vertices is connected

Here $d_i = (n-1)$, $\forall i \in \{1, \cdots, n\}$

Random Graphs: Regular Graph (Dirac)



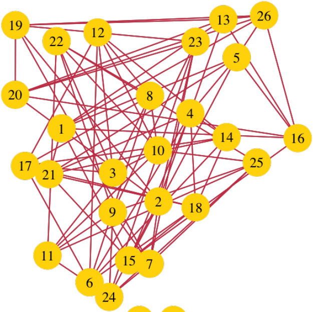


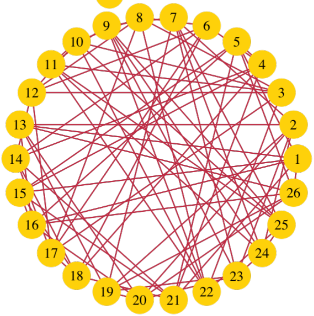
Definition 3.18: (r) Regular graph

a regular graph is a graph where each vertex has the same number of neighbors; i.e. every vertex has the same degree.

Here $d_i = r$, $\forall i \in \{1, \cdots, n\}$

Random Graphs: Regular Graph (Dirac)





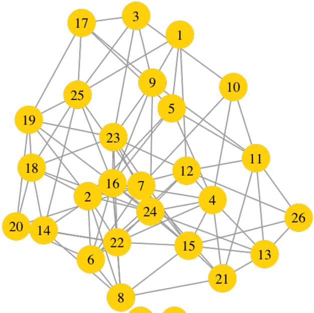
Definition 3.19: (r) Regular graph

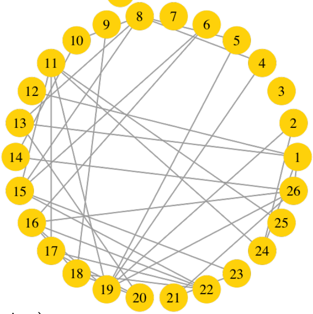
a regular graph is a graph where each vertex has the same number of neighbors; i.e. every vertex has the same degree.

Here $d_i = r$, $\forall i \in \{1, \cdots, n\}$

See Bollobás (1998) for regular random graphs

Random Graphs: Erdös-Rényi (Binomial-Poisson)





From Gilbert (1959), $d_i \leftarrow D_i \sim \mathcal{B}(n-1, p)$

Definition 3.20: Erdös-Rényi graph

$A_{i,j} = A_{j,i} \leftarrow X_{i,j}$ where $X_{i,j}$ are i.i.d. $\mathcal{B}(p)$ random variables (each edge has a fixed probability of being present or absent, independently of the other edges).

$$\mathbb{P}(D_i = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k},$$

$$\mathbb{P}(D_i = k) \rightarrow \frac{(np)^k \mathrm{e}^{-np}}{k!} \quad \text{as } n \rightarrow \infty \text{ and } np = \text{constant}.$$

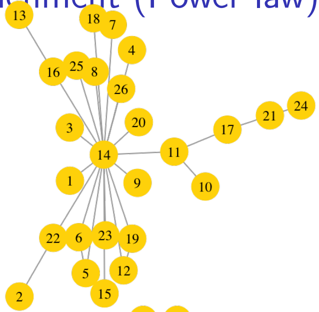
Random Graphs: Barabási–Albert, preferential attachment (Power law)

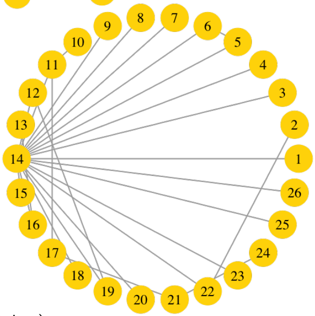
From Barabási and Albert (1999),

Definition 3.21: Barabási–Albert

Let $m \geq 1$. The network initializes with a network of $m_0 \geq m$ nodes. At each step, add 1 new node, then sample m existing vertices from the network, with a probability that is proportional to the number of links that the existing nodes already have.

(heavily linked nodes ("hubs") tend to quickly accumulate even more links)





Networks Generation

> **Havel–Hakimi algorithm**
> The Havel–Hakimi algorithm is an algorithm in graph theory solving the graph realization problem. That is, it answers the following question: *Given a finite list of nonnegative integers in non-increasing order, is there a simple graph such that its degree sequence is exactly this list?* A simple graph contains no double edges or loops. W

Suppose that the sum of degrees is even, random networks can then be generated with the algorithm of Havel (1955) and Hakimi (1962) (see also Viger and Latapy (2005)).

```
> degs = sort(round(1+rexp(100, 1/10)), decreasing=TRUE)
> if (sum(degs) %% 2 != 0) {
+   degs[1] <- degs[1] + 1
+ }
> g = realize_degseq(degs, allowed.edge.types = "all")
```

Majorization on networks

From Marshall and Olkin (1979),

B.2. Theorem (Hardy, Littlewood, and Pólya, 1929). A necessary and sufficient condition that $x \prec y$ is that there exist a doubly stochastic matrix P such that $x = yP$.

The lemma involves a special kind of linear transformation called a T-*transformation*, or more briefly a T-*transform*. The matrix of a T-transform has the form

$$T = \lambda I + (1-\lambda)Q,$$

where $0 \leq \lambda \leq 1$ and Q is a permutation matrix that just interchanges two coordinates. Thus xT has the form

$$xT = (x_1, \dots, x_{j-1}, \lambda x_j + (1-\lambda)x_k, x_{j+1}, \dots, x_{k-1}, \lambda x_k + (1-\lambda)x_j, x_{k+1}, \dots, x_n).$$

B.1. Lemma (Muirhead, 1903; Hardy, Littlewood, and Pólya, 1934, 1952, p. 47). If $x \prec y$, then x can be derived from y by successive applications of a finite number of T-transforms.

D Majorization in Integers

Consider the basic Lemma 2.B.1, which states that if $x \prec y$, then x can be derived from y by successive applications of a finite number of "T-transforms." Recall that a T-transform leaves all but two components of a vector unchanged, and replaces these two components by averages. If $a_1, \dots, a_n$ and $b_1, \dots, b_n$ are integers and $a \prec b$, can a be derived from b by successive applications of a finite number of T-transforms in such a way that after the application of each T-transform a vector with integer components is obtained? An affirmative answer was given by Muirhead (1903) and by Folkman and Fulkerson (1969). Using the same term as Dalton (1920), Folkman and Fulkerson (1969) define an operation called a *transfer*. If $b_1 \geq \cdots \geq b_n$ are integers and $b_i > b_j$, then the transformation

$$b_i' = b_i - 1,$$
$$b_j' = b_j + 1,$$
$$b_k' = b_k, \qquad k \neq i, j,$$

is called a *transfer from i to j*. This transfer is a T-transform, because

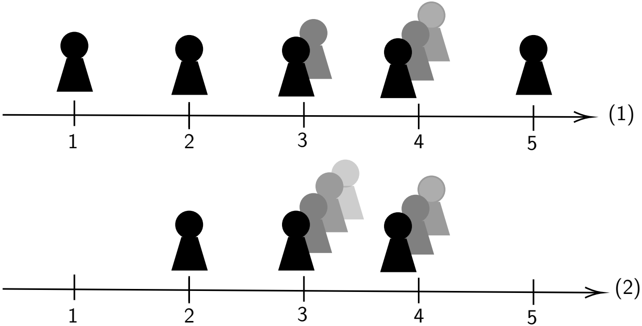
$$b_i' = \alpha b_i + (1-\alpha)b_j, \qquad b_j' = (1-\alpha)b_i + \alpha b_j,$$

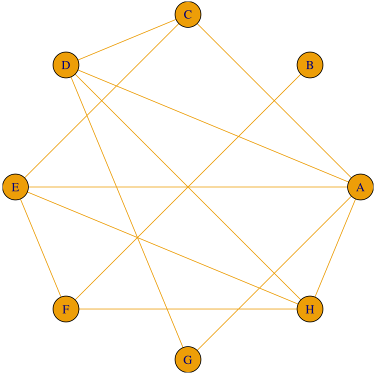
where $\alpha = (b_i - b_j - 1)/(b_i - b_j)$.

D.1. Lemma (Muirhead, 1903). If $a_1, \dots, a_n, b_1, \dots, b_n$ are integers and $a \prec b$, then a can be derived from b by successive applications of a finite number of transfers.

Majorization on networks

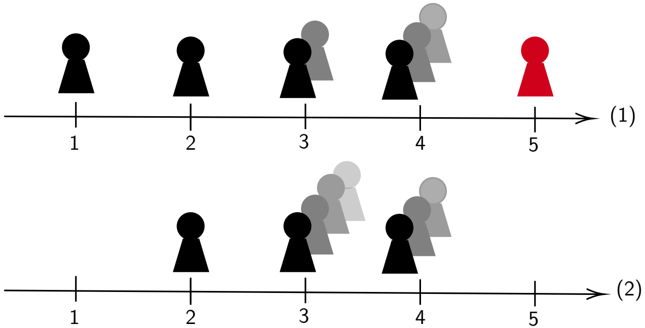
$$\boldsymbol{y}^{(1)} = (5,4,4,4,3,3,2,1) \to (4,4,4,3,3,3,3,2) = \boldsymbol{y}^{(2)}$$

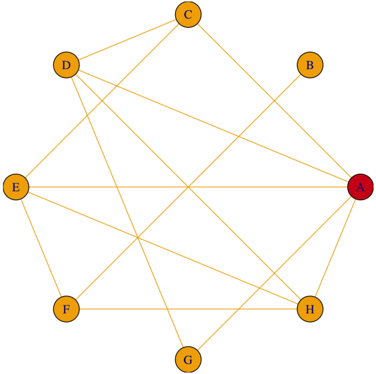




Majorization on networks

$$\boldsymbol{y}^{(1)} = (5, 4, 4, 4, 3, 3, 2, 1) \rightarrow (4, 4, 4, 3, 3, 3, 3, 2) = \boldsymbol{y}^{(2)}$$

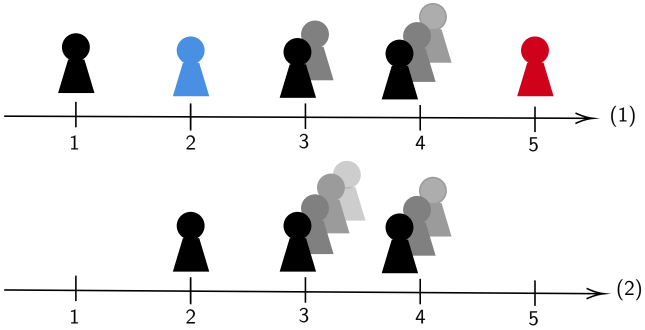


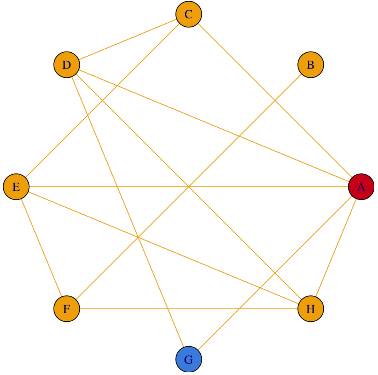


Find largest j such that $\displaystyle\sum_{i=1}^{j} y_i^{(1)} < \sum_{i=1}^{j} y_i^{(2)}$

Majorization on networks

$$\boldsymbol{y}^{(1)} = (5,4,4,4,3,3,2,1) \rightarrow (4,4,4,3,3,3,3,2) = \boldsymbol{y}^{(2)}$$

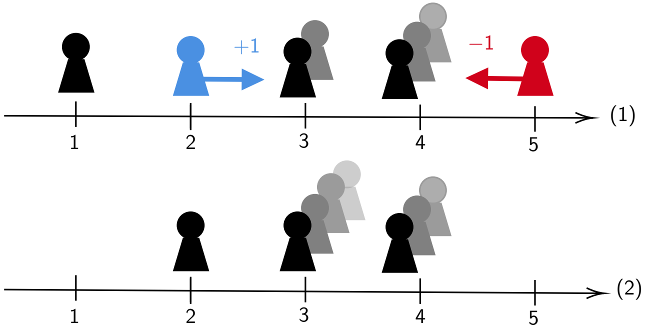


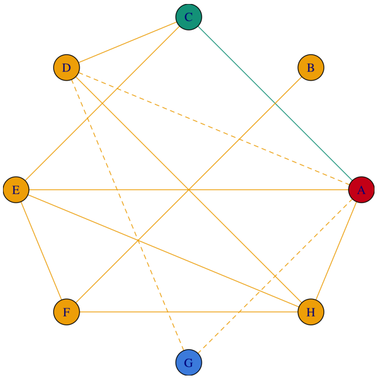


Find largest $k < j$ such that $y_k^{(2)} > y_k^{(1)} > y_{j+1}^{(2)} > y_{j+1}^{(1)}$

Majorization on networks

$$\boldsymbol{y}^{(1)} = (5,4,4,4,3,3,2,1) \to (4,4,4,3,3,3,3,2) = \boldsymbol{y}^{(2)}$$

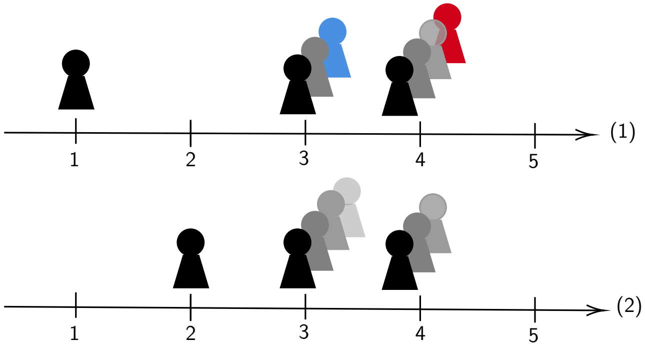


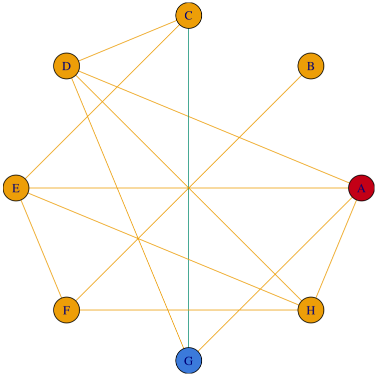


To get a transfer, we need to select ℓ

Majorization on networks

$$\boldsymbol{y}^{(1)} = (5,4,4,4,3,3,2,1) \to (4,4,4,3,3,3,3,2) = \boldsymbol{y}^{(2)}$$

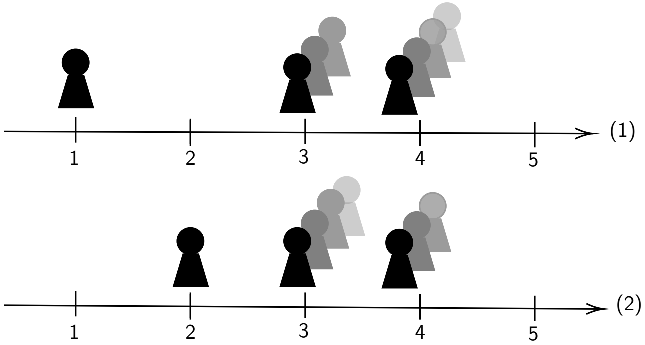


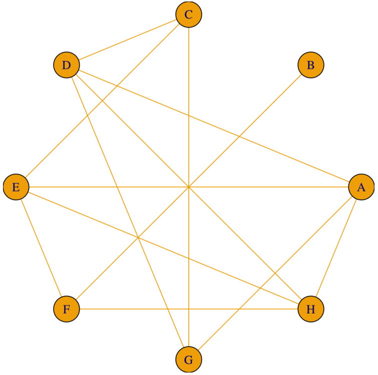


Remove edge (ℓ, j) and create (ℓ, k)

Majorization on networks

$$y^{(1)} = (5,4,4,4,3,3,2,1) \to (4,4,4,3,3,3,3,2) = y^{(2)}$$

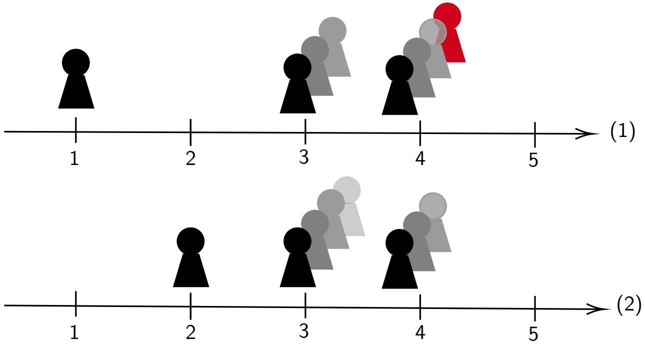


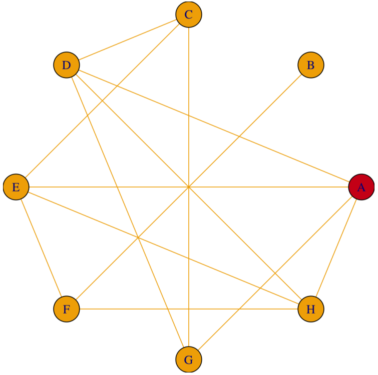


... we're back at stage 1...

Majorization on networks

$$\boldsymbol{y}^{(1)} = (5,4,4,4,3,3,2,1) \rightarrow (4,4,4,3,3,3,3,2) = \boldsymbol{y}^{(2)}$$

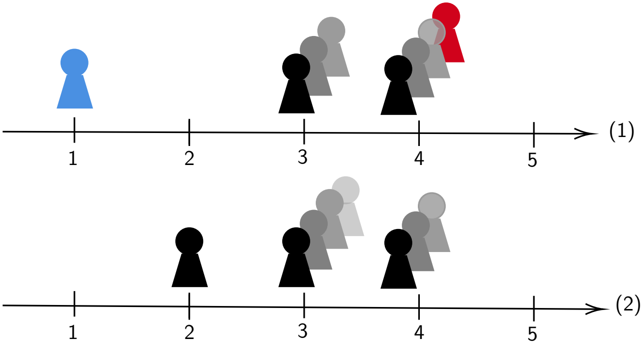


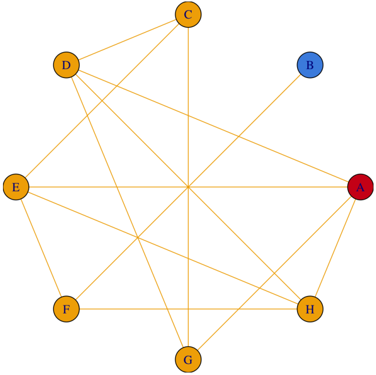


Find largest j such that $\displaystyle\sum_{i=1}^{j} y_i^{(1)} < \sum_{i=1}^{j} y_i^{(2)}$

Majorization on networks

$$\boldsymbol{y}^{(1)} = (5,4,4,4,3,3,2,1) \rightarrow (4,4,4,3,3,3,3,2) = \boldsymbol{y}^{(2)}$$

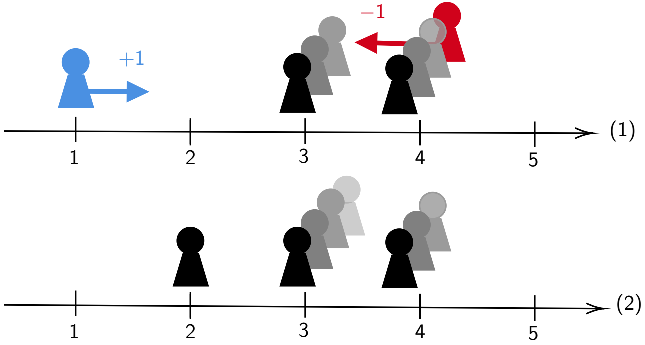


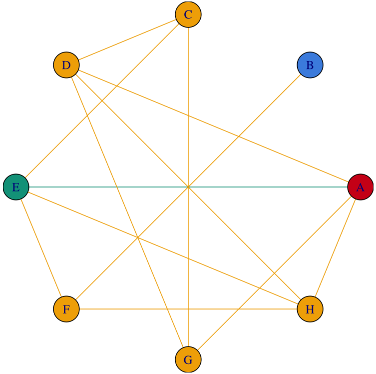


Find largest $k < j$ such that $y_k^{(2)} > y_k^{(1)} > y_{j+1}^{(2)} > y_{j+1}^{(1)}$

Majorization on networks

$$\boldsymbol{y}^{(1)} = (5,4,4,4,3,3,2,1) \rightarrow (4,4,4,3,3,3,3,2) = \boldsymbol{y}^{(2)}$$

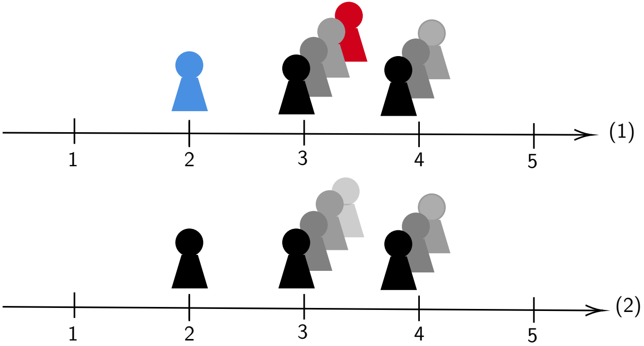


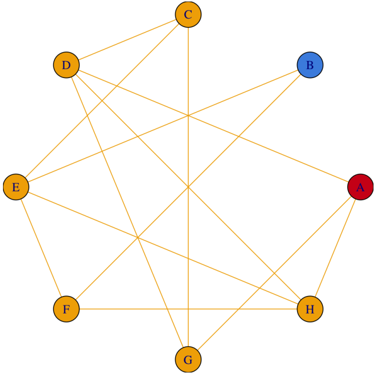


To get a transfer, we need to select ℓ

Majorization on networks

$$\boldsymbol{y}^{(1)} = (5,4,4,4,3,3,2,1) \to (4,4,4,3,3,3,3,2) = \boldsymbol{y}^{(2)}$$

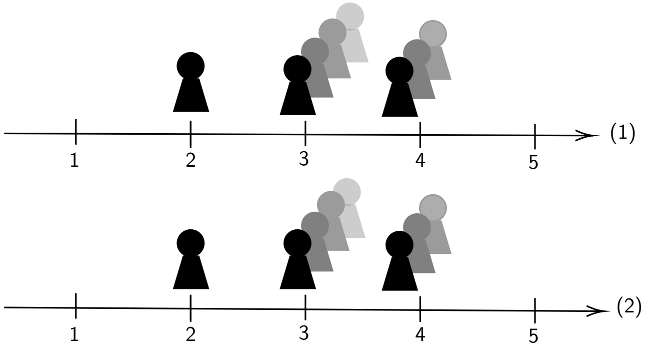


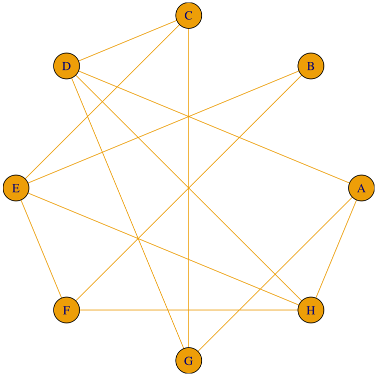


To get a transfer, we need to select ℓ

Majorization on networks

$$\boldsymbol{y}^{(1)} = (5,4,4,4,3,3,2,1) \to (4,4,4,3,3,3,3,2) = \boldsymbol{y}^{(2)}$$





Remove edge (ℓ,j) and create (ℓ,k)

Majorization on networks

djacency matrix $\boldsymbol{A}_1$ degree vector $\boldsymbol{d}_1$

$$\boldsymbol{A}_1 = \begin{bmatrix} . & . & 1 & 1 & 1 & . & 1 & 1 \\ . & . & . & . & . & 1 & . & . \\ 1 & . & . & 1 & 1 & . & . & . \\ 1 & . & 1 & . & . & . & 1 & 1 \\ 1 & . & 1 & . & . & 1 & . & 1 \\ . & 1 & . & . & 1 & . & . & 1 \\ 1 & . & . & 1 & . & . & . & . \\ 1 & . & . & 1 & 1 & 1 & . & . \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \\ 3 \\ 4 \\ 4 \\ 3 \\ 2 \\ 4 \end{bmatrix}, \boldsymbol{A}_2 = \begin{bmatrix} . & . & . & 1 & . & . & 1 & 1 \\ . & . & . & . & 1 & 1 & . & . \\ . & . & . & 1 & 1 & . & 1 & . \\ 1 & . & 1 & . & . & . & 1 & 1 \\ . & 1 & 1 & . & . & 1 & . & 1 \\ . & 1 & . & . & 1 & . & . & 1 \\ 1 & . & 1 & 1 & . & . & . & . \\ 1 & . & . & 1 & 1 & 1 & . & . \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 3 \\ 4 \\ 4 \\ 3 \\ 3 \\ 4 \end{bmatrix}$$

Note that $\boldsymbol{d}_2 = \boldsymbol{y}^{(2)} \preceq \boldsymbol{y}^{(1)} = \boldsymbol{d}_1$ and $A_2 \preceq A_1$.

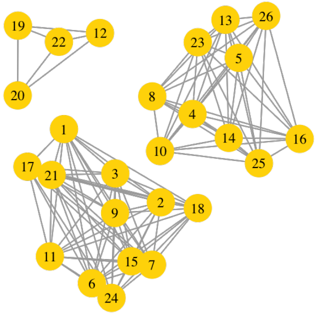
Risk Sharing on Networks

References: Gibbons (1985); Trudeau (1993); Barabási (2002); Diestel (2006); Harris (2008); Easley et al. (2010); Evans (2017); Hougaard and Moulin (2018); Pass (2019) (etc.)

Risk sharing on cliques

Definition 3.22: Cliques

A clique of a graph $\mathcal{G} = (V, E)$ is an induced subgraph of $\mathcal{G}$ that is complete. Formally, a clique C is a subset of the vertices, $C \subset V$, such that every two distinct vertices are adjacent.



With disconnected cliques, $\boldsymbol{A}$ is a diagonal block matrix

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_k \end{bmatrix}, \begin{cases} \mathbf{A}_k : n_k \times n_k \text{ matrix} \\ \mathbf{A}_k = (\mathbf{1}_k \mathbf{1}_k^\top - \mathbb{I}_k) \end{cases}$$

(up to some rearrangement of indices)

Risk sharing on cliques

Let $\mathcal{G}$ be a collection of disconnected cliques (partition of cliques), set

$$\boldsymbol{M} = \begin{bmatrix} \mathbf{M}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}_k \end{bmatrix}, \begin{cases} \mathbf{M}_k : n_k \times n_k \text{ matrix} \\ \mathbf{M}_k = \dfrac{1}{n_k - 1}\mathbf{M}_k = \dfrac{1}{n_k - 1}(\mathbf{1}_k \mathbf{1}_k^\top - \mathbb{I}_k) \end{cases}$$

(up to some rearrangement of indices).

Proposition 3.3: Risk Sharing on Cliques

$\boldsymbol{M}$ is a doubly-stochastic matrix, and $\xi = \boldsymbol{M}\boldsymbol{X}$ is a linear risk sharing of $\boldsymbol{X}$, (strictly) preferable in the sense that $\xi \prec_{CCX} \boldsymbol{X}$.

Risk sharing on networks

With our disconnected cliques, if $A_{n_k} = \mathbf{1}_{n_k}\mathbf{1}_{n_k}^\top - \mathbb{I}_{n_k}$,

$$\boldsymbol{M} = \begin{bmatrix} (n_1-1)\mathbb{I}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (n_2-1)\mathbb{I}_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (n_k-1)\mathbb{I}_{n_k} \end{bmatrix}^{-1} \begin{bmatrix} A_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & A_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & A_{n_k} \end{bmatrix}$$

i.e. $\boldsymbol{M}$ is the normalized adjacency matrix of $\boldsymbol{A}$, $\boldsymbol{D}^{-1}\boldsymbol{A}$, where $\boldsymbol{D} = \text{diag}(\boldsymbol{d})$.

For example, if $\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, then $\boldsymbol{M} = \boldsymbol{D}^{-1}\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$.

Note that here $\boldsymbol{M}$ is a row stochastic matrix, but not doubly stochastic.

Risk sharing on networks

Definition 3.23: Linear Sharing Risks with 'Friends' on a Network

Given a network $\mathcal{G} = (V, E)$ with adjacency matrix $\boldsymbol{A}$, the Linear Sharing Risks with 'Friends' is

$$\boldsymbol{\xi}_A = (\boldsymbol{D}^{-1}\boldsymbol{A})\boldsymbol{X}, \text{ i.e. } \xi_{A,j} = \frac{1}{d_j}\sum_{j\in N_i} X_i$$

$$\begin{cases} \mathcal{G}_1 = (\mathcal{I}_n, E_1), \text{ with adjacency matrix } A_1, \text{ with degree vectors } \boldsymbol{d}_1 \\ \mathcal{G}_2 = (\mathcal{I}_n, E_2), \text{ with adjacency matrix } A_2, \text{ with degree vectors } \boldsymbol{d}_2 \end{cases}$$

Risk sharing on networks

Definition 3.24: Network Ordering based on $\boldsymbol{d}$

Consider $\mathcal{G}_1 = (\mathcal{I}_n, E_1)$ and $\mathcal{G}_2 = (\mathcal{I}_n, E_2)$, write $\mathcal{G}_1 \preceq_D \mathcal{G}_2$ if $\boldsymbol{d}_1 \preceq_M \boldsymbol{d}_2$

Definition 3.25: Network Ordering based on $\boldsymbol{A}$

Consider $\mathcal{G}_1 = (\mathcal{I}_n, E_1)$ and $\mathcal{G}_2 = (\mathcal{I}_n, E_2)$, write $\mathcal{G}_1 \preceq_A \mathcal{G}_2$ if $A_1 \preceq_M A_2$

Proposition 3.4: Risk Sharing on Networks

If $\mathcal{G}_1 \preceq_A \mathcal{G}_2$, then $\boldsymbol{\xi}_{A_1} \preceq \boldsymbol{\xi}_{A_2} \preceq \boldsymbol{x}$

Risk sharing on networks

As mentioned in Charpentier et al. (2021), there are several problems,

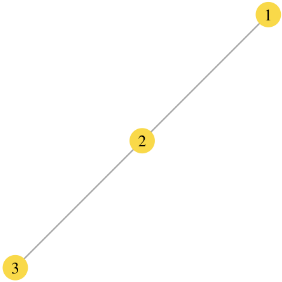
1. suppose that $\boldsymbol{A}$ is a permutation matrix, then $\boldsymbol{M} = \boldsymbol{A}$ is doubly stochastic

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \text{ then } \boldsymbol{M} = \boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Related to the anonymicity principle in Pigou Dalton transfer (and law based properties).

2. such risk sharing schemes are not, per se, reciprocal.

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \text{ then } \boldsymbol{M} = \boldsymbol{D}^{-1}\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}.$$



3. diag$(\boldsymbol{M}) = \mathbf{0}$ is problematic...

Centrality and Algebra of Networks

References: Gibbons (1985); Trudeau (1993); Barabási (2002); Diestel (2006); Harris (2008); Easley et al. (2010); Evans (2017); Hougaard and Moulin (2018); Pass (2019) (etc.)

Networks Centrality

Definition 3.26: Degree Centrality

Degree centrality of node i is $c_d(i) = d_i$, and $\boldsymbol{c}_d = \boldsymbol{d}$.

Definition 3.27: Eigenvector Centrality

Eigenvector centrality of node i is solution of $c_e(i) = \dfrac{1}{\lambda}\sum_{j=1}^{n} A_{i,j}c_e(j)$, or $\boldsymbol{c}_e = \dfrac{1}{\lambda}\boldsymbol{A}^\top\boldsymbol{c}_e$, for some fixed constant $\lambda > 0$.

Equation $\boldsymbol{A}^\top\boldsymbol{c}_e = \lambda\boldsymbol{c}_e$ means that $\boldsymbol{c}_e$ is some eigenvector associates with $\boldsymbol{A}^\top$ (or $\boldsymbol{A}$ if $\mathcal{G}$ is undirected).

Networks Centrality

Definition 3.28: PageRank Centrality

PageRank centrality of node i is solution of $c_p(i) = \alpha \sum_{j=1}^{n} A_{i,j}\frac{c_p(j)}{d_j} + \beta$, or $\boldsymbol{c}_p = \alpha \boldsymbol{A}^\top \boldsymbol{D}^{-1}\boldsymbol{c}_p + \beta\boldsymbol{1}$, for some fixed constant α and β.

```
> eigen_centrality(g)
 [1] 0.527 0.821 0.732 0.544 0.060
 [6] 0.702 0.671 0.833 1.000 0.788
[11] 0.310 0.864 0.286 0.298 0.458
[16] 0.019 0.075 0.657 0.030 0.030
[21] 0.162 0.160 0.544 0.345 0.060
[26] 0.046
> eigen(t(get.adjacency(g1)))
    $vectors[,1]
```

```
> page_rank(g)
 [1] 0.030 0.049 0.043 0.045 0.033
 [6] 0.060 0.036 0.049 0.070 0.049
[11] 0.036 0.058 0.036 0.044 0.031
[16] 0.024 0.023 0.052 0.026 0.026
[21] 0.020 0.044 0.045 0.025 0.033
[26] 0.014
>
>
```

Networks Centrality

Definition 3.29: Closeness Centrality

Closeness centrality of node i is $c_c(i) = \dfrac{n}{\displaystyle\sum_{j=1}^{n} d_{sp}(i,j)}$.

```
> closeness(g1)
 [1] 0.014 0.018 0.017 0.017 0.011
 [6] 0.016 0.017 0.019 0.019 0.019
[11] 0.014 0.019 0.013 0.014 0.015
[16] 0.009 0.012 0.018 0.011 0.011
[21] 0.012 0.014 0.017 0.014 0.011
[26] 0.010
```

```
>
```

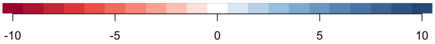
See Freeman et al. (1979)

Networks Centrality

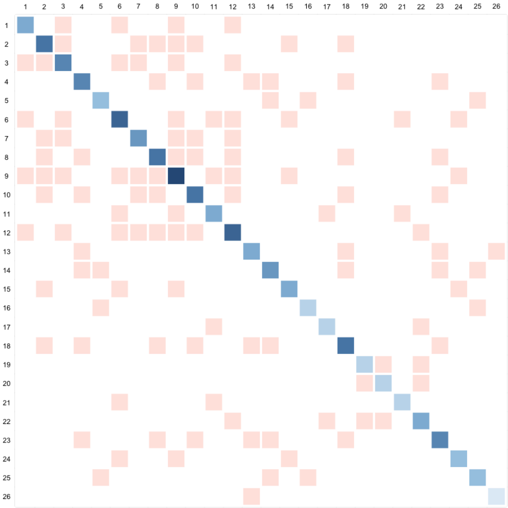
Definition 3.30: Laplacian

$$\boldsymbol{L} = \text{diag}(\boldsymbol{d}) - \boldsymbol{A},$$

$$L_{i,j} := \begin{cases} d_i & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and } (i,j) \in E \\ 0 & \text{otherwise,} \end{cases}$$




```
> L = laplacian_matrix(g)
```



Networks Centrality

Proposition 3.5: Alternative expression for L

Let $e_i = (0, \cdots, 0, 1, , 0 \cdots, 0) \in \{0,1\}^n$,

$$\boldsymbol{L} = \sum_{(i,j)\in E} \underbrace{(e_i - e_j)(e_i - e_j)^\top}_{\ell_{i,j}}$$

$$\ell_{i,j}\ n \times n \text{ matrix}, \ell_{i,j} = \begin{array}{cc} \begin{array}{ccccc} & i & & j & \end{array} & \\ \begin{pmatrix} (0) & \vdots & (0) & \vdots & (0) \\ \cdots & 1 & \cdots & -1 & \cdots \\ (0) & \vdots & (0) & \vdots & (0) \\ \cdots & -1 & \cdots & 1 & \cdots \\ (0) & \vdots & (0) & \vdots & (0) \end{pmatrix} & \begin{array}{c} \\ i \\ \\ j \\ \\ \end{array} \end{array}$$

Sidenote on quadratic forms

Definition 3.31: Normalized Laplacian Matrix

$\boldsymbol{L}_0 = \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2} = \mathbb{I} - \boldsymbol{A}_0$ is the normalized adjacency matrix.

$\boldsymbol{L}$ and $\boldsymbol{L}_0$ are symmetric positive semidefinite matrices.

Proposition 3.6: Laplacian and quadratic form

$\boldsymbol{L} = \text{diag}(\boldsymbol{d}) - \boldsymbol{A}$,

$$\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \frac{1}{2}\sum_{(i,j)\in E}(x_i - x_j)^2 = \frac{1}{2}\sum_{i,j=1}^{n} A_{i,j}(x_i - x_j)^2$$

Sidenote on quadratic forms

Proof.

$$A_{i,\cdot} = \sum_{j=1}^{n} A_{i,j} = d_i$$

$$\sum_{i,j=1}^{n} A_{i,j}(x_i - x_j)^2 = \sum_{i,j=1}^{n} A_{i,j}(x_i^2 - 2x_ix_j + x_j^2) = \sum_{i=1}^{n} \boxed{A_{i,\cdot}}\, x_i^2 - \sum_{i,j=1}^{n} 2A_{i,j}x_ix_j + \sum_{j=1}^{n} A_{\cdot,j}x_j^2$$

$$\sum_{i,j=1}^{n} A_{i,j}(x_i - x_j)^2 = 2\underbrace{\boxed{\sum_{i=1}^{n} d_i x_i^2}}_{\boldsymbol{x}^\top \boldsymbol{D}\boldsymbol{x}} + 2\underbrace{\boxed{\sum_{i,j=1}^{n} A_{i,j}x_ix_j}}_{\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}} = 2\boldsymbol{x}^\top(\boldsymbol{D} - \boldsymbol{A})\boldsymbol{x} = 2\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}$$

□

Since $\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x}$, $\boldsymbol{L}$ is symmetric positive semidefinite matrices.

Sidenote on quadratic forms

Let $\lambda_n \geq \lambda_{n-1} \geq \cdots \geq \lambda_2 \geq \lambda_1 \geq 0$ denote $\boldsymbol{L}$'s eigenvalues.

Proposition 3.7: Spectrum of $\boldsymbol{L}$ and λ_1

The n-vector of one's, $\mathbf{1}$, is an eigenvector of $\boldsymbol{L}$ associated with eigenvalue $\lambda_1 = 0$.

Proof.

$$\boldsymbol{L}\mathbf{1} = \sum_{(i,j)\in E} (\mathbf{1}_i - \mathbf{1}_j) \underbrace{(\mathbf{1}_i - \mathbf{1}_j)^\top \mathbf{1}}_{0} = \sum_{(i,j)\in E} (\mathbf{1}_i - \mathbf{1}_j) 0 = 0.$$

□

Proposition 3.8: Spectrum of $\boldsymbol{L}$ and λ_2

Network $\mathcal{G} = (E, V)$ is disconnected in two groups if and only if $\lambda_2 = 0$.

Sidenote on quadratic forms

Proposition 3.9: Spectrum of L and λ_2

Network $\mathcal{G} = (E, V)$ is disconnected in at least k groups if and only if $\lambda_k = 0$.

Proposition 3.10: Laplacian and quadratic form

$\boldsymbol{L}_0 = \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}$,

$$\boldsymbol{x}^\top \boldsymbol{L}_0 \boldsymbol{x} = \frac{1}{2}\sum_{(i,j)\in E}\left(\frac{x_i}{d_i} - \frac{x_j}{d_j}\right)^2 = \frac{1}{2}\sum_{i,j=1}^{n} A_{i,j}\left(\frac{x_i}{d_i} - \frac{x_j}{d_j}\right)^2$$

Networks Homophily and Assortative Mixing

Reference: Shrum et al. (1988); Yamaguchi (1990); McPherson et al. (2001); Jackson (2008); Newman (2018) (etc.)

Networks Homophily and Assortative Mixing

Definition 4.1: Homophily

Homophily is the tendency of individuals to form relations with others similar to them.

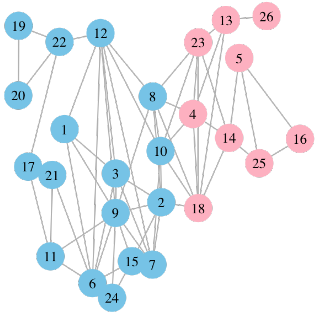
Definition 4.2: Community, Newman and Girvan (2004), Newman (2018)

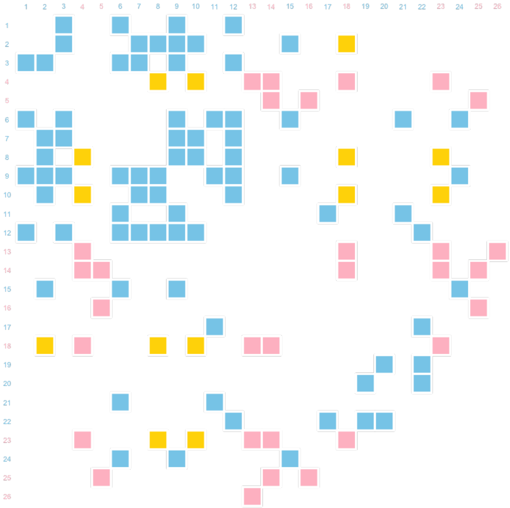
Communities are partitions of nodes.

The total number of edges that run between nodes of the same type is

$$\sum_{(i,j)\in E}\delta(c_i,c_j)=\frac{1}{2}\sum_{i,j}A_{i,j}\delta(c_i,c_j)\text{ where }\delta(c_i,c_j)=\begin{cases}1\text{ if }c_i=c_j\\0\text{ otherwise.}\end{cases}$$

Networks Homophily and Assortative Mixing





Networks Homophily and Assortative Mixing

The expected number of edges between nodes if edges are placed at random is

$$\frac{1}{2}\sum_{i,j}\frac{d_i d_j}{2m}\delta(c_i, c_j)$$

and the difference between the actual and expected number of edges in the network that join nodes of the same type is mQ where Q is the modularity measure,

Definition 4.3: Modularity measure, Newman (2003)

The modularity measure of a partition (c) of a network (E, V) is

$$Q = \frac{1}{2m}\sum_{i,j}\overbrace{\left(A_{i,j} - \frac{d_i d_j}{2m}\right)}^{B_{i,j}}\delta(c_i, c_j)$$

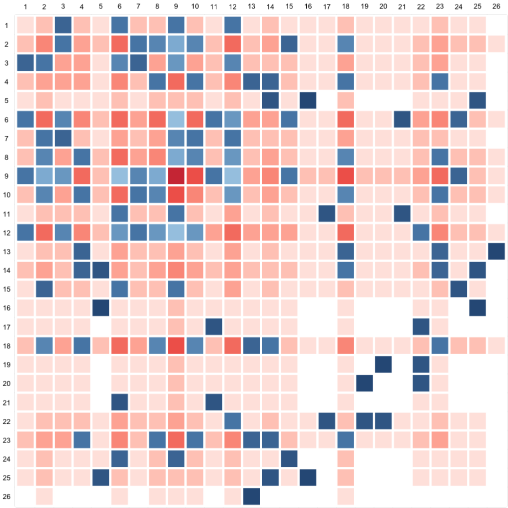
where m is the total number of links. $\boldsymbol{B}$ is coined "modularity matrix".

Networks Homophily and Assortative Mixing

A network is said to be assortative if a significant portion of its links are between nodes that belong to the same community

$$\boldsymbol{B} = \boldsymbol{A} - \frac{\boldsymbol{d}^\top \boldsymbol{d}}{2m}, \text{ i.e. } B_{i,j} = A_{i,j} - \frac{d_i d_j}{2m},$$


```
> A = get.adjacency(g)
> m = sum(A)/2
> d = apply(A,1,sum)
> B = A - d %*% t(d)/(2*m)
```



Networks Homophily and Assortative Mixing

Since $B_{i,j} = A_{i,j} - \dfrac{d_i d_j}{2m}$, and

$$\sum_{i=1}^{n} B_{i,j} = \underbrace{\boxed{\sum_{i=1}^{n} A_{i,j}}}_{d_j} - \frac{d_j}{2m} \underbrace{\boxed{\sum_{i=1}^{n} d_i}}_{2m} = d_j - \frac{d_j}{2m} 2m = 0$$

$$\sum_{j=1}^{n} B_{i,j} = \underbrace{\boxed{\sum_{j=1}^{n} A_{i,j}}}_{d_i} - \frac{d_i}{2m} \underbrace{\boxed{\sum_{j=1}^{n} d_j}}_{2m} = d_i - \frac{d_i}{2m} 2m = 0$$

In the case where were two communities, A and B, set

$$s_i^A = \begin{cases} +1 \text{ if } i \in A \\ -1 \text{ if } i \in B \end{cases} \quad \text{and } s_i^B = -s_i^A = \begin{cases} +1 \text{ if } i \in B \\ -1 \text{ if } i \in A \end{cases}.$$

Networks Homophily and Assortative Mixing

then

$$\delta(c_i, c_j) = \frac{1}{2}(s_i s_j + 1)$$

so that

$$Q = \frac{1}{2m}\sum_{i,j} B_{i,j}\delta(c_i, c_j) = \frac{1}{4m}\sum_{i,j} B_{i,j}(s_i s_j + 1) = \frac{1}{4m}\sum_{i,j} B_{i,j}s_i s_j = \frac{1}{4m}\boldsymbol{s}^\top \boldsymbol{B}\boldsymbol{s},$$

(whatever the reference group).

Proposition 4.1

The modularity measure can be written

$$Q = \frac{1}{4m}\boldsymbol{s}^\top \boldsymbol{B}\boldsymbol{s},\ \text{where } \boldsymbol{s} = \mathbf{1}_A - \mathbf{1}_B,\ \text{i.e. } s_i^A = \begin{cases} +1 \text{ if } i \in A \\ -1 \text{ if } i \in B \end{cases}$$

Networks Homophily and Assortative Mixing

When is Q maximal (in $\boldsymbol{s}$) ? see "modularity maximization," in Newman (2012)

Recall that $\boldsymbol{s} \in \{\pm 1\}^n$, so that $\boldsymbol{s}^\top \boldsymbol{s} = n$. Our problem is

$$\max_{\boldsymbol{s} \in \{\pm 1\}^n} \{\boldsymbol{s}^\top \boldsymbol{B} \boldsymbol{s}\}, \text{ subject to } \boldsymbol{s}^\top \boldsymbol{s} = n.$$

Using the Legrangian, our optimization problem has the following first order condition

$$\frac{\partial}{\partial \boldsymbol{s}} (\boldsymbol{s}^\top \boldsymbol{B} \boldsymbol{s} + \lambda(n - \boldsymbol{s}^\top \boldsymbol{s})) = \boldsymbol{0}$$

i.e.

$$\frac{\partial}{\partial s_k} \Big(\sum_{i,j} B_{i,j} s_i s_j + \lambda \Big(n - \sum_j s_j^2 \Big) \Big) = \sum_{i=1}^n B_{i,k} s_i - \lambda s_k = 0, \ \forall k$$

or, with matrix notations, $\boldsymbol{B} \boldsymbol{s}_\star = \lambda \boldsymbol{s}_\star$, i.e. $\boldsymbol{s}_\star$ is an eigenvector of $\boldsymbol{B}$. Thus

$$Q^\star = \frac{1}{4m} \boldsymbol{s}_\star^\top \boldsymbol{B} \boldsymbol{s}_\star = \frac{1}{4m} \boldsymbol{s}_\star^\top \lambda \boldsymbol{s}_\star = \frac{n}{4m} \lambda.$$

Networks Homophily and Assortative Mixing

```
> modularity(g1, 1+(V(g1)$gender=="female"))
[1] 0.3078474
```

Networks, without networks

Following Morris (1995), from AMEN (AIDS in Multi-Ethnic Neighborhoods) Study

Men: \ Women:	Black	Latina	White	Other	Pairs	Contact Rate
Black	506	32	69	26	633	2.61
Latino	23	308	114	38	483	1.61
White	26	46	599	68	739	2.01
Other	10	14	47	32	103	1.56
Pairs	565	400	829	164	1958	
Contact rate	1.58	2.00	2.53	2.73	2.07	2.04
Initial prevalence:						
males:	1.8%	2.4%	3.7%	2.2%		
females:	0.5%	0.7%	0.3%	0.0%		

	Black	Latina	White	Other	Margins
	36.23				1.00
		8.24			3.16
			3.97		5.70
				1.75	12.24
	1.00	1.21	1.36	2.32	0.31

Figure 3: Race and ethnicity matching among heterosexuals. The first ta-

Networks, without networks

Here we have 4 sensitive groups, on a bipartite network (heterosexual relationships)

Consider a discrete copula representation

$e_{i,j}$	Black	Hispanic	White	Other	a_i
Black	0.258	0.016	0.035	0.013	0.323
Hispanic	0.012	0.157	0.058	0.019	0.247
White	0.013	0.023	0.306	0.035	0.377
Other	0.005	0.007	0.024	0.016	0.053
b_j	0.289	0.204	0.423	0.084	

$$a_i = \sum_{j=1}^{K} e_{i,j} = \boldsymbol{E}_{i,\cdot}^{\top}\boldsymbol{1} \text{ and } b_j = \sum_{i=1}^{K} e_{i,j} = \boldsymbol{E}_{\cdot,j}^{\top}\boldsymbol{1}$$

Further

$$\boldsymbol{a}^{\top}\boldsymbol{b} = \sum_{k=1}^{K} a_k b_k = \sum_{k=1}^{K}\left(\sum_{i=1}^{K} e_{i,k}\right)\left(\sum_{j=1}^{K} e_{k,j}\right) = \sum_{i,j}(\boldsymbol{E}^2)_{i,j} = \|\boldsymbol{E}^2\|$$

Networks, without networks

Thus, we recover the coefficient introduced in Gupta et al. (1989),

Definition 4.4: Assortativity coefficient, Gupta et al. (1989)

With K communities

$$r = \frac{\displaystyle\sum_{k=1}^{K} e_{k,k} - \sum_{k=1}^{K} a_k b_k}{\displaystyle 1 - \sum_{k=1}^{K} a_k b_k} = \frac{\text{trace}[\boldsymbol{E}] - \|\boldsymbol{E}^2\|}{1 - \|\boldsymbol{E}^2\|}$$

More generally, when dealing with data with a network topology, we should be careful...

Statistics with a Network Topology

References: Peel et al. (2022)

Statistics with a Network Topology

sample data $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{S})$

$$\overline{y} = \frac{1}{n}\sum_{j=1}^{n} y_j = \frac{\mathbf{1}^\top \boldsymbol{y}}{\mathbf{1}^\top \mathbf{1}}$$

$\rightarrow$ sample version of $\mathbb{E}[Y]$.

$$\overline{y}_s = \frac{1}{n_s}\sum_{j=1}^{n} \mathbf{1}(s_j = s) y_j = \frac{\mathbf{1}_s^\top \boldsymbol{y}}{\mathbf{1}_s^\top \mathbf{1}}$$

$\rightarrow$ sample version of $\mathbb{E}[Y|S = s]$.

network data $(V, E, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{S})$

for a node $i \in V$,

$$\overline{y}(i) = \frac{1}{d_i}\sum_{j \in N_i} y_j = \frac{1}{d_i}\sum_{i=1}^{n} A_{i,j} y_j = \frac{\boldsymbol{A}_{i\cdot}^\top \boldsymbol{y}}{\boldsymbol{A}_{i\cdot}^\top \mathbf{1}}$$

$\rightarrow$ sample version of $E_i[Y]$.

$$\overline{y}_s(i) = \frac{1}{d_{i:s}}\sum_{j \in N_i} \mathbf{1}(s_j = s) y_j = \frac{(\boldsymbol{A}_{i\cdot} \cdot \mathbf{1}_s)^\top \boldsymbol{y}}{(\boldsymbol{A}_{i\cdot} \cdot \mathbf{1}_s)^\top \mathbf{1}}$$

$\rightarrow$ sample version of $E_i[Y|S = s]$.
where $\boldsymbol{a} \cdot \boldsymbol{b}$ is the element-wise product.

Statistics with a Network Topology

Given sample $\{x_1, \cdots, x_n\}$, the empirical variance,

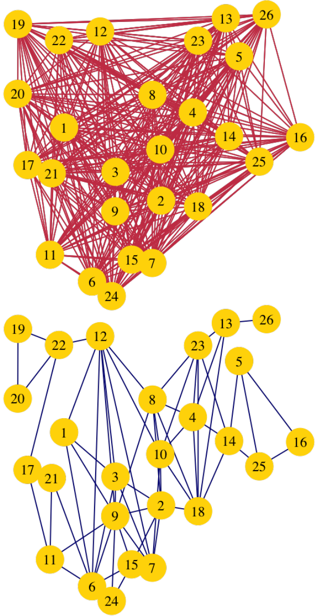
$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \text{, where } \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

could be written as a U-stat, Lee (2019)

$$\sigma^2 = \frac{1}{2n^2}\boxed{\sum_{i,j=1}^{n}(x_i - x_j)^2}.$$

On a network, with adjacency matrix A,

$$\sigma^2_{\mathcal{G}} = \frac{1}{4e}\boxed{\sum_{i,j=1}^{n} A_{i,j}(x_i - x_j)^2}\text{, where } 2e = \sum_{i,j=1}^{n} A_{i,j}$$



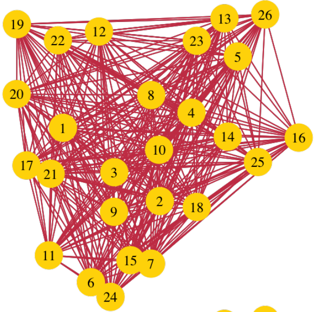
Statistics with a Network Topology

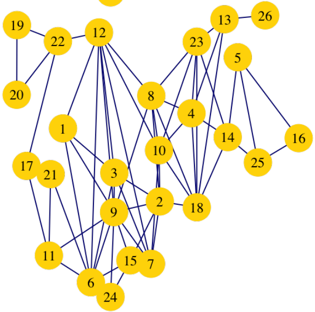
Given sample $\{(x_1, y_1), \cdots, (x_n, y_n)\}$, the empirical co-variance could be written as a U-stat,

$$cv = \frac{1}{2n^2} \boxed{\sum_{i,j=1}^{n} (x_i - x_j)(y_i - y_j)}$$

and if observations are nodes on a network

$$cv_{\mathcal{G}} = \frac{1}{4e} \boxed{\sum_{i,j=1}^{n} A_{i,j} (x_i - x_j)(y_i - y_j)}$$



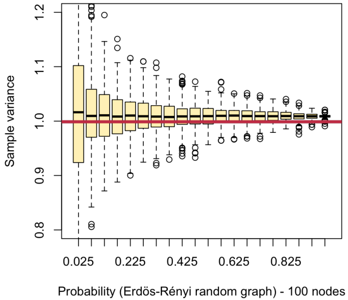


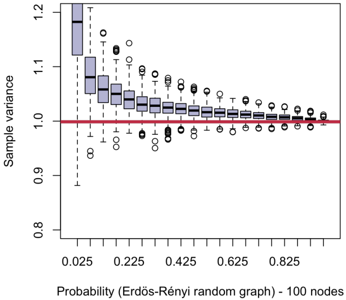
Statistics with a Network Topology

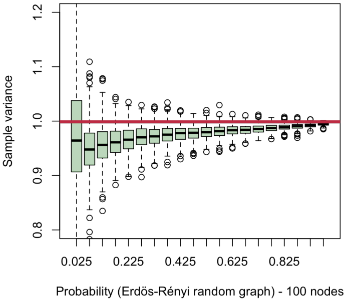
If x is independent of the topology of the network (summarized by A),

$$\boxed{\sigma^2} = \frac{1}{2n^2}\boxed{\sum_{i,j=1}^{n}(x_i - x_j)^2} \approx \frac{1}{4e}\boxed{\sum_{i,j=1}^{n}A_{i,j}\,(x_i - x_j)^2} = \boxed{\sigma^2_{\mathcal{G}}}$$

otherwise, the topology of the network is not neutral...



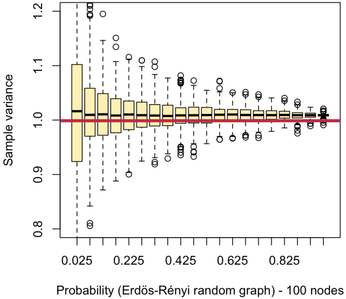




Statistics with a Network Topology

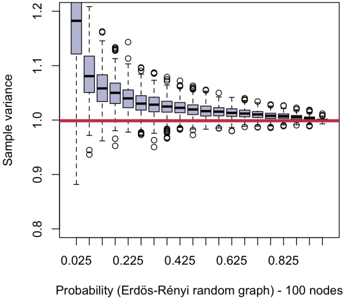
$\sigma^2_{\mathcal{G}} \approx \sigma^2$

"$\text{cor}(A, x) \approx 0$"



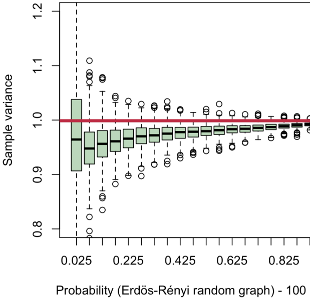
$\sigma^2_{\mathcal{G}} \geq \sigma^2$

"$\text{cor}(A, x) \geq 0$"



$\sigma^2_{\mathcal{G}} \leq \sigma^2$

"$\text{cor}(A, x) \leq 0$"



Erdös-Rényi network with $n = 100$ nodes, probability p (drawn randomly in $[0, 1]$)

Statistics with a Network Topology

Following Hall (1970), write

$$s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^n (\,x_i - x_j + x_j - \bar{x}\,)^2$$

$$\Longrightarrow\ (n-1)s^2 = \sum_{i=1}^n (x_i-x_j)^2 + 2\sum_{i=1}^n (x_i-x_j)(x_j-\bar{X}) + \sum_{i=1}^n (x_j-\bar{x})^2.$$

$$\Longrightarrow\ n(n-1)s^2 = \sum_{j=1}^n\sum_{i=1}^n (x_i-x_j)^2 + 2\sum_{j=1}^n\sum_{i=1}^n (x_i-x_j)(x_j-\bar{x}) + \sum_{j=1}^n\sum_{i=1}^n (x_j-\bar{x})^2.$$

$$\sum_{i=1}^n\sum_{j=1}^n (x_i-x_j)^2 = -2\sum_{i=1}^n\sum_{j=1}^n (x_i-x_j)(x_j-\bar{x}) = 2\sum_{i=1}^n\sum_{j=1}^n (x_j-\bar{x}+\bar{x}-x_i)(x_j-\bar{x})$$

$$\Longrightarrow\ \sum_{i=1}^n\sum_{j=1}^n (x_i-x_j)^2 = 2\sum_{i=1}^n\sum_{j=1}^n (x_j-\bar{x})^2 + 2\sum_{i=1}^n (\bar{x}-x_i)\underbrace{\sum_{j=1}^n (x_j-\bar{x})}_{0} = 2n(n-1)s^2.$$

Statistics with a Network Topology

Thus, for any m, write

$$2n(n-1)s^2 = \sum_{i,j=1}^{n} (x_i - x_j)^2 = \sum_{i,j=1}^{n} \Big(\underbrace{x_i - m}_{u_i} - \underbrace{x_j - m}_{u_j} \Big)^2 = \sum_{i,j=1}^{n} u_i^2 + u_j^2 + 2u_i u_j$$

If $m = \overline{x}$,

$$\sum_{i,j=1}^{n} u_i^2 = n\sum_{i=1}^{n} u_i^2 = n(n-1)S^2 \text{ and therefore } \sum_{i,j=1}^{n} u_i u_j = 0.$$

Hence,

$$\frac{1}{2n^2}\sum_{i,j=1}^{n} (x_i - \overline{x})(x_j - \overline{x}) = 0 \text{ but possibly } \frac{1}{4e}\sum_{i,j=1}^{n} A_{i,j}(x_i - \overline{x})(x_j - \overline{x}) \neq 0.$$

Paradoxes in Networks

"*on average your friends have more friends than you do.*"

Proposition 4.2: Friendship Paradox

The average number of friends of the collection of friends of individuals in a social network will be higher than the average number of friends of the collection of the individuals themselves. More formally

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{d_i}\sum_{j=1}^{n}A_{ij}d_j\right) \geq \frac{1}{n}\sum_{i=1}^{n}d_i.$$

Define differences Δ_i's between the average of its neighbours' degrees and its own degree, in the sense that

$$\Delta_i = \frac{1}{d_i}\sum_{j=1}^{n}A_{ij}d_j - d_i.$$

Paradoxes in Networks

Write the average as

$$\frac{1}{n}\sum_{i=1}^{n}\Delta_i=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{d_i}\sum_{j=1}^{n}A_{ij}d_j-d_i\right)=\frac{1}{n}\sum_{ij=1}^{n}\left(A_{ij}\frac{d_j}{d_i}-A_{ij}\right),$$

that yields

$$\frac{1}{n}\sum_{i=1}^{n}\Delta_i=\frac{1}{n}\sum_{ij=1}^{n}A_{ij}\left(\frac{d_j}{d_i}-1\right)\text{ but also }\frac{1}{n}\sum_{ij=1}^{n}A_{ij}\left(\frac{d_i}{d_j}-1\right),$$

by exchanging the summation indices, and because $\boldsymbol{A}$ is a symmetric matrix. By adding the two, we can write

$$\frac{2}{n}\sum_{i=1}^{n}\Delta_i=\frac{1}{n}\sum_{ij}A_{ij}\left(\frac{d_j}{d_i}+\frac{d_i}{d_j}-2\right)=\frac{1}{2n}\sum_{ij}A_{ij}\left(\sqrt{\frac{d_j}{d_i}}-\sqrt{\frac{d_i}{d_j}}\right)^2\geq 0.$$

(the exact equality holds only when $d_i=d_j$ for all pairs of neighbors)

Attributed Networks

Definition 4.5: Attributed Network

An attributed (directed) network $\mathcal{G}_{\boldsymbol{x}} = (V, E, \boldsymbol{X})$ is a network (V, E) where $\boldsymbol{X}$ is a node attributes matrix, $n \times k$, where each row is a feature vectors, for each node in $V = \mathcal{I}_n$.

If $\boldsymbol{X} = (x_1, \cdots, x_n)$, the classical average is

$$\mu(\boldsymbol{x}) = \overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Given an attributed (directed) network $\mathcal{G}_{\boldsymbol{x}} = (V, E, \boldsymbol{X})$, where $\boldsymbol{X} = (x_1, \cdots, x_n)$,

$$\mu_{\mathcal{G}}(\boldsymbol{x}) = \frac{1}{\displaystyle\sum_{i,j} A_{i,j}}\sum_{i,j} A_{i,j} x_i = \frac{1}{2m}\sum_{i=1}^{n} d_i x_i$$

Attributed Networks

Similarly, the variance of $\boldsymbol{X}=(x_1,\cdots,x_n)$ is

$$\text{Var}(\boldsymbol{x})=\frac{-1}{n-1}\sum_{i\neq j}(x_i-\mu(\boldsymbol{x}))(x_j-\mu(\boldsymbol{x}))$$

while variance over edges

$$\text{Var}_{\mathcal{G}}(\boldsymbol{x})=\frac{1}{\sum_{i,j}A_{i,j}}\sum_{i,j}A_{i,j}(x_i-\mu_{\mathcal{G}}(\boldsymbol{x}))(x_j-\mu_{\mathcal{G}}(\boldsymbol{x}))=\frac{1}{2m}\sum_{i,j}\left(A_{i,j}-\frac{d_id_j}{2m}\right)x_ix_j.$$

This leads to an other modularity measure, after anoter renormalization, so that it takes the value 1 in a network with perfect assortative mixing—one in which all edges fall between nodes with precisely equal values of x_i,

$$\overline{Q}=\frac{1}{2m}\sum_{i,j}\left(A_{i,j}x_i^2-\frac{d_id_j}{2m}x_ix_j\right)=\frac{1}{2m}\sum_{i,j}\left(d_i\mathbf{1}_{i=j}-\frac{d_id_j}{2m}\right)x_ix_j$$

Attributed Networks

Definition 4.6: Modularity measure for attributed networks

For some categorical variable x, the modularity measure is

$$Q = \frac{1}{2m}\sum_{i,j}\left(A_{i,j} - \frac{d_i d_j}{2m}\right)\delta(x_i, x_j).$$

If x is a numerical variable, a different normalization is considered

$$Q = \frac{1}{\kappa}\sum_{i,j}\left(A_{i,j} - \frac{d_i d_j}{2m}\right)x_i x_j, \text{ where } \kappa = \sum_{k,l}\left(d_k \mathbf{1}_{k=l} - \frac{d_k d_l}{2m}\right)x_k x_l$$

also coined "assortativity coefficient".

Attributed Networks

"*you apply for a loan and your would-be lender somehow examines the credit ratings of your Facebook friends. If the average credit rating of these members is at least a minimum credit score, the lender continues to process the loan application. Otherwise, the loan application is rejected,*" Bhattacharya (2015)

"*il ne faut jamais juger les gens sur leurs fréquentations. Tenez, Judas, par exemple, il avait des amis irréprochables,*" Paul Verlaine

For the generalized friendship paradox, which considers attributes other than degree, as in Cantwell et al. (2021), one can define an analogous quantity, $\Delta_i^{(x)}$, for some attribute x (such as the wealth) is defined as

$$\Delta_i^{(x)} = \frac{1}{d_i}\sum_j A_{ij}x_j - x_i,$$

Attributed Networks

which measures the difference between the average of the attribute for node i's neighbours and the value for i itself. When the average of this quantity over all nodes is positive one may say that the generalized friendship paradox holds. In contrast to the case of degree, this is not always true – the value of $\Delta_i^{(x)}$ can be zero or negative – but we can write the average as

$$\frac{1}{n}\sum_i \Delta_i^{(x)} = \frac{1}{n}\sum_i \left(\frac{1}{d_i}\sum_j A_{ij}x_j - x_i\right) = \frac{1}{n}\sum_i \left(x_i \sum_j \frac{A_{ij}}{d_j} - x_i\right),$$

where the second line again follows from interchanging summation indices. Defining the new quantity

$$\delta_i = \sum_j \frac{A_{ij}}{d_j},$$

and noting that

$$\frac{1}{n}\sum_i \delta_i = \frac{1}{n}\sum_{ij} \frac{A_{ij}}{d_j} = \frac{1}{n}\sum_j \frac{1}{d_j}\sum_i A_{ij} = 1,$$

Attributed Networks

we can then write

$$\frac{1}{n}\sum_i \Delta_i^{(x)} = \frac{1}{n}\sum_i x_i\delta_i - \frac{1}{n}\sum_i x_i \frac{1}{n}\sum_i \boldsymbol{\delta}_i = \text{Cov}(\boldsymbol{x},\boldsymbol{\delta}).$$

Thus, we will have a generalized friendship paradox in the sense defined here if (and only if) $\boldsymbol{x}$ and $\boldsymbol{\delta}$ are positively correlated. But this is not always the case

$$\left.\begin{array}{l}\text{Cov}(\boldsymbol{d},\boldsymbol{\delta}) \geq 0\\ \text{Cov}(\boldsymbol{x},\boldsymbol{\delta}) \geq 0\end{array}\right\} \not\Longrightarrow \text{Cov}(\boldsymbol{d},\boldsymbol{x}) \geq 0.$$

Group Fairness

References: Kearns and Roth (2019); Barocas et al. (2023); Charpentier (2024) (etc.)

Group Fairness

Fairness (machine learning)

Fairness in machine learning refers to the various attempts at correcting algorithmic bias in automated decision processes based on machine learning models. Decisions made by computers after a machine-learning process may be considered unfair if they were based on variables considered sensitive. For example gender, ethnicity, sexual orientation or disability W

$$\begin{cases} \boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d : \text{'explanatory' variables} \\ s \in \{\texttt{A}, \texttt{B}\} : \text{(binary) "sensitive variable"} \\ y \in \{0,1\} : \text{classification problem} \\ \widehat{y} \in \{0,1\} : \text{prediction, classically } \underbrace{\widehat{y}}_{\text{class } \in \{0,1\}} = \mathbf{1}(\underbrace{m(\boldsymbol{x}, s)}_{\text{score } \in [0,1] \subset \mathbb{R}} > t) \end{cases}$$

$$\begin{cases} \text{Independence (Definition 5.1)} : \; m(\boldsymbol{Z}) \perp\!\!\!\perp S \\ \text{Separation (Definition 5.2)} : \; m(\boldsymbol{Z}) \perp\!\!\!\perp S \mid Y \\ \text{Sufficiency (Definition 5.6)} : \; Y \perp\!\!\!\perp S \mid m(\underbrace{\boldsymbol{Z}}_{\boldsymbol{z} = (\boldsymbol{x}, s)}) \end{cases}$$

Group Fairness

Definition 5.1: Independence, Barocas et al. (2017)

A model m satisfies the independence property if $m(\boldsymbol{Z}) \perp\!\!\!\perp S$, with respect to the distribution $\mathbb{P}$ of the triplet $(\boldsymbol{X}, S, Y)$.

Definition 5.2: Separation, Barocas et al. (2017)

A model $m : \mathcal{Z} \to \mathcal{Y}$ satisfies the separation property if $m(\boldsymbol{Z}) \perp\!\!\!\perp S \mid Y$, with respect to the distribution $\mathbb{P}$ of the triplet $(\boldsymbol{X}, S, Y)$.

Group Fairness

Definition 5.3: True positive equality, (Weak) Equal Opportunity, Hardt et al. (2016)

A decision function $\widehat{y}$ – or a classifier $m_t(\cdot)$, taking values in $\{0,1\}$ – satisfies equal opportunity, with respect to some sensitive attribute S if

$$\begin{cases}\mathbb{P}[\widehat{Y}=1|S=\texttt{A},Y=1]=\mathbb{P}[\widehat{Y}=1|S=\texttt{B},Y=1]=\mathbb{P}[\widehat{Y}=1|Y=1]\\ \mathbb{P}[m_t(\boldsymbol{Z})=1|S=\texttt{A},Y=1]=\mathbb{P}[m_t(\boldsymbol{Z})=1|S=\texttt{B},Y=1]=\mathbb{P}[m_t(\boldsymbol{Z})=1|Y=1],\end{cases}$$

which corresponds to parity of true positives, in the two groups, $\{\texttt{A},\texttt{B}\}$.

Group Fairness

Definition 5.4: False positive equality, Hardt et al. (2016)

A decision function $\widehat{y}$ – or a classifier $m_t(\cdot)$, taking values in $\{0,1\}$ – satisfies parity of false positives, with respect to some sensitive attribute s, if

$$\begin{cases}\mathbb{P}[\widehat{Y}=1|S=\mathtt{A},Y=0]=\mathbb{P}[\widehat{Y}=1|S=\mathtt{B},Y=0]=\mathbb{P}[\widehat{Y}=1|Y=0]\\ \mathbb{P}[m_t(\boldsymbol{Z})=1|S=\mathtt{A},Y=0]=\mathbb{P}[m_t(\boldsymbol{Z})=1|S=\mathtt{B},Y=0]=\mathbb{P}[m_t(\boldsymbol{Z})=1|Y=0].\end{cases}$$

Group Fairness

Definition 5.5: Equalized Odds, Hardt et al. (2016)

A decision function $\widehat{y}$ – or a classifier $m_t(\cdot)$ taking values in $\{0,1\}$ – satisfies equal odds constraint, with respect to some sensitive attribute S, if

$$\begin{cases} \mathbb{P}[\widehat{Y}=1|S=\mathtt{A},Y=y]=\mathbb{P}[\widehat{Y}=1|S=\mathtt{B},Y=y]=\mathbb{P}[\widehat{Y}=1|Y=y],\ \forall y\in\{0,1\} \\ \mathbb{P}[m_t(\boldsymbol{Z})=1|S=\mathtt{A},Y=y]=\mathbb{P}[m_t(\boldsymbol{Z})=1|S=\mathtt{B},Y=y],\ \forall y\in\{0,1\}, \end{cases},$$

which corresponds to parity of true positive and false positive, in the two groups.

Definition 5.6: Sufficiency, Barocas et al. (2017)

A model $m:\mathcal{Z}\to\mathcal{Y}$ satisfies the sufficiency property if $Y \perp\!\!\!\perp S \mid m(\boldsymbol{Z})$, with respect to the distribution $\mathbb{P}$ of the triplet $(\boldsymbol{X},S,Y)$.

Individual Fairness

References: Dwork et al. (2012); Kusner et al. (2017); Charpentier (2024) (etc.)

Individual Fairness

"*We capture fairness by the principle that any two individuals who are similar with respect to a particular task should be classified similarly*," Dwork et al. (2012)

Definition 5.7: Similarity Fairness, Dwork et al. (2012)

We have similarity fairness if for all $i \neq j$ such that $\boldsymbol{x}_i = \boldsymbol{x}_j$, and $s_i \neq s_j$, then

$$m(\boldsymbol{x}_i, s_i = A) = m(\boldsymbol{x}_j, s_j = B).$$

"*In order to accomplish this individual- based fairness, we assume a distance metric that defines the similarity between the individuals*," Dwork et al. (2012)

Individual Fairness

Definition 5.8: Similarity Fairness, Luong et al. (2011), Dwork et al. (2012)

Consider two metrics, one on $\mathcal{Y}$ (for a classifier $[0,1]^2$ and not $\{0,1\}^2$) noted D_y, and one on $\mathcal{X}$ noted D_x. We will have similarity fairness on a database of size n if we have the following property (called Lipschitz property)

$$D_y(m(\boldsymbol{x}_i, s_i), m(\boldsymbol{x}_j, s_j)) \leq L \cdot D_x(\boldsymbol{x}_i, \boldsymbol{x}_j),\ \forall i, j = 1, \cdots, n,$$

for some $L < \infty$.

Individual Fairness

	Gender	Name	Treatment	Outcome (Weight)				Height	$\cdots$
			t_i	y_i	$y^\star_{i,T\leftarrow A}$	$y^\star_{i,T\leftarrow B}$	TE	x_i	$\cdots$
1	H	Alex	A	75	75	64	11	172	$\cdots$
2	F	Betty	B	52	67	52	15	161	$\cdots$
3	F	Beatrix	B	57	71	57	14	163	$\cdots$
4	H	Ahmad	A	78	78	61	17	183	$\cdots$

Different notations are used $y(1)$ and $y(0)$ in Imbens and Rubin (2015), y^1 and y^0 in Cunningham (2021), or $y_{t=1}$ and $y_{t=0}$ in Pearl and Mackenzie (2018)

Define potential outcomes to quantify the treatment effect, $\text{TE} = y^\star_{i,T\leftarrow B} - y^\star_{i,T\leftarrow A}$

$$\begin{cases} \text{observation} & : \; y^\star_{i,T\leftarrow B} \text{ when } t_i = B \text{ is observed, and } \boldsymbol{x}_i \\ \text{counterfactual} & : \; y^\star_{i,T\leftarrow A} \text{ when } t_i = B \text{ is observed, and } \boldsymbol{x}_i \end{cases}$$

Individual Fairness

Definition 5.9: Fairness on Average Treatment Effect, Kusner et al. (2017)

We achieve counterfactual fairness on average if

$$\text{ATE} = \mathbb{E}\big[Y^\star_{S\leftarrow A} - Y^\star_{S\leftarrow B}\big] = 0.$$

A decision satisfies counterfactual fairness if "*had the protected attributes (e.g., race) of the individual been different, other things being equal, the decision would have remained the same*" Kusner et al. (2017)

Definition 5.10: Counterfactual Fairness, Kusner et al. (2017)

We achieve counterfactual fairness for an individual with characteristics $\boldsymbol{x}$ if

$$\text{CATE}(\boldsymbol{x}) = \mathbb{E}\big[Y^\star_{S\leftarrow A} - Y^\star_{S\leftarrow B}\big|\boldsymbol{X} = \boldsymbol{x}\big] = 0.$$

Individual Fairness

If we get back to the "similarity approach", Zemel et al. (2013) suggested to use a consistency index

$$\frac{1}{n}\sum_{i=1}^{n}\left|m(\boldsymbol{x}_i,s_i)-\frac{1}{k}\sum_{j\in N_{i:k}}m(\boldsymbol{x}_j,s_j)\right|$$

where $N_k(i)$ at the k closest neighbor of $\boldsymbol{x}_i$ in $\mathcal{X}$. Why not consider

$$\frac{1}{n}\sum_{i=1}^{n}\left|m(\boldsymbol{x}_i,s_i)-\frac{1}{n_i}\sum_{j\in N(i)}m(\boldsymbol{x}_j,s_j)\right|$$

or

$$\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{n_i^A}\sum_{j\in N^A(i)}m(\boldsymbol{x}_j,s_j)-\frac{1}{n_i^B}\sum_{j\in N^B(i)}m(\boldsymbol{x}_j,s_j)\right|$$

corresponding to some network centric approach (perceived discrimination Pascoe and Richman (2009), Schmitt et al. (2014)).

Fairness Perception on Networks

"*Network and data analyses compound and reflect discrimination embedded within society*," Bernstein (2007)

References: Shalmani (2021)

Network Centric Fairness Perception

Definition 5.11: *d*-Neighbors

Given $d \in \mathbb{N}_\star$, let $N_d : V \to \mathcal{P}(V)$ defined as $\overline{N}_d(i) = \{j \in V : \exists k \leq d, (A^k)_{i,j} > 0\}$. $N_1(i) = N_i$ corresponds to (standard) neighbors of node i.

Definition 5.12: *d*-centered subgraph

Given $d \in \mathbb{N}_\star$, and a node i, the subgraph centered on node i (of order d) is $\mathcal{G}_i^d = (\overline{N}_d(i), E_d(i))$ where $E_d(i) = \{(j,j') \in E : j,j' \in \overline{N}_d(i)\}$.

Suppose that y is binary, $y_i \in \{0,1\}$.

Instead of a "model" $m : \mathcal{X} \to [0,1]$, consider a decision function $h : V \to [0,1]$ decision function.

Network Centric Fairness Perception

Definition 5.13: Isomorphic Networks

Two subgraphs $\mathcal{G}_1 = (V_1, E_1)$ and $\mathcal{G}_2 = (V_2, E_2)$ of $\mathcal{G}$ are isomorphic with respect to $h : V \to \mathbb{R}$ if there exists a one-to-one mapping $\psi : V_1 \to V_2$ such that

- $\forall (k, l) \in E_1$, $(\psi(k), \psi(l)) \in E_2$,
- $\forall k \in V_1$, $h(k) = h(\psi(k))$.

Network Centric Fairness Perception

Definition 5.14: Isomorphic Attributed Networks

Two attributed subgraphs $\mathcal{G}_1 = (V_1, E_1, \boldsymbol{X}_1)$ and $\mathcal{G}_2 = (V_2, E_2, \boldsymbol{X}_2)$ of $\mathcal{G}$ are isomorphic with respect to $h : V \to \mathbb{R}$ if there exists a one-to-one mapping $\psi : V_1 \to V_2$ such that

- $\forall (k, l) \in E_1$, $(\psi(k), \psi(l)) \in E_2$,
- $\forall k \in V_1$, $h(k) = h(\psi(k))$,and $\boldsymbol{x}_k = \mathbf{1}, \boldsymbol{x}_{2,\psi(k)}$.

Network Centric Fairness Perception

Definition 5.15: Fairness Perception Function

$\mathcal{F}(i,h)$ associate with decision h, for some node i (on a given network $\mathcal{G}$), "fairness perception function" if

- local axiom, if $h(i)=h'(i)$ and $\forall j\in N(i)$, $h(j)=h'(j)$, then $\mathcal{F}(i,h)=\mathcal{F}(i,h')$,
- monotonicty axiom, if $h(i)<h'(i)$ and $\forall j\in N(i)$, $h(j)=h'(j)$, then $\mathcal{F}(i,h)\leq\mathcal{F}(i,h')$,
- neighborhood expectation axiom, if $h(i)=h'(i)$ and $\forall j\in N(i)$, $h(j)\leq h'(j)$, then $\mathcal{F}(i,h)\geq\mathcal{F}(i,h')$,
- homogeneity axiom, let $\mathcal{G}_i=(E_i,V_i)$ and $\mathcal{G}_j=(E_j,V_j)$ be two subgraphs, if $\mathcal{G}_i$ and $\mathcal{G}_j$ are isomorphic with decision function h, then $\mathcal{F}(i,h)=\mathcal{F}(j,h)$

Network Centric Fairness Perception

Definition 5.16: Neighborhood Peer Expectation

Given an network $\mathcal{G}$, a decision function $h : V \to [0,1]$, and a node i

$$E_i[h] = \frac{y_i}{\sum_{j \in N_i} y_j} \sum_{j \in N_i} y_j h(j) + \frac{1-y_i}{\sum_{j \in N_i} 1-y_j} \sum_{j \in N_i} (1-y_j) h(j)$$

where actually, if $y_i = 1$, $E_i[h] = \frac{1}{\sum_{j \in N_i} y_j} \sum_{j \in N_i} y_j h(j)$,

while if $y_i = 0$, $E_i[h] = \frac{1}{\sum_{j \in N_i} 1-y_j} \sum_{j \in N_i} (1-y_j) h(j)$.

Network Centric Fairness Perception

The Neighborhood Peer Expectation considers the average decision of all neighbors with the same output y.

$$E_i[h] = \frac{y_i}{\sum_{j\in N_i} y_j} \sum_{j\in N_i} y_j h(j) + \frac{1-y_i}{\sum_{j\in N_i} 1-y_j} \sum_{j\in N_i} (1-y_j)h(j)$$

can we extended when considered larger networks, with $d \geq 1$,

$$E_{i,d}[h] = \frac{y_i}{\sum_{j\in \overline{N}_d(i)} y_j} \sum_{j\in \overline{N}_d(i)} y_j h(j) + \frac{1-y_i}{\sum_{j\in \overline{N}_d(i)} 1-y_j} \sum_{j\in \overline{N}_d(i)} (1-y_j)h(j)$$

Network Centric Fairness Perception

Proposition 5.1: Network-Centric Fairness Perception

Given a network $\mathcal{G} = (V, E)$, and a decision function h, the network-centric fairness perception function is defined as

$$\mathcal{F}(i, h) = \begin{cases} 1 \text{ if } E_i[h] \leq h(i) \\ 0 \text{ otherwise} \end{cases}$$

satisfies the locality, monotonicity, neighborhood expectation, and homogeneity axioms, i.e. it is a fairness perception function.

More generally, function $E_i[h]$ should satisfy

- if $\forall j \in N_i$, such that $h(j) = h'(j)$, then $E_j[h] = E_j[h']$,
- if $\forall j \in N_i$, such that $h(j) \leq h'(j)$, then $E_j[h] \leq E_j[h']$,
- if $\mathcal{G}_i$ and $\mathcal{G}_j$ are isomorphic, with respect to h, $E_i[h] = E_j[h]$

Network Centric Fairness Perception

Consider an attributed network $\mathcal{G}_{\boldsymbol{s}} = (V, E, \boldsymbol{S})$

Definition 5.17: Fairness Visibility

Let $V_{\boldsymbol{s}} = \{i \in V : \boldsymbol{S}_i = \boldsymbol{s}\}$, then fairness visibility of h for group $\boldsymbol{s}$ is

$$\overline{\mathcal{F}}_d(\boldsymbol{s}, h) = \frac{1}{\# V_{\boldsymbol{s}}} \sum_{i \in V_{\boldsymbol{s}}} \mathcal{F}_d(i, h)$$

Definition 5.18: Fairness Visibility Parity

h satisfies fairness visibility parity, with respect to $\boldsymbol{S}$, if

$$\overline{\mathcal{F}}_d(\boldsymbol{s}, h) = \overline{\mathcal{F}}_d(\boldsymbol{s}', h).$$

Network Centric Fairness Perception

Consider some binary decision rule $h: V \to \{0,1\}$,

Proposition 5.2: Asymptotic Fairness Visibility

Assuming the network graph is connected, and the decision function h has non-zero true positive and false positive rates, the fairness visibility of group $V_{\boldsymbol{s}}$, based on the neighborhood peer expectation, converges to the acceptance probability for $V_{\boldsymbol{s}}$ as the d-neighborhood size increases,

$$\overline{\mathcal{F}}_d(\boldsymbol{s}, h) = \frac{1}{\# V_{\boldsymbol{s}}} \sum_{i \in V_{\boldsymbol{s}}} \mathcal{F}(i,h) \to \mathbb{P}[h(i) = 1 | i \in V_{\boldsymbol{s}}), \text{ as } d \to \infty.$$

Heuristically, since the graph is connected, $N_i^{(d)} \to V$ as d increases.

For any i, ultimately, $\begin{cases} \mathcal{F}_d(i,h) = 1 \text{ if } h(i) = 1 \\ \mathcal{F}_d(i,h) = 0 \text{ if } h(i) = 0 \end{cases}$, thus consider only $i \in V_{\boldsymbol{s}}$

Network Centric Fairness Perception

For non-relational data, standard definition of demographic parity is

Definition 5.19: Demographic Parity

Decision function h satisfies demographic parity if

$$\underbrace{\mathbb{P}[h(i)=1|i\in V_{\boldsymbol{s}})}_{\text{acceptance probability for group } s} = \mathbb{P}[h(i)=1|i\in V_{\boldsymbol{s}'}).$$

Again, this definition ignores the neighborhood structure of a node.

Network Centric Fairness Perception

Proposition 5.3: Local vs. Asymptotic Fairness Visibility

Even if decision function h satisfies demographic parity,

$$\mathbb{P}[h(i)=1|i\in V_{\boldsymbol{s}})=\mathbb{P}[h(i)=1|i\in V_{\boldsymbol{s}'}),$$

there can still be non-parity w.r.t. fairness visibility, for some d,

$$\overline{\mathcal{F}}_d(\boldsymbol{s},h)\neq\overline{\mathcal{F}}_d(\boldsymbol{s}',h).$$

Wrap-up (?)

- Group fairness is interesting, but only "distribution based", with distributions of $m(\boldsymbol{X}, S)$ conditional on S
- Individual fairness is more interesting, by comparing outcomes from model m with counterfactuals version of $(\boldsymbol{x}_i, s_i)$
- Perceived discrimination is more complex to analyze, individual $(\boldsymbol{x}_i, s_i)$ comparing with neighbors on some attributed network $(\mathcal{I}_n, E, \boldsymbol{X}, S)$
- It becomes even more complex without centralized insurance, in collaborative insurance schemes, with peer-to-peer risk sharing

UNIWERSYTET WARSZAWSKI | WYDZIAŁ NAUK EKONOMICZNYCH

Polskie Stowarzyszenie Aktuariuszy

Warsaw Actuarial School 2024

The Polish Society of Actuaries
and University of Warsaw, Faculty of Economic Sciences
invite to the course

Insurance, biases, discrimination and fairness

by Prof. Arthur Charpentier (UQAM, Montréal, Canada)
Warsaw, 18-19 September 2024

- Registration will start in the middle of August
- You can participate on-line or on-site
- The number of on-site participants is limited to those who register first
- For more information visit the site of the Polish Society of Actuaries: https://actuary.org.pl/

References

Arrow, K. J. (1974). Optimal insurance and generalized deductibles. *Scandinavian Actuarial Journal*, 1974(1):1–42.

Atkinson, A. B. (2015). *Inequality*. Harvard University Press.

Barabási, A.-L. (2002). The new science of networks. *Cambridge University Press*.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439):509–512.

Barocas, S., Hardt, M., and Narayanan, A. (2017). Fairness in machine learning. *Nips tutorial*, 1:2017.

Barocas, S., Hardt, M., and Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT press.

Bernstein, E. (2007). *Temporarily Yours: Intimacy, Authenticity, and the Commerce of Sex*. University of Chicago Press.

Bhattacharya, A. (2015). Facebook patent: Your friends could help you get a loan - or not. *CNN Business*, 2015/08/04.

Bollobás, B. (1998). *Random graphs*. Springer.

References

Brown, C., Matthews, K. A., Bromberger, J. T., and Chang, Y. (2006). The relation between perceived unfair treatment and blood pressure in a racially/ethnically diverse sample of women. *American Journal of Epidemiology*, 164(3):257–262.

Burkart, N. and Huber, M. F. (2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, 70:245–317.

Cantwell, G. T., Kirkley, A., and Newman, M. E. (2021). The friendship paradox in real and model networks. *Journal of Complex Networks*, 9(2):cnab011.

Carlier, G., Dana, R.-A., and Galichon, A. (2012). Pareto efficiency for the concave order and multivariate comonotonicity. *Journal of Economic Theory*, 147(1):207–229.

Charpentier, A. (2024). *Insurance: biases, discrimination and fairness*. Springer Verlag.

Charpentier, A., Kouakou, L., Löwe, M., Ratz, P., and Vermet, F. (2021). Collaborative insurance sustainability and network structure. *arXiv*, 2107.02764.

Cheung, K. C., Chong, W. F., and Yam, S. C. P. (2015). Convex ordering for insurance preferences. *Insurance: Mathematics and Economics*, 64:409–416.

Cunningham, S. (2021). *Causal inference*. Yale University Press.

Dalton, H. (1920). The measurement of the inequality of incomes. *The Economic Journal*, 30(119):348–361.

References

Denuit, M. and Dhaene, J. (2012a). Convex order and comonotonic conditional mean risk sharing. *Insurance: Mathematics and Economics*, 51(2):265–270.

Denuit, M. and Dhaene, J. (2012b). Convex order and comonotonic conditional mean risk sharing. *Insurance: Mathematics and Economics*, 51(2):265–270.

Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. Wiley.

Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2006). *Actuarial theory for dependent risks: measures, orders and models*. John Wiley & Sons.

Diestel, R. (2006). *Graph Theory*. Springer Verlag.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 214–226.

Easley, D., Kleinberg, J., et al. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*, volume 1. Cambridge University Press.

Evans, J. (2017). *Optimization algorithms for networks and graphs*. CRC Press.

Feng, R., Liu, C., and Taylor, S. (2023). Peer-to-peer risk sharing with an application to flood risk pooling. *Annals of Operations Research*, 321(1):813–842.

References

Freeman, L. C. et al. (1979). Centrality in social networks: Conceptual clarification. *Social network: critical concepts in sociology. Londres: Routledge*, 1:238–263.

Gibbons, A. (1985). *Algorithmic graph theory*. Cambridge university press.

Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144.

Gollier, C. and Schlesinger, H. (1996). Arrow's theorem on the optimality of deductibles: a stochastic dominance approach. *Economic Theory*, 7:359–363.

Gonzalez, D., McDaniel, M., Kenney, G. M., and Skopec, L. (2021). Perceptions of unfair treatment or judgment due to race or ethnicity in five settings. *Washington, DC: Urban Institute*.

Gupta, S., Anderson, R. M., and May, R. M. (1989). Networks of sexual contacts: implications for the pattern of spread of hiv. *Aids*, 3(12):807–818.

Hakimi, S. L. (1962). On realizability of a set of integers as degrees of the vertices of a linear graph. i. *Journal of the Society for Industrial and Applied Mathematics*, 10(3):496–506.

Hall, K. M. (1970). An r-dimensional quadratic placement algorithm. *Management science*, 17(3):219–229.

Hardt, M., Price, E., and Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in neural information processing systems*, 29:3315–3323.

References

Hardy, G., Littlewood, J., and G. Polya, . (1929). Some simple inequalities satisfied by convex functions. *British Mathematics Journal*, 38:145–152.

Hardy, G., Littlewood, J., and Polya, G. (1934). *Inequalities*. Cambridge University Press.

Harris, J. M. (2008). *Combinatorics and graph theory*. Springer.

Havel, V. (1955). A remark on the existence of finite graphs. *Časopis pro pěstování' matematiky*, 80:477–480.

Hougaard, J. L. and Moulin, H. (2018). *Allocation in networks*. MIT Press.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press.

Kaas, R., Dhaene, J., Vyncke, D., Goovaerts, M., and Denuit, M. (2002). A simple geometric proof that comonotonic risks have the convex-largest sum. *ASTIN Bulletin*, 32(1):71–80.

Kaas, R., Goovaerts, M., and Dhaene, J. (2008). *Modern actuarial risk theory*. Springer Science & Business Media.

Kaas, R., van Heerwaarden, A. E., Goovaerts, M. J., et al. (1994). *Ordering of actuarial risks*. Kluwer.

References

Karimi, A.-H., Barthe, G., Balle, B., and Valera, I. (2020). Model-agnostic counterfactual explanations for consequential decisions. In *International conference on artificial intelligence and statistics*, pages 895–905. PMLR.

Karimi, A.-H., Schölkopf, B., and Valera, I. (2021). Algorithmic recourse: from counterfactual explanations to interventions. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 353–362.

Karlin, S. and Novikoff, A. (1963). Generalized convex inequalities. *Pacific Journal of Mathematics*, 13:1251–1279.

Kearns, M. and Roth, A. (2019). *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.

Kusner, M. J., Loftus, J., Russell, C., and Silva, R. (2017). Counterfactual fairness. In *Advances in Neural Information Processing Systems*, pages 4066–4076.

Lee, A. J. (2019). *U-statistics: Theory and Practice*. CRC Press.

Luong, B. T., Ruggieri, S., and Turini, F. (2011). *k*-nn as an implementation of situation testing for discrimination discovery and prevention. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 502–510.

References

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and its Applications*, volume 143. Academic Press.

Martínez Pería, F. D., Massey, P. G., and Silvestre, L. E. (2005). Weak matrix majorization. *Linear Algebra and its Applications*, 403:343–368.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444.

Meyer, R. (2015). Could a bank deny your loan based on your facebook friends? *The Atlantic*, 2015/09/25.

Morris, M. (1995). Data driven network models for the spread of infectious disease. *Epidemic models: their structure and relation to data*, 5:302–22.

Muirhead, R. F. (1903). Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. *Proceedings of the Edinburgh Mathematical Society*, 21:144–162.

Müller, A. and Stoyan, D. (2002). *Comparison methods for stochastic models and risks*.

Newman, M. (2018). *Networks*. Oxford university press.

Newman, M. E. (2003). Mixing patterns in networks. *Physical review E*, 67(2):026126.

References

Newman, M. E. (2012). Communities, modules and large-scale structure in networks. *Nature physics*, 8(1):25–31.

Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113.

Ohlin, J. (1969). On a class of measures of dispersion with application to optimal reinsurance. *ASTIN Bulletin: The Journal of the IAA*, 5(2):249–266.

Pascoe, E. A. and Richman, L. S. (2009). Perceived discrimination and health: a meta-analytic review. *Psychological Bulletin*, 135(4):531.

Pass, R. (2019). *A Course in Networks and Markets: Game-theoretic Models and Reasoning*. MIT Press.

Pearl, J. and Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic books.

Peel, L., Peixoto, T. P., and De Domenico, M. (2022). Statistical inference links data and theory in network science. *Nature Communications*, 13(1):6794.

Rothschild, M. and Stiglitz, J. E. (1978). Increasing risk: I. a definition. In *Uncertainty in economics*, pages 99–121. Elsevier.

References

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215.

Rüschendorf, L. (2013). *Mathematical risk analysis*. Springer.

Schmitt, M. T., Branscombe, N. R., Postmes, T., and Garcia, A. (2014). The consequences of perceived discrimination for psychological well-being: a meta-analytic review. *Psychological bulletin*, 140(4):921.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. Springer.

Shalmani, F. M. (2021). *Fairness in Social Network Analysis: Measures and Algorithms*. Michigan State University PhD Thesis.

Shrum, W., Cheek Jr, N. H., and MacD, S. (1988). Friendship in school: Gender and racial homophily. *Sociology of Education*, pages 227–239.

Trudeau, R. J. (1993). *Introduction to graph theory*. Dover Publications.

Van Houtven, C. H., Voils, C. I., Oddone, E. Z., Weinfurt, K. P., Friedman, J. Y., Schulman, K. A., and Bosworth, H. B. (2005). Perceived discrimination and reported delay of pharmacy prescriptions and medical tests. *Journal of general internal medicine*, 20:578–583.

Viger, F. and Latapy, M. (2005). Fast generation of random connected graphs with prescribed degrees. *arXiv*, 0502085.

References

Wu, X.-Z., Percus, A. G., and Lerman, K. (2017). Neighbor-neighbor correlations explain measurement bias in networks. *Scientific Reports*, 7(1):5576.

Yamaguchi, K. (1990). Homophily and social distance in the choice of multiple friends an analysis based on conditionally symmetric log-bilinear association models. *Journal of the American Statistical Association*, 85(410):356–366.

Zemel, R., Wu, Y., Swersky, K., Pitassi, T., and Dwork, C. (2013). Learning fair representations. In *International Conference on Machine Learning*, pages 325–333.