Complex Networks

- reprezentace reálného systému pomocí grafu
- sociální síť
  - vrcholy = aktéři – uživatelé, skupiny, komunity
  - hrany = interakce – friendship, collaboration

Facebook = (users, friendship) ... neorientovaný graf bez smyček

- Milgram experiment 1963 v Omaha
  - ↳ structure of social networks
  - ↳ kolik skoků je potřeba aby kdokoli poslal dopis komukoli?
    - ⇒ six degrees of separation ... v průměru stačilo 6
  - ↳ small world phenomenon
    - ↳ vágní definice:
      1. vznikají clustery vrcholů
      2. krátké vzdálenosti mezi vrcholy
      3. ta síť je řídká

- Citation network: (papers, references) ... wikipedia

- Dynamic systems
  - → systém se skládá z množiny menších subsystémů co spolu interagují
  - → vnitřní dynamika těch subsystémů většinou popsaná diff. rovnicema
  - → synchronizace = interakce mezi páry subsystémů
  - → příklad: Světové klima
    - → simulovat celý svět naráz je nemožné
    - → řeší se menší regiony / fenomény a jak spolu interagují
  - → pokud známe vnitřní dynamiku a synchronizace & máme dost dat
    - ⇒ je to řešitelné, ale problémy s výpočetní náročností a stabilitou
  - → reálné systémy jsou vosty
    - klima – není dost dat
    - mozek – neznáme vnitřní procesy

- Strukturální × funkční konektivita
  - příklad: regiony mozku
    - mozek je nějaká fyzická síť → můžu udělat strukturální síť
    - magnetická rezonance umí určit jak silně spolu jednotlivé regiony komunikují
      ↳ funkční síť

  ! ale ty dvě sítě si u lidského mozku neodpovídají

  → zjistilo se, že mozek je small world
    ↳ regiony mozku = clustery, komunikace mezi nimi je velmi rychlá

  → Alzheimer: denser clusters, longer distances
    ↳ těžce porušuje small world
    ↳ přestává být efektivní přenos informace mezi regiony

- Protein interactions
  vrcholy: proteiny
  hrany: jak spolu ty proteiny interagují v daném prostředí (uvnitř buňky,...)
  → cíl: zjistit funkci proteinu podle jeho sousedů
    ↳ PPIN = Protein-Protein Interaction Network

  - neighborhood methods: funkce proteinu ~ funkce jeho sousedů
  - communities: vrcholy rozložíme na communities $V = C_1 \dot\cup C_2 \dot\cup \dots \dot\cup C_n$
    ↳ jednotlivé communities by uvnitř měly být dense
    ↳ každá komunita by snad měla mít nějakou hlavní funkci

- Internet – (IP adresy, kabely)
  - percolation theory = resistance sítě vůči útokům
    ↳ kolik vrcholů/hran (DNS servery...) můžu smazat aby síť zůstala souvislá

- WWW – (web pages, references)
  - řeší se spread of fake news, communities, business analytics

- Large networks ↑ některé sítě jsou opravdu obrovské
  - WWW, internet není reálně možné analyzovat jako celek nebo někam uložit
  → dělá se sampling a modeluje se to přes random grafy
  → jsou metody které se hranám přistupují víc jako k pravděpodobnosti
  → idea grafových limit: $P[\text{event}] \to 0$ as $|V(G)| \to \infty$

- Stupně vrcholů reálných sítí

→ chceme aby to reprodukovaly ty random grafy kterými to budeme modelovat

→ reálné sítě se chovají jako:

- few very high degree nodes
- many low degree nodes

} Scale-free property

random attacks jsou OK → většina vrcholů není důležitá

targeted attacks jsou problém → potřeba chránit high degree nodes = Hubs

- Měření důležitosti vrcholu

→ různé způsoby na základě aplikace

1) degree centrality = normalized vertex degree $= \frac{deg(v)}{n-1}$ ... $n = |V|$

↳ searching influencers in social networks

2) eccentricity = max. distance of a vertex to any other

↳ finding the airport with best routes to any other

↳ důležitým vrcholům se říká hubs nebo centralities

- Super centralities

→ nejenom, že má hodně friends, ale má i hodně good friends

→ nejen že z tohodle letiště je to všude nejblíž, ale zároveň je většina spojů přes tohle letiště efektivní

Small-World Graphs

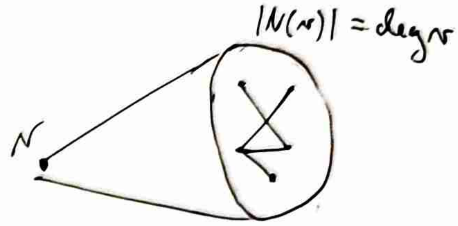
- locally dense subgraphs $\rightarrow$ nodes create clusters ... $\rho(\text{cluster}) \in \mathcal{O}(1)$
- globally the graph is sparse ... $\rho(G) \in \mathcal{O}(\frac{1}{n})$, $\mathcal{O}(\frac{\log n}{n})$ ... $\lim_{n\to\infty} \rho = 0$
- short average distance between nodes ... $d(u,v) \ll n$

Def: Density of $G=(V,E)$, $|V|=n$, $|E|=m$ is $\rho(G) := \frac{m}{\binom{n}{2}}$

Def: For $v \in V$, define

$N_G(v) := \{u \in V \mid \{u,v\} \in E\}$ ... neighbors

$G[N_G(v)]$ ... induced graph



$C(v) := \rho(G[N_G(v)]) = \frac{|E(G[N(v)])|}{\binom{\deg v}{2}}$ ... vertex clustering coefficient

$\hookrightarrow$ for $\deg(v) \in \{0,1\}$, $C(v) = 0$

? When do we have $C(v)=0 \ \forall v \in V$? $\rightarrow$ paths, cycles, trees, bipartite graphs

$\forall v \in V: C(v) = 0 \iff$ no $\triangle$ in G

Def: The average clustering coefficient is

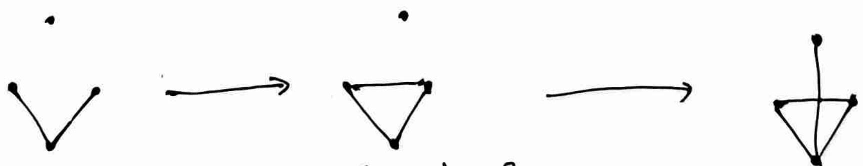
$$C(G) := \frac{1}{n}\sum_{v\in V} C(v)$$

$C(G) = 0 \iff$ no $\triangle$ in G

$C(G) = 1 \iff G$ is a disjoint union of complete graphs

Note: $C(v)$ can be interpreted as the probability that two neighbours of v are connected.

! C is not a monotone property with respect to adding edges



$C=0$ $\qquad$ $C=\frac{1}{4}(0+3)=\frac{3}{4}$ $\qquad$ $C=\frac{1}{4}(0+2+\frac{1}{3}) < \frac{3}{4}$

• **Turan graphs**

Problem: What is the max. number of edges a graph of size n can have for $C(G)=0$?

↳ famous problem from extremal graph theory

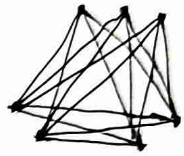
→ use the fact that $\Delta \notin G$

Def: $ex(n, K_{r+1}) :=$ max # of edges in a graph of size n not containing K_{r+1}

→ the answer for our problem is $ex(n, K_3)$

r-partite graphs can have a lot of edges but can not contain K_{r+1}

Def: Turán graph $T_r(n)$: complete r-partite graph of size n with all partites having sizes as similar as possible.



$T_3(7)$

Denote $t_r(n) := |E(T_r(n))|$

Theorem: For $n = q\cdot r + s$, $s<r$: $t_r(n) = \left(1-\frac{1}{r}\right)\cdot\frac{n^2}{2} - \frac{s(r-s)}{2r}$ → 0 for $s=0$

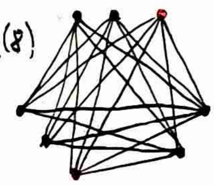
Pf: Consider $n = q\cdot r$

↳ r partites, each of size q

⇒ between any pair: complete bipartite graph $K_{\frac{n}{r},\frac{n}{r}}$ ⇒ $t_r(n) = \binom{r}{2}\cdot\left(\frac{n}{r}\right)^2 = \frac{r-1}{r}\cdot\frac{n^2}{2}$

Consider $n = q\cdot r + s$

$T_3(8)$



↳ $T(q\cdot r+s) \sim T(q\cdot r)$ and extra s vertices, each one in a different partite of $T(q,r)$

⇒ for $T(q\cdot r)$ we know $\frac{r-1}{r}\cdot\frac{(q\cdot r)^2}{2}$ edges

- connect extra vertices between each other → $\binom{s}{2}$
- connect them to partites of $T(q,r)$ → $s\cdot(r-1)\cdot q$

⇒ in total:

$$t_r(n) = \frac{r-1}{r}\cdot\frac{(q\cdot r)^2}{2} + \binom{s}{2} + s\cdot(r-1)\cdot q$$

$$= \frac{r-1}{r}\cdot\frac{(n-s)^2}{2} + s\cdot(r-1)\cdot\frac{n-s}{r} + \binom{s}{2} = \frac{r-1}{r}(n-s)\left(\frac{n-s}{2}+s\right) + \binom{s}{2}$$

$$= \frac{r-1}{r}(n-s)\frac{n+s}{2} + \binom{s}{2} = \frac{r-1}{r}\cdot\frac{n^2-s^2}{2} + \frac{s(s-1)}{2} = \frac{r-1}{r}\cdot\frac{n^2}{2} - \frac{s^2}{2}\cdot\frac{r-1}{r} + \frac{s(s-1)}{2}$$

$$= \frac{r-1}{r}\cdot\frac{n^2}{2} - \frac{s}{2r}\left(s(r-1) - r(s-1)\right) = \left(1-\frac{1}{r}\right)\cdot\frac{n^2}{2} - \frac{s(r-s)}{2r}$$ ∎

Corollary: For large n: $t_r(n) \sim \left(1-\frac{1}{r}\right)\cdot\frac{n^2}{2}$ → for $r=2$ we have $t_2(n) \sim \frac{n^2}{4}$

Theorem (Turán 1940): G is of size n & $|E| > t_r(n) \Rightarrow G$ contains K_{r+1}.

↳ this means that Turán graphs solve the max-edge problem

↳ max # edges so that $C(G)=0$ is $\sim \frac{n^2}{4}$

• Average path length

$$L(G) := \frac{1}{\binom{n}{2}} \sum_{\substack{\{u,v\}\subseteq V \\ u\neq v}} d(u,v) = \frac{1}{2}\cdot\frac{1}{\binom{n}{2}} \sum_{\substack{u,v\in V \\ u\neq v}} d(u,v) \quad , \quad \frac{1}{\binom{n}{2}} = \frac{2}{n(n-1)}$$

Examples:

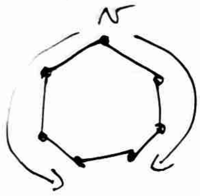
• Complete graphs $L(K_n) = 1$

• Stars, S_n

$$L(S_n) = \frac{2}{(n+1)n}\left(n\cdot 1 + \binom{n}{2}\cdot 2\right) = \frac{2}{n+1}\left(1 + \frac{n-1}{2}\cdot 2\right) = \frac{2n}{n+1} \longrightarrow 2$$

→ 1 center + n nodes

• Cycles, C_n



$$n = 2k+1 \Rightarrow L(C_n) = \frac{1}{n(n-1)}\cdot\Big(\underbrace{n}_{u,v\in V \text{ / } n \text{ vertices}}\cdot\big(2\cdot\overbrace{(1+2+\dots+k)}^{\frac{k(k+1)}{2}}\big)\Big) = \frac{1}{n(n-1)}\cdot n\cdot\frac{n-1}{2}\cdot\frac{n+1}{2} = \frac{n+1}{4}, \quad k = \frac{n-1}{2}$$

$$n = 2k+2 \Rightarrow L(C_n) = \frac{1}{n(n-1)}\cdot n\cdot\left(\frac{n}{2} + 2\cdot(1+2+\dots+k)\right) = \frac{1}{n-1}\left(\frac{n}{2} + \left(\frac{n}{2}-1\right)\left(\frac{n}{2}\right)\right) = \frac{n^2}{4(n-1)}$$

$k = \frac{n}{2} - 1$

$$L(C_n) \to \frac{n}{4}$$

• Small world comparison of simple graphs

• Stars S_n: $L(S_n) \sim 2$, $C(S_n) = 0$, $\rho(S_n) = \frac{n}{\binom{n+1}{2}} = \frac{2n}{n(n+1)} = \frac{2}{n+1} \to 0$

↳ short distances, no clustering, low density

• Cycles C_n: $L(C_n) \sim \frac{n}{4}$, $C(C_n) = 0$, $\rho(C_n) = \frac{n}{\binom{n}{2}} = \frac{2}{n-1} \to 0$

↳ long distances, no clustering, low density

• Complete graphs K_n: $L(K_n) = 1$, $C(K_n) = 1$, $\rho(K_n) = 1$

↳ short distances, absolute clustering, full density

• Wheels W_n: $L(W_n) \sim 2$, $C(W_n) \sim \frac{2}{3}$, $\rho(W_n) = \frac{2n}{\binom{n+1}{2}} = \frac{4}{n+1} \to 0$

↳ short distances, high clustering, low density

$$C(W_n) = \frac{1}{n+1}\left(\frac{n}{\binom{n}{2}} + n\cdot\frac{2}{3}\right) = \frac{1}{n+1}\left(\frac{2}{n-1} + \frac{2n}{3}\right) \longrightarrow \frac{2}{3}$$

$$L(W_n) = \frac{1}{n(n+1)}\left(n\cdot 1 + n\cdot(3 + (n-3)\cdot 2)\right) = \frac{1+3+2n-6}{n+1} = \frac{2(n-1)}{n+1} \longrightarrow 2$$

$C_n + v$

• <u>Complete bi-graph</u> $K_{m,m}$: $L(K_{m,m}) \sim \frac{3}{2}$, $C(K_{m,m}) = 0$, $\rho(K_{m,m}) \sim \frac{1}{2}$

↳ short distances, no clustering, high density

$m \quad m$

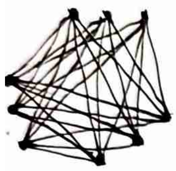
$$L(K_{m,m}) = \frac{2}{(2m)(2m-1)}\left(m \cdot m + 2 \cdot \binom{m}{2} \cdot 2\right) = \frac{m^2 + 2m(m-1)}{m(2m-1)} = \frac{m + 2(m-1)}{2m-1} = \frac{3m-2}{2m-1} \to \frac{3}{2}$$

$$C(K_{m,m}) = 0 \because \triangle \notin K_{m,m}, \quad \rho(K_{m,m}) = \frac{m^2}{\binom{2m}{2}} = \frac{2m^2}{2m(2m-1)} = \frac{m}{2m-1} \to \frac{1}{2}$$

• <u>Complete k-partite graphs</u> K_{k*m}: $L(K_{k*m}) \sim \frac{k+1}{k}$, $C(K_{k*m}) \sim \frac{k-2}{k-1}$, $\rho(K_{k*m}) \sim \frac{k-1}{k}$

↳ short distances, high clustering $(k>2)$, high density

3 partite



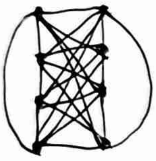
$$L(K_{k*m}) = \frac{1}{(km)(km-1)}(km)\left(1 \cdot (k-1)m + 2 \cdot (m-1)\right) = \frac{km - m + 2m - 2}{km-1} = \frac{m(k+1)-2}{m \cdot k - 1} \to \frac{k+1}{k}$$

$$C(K_{k*m}) = \frac{1}{km} \cdot (km) \cdot \rho(K_{(k-1)*m}) = \frac{m(k-2)}{m(k-1)-1} \to \frac{k-2}{k-1}$$

$$\rho(K_{k*m}) = \frac{1}{\binom{km}{2}} \cdot \frac{1}{2}(k \cdot m \cdot (k-1)m) = \frac{km(k-1)m}{km(km-1)} = \frac{m(k-1)}{m \cdot k - 1} \to \frac{k-1}{k}$$

• <u>$B_{m,m}$ = 2 cycles + K_{m*m}</u>: $L(B_{m,m}) \sim \frac{3}{2}$, $C(B_{m,m}) \sim 0$, $\rho(B_{m,m}) \sim \frac{1}{2}$

↳ short distances, no clustering, high density



$$L(B_{m,m}) = \frac{1}{2m(2m-1)} \cdot 2m\left(2 \cdot 1 + m \cdot 1 + (m-3) \cdot 2\right) = \frac{m + 2 + 2m - 6}{2m-1} = \frac{3m-4}{2m-1} \to \frac{3}{2}$$

$$C(B_{m,m}) = \frac{1}{2m} \cdot 2m \cdot \frac{3m}{\binom{m+2}{2}} = \frac{6m}{(m+2)(m+1)} \to 0$$

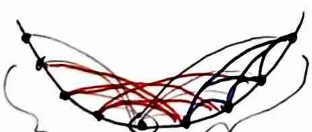
$$\rho(B_{m,m}) = \frac{m \cdot m + 2m}{\binom{2m}{2}} = \frac{2(m^2+2m)}{2m(2m-1)} = \frac{m+2}{2m-1} \to \frac{1}{2}$$

• <u>Ring Lattice</u> $C_{m,k}$: $L(C_{m,k}) \sim \frac{m}{2k}$, $C(C_{m,k}) \sim \frac{3}{4}$, $\rho(C_{m,k}) = \frac{k}{m-1}$

↳ C_m + each vertex is connected to k nearest neighbors

→ usually $k = 2l$ & $m > \frac{3k}{2}$

• <u>clustering</u>



$k/2 \quad v \quad k/2$

$$2 \cdot \binom{k/2}{2} + \overbrace{\left(1 + 2 + \dots + (k/2 - 1)\right)}^{\frac{(k/2-1)(k/2)}{2} = \binom{k/2}{2}} = 3 \cdot \binom{k/2}{2}$$

$$\Rightarrow C(C_{m,k}) = C(v) = \frac{3 \cdot \binom{k/2}{2}}{\binom{k}{2}} = \frac{3\left(\frac{k}{2}\right)\left(\frac{k}{2}-1\right)}{k(k-1)} = \frac{3}{4} \cdot \frac{k-2}{k-1} \to \frac{3}{4}$$

• <u>average path length</u> $m = 2l + 1$

$$L(C_{m,k}) \approx \frac{1}{m(m-1)} \cdot m \cdot 2\left(\frac{k}{2} \cdot 1 + \frac{k}{2} \cdot 2 + \frac{k}{2} \cdot 3 + \dots + \frac{k}{2} \cdot \frac{m}{k}\right) = \frac{1}{m-1} \cdot k \cdot \frac{\frac{m}{k}\left(\frac{m}{k}+1\right)}{2}$$

$$= \frac{m}{m-1} \cdot \left(\frac{m}{2k} + \frac{1}{2}\right) \to \frac{m}{2k}$$

$l = q \cdot \frac{k}{2} + r$

$q \approx \frac{2l}{k} \approx \frac{m}{k}$

• <u>density</u>: $\rho = \frac{(mk)/2}{\binom{m}{2}} = \frac{k}{m-1}$

Generating Small World Graphs

→ ring lattice has good clustering and low density but long distances
→ we randomly rewire some edges to improve distances

• **Watts – Strogatz model**

1. Start with ring lattice $C_{n,k}$
2. For $v \in V(C_{n,k})$:
3. For $\forall$ rightwards edge $\{v,u\}$: # $k/2$ of them in total
4. rewire $\{v,u\}$ with probability π

↳ change $\{v,u\}$ to $\{v,r\}$ where r is chosen uniformly from $\{r \in V \mid r \neq v \;\&\; \{v,r\} \notin E\}$ ... avoiding self loops and multiedges

Properties: keeps good clustering and introduces short distances

• **Erdős – Renyi random graph** $G_{n,p}$

1. take vertices $V = \{1, \dots n\}$
2. for $\forall \{i,j\} \subset V$: insert an edge wit probability $p \approx$ density of final graph

$WS(\pi \to 1) \rightsquigarrow ER\left(p = \frac{k}{n-1}\right)$

Properties: $k :=$ average degree $\Rightarrow p \approx \frac{k}{n}$

$L(G_{n,p}) \sim \frac{\log(n)}{k} \sim \frac{1}{p} \cdot \frac{\log(n)}{n}$ ⇒ short distances

$\left.\begin{array}{l} C(G_{n,p}) \sim \frac{k}{n} \sim p \\ \rho(G_{n,p}) \sim \frac{k}{n-1} \sim p \end{array}\right\}$ clustering $\approx$ density ⇒ bad for small world

Small World Coefficient

→ all small world notions are somewhat relative ⇒ we need a reference model

↳ simplest model: for graph G generate random graph R_G with the same number of vertices and edges

→ we can say that G is small world $\iff L(G) \gtrsim L(R_G)$ & $C(G) \gg C(R_G)$

→ there are many ways of measuring "small-worldliness" of G

Def: The small-world coefficient of G is $\sigma(G) := \dfrac{C(G)/\mathbb{E}[C(R_G)]}{L(G)/\mathbb{E}[L(R_G)]}$, want $\sigma > 1$.

Centralities

↳ important nodes in a network

- Degree centrality: $d(v) := \frac{\deg(v)}{n}$ ... higher = better

- Excentricity: $\epsilon(v) := \max_{u \in V} d(v,u)$ ... goes in the other direction

- Closeness centrality

$$\ell(v) := \frac{1}{n-1} \sum_{u \in V} d(v,u)$$ ... average distance from v to other vertices

$$C_C(v) := \frac{1}{\ell(v)} = \frac{n-1}{\sum_u d(v,u)}$$ ... higher = better

- Betweenness centrality

→ expresses how much communication goes throug v

v = inner vertex

↑

$$C_B(v) = \frac{\#\text{shortest paths in } G \text{ through } v}{\#\text{shortest paths in } G}$$

$$C_B(v) := \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}, \quad \sigma_{st}(v) = \#\text{ shortest paths between } s \text{ and } t \text{ through } v$$
$$\sigma_{st} = \# \text{ ——— " ———}$$

↳ $\sigma_{ss} = 1$, $\sigma_{st}(s) = \sigma_{st}(t) = 0$

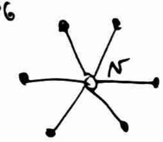
↳ $\{s,t\} \subseteq V$, $s \neq t$, $s \neq v \neq t$

→ iterating pairs ⇒ $\binom{n}{2}$ summands

Examples

- Star S_n

S_6



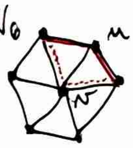
leafs: BC = 0

center: $C_B(v) = \sum_{s \neq v \neq t} \frac{1}{1} = \binom{n}{2}$

$C_B(v) = 0 \iff G[N(v)] \cong K_{\deg(v)}$

- Wheel W_n

W_6

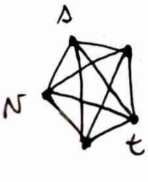


outer: $C_B(u) = \frac{1}{2}$

center: • adjacent vertices contribute 0 → n of these
- nonadjacent vertices with a common neighbor: $\frac{1}{2}$
- more distant contribute 1 → $\binom{n}{2} - n - n$ of these

$$\Rightarrow C_B(v) = n \cdot \frac{1}{2} + \left(\binom{n}{2} - 2n\right) \cdot 1 = \frac{n}{2} + \frac{n(n-1)}{2} - 2n = \frac{n}{2}(1 + (n-1) - 4)$$
$$= \frac{1}{2} n(n-4)$$

- Complete graph K_n



$\forall v : C_B(v) = 0$

↳ all vertices are important ⇒ none of them is

• Global betweenness centrality

$$\overline{C}_B(G) := \frac{1}{n} \sum_{v \in V} C_B(v) \quad \ldots \text{ average of vertex b. centralities}$$

Recall: $L(G) = \frac{1}{\binom{n}{2}} \sum_{\substack{\{u,v\} \subseteq V \\ u \neq v}} d(u,v)$

Theorem: Let G be an undirected graph and $|V(G)| = n$. Then

$$\overline{C}_B(G) = \frac{n-1}{2}\left(L(G) - 1\right)$$

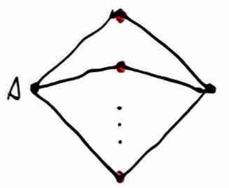
Proof: Define overall BC as $C_{BS} := n\,\overline{C}_B(G) = \sum_{v \in V} C_B(v)$

↳ how much does a pair $\{s,t\} \subseteq V$, $s \neq t$ contribute to C_{BS}?

1) $d(s,t) = 1$: s —• t

↳ no shortest path $s \leftrightarrow t$ uses any other edges ⇒ contribution $= 0$

2) $d(s,t) = 2$:



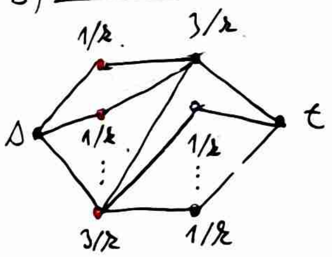
→ exactly $k \geq 1$ vertices between s and t

⇒ exactly k shortest paths

⇒ so each vertex contributes $\frac{1}{k}$

$\Big\}$ contrib. $= k \cdot \frac{1}{k} = 1$

3) $d(s,t) \geq 3$



→ exactly k shortest paths between s and t

→ imagine vertices in layer ℓ

↳ each shortest path uses exactly 1 vertex from ℓ

⇒ overall contrib. from layer $\ell = 1$

⇒ contrib of s,t = # layers = $d(s,t) - 1$

$$\Rightarrow \overline{C}_B(G) = \frac{1}{n} C_{BS} = \frac{1}{n} \sum_{\substack{\{u,v\} \subseteq V \\ u \neq v}} \left(d(u,v) - 1\right) \quad \ldots \text{ rename } \{s,t\} \rightsquigarrow \{u,v\}$$

$$= \frac{1}{n} \sum d(u,v) - \frac{1}{n} \cdot \binom{n}{2} = \frac{1}{n} \sum d(u,v) - \frac{n-1}{2}$$

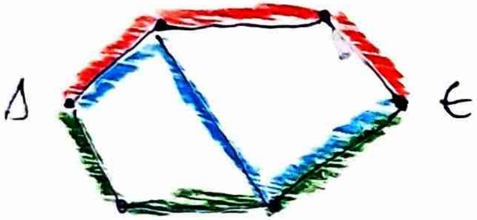
$$\Rightarrow \frac{2\overline{C}_B(G)}{n-1} = \frac{1}{\binom{n}{2}} \sum d(u,v) - 1 = L(G) - 1 \qquad \blacksquare$$

Note: We always assume G connected.

Alternative proof

Lemma: Let G be a connected graph and $s \neq t$ vertices. Then $\sum_{\substack{v \in V \\ s \neq v \neq t}} \sigma_{st}(v) = (d(s,t)-1)\,\sigma_{s,t}$

Pf:



LHS: # shortest paths $s \leftrightarrow t$ using v for all $v \in V \setminus \{s,t\}$

RHS: Enumerate all shortest paths between $s \leftrightarrow t$

↳ each path contributes by all its inner vertices … $d-1$

$\Rightarrow$ together $(d(s,t)-1)\cdot\sigma_{s,t}$ ∎

Theorem: $\overline{C_B}(G) = \frac{n-1}{2}\left(L(G)-1\right)$

Pf:
$$\overline{C_B}(G) = \frac{1}{n}\sum_{v \in V} C_B(v) = \frac{1}{n}\sum_{v \in V}\sum_{s \neq v \neq t}\frac{\sigma_{st}(v)}{\sigma_{st}} = \frac{1}{n}\sum_{s \neq t}\frac{1}{\sigma_{st}}\sum_{v \in V\setminus\{s,t\}}\sigma_{st}(v)$$

$$= \frac{1}{n}\sum_{s \neq t}\frac{1}{\sigma_{st}}\left(d(s,t)-1\right)\cdot\sigma_{st} = \frac{1}{n}\left(\sum_{s \neq t} d(s,t) - \sum_{s \neq t} 1\right)$$

$$= \frac{1}{n}\left(\binom{n}{2}L(G) - \binom{n}{2}\right) = \frac{n-1}{2}\left(L(G)-1\right)$$ ∎

- Betweenness and adding edges
  → is BC monotone?

Lemma: Global BC decreases with adding edges

Pf: $\overline{C_B}(G) = \frac{n-1}{2}(L(G)-1)$ and $L(G)$ decreases when adding edges

Lemma: Let G be a graph of size n and $G' = G + uv$, where $uv \notin E(G)$. Then

$$\overline{C_B}(G') \leq \overline{C_B}(G) - \frac{d-1}{n}, \quad d = d_G(u,v) > 1$$

↳ Global BC decreases at least by $\frac{d-1}{n}$

Pf: $1 = d_{G'}(u,v) < d_G(u,v) = d$

$d_{G'}(x,y) \leq d_G(x,y)$ for $\forall (x,y) \neq (u,v)$

$$\Rightarrow \sum_{\{x,y\} \subseteq V} d_{G'}(x,y) \leq \sum_{\{x,y\} \subseteq V} d_G(x,y) - (d-1)$$

$$\binom{n}{2} \cdot L(G') \leq \binom{n}{2} \cdot L(G) - (d-1)$$

$$\frac{n-1}{2} L(G') \leq \frac{n-1}{2} L(G) - \frac{d-1}{n}$$ & recall $\frac{n-1}{2} L(G) = \overline{C_B}(G) + \frac{n-1}{2}$ ■

- Global BS of spanning trees

Theorem: Let $G = (V,E)$ be a graph and $G' = (V,E')$ its spanning subgraph. Then

$$\overline{C_B}(G) \leq \overline{C_B}(G') - \frac{r}{n}, \quad r = |E \setminus E'|$$

Pf: Add edges of $E \setminus E'$ iteratively r times

↳ each pair x,y has distance $d = d(x,y) \geq 2$

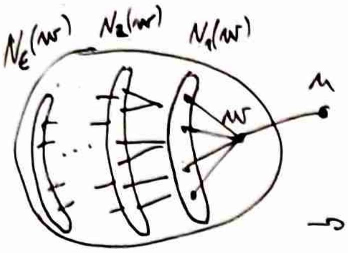
$\Rightarrow$ BC decreased by $\frac{d-1}{n} \geq \frac{1}{n}$ each step (previous lemma) ■

Corollary: Let $G = (V,E)$ be a graph and T its spanning tree. Then

$$\overline{C_B}(G) \leq \overline{C_B}(T) - \frac{m-n+1}{n}$$ ... ∵ $|E(T)| = n-1$

• Betweenness and adding vertices

→ lets just try adding an extra leaf to the graph



→ there are layers of k-neighborhoods

$$N_k(w) := \{v \in V \mid d(u,v) = k\}$$

→ adding a leaf does not affect the shortest paths between old nodes

⇒ but we need to consider shortest paths which have u as the starting node

Theorem: Let G be a graph of size n, $w \in V$ and let $G' = G + u + \{u,w\}$. Then

$$\overline{C}_B(G') = \frac{1}{n+1}\left(n \cdot \overline{C}_B(G) + \sum_{k=1}^{\varepsilon(w)} k \cdot |N_k(w)|\right)$$

, $\varepsilon(w)$ = eccentricity of w in G

Pf: By adding u we add n new $d_{G'}(u,v)$

$$\sum_{\{x,y\} \subseteq V'} d_{G'}(x,y) = \sum_{\{x,y\} \subseteq V} d_G(x,y) + \sum_{k=1}^{\varepsilon(w)} (k+1)|N_k(w)| + \underbrace{1}_{= d(u,w)}$$

$$= \sum d_G(x,y) + \underline{\sum_k k \cdot |N_k(w)|} + n = \sum d_G(x,y) + \underline{\Delta} + n$$

$$\Rightarrow \frac{1}{\binom{n+1}{2}} \sum d_{G'}(x,y) = L(G') = \frac{2}{n(n+1)} \sum d_G(x,y) + \frac{2}{n(n+1)} \Delta + \frac{2}{n+1}$$

$$= \frac{n-1}{n+1} \cdot \frac{1}{\binom{n}{2}} \cdot \sum + \ldots = \frac{n-1}{n+1} L(G) + \frac{2}{n(n+1)} \Delta + \frac{2}{n+1}$$

• $$\overline{C}_B(G') = \frac{n}{2}\left(L(G') - 1\right) = \frac{n}{2}\left(\frac{n-1}{n+1} L(G) + \frac{2}{n(n+1)} \Delta + \frac{2}{n+1} - 1\right)$$

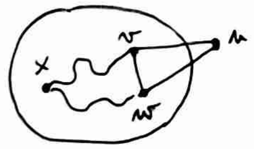
$$\Rightarrow \overline{C}_B(G') = \frac{n}{n+1} \cdot \frac{n-1}{2}\left(L(G) - 1 + 1\right) + \frac{1}{n+1} \Delta + \frac{n}{2} \cdot \frac{1-n}{n+1}$$

$$= \frac{n}{n+1} \cdot \overline{C}_B(G) + \frac{1}{n+1} \Delta + \frac{n(n-1)}{2(n+1)} + \frac{n(1-n)}{2(n+1)} = \frac{1}{n+1}\left(n \overline{C}_B(G) + \Delta\right) \quad \blacksquare$$

Theorem: Let G be a graph of size n and $G' = G + u + uv + uw$ s.t. $1 \leq d(v,w) \leq 2$.

$$\overline{C}_B(G') \geq \frac{1}{n+1}\left(n \cdot \overline{C}_B(G) + n - 2\right)$$

- Pf: For $d(v,w) = 1$ introduce a triangle u, v, w



→ how much do v and w contribute to overall BC ?

↳ there are $n-2$ vertices $x \in V(G) \setminus \{v,w\}$

→ paths between x and u must use v or w ⇒ $+1$

$C_B'(u) = 0$

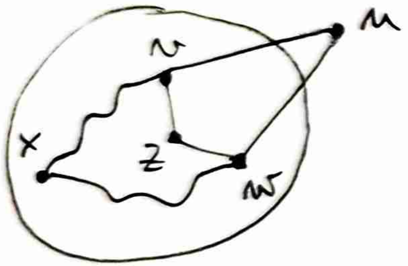
$C_B'(v) + C_B'(w) = C_B(v) + C_B(w) + n - 2$

→ other vertices (not only v, w) can also contribute

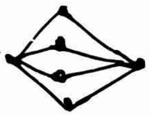
$$\Rightarrow \sum_{y \in V(G')} C_B'(y) \geq \sum_{x \in V(G)} C_B(x) + n - 2$$

NOT END

For $d(v,w) = 2$ there $\exists\, z \in V(G)$ connecting v and w



BC of z will become smaller $\rightarrow$ does that change stuff?

↳ uw: previously v  w contributed 1, now also 1

↳ see proof of relationship of $\overline{C_B}(G)$ and $L(G)$

again, only $n-2$ paths via v and w are influenced $\Rightarrow \sum_i C'_B(y) \geq \sum_i C_B(x) + n-2$ ■

Calculating Betweenness

→ assume G undirected, without weights ... can be modified

Algorithm:

Input: $G = (V, E)$

Output: $C_B(v)$ for $\forall v \in V$

$S(v,t) := \sigma_{s,t}(v)$ represented as a 2D array

$P_s(v) := \{u \in V \mid uv \in E \;\&\; d(s,v) = d(s,u)+1\}$

↳ predecessors

1. $\forall v$: $C_B(v) \leftarrow 0$
2. For $\forall s \in V$ do:
3. $\quad S(v,t) \leftarrow 0 \quad \forall t, v \in V$
4. $\quad$ Run BFS to find $P_s(v)$ for $\forall v \neq s$
5. $\quad$ For $\forall t \neq s$ do:
6. $\qquad$ Iterate back to s from t using $P_s(\cdot)$ and increment $S(v,t)$ when we encounter v
7. $\quad$ Update $C_B(v) \leftarrow C_B(v) + \frac{S(v,t)}{S(s,t)}$ for $\forall s \neq v \neq t$
8. Return $C_B(v)/2$

↳ σ_{st}

↳ we calculated both $s \rightsquigarrow v \rightsquigarrow t$ and $t \rightsquigarrow v \rightsquigarrow s$

↳ but $C_B(v) = \sum\limits_{\substack{\{s,t\} \subseteq V \\ s \neq v \neq t}} \frac{\sigma_{s,t}(v)}{\sigma_{st}}$ ... unordered pairs

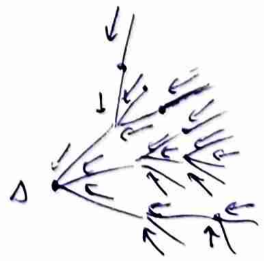
Idea: Calculate $\frac{S(v,t)}{S(s,t)}$ directly when iterating back from t to s.

↳ assume there exists some $\delta(v)$ s.t.

1. $\delta(t) = 0$ if t is a terminal vertex in BFS from s
2. We can calculate $\delta(v)$ from deltas of the succesors of v
3. We can acumulate $\delta(v)$ to calculate $C_B(v)$

→ then in the BFS step, we can create a Stack and add vertices to it as we encounter them

→ 5. 6. 7. would be then replaced by



$\delta(v) \leftarrow 0 \quad \forall v \in V$

While Stack is not empty:

$\quad$ pop $v \leftarrow$ Stack

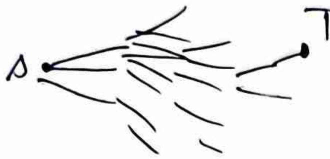
$\quad$ for $\forall u \in P_s(v)$: $\delta(u) \leftarrow \delta(u) + \underbrace{UPDATE(\delta(v))}_{\text{dont know}}$

$\quad$ if $v \neq s$: $C_B(v) \leftarrow C_B(v) + \delta(v)$

Def: $\delta_{s,t}(v) := \dfrac{\sigma_{st}(v)}{\sigma_{st}}$, $\delta_{s*}(v) := \sum_{t \in V} \delta_{st}(v) = \sum_{\substack{t \in V \\ t \neq v}} \delta_{st}(v)$

$\hookrightarrow \delta_{s,t}(s) = \delta_{st}(t) = 0 \because \sigma_{st}(s) = \sigma_{st}(t) = 0 \;\&\; \sigma_{ss} = 1$

① $\delta_{s*}(T) = 0$ if T is a terminal vertex in BFS from s



... $\delta_{s*}(T) = \sum_{t \neq v} \delta_{st}(T) = 0 \because$ all $\sigma_{st}(T) = 0$

$\hookrightarrow$ else T not terminal ↯

③ $C_B(v) = \sum_{s \neq v \neq t} \delta_{st}(v) = \sum_{s \neq v} \sum_{t \neq v} \delta_{st}(v) = \sum_{s \neq v} \delta_{s*}(v)$

② __Want:__ $\delta_{s*}(v) = \sum_{w \in Succ(v)} f(\delta_{s*}(w))$, $v \neq s$

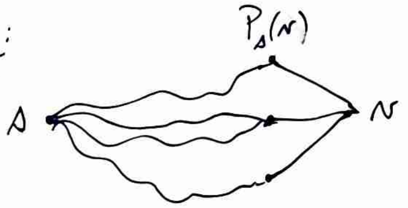
__Lemma 1:__ Let $s, v, t \in V$. From the Δ-ineq ... two options

- $d(s,t) = d(s,v) + d(v,t) \Rightarrow \sigma_{st}(v) = \sigma_{sv} \cdot \sigma_{st}$
- $d(s,t) < d(s,v) + d(v,t) \Rightarrow \sigma_{st}(v) = 0$

__Pf:__ 👁 v lies on a shortest path $s \leftrightarrow t \iff d(s,t) = d(s,v) + d(v,t)$. ■

__Lemma 2:__ For $s \neq v$: $\sigma_{sv} = \sum_{u \in P_s(v)} \sigma_{su}$. $P_s(v)$ ... predecesors

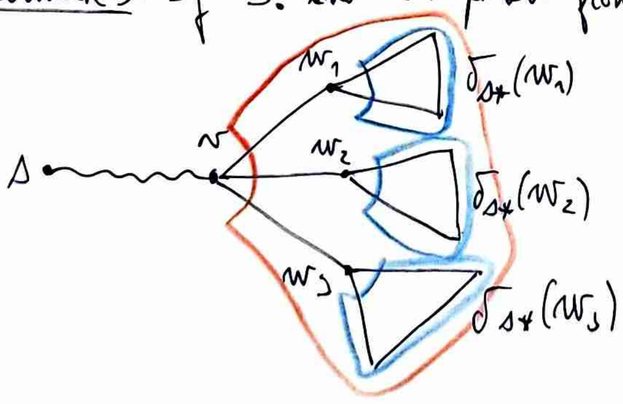
__Pf:__



__Clearly:__ $\sigma_{sv} \geq \sum_{u \in P_s(v)} \sigma_{su}$

$\hookrightarrow$ what if $s \rightsquigarrow \overset{\in P_s(v)}{u} \to w \to v$ $\nearrow$ wouldn't be shortest path ■

__Lemma 3:__ If $\exists!$ shortest path from s to every t, then $\delta_{s*}(v) = \sum_{w \in Succ(v)} (1 + \delta_{s*}(w))$



👁 $\forall t: \sigma_{st} = 1 \;\&\; \sigma_{st}(v) \in \{0,1\} \Rightarrow \delta_{st}(v) \in \{0,1\}$

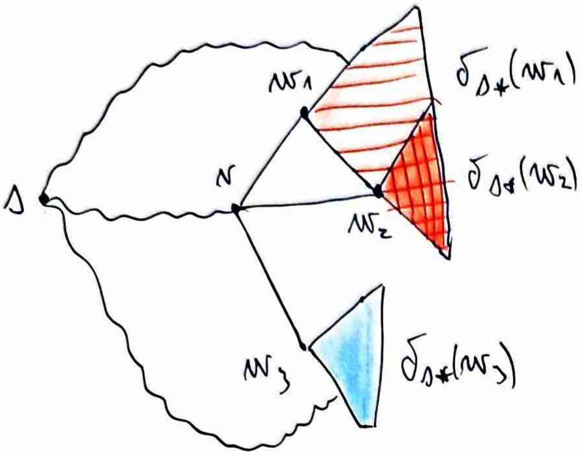
$\Rightarrow \delta_{s*}(v) = \sum_{t \neq v} \delta_{st}(v) = \#t: d(s,t) > d(s,v)$

$\Rightarrow$  $= \sum_{w} (1 +$  $)$ ■

$\to$ this is a special case of ②, we want this for general graphs.

Theorem: Let $s \in V$ and $v \neq s$. Then $\delta_{s*}(v) = \sum_{w \in Succ(v)} \frac{\sigma_{sv}}{\sigma_{sw}} (1 + \delta_{s*}(w))$.

Pf: $\delta_{s*}(v) = \sum_{t \neq v} \delta_{st}(v)$, v is fixed, change $t \neq v$



(o) $\delta_{st}(v) > 0 \iff v \in$ some $\overbrace{\text{shortest path}}^{SP}$ $P: s \to t$

$\hookrightarrow$ for $\forall$ such P $\exists!$ edge $vw \in P$ s.t. $w \in Succ(v)$

$\Rightarrow$ extend $\sigma_{st}(v)$ to contain edge info

$\sigma_{st}(v, e) = \#$ SPs $s \to t$ through both v and e

$\delta_{st}(v, e) := \frac{\sigma_{st}(v, e)}{\sigma_{st}}$

$$\Rightarrow \delta_{s*}(v) = \sum_{t \neq v} \delta_{st}(v) = \sum_{t \neq v} \sum_{w \in Succ(v)} \delta_{st}(v, \{v, w\}) = \sum_{w \in Succ(v)} \sum_{t \neq v} \delta_{st}(v, vw)$$

$\to$ How much is $\delta_{st}(v, vw) = \frac{\sigma_{st}(v, vw)}{\sigma_{st}}$? $\to$ even for $t = w$: $\sigma_{tt} = 1$

$s \sim\sim v - w \sim\sim t$ (o) $\sigma_{st}(v, vw) = \sigma_{sv} \cdot \sigma_{wt}$ | $\sigma_{st}(w) = \sigma_{sw} \cdot \sigma_{wt}$ for $w \neq t$ (Lemma 1)

$t = w$: $\delta_{st}(v, vw) = \frac{\sigma_{sv} \cdot 1}{\sigma_{st}} = \frac{\sigma_{sv}}{\sigma_{sw}}$

$t \neq w$: $\delta_{st}(v, vw) = \frac{\sigma_{sv}}{\sigma_{st}} \cdot \sigma_{wt} = \frac{\sigma_{sv}}{\sigma_{st}} \cdot \frac{\sigma_{st}(w)}{\sigma_{sw}} = \frac{\sigma_{sv}}{\sigma_{sw}} \cdot \delta_{st}(w)$

$$\Rightarrow \delta_{s*}(v) = \sum_{w \in Succ(v)} \left(\frac{\sigma_{sv}}{\sigma_{sw}} + \sum_{\substack{t \neq v \\ t \neq w}} \frac{\sigma_{sv}}{\sigma_{sw}} \cdot \delta_{st}(w) \right) = \sum_{w \in Succ(v)} \frac{\sigma_{sv}}{\sigma_{sw}} (1 + \delta_{s*}(w)) \quad \blacksquare$$

Conclusion: The quantity $\delta_{s*}(\cdot)$ satisfies all 1, 2, 3 $\Rightarrow$ we can use it.

Brandes Algorithm

Input: undirected connected $G=(V,E)$, $|V|=n$, $|E|=m$

Output: $C_B(v)$ for $\forall v \in V$

$\forall v: C_B(v) \leftarrow 0$

For $\forall s \in V$ do:

$\delta_{s*}(v)$ $\quad \sigma_{sv}$ $\quad d(s,v)$ $\quad P_s(v)$

$\quad \forall v: \delta(v) \leftarrow 0,\ \sigma(v) \leftarrow 0,\ d(v) \leftarrow \infty,\ P(v) \leftarrow \emptyset$

$\quad \sigma(s) \leftarrow 1,\ d(s) \leftarrow 0$

$\quad Q \leftarrow \text{Queue}(s),\ S \leftarrow \text{empty stack}$

init $\Theta(n)$

$\quad$ While Q is not empty do: $\qquad$ BFS is $\Theta(n+m) = \Theta(m)$ ↳ G connected

$\qquad v \leftarrow Q.\text{dequeue}()$

$\qquad S.\text{push}(v)$

$\qquad$ For $\forall w \in \text{Neighbours}(v)$ do:

$\qquad\quad$ if $d(w) = \infty$: $d(w) \leftarrow d(v)+1$ & $Q.\text{enque}(w)$

$\qquad\quad$ if $d(w) = d(v)+1$: $\sigma(w) \leftarrow \sigma(w) + \sigma(v)$ & $P(w).\text{add}(v)$

$\quad$ While S is not empty do: $\qquad$ Total # predecessors is $\Theta(m)$

$\qquad w \leftarrow S.\text{pop}()$

$\qquad$ For $\forall v \in P(w)$ do: $\delta(v) \leftarrow \delta(v) + \frac{\sigma(v)}{\sigma(w)} \cdot (1+\delta(w))$

$\qquad$ if $w \neq s$: $C_B(w) \leftarrow C_B(w) + \delta(w)$ $\qquad \rightarrow C_B(w) = \sum_{s \neq w} \delta_{s*}(w)$

end for

return $C_B/2$ ... counted both $s \to t$ and $t \to s$

Time complexity

→ for $\forall s \in V$: we have $\Theta(m)$ $\Rightarrow$ $\Theta(n \cdot m)$

Generalizations:

- directed graphs: $C_B(v) := \sum_{s \neq v} \sum_{t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}} = 2 \cdot C_B$ for connected G

  ↳ simply dont divide by 2 at the end

- weighted graphs:

  ↳ use Dijkstra instead of BFS

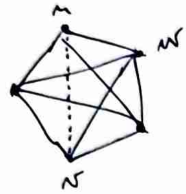
Bounds on betweenness

$0 \le C_B(v) \le \binom{n-1}{2}$ ... K_n  $\rightsquigarrow 0$ $\quad$ S_n  $\rightsquigarrow \binom{n-1}{2}$

$C_B(v) = 0 \iff G[N(v)] \cong K_{\deg(v)}$

Ex:

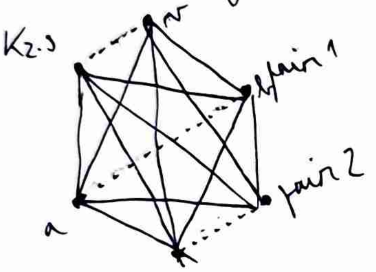
1) Almost complete: $K_n \setminus \{u,v\}$

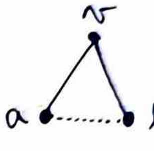


$C_B(u) = C_B(v) = 0$

$C_B(w)$ ... now $\sigma_{uv}(w) = 1$, $\sigma_{uv} = n-2 \Rightarrow C_B(w) = \frac{1}{n-2}$

2) Cocktail graph: K_{2n} − {Perfect matching of size n}



$C_B(v)$ ... all pairs $\{a, b\}$ s.t.  contribute $\frac{1}{2n-2}$

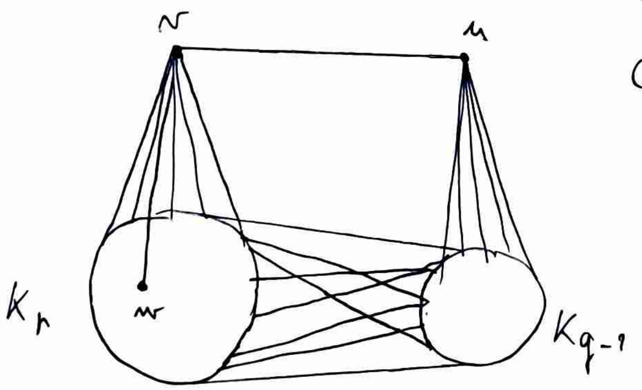
↳ $\sigma_{ab}(v) = 1$, $\sigma_{ab} = 2n-2$

↳ there are $n-1$ such pairs

$\Rightarrow C_B(v) = \frac{n-1}{2n-2} = \frac{1}{2}$

3) Bicycle graph

↳ we want G s.t. $\exists v \in V(G) : C_B(v) = \frac{p}{q} \in \mathbb{Q}^+$



$C_B(v)$ ... only pairs $\{u, w\}$, $w \in K_p$ contribute

- $\sigma_{uw}(v) = 1$, $\sigma_{uw} = 1 + (q-1) = q$

$\Rightarrow C_B(v) = \frac{p}{q}$ $\rightarrow$ $|K_p| = p$

Open problems

1) For given n determine all possible values of BC for graphs of size n

2) Def: Score is a sequence of centrality values of vertices of the given graph

↳ Given a BC-score ... does a graph with this score exist?

... can we generate it?

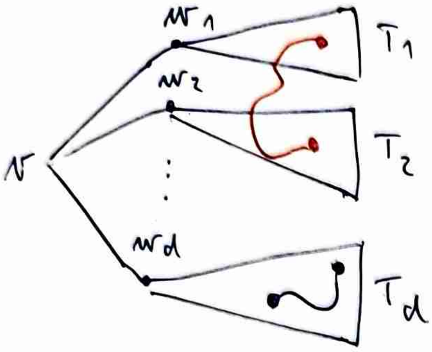
... can we create a model which generates random graphs with this score?

Theorem: Let G be a graph with max. deg Δ. Then

$$\forall v \in V(G): \quad C_B(v) \leq \frac{\Delta-1}{2\Delta}(n-1)^2$$

Pf: Let $v \in V(G)$ and $d := \deg(v)$.

→ let T be a spanning tree of G obtained using BFS from v



→ for neighbours w_i of v we have trees T_i

$$t_i := |V(T_i)| \quad , \quad \sum t_i = n-1$$

→ $C_B(v) = ?$ consider the contributions:

- $x, y \in T_i$: $\sigma^G_{xy}(v) = \sigma^T_{xy} = 0$ ... BFS

$i \neq j$ • $x \in T_i, y \in T_j$: $\frac{\sigma^T_{xy}(v)}{\sigma^T_{xy}} = 1 \geq \frac{\sigma^G_{xy}(v)}{\sigma^G_{xy}}$

$$\Rightarrow C^G_B(v) \leq C^T_B(v) = \sum_{\substack{i,j=1 \\ i<j}}^{d} t_i t_j \underset{\circledast}{\leq} \sum_{i<j}^{d} \tilde{t} \cdot \tilde{t} = \binom{d}{2}\tilde{t}^2 \underset{d \leq \Delta}{\leq} \binom{\Delta}{2} \cdot \left(\frac{n-1}{\Delta}\right)^2$$

$\circledast$: Sum is maximized when all t_i are equal: $\forall i: t_i = \frac{n-1}{\Delta} =: \tilde{t}$

↳ detail: what if $\frac{n-1}{\Delta} \notin \mathbb{N}$? ... divide best as possible, ineq. holds ∎

Maximal betweenness

Def: For graph G define $C_{B\max}(G) := \max\limits_{v \in V(G)} C_B(v)$

Def: For a family of graphs $\mathcal{H}$ define $C_{B\max}(\mathcal{H}) := \max\limits_{G \in \mathcal{H}} C_B(\mathcal{H})$

- What we know:

$C_{BM}(\mathcal{G}_n) = \binom{n-1}{2}$ ... $\mathcal{G}_n$ = graphs of size n

$C_{BM}(\mathcal{G}^\Delta_n) = \frac{\Delta-1}{2\Delta}(n-1)^2$ ... $\mathcal{G}^\Delta_n$ = graphs of size n with max. deg Δ

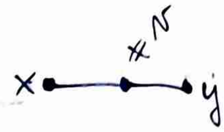
- We will show:

$C_{BM}(\mathcal{G}^2_n) = \frac{(n-3)^2}{2}$ ... $\mathcal{G}^2_n$ = 2-connected graphs of size n

Theorem: Let G be a graph and $v \in V(G)$. Let $H := G - v$ and denote

$e_1(H) :=$ # adjacent pairs in H $= |E(H)|$

$e_2(H) :=$ # pairs with distance 2 in H $= \#\{x,y\}$, $x \neq y \in V(G)-v$ s.t.



Then $C_B(v) \leq \binom{n-1}{2} - e_1(H) - \frac{1}{2} e_2(H)$.

Pf: $V_2 :=$ all pairs $\{x,y\}$, $x \neq y \in V(H)$

$E_1 = E(H) \subseteq V_2$

$E_2 \subseteq V_2 \setminus E_1$ ... pairs with dist 2 in H

$R = V_2 \setminus (E_1 \cup E_2)$

(*) $\sigma_{xy}(v) = 0$ ... $\leq \frac{1}{2}$ ✓

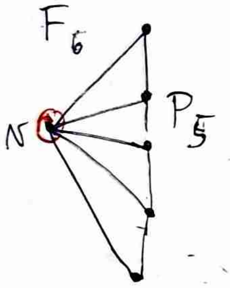
$\sigma_{xy}(v) = 1$

$\Rightarrow \sigma_{xy} \geq 2$ — first path, second path via v

$$\Rightarrow C_B(v) = \underbrace{\sum_{\{x,y\} \in E_1} \frac{\sigma_{xy}(v)}{\sigma_{xy}}}_{0} + \sum_{\{x,y\} \in E_2} \underbrace{\frac{\sigma_{xy}(v)}{\sigma_{xy}}}_{(*)\ \leq \frac{1}{2}} + \sum_{\{x,y\} \in R} \underbrace{\frac{\sigma_{xy}(v)}{\sigma_{xy}}}_{\leq 1}$$

$$\leq 0 \cdot e_1(H) + \frac{1}{2} e_2(H) + 1 \cdot \left(\binom{n-1}{2} - e_1(H) - e_2(H) \right) = \binom{n-1}{2} - e_1(H) - \frac{1}{2} e_2(H)$$

Ex: Broken wheel graph $F_n = P_{n-1}$ and 1 extra vertex connected to all



$C_B(v) = ?$ ... contributing pairs, $x,y \in P_{n-1}$

- $d^P(x,y) = 1$ .... contributes 0
- $d^P(x,y) = 2$ ... contributes $\frac{1}{2}$, overall $\frac{1}{2} \cdot (|P| - 2) = \frac{n-3}{2}$
- $d^P(x,y) \geq 3$ ... contributes 1, overall $\underbrace{\binom{n-1}{2}}_{total} - \underbrace{(n-2)}_{d=1} - \underbrace{(n-3)}_{d=2}$

$$\Rightarrow C_B(v) = \frac{n-3}{2} + \binom{n-1}{2} - 2n + 5 = \dots = \frac{(n-3)^2}{2}$$

Theorem: Let G be 2-connected with size n. Then $\forall v$: $C_B(v) \leq \frac{(n-3)^2}{2}$

Pf: Let $v \in V(G)$ and consider the decomposition of $H := G - v$ from the previous theorem

$\to$ define $p(H) := e_1(H) + \frac{1}{2} e_2(H)$

(*) it holds that $p(H) \geq \frac{3n-7}{2}$ ... proof later

$\to$ using previous theorem

$$C_B(v) \leq \binom{n-1}{2} - p(H) \leq \frac{(n-1)(n-2)}{2} - \frac{3n-7}{2} = \dots = \frac{(n-3)^2}{2}$$

$\to$ proof of (*):

G is 2-connected $\Rightarrow$ H is connected on $n-1$ vertices

$\to$ minimal such graph is P_{n-1}, then $p(P_{n-1}) = n - 2 + \frac{1}{2}(n-3) = \frac{3n-7}{2}$

$\to$ intuitively for "bigger" graphs, p should be larger

$\Rightarrow$ proof of $(*)$

min deg

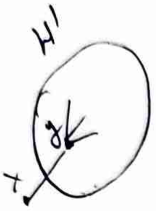
① $\delta(H) \geq 3$: $e_1(H) = |E(H)| = \frac{1}{2} \sum_{v \in H} \deg(v) \geq \frac{1}{2} \cdot 3 \cdot (n-1) = \frac{3n-3}{2} > \frac{3n-7}{2}$

② $\delta(H) < 3$: $\exists\, x \in H$ with $\deg(x) \in \{1,2\}$

$\Rightarrow$ induction over $|H|$, base of induction = paths P_n

$\Rightarrow$ remove x from H and examine $H' := H - x$

a) $\deg(x) = 1$: $\Rightarrow$ H' connected on $n-2$ vertices



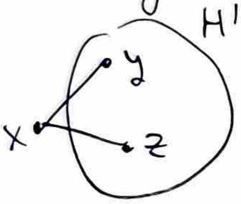
I.H.: $p(H') \geq \frac{3(n-1)-7}{2} = \frac{3n-10}{2}$, we add x back to H'

want: $p(H) \geq \frac{3n-7}{2}$, meaning $\Delta p \geq \frac{3}{2}$

- $e_1(H) = e_1(H') + 1$ ... $\Delta e_1 = 1$
- $e_2(H) = e_2(H') + \deg_{H'}(y) \geq e_2(H') + 1$ ... $\Delta e_2 \geq 1$

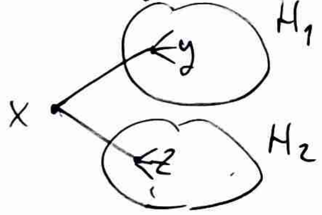
$\Big\} \ \Delta p \geq \underbrace{1}_{\Delta e_1} + \frac{1}{2} \cdot \underbrace{1}_{\Delta e_2} = \frac{3}{2}$ ✓

b) $\deg(x) = 2$ & H' connected



- $e_1(H) = e_1(H') + 2$ ... $\Delta e_1 = 2$ $\Rightarrow$ $\Delta p \geq 2$ ✓

c) $\deg(x) = 2$ & H' disconnected $= H_1 \cup H_2$



- H_1 size n_1: $p(H_1) \geq \frac{3(n_1+1)-7}{2}$
- H_2 size n_2: $p(H_2) \geq \frac{3(n_2+1)-7}{2}$

$n_1 + n_2 = n-2$

$\oplus \quad \frac{3(\overbrace{n_1+n_2}^{n-2}+2)-14}{2} = \frac{3n-14}{2}$

$\Rightarrow$ need $\Delta p \geq \frac{7}{2}$

- $e_1(H) = e_1(H_1) + e_1(H_2) + 2$ ... $\Delta e_1 = 2$
- $e_2(H) = e_2(H_1) + e_2(H_2) + 1 + \deg_{H_1}(y) + \deg_{H_2}(z)$ ... $\Delta e_2 \geq 3$

$\Rightarrow \Delta p = \Delta e_1 + \frac{1}{2}\Delta e_2 = 2 + \frac{3}{2} = \frac{7}{2}$ ✓

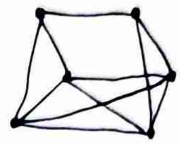
• Betweenness uniform graphs   BUGs

Def: G is vertex transitive $\equiv \forall x,y \in V(G)$: $\exists$ automorphism σ mapping $x \to y$.

Intuition: All vertices are indistinguishable based on properties alone.

Def (BUG): G is betweenness uniform $\equiv \forall x,y \in V(G)$: $C_B(x) = C_B(y)$.

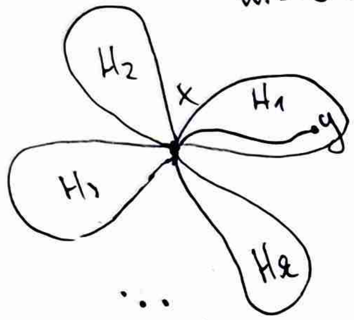
👁 Vertex transitive graphs are BUGs.



! $\exists$ non vertex transitive, non regular BUGs

Theorem: Each BUG is 2-connected.

Pf: For contradiction, let G be a BUG with a cut-vertex x and let $G-x$ have $k>1$ components $H_1, \dots H_k$ with sizes $h_1, \dots h_k$,
W.L.O.G let h_1 be minimal: $\forall i: h_1 \le h_i$ (*)



👁 $C_B(x) \ge \sum_{\substack{i,j=1\\ i<j}}^{k} h_i h_j$

$\to C_B(y) = ?$

Let $y \in H_1$ be a vertex from H_1 with maximal distance from x

$\to$ pairs of u,v contributing to $C_B(y)$

- $u \in H_i, v \in H_j, \; i \ne 1 \ne j$: none
- $u \in H_i, v \in H_1, \; i \ne 1$: none
- $u,v \in (H_1 + x)$: some

$|H_1 + x| = h_1 + 1$

$$\Rightarrow C_B(y) \le \frac{h_1(h_1-1)}{2} < \frac{h_1^2}{2} \overset{(*)}{<} \sum_{\substack{i,j=1\\ i<j}}^{k} h_i h_j \le C_B(x) \quad \Rightarrow G \text{ is not a BUG } ↯ \; ■$$

$|V|$

Theorem: Let $G=(V,E)$ be a BUG on n vertices with $n - \Delta(G) = k$. Then $d(G) \le k$.
The equality holds for $k = 1, 2$.

$d(G) := \max_{x,y \in V} d(x,y)$

diameter

$\to$ we will show:

- $k=1$ ... equivalent to $G \cong K_n$
- $k=2$ ... show something about BUGs with diam $= 2$

$\to$ couldn't equality hold for $k \ge 3$?

$\to$ for $k=3$ consider for $n \ge 5$ the complement of cycle C_n

$\overline{C_6}$

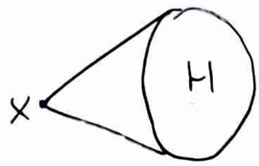
👁 $G = \overline{C_n}$ is vertex transitive $\Rightarrow$ it is a BUG

$\Delta(G) = n-3 = n-k$

$d(G) = 2 < 3 = k$

Theorem: If G is a BUG with $\Delta(G) = n-1$, then $G \cong K_n$.

Pf: Let $x \in V$ have $\deg(x) = \Delta = n-1$ and let $H = G - x$.



Note: x is connected to all $y \in V(H)$

$\Rightarrow$ for $\forall\, u,v \in H$ $\exists\, P_{u,v}(x)$ of length 2 ... path $u \to v$ through x

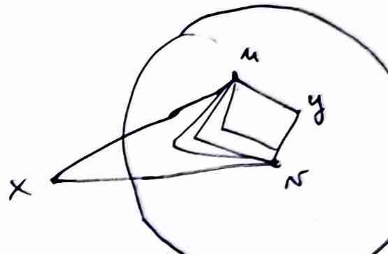
$V_2 := \{\{u,v\} \mid u,v \in H\}$, $T := \{uv \in V_2 \mid d_H(u,v) = 2\}$, $S := \{uv \in V_2 \mid d_H(u,v) \geq 3\}$

$\{u,v\} \in T \Rightarrow \sigma_{uv}(x) = 1$ ... only one SP of length 2 visits x

$\{u,v\} \in S \Rightarrow \sigma_{uv} = \sigma_{uv}(x) = 1$ ... of length 2 via x

Denote: $\sigma'_{u,v} := \#SP\ u \to v$ in H

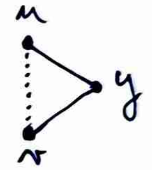
$$C_B(x) = \sum_{uv \in T} \frac{\sigma_{uv}(x)}{\sigma_{uv}} + \sum_{uv \in S} \frac{\sigma_{uv}(x)}{\sigma_{uv}} = \sum_{uv \in T} \frac{1}{1+\sigma'_{uv}} + \sum_{uv \in S} 1$$



$$y \in H: \ C_B(y) = \sum_{uv \in T} \frac{\sigma_{uv}(y)}{\sigma_{uv}} + \sum_{uv \in S} \frac{\sigma_{uv}(y)}{\sigma_{uv}} = \sum_{\substack{uv \in T \\ u \neq y \neq v \\ uy,\, yv \in E}} \frac{1}{1+\sigma'_{uv}} + \sum_{uv \in S} 0$$

Since G is BU, we must have $C_B(x) = C_B(y)$ for $\forall y$

$$\Rightarrow \sum_{uv \in T} \frac{1}{1+\sigma'_{uv}} + |S| = \sum_{\substack{uv \in T \\ u \neq y \neq v \ \&\ uy,\, yv \in E}} \frac{1}{1+\sigma'_{uv}}$$



This is possible only if $|S| = 0$ & (*) holds for $\forall\, u,v \in T = V(H)$

We want to show that G is complete. $\Rightarrow$ let $y \in H$ have $\deg(y) < n-1$

$\sigma'_{uv} = n-3$ ... for $\forall y \in H$, $y \notin \{u,v\}$: $uy, yv \in E$ using (*)

$$C_B(x) = \sum_{uv \in T} \frac{1}{1+\sigma'_{uv}} = \sum_{uv \in T} \frac{1}{n-2} = \frac{1}{n-2} \cdot |E(\overline{H})|$$ $\rightarrow$ # non-edges

$$C_B(y) = \sum_{\substack{uv \in T \\ (*)}} \frac{1}{n-2} < \sum_{uv \in T} \frac{1}{n-2} = C_B(x)$$ ↯ G is not a BUG

Therefore $\forall y \in H: \deg(y) = n-1 \Rightarrow G \cong K_n$

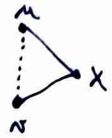
because $\deg(y) < n-1$, $\exists\, w \in V(H)$ s.t. $yw \notin E$

$\rightarrow$ yw is a non-edge in H $\Rightarrow$ $yw \in T$

! but yw does not satisfy (*)

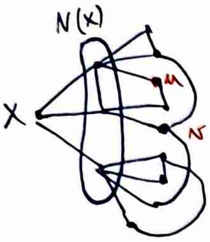
∎

Lemma: Let $G=(V,E)$ have diameter 2. Then $\forall x \in V$: $C_B(x) = \sum\limits_{\substack{u,v \in N(x) \\ uv \notin E}} \frac{1}{\sigma_{uv}}$



Pf: $uv \in E \Rightarrow \sigma_{uv}(x) = 0$

$$C_B(x) = \sum_{u \neq x \neq v} \frac{\sigma_{uv}(x)}{\sigma_{uv}} = \sum_{\substack{u\neq x \neq v \\ uv \notin E}} \frac{\sigma_{uv}(x)}{\sigma_{uv}} = \underbrace{\sum_{\substack{u,v \in N(x) \\ uv \notin E}} \frac{\sigma_{uv}(x)}{\sigma_{uv}}}_{1} + \underbrace{\sum_{u,v \notin N(x)} \frac{\sigma_{uv}(x)}{\sigma_{uv}}}_{2} + \underbrace{\sum_{\substack{u \in N(x) \\ v \notin N(x)}} \frac{\sigma_{uv}(x)}{\sigma_{uv}}}_{3}.$$



2 $P_{u,v}(x)$ has length 4 but $d(u,v) \leq \text{diam} = 2 \Rightarrow \sigma_{uv}(x) = 0$

3 $P_{u,v}(x)$ has length 3 but $d(u,v) \leq \text{diam} = 2 \Rightarrow \sigma_{uv}(x) = 0$

1 $uv \notin E \Rightarrow d(u,v) = 2 \Rightarrow \sigma_{uv}(x) = 1$ ∎

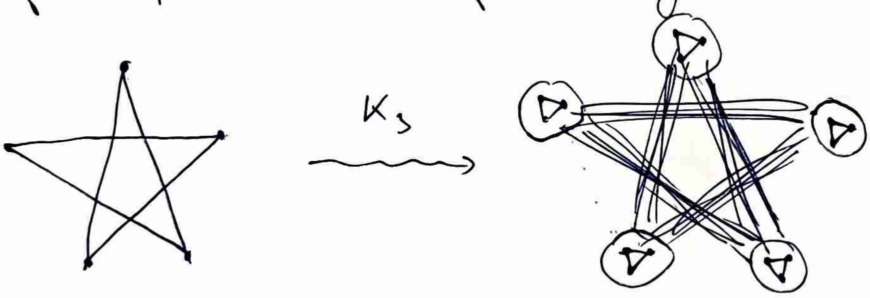
Theorem: Let $G=(V,E)$ be a BUG with diam $=2$. Then $\forall x$: $C_B(x) = \frac{n-1}{2} - \frac{m}{n}$ $\qquad n = |V|$, $m = |E|$

Pf Use previous lemma to calculate $\sum C_B(x)$ and divide by $n = |V|$.

$$\sum_{x \in V} C_B(x) = \sum_{x \in V} \sum_{\substack{u,v \in N(x) \\ uv \notin E}} \frac{1}{\sigma_{uv}} = \sum_{\substack{u,v \in V \\ uv \notin E}} \sum_{\substack{x \in V \\ u,v \in N(x)}} \frac{1}{\sigma_{uv}} = \sum_{\substack{u,v \in V \\ uv \notin E}} \sigma_{uv} \cdot \frac{1}{\sigma_{uv}} = \binom{n}{2} - m$$

Since G is BU we have $\forall x$: $C_B(x) = \frac{1}{n}\left[\binom{n}{2} - m\right] = \frac{n-1}{2} - \frac{m}{n}$ ∎

Fact: If we replace each vertex of a BUG by K_r, $r \geq 1$, then we get a BUG as well.



$v \mapsto \{v_1, \dots v_r\}$

$vu \mapsto \{v_i u_j \mid i,j = 1, \dots n\}$

blow-up graph

Fact: For $\forall$ automorphism group A we can create a BUG G s.t. $\text{Aut}(G) = A$.

Open problem: Create a BUG G with $C_B(x) = \frac{p}{q} \in \mathbb{Q}$ for arbitrary $\frac{p}{q} > \frac{1}{2}$.

Spectral characteristics of networks

Theorem: $G=(V,E)$ graph with vertices $V=\{v_1, v_2, \dots v_n\}$. Define its adjencency matrix A_G as $a_{ij} = [v_i v_j \in E]$. Clearly a_{ij} = number of walks of lenght 1 between v_i, v_j. In general it holds that

$$(A_G^k)_{ij} = \#\text{ of } v_i - v_j \text{ walks of lenght } k \text{ in } G \geq 0$$

Proof: By induction using the definition of matrix multiplication. ∎

Proposition: If G is bipartite, then never $A_G^k > 0$. $\rightarrow \forall a_{ij} > 0$

Proof: Partits V_1, V_2

- walks $V_1 - V_2$ have odd lenght $\Rightarrow (A_G^{EVEN})_{V_1, V_2} = 0$
- walks $V_i - V_i$ have even lenght $\Rightarrow (A_G^{ODD})_{V_i, V_i} = 0$ ∎

Fact: If G is not bipartite, then at some point $A_G^k > 0$.

Linear Algebra Recap

Def: The spectrum of a matrix is the set of its eigenvalues. $A \in \mathbb{C}^{n \times n}$

$$\sigma(A) := \{\lambda \in \mathbb{C} \mid \exists\, 0 \neq v \in \mathbb{C}^n : Av = \lambda v\}$$

The spectral radius of a matrix is the largest size of an eigenvalue

$$\rho(A) := \max\{|\lambda| \mid \lambda \in \sigma(A)\}$$

Def: The maxim norm of a matrix $A \in \mathbb{C}^{n \times n}$ is

$$\|A\| := \max\{|a_{ij}| \mid 1 \leq i, j \leq n\}$$

Theorem: If $A \in \mathbb{R}^{n \times n}$ is symmetric, then all of its eigenvalues are real.

Theorem (Spectral decomp.): $A \in \mathbb{R}^{n \times n}$ symmetric $\Rightarrow A = Q \Lambda Q^T$ where

- Λ is a diagonal matrix containing the (real) eigenvalues of A
- Q is orthogonal $\equiv Q^T = Q^{-1} \equiv$ colums form an orthonormal system
- colums of Q are the normalized eigenvectors of A

☼ G undirected graph $\Rightarrow A_G$ symmetric and $A_G \geq 0$.

Theorem: $\lim_{n\to\infty} \|A^n\| = \begin{cases} 0 & , \rho(A) < 1 \\ \infty & , \rho(A) > 1 \end{cases}$ don't know for $\rho(A) = 1$

Proof: If A is diagonalizable write $A = Q \Lambda Q^{-1}$

$\to A^n = Q \Lambda^n Q^{-1}$ ... Q, Q^{-1} constant Λ^n diagonal eigenvalues of A

$\Rightarrow \|\Lambda^n\| = \rho(A)^n \Rightarrow \|A^n\| \approx |\text{constant}| \cdot \rho(A)^n$

If A is not diagonalizable, use Jordan normal form. ∎

Properties of eigenvalues

$A \in \mathbb{R}^{n\times n}$ with eigenvalues $\lambda_1, \dots \lambda_n$, eigenvectors $v_1, \dots v_n$

1. $\alpha A \Rightarrow \alpha\lambda_1, \dots \alpha\lambda_n$ & $v_1, \dots v_n$
2. $A + \alpha I_n \Rightarrow \lambda_1 + \alpha, \dots \lambda_n + \alpha$ & $v_1, \dots v_n$
3. $A^2 \Rightarrow \lambda_1^2, \dots \lambda_n^2$ & $v_1, \dots v_n$
4. regular $A^{-1} \Rightarrow \frac{1}{\lambda_1}, \dots \frac{1}{\lambda_n}$ & $v_1, \dots v_n$

Def: $A \in \mathbb{R}^{n\times n}$ is positive $\equiv \forall a_{ij} > 0$. We write $A > 0$

Def: $A, B \in \mathbb{R}^{n\times n}$. We write $A \leq B \equiv \forall a_{ij} \leq b_{ij}$. Similarly $<$.

Lemma: Let $A \in \mathbb{R}^{n\times n}$ s.t. $A > 0, \rho(A) = 1$. Then $\lambda \in \sigma(A)$ & $|\lambda| = \rho \Rightarrow Re(\lambda) > 0$.

Proof: Suppose $Re(\lambda) \leq 0$.

$\to$ choose $\varepsilon > 0$ s.t. $A - \varepsilon I_n > 0$ ... $A > 0$.

$|\lambda - \varepsilon| > 1$ & for $\delta < 1$ close enough to 1 we still have $\delta|\lambda - \varepsilon| > 1$

$\Rightarrow$ define matrices

$A_1 := \delta(A - \varepsilon I_n) \Rightarrow \delta(\lambda - \varepsilon) \in \sigma(A_1) \Rightarrow \rho(A_1) > 1$

$A_2 := \delta A \Rightarrow$ since $0 < \delta < 1 \Rightarrow \rho(A_2) < 1$

$\Rightarrow$ therefore $\|A_1^n\| \to \infty$ & $\|A_2^n\| \to 0$ as $n \to \infty$

$\to$ but both $A_1 > 0, A_2 > 0$ and $A_1 \leq A_2$. Therefore $\|A_1^n\| \leq \|A_2^n\|$ ↯ ∎

Theorem (Perron-Frobenius): $A \in \mathbb{R}^{n\times n}$ s.t. $A \geq 0$. If at some point $A^k > 0$, then

1. $\rho(A) \in \mathbb{R}_0^+$ is an eigenvalue of A
2. ρ is strictly dominant $\equiv \lambda \in \sigma(A),\ \lambda \neq \rho \Rightarrow |\lambda| < |\rho| = \rho$

Proof: If $\rho = 0$, then all eigenvalues are zeroes 0, so the claim holds.

$\Rightarrow$ assume $\rho(A) > 0$ and let $A^{k_0} > 0$, $\forall k \geq k_0 \Rightarrow A^k > 0$ $\because$ $A \geq 0$.

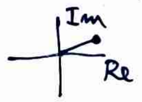
$\rightarrow$ let $\lambda \in \sigma(A)$ s.t. $|\lambda| = \rho(A)$, we want to show $\lambda = \rho(A) \in \mathbb{R}^+$

$\Rightarrow$ define $A_1 := A/\rho(A) \Rightarrow \rho(A_1) = 1$

$\forall\ \lambda_1 := \lambda/\rho(A) \in \sigma(A_1)$ and $|\lambda_1| = \frac{|\lambda|}{\rho(A)} = 1$

claim: $\lambda_1 = 1$, from that $\lambda = \lambda_1 \cdot \rho(A) = \rho(A)$ which proofs ①

↳ since we didn't assume anything special about λ, it also means that for all λ' s.t. $|\lambda'| = \rho(A)$ we have $\lambda' = \rho(A) \Rightarrow$ ②

↳ proof: assume $\lambda_1 \neq 1 \Rightarrow$ 

$\Rightarrow \exists k \geq k_0$ s.t. $Re(\lambda_1^k) \leq 0$

↳ but: $A_1^k > 0$, $\lambda_1^k \in \sigma(A_1^k)$, $\rho(A_1^k) = |\lambda_1^k| = 1 \xrightarrow{\text{LEMMA}} Re(\lambda_1^k) > 0$ ↯ ■

Fact: This can be made even stronger. ② shows $Geom(\rho) = 1$. But in fact

3. ρ is a simple eigenvalue $\equiv 1 = Geom(\rho) = Alg(\rho)$

↳ in general $1 \leq G \leq A$

What is this good for?

$\rightarrow$ if $G = (V, E)$ is not bipartite, then $A_G \geq 0$ is at some point $A_G^k > 0$

$\rightarrow$ moreover A_G is symmetric $\Rightarrow$ diagonalizable

$\rightarrow$ this theorem says that the leading eigenvalue of A_G has multiplicity 1

$\Rightarrow$ the limit behavior of A_G^k is fully determined by this single leading eigenvalue (which is positive $= \rho(A)$ and its eigenvector

Eigenvector vertex centrality – motivation on undirected graphs

recall: degree centrality $= \frac{\deg(v)}{n-1}$

↳ social networks: says to how many people you are connected

→ but also important: what is the score/centrality of your connections?

⇒ idea: award score proportional to the score of neighbors

↳ vertices $v_1, \dots v_n$, centralities $x_1, \dots x_n$, adjacency matrix A

1. $x_i \leftarrow 1$ for $\forall i$
2. $x_i' \leftarrow \sum_j A_{ij} x_j \quad = \deg(v_i)$
3. $x_i'' \leftarrow \sum_j A_{ij} x_j' \quad = \sum_{v_j \in N(v_i)} \deg(v_j)$ ... sum of degrees of neigbors

→ vectors $\tilde{x}, \tilde{x}', \tilde{x}''$ ⊛ $\tilde{x}' = A\tilde{x}$, $\tilde{x}'' = A\tilde{x}'$

Def (eigenvector centrality): Undirected graph G with $V = \{v_1, \dots v_n\}$. Define

$\tilde{x}(0) \in \mathbb{R}^n$ – initialization, usually $\tilde{x}(0) = (1, 1, \dots 1)$.

→ $\tilde{x}(t) := A_G \tilde{x}(t-1) = A_G^t \cdot \tilde{x}(0)$

→ $\tilde{x}(t)_i$ = score of v_i after t iterations

Idea: A_G is symmetric ⇒ $A_G = Q \Lambda Q^T$ where colums of Q are the normalized eigenvectors of A_G and they form an orthonormal basis of $\mathbb{R}^n$

⇒ let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ be the eigenvalues of A_G, $v_1, \dots v_n$ eigenvectors

$\tilde{x}(0) = \sum c_i v_i$ for some $\tilde{c} \in \mathbb{R}^n$ ... $\{v_i\}$ are a basis

⇒ $\lambda_1^t \geq \lambda_2^t \geq \dots \geq \lambda_n^t$ are eigenvalues of A_G^t, $v_1, \dots v_n$ eigenvectors

$\tilde{x}(t) = A_G^t \sum c_i v_i = \sum c_i A_G^t v_i = \sum c_i \lambda_i^t v_i = \rho(A)^t \sum c_i \left(\frac{\lambda_i}{\rho(A)}\right)^t v_i$

⇒ if G is not bipartite, then Perron Frobenius: $\rho(A) = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_n$

$\tilde{x}(t) \to c_1 \lambda_1^t v_1$

→ because $\tilde{x}(t) \geq 0$, we can guess that probably $v_1 \geq 0$ and $c_1 > 0$
or $v_1 \leq 0$ and $c_1 < 0$
↳ $-v_1 \geq 0$

⇒ we will show that this works even for bipartite graphs

Lemma: Let G be an undirected graph and let A_G have (real) eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Then it holds

① $\lambda_1 \geq 0$ ② $\sum \lambda_i = 0$ ③ $\prod \lambda_i = \det(A_G)$

Proof: To show ② we will prove $\sum \lambda_i = tr(A_G) = \sum a_{ii} = 0$.

- A_G symmetric $\Rightarrow A = Q \Lambda Q^{-1}$

👁 $tr(AB) = tr(BA) = \sum a_{ij} b_{ij}$

$\Rightarrow tr(A) = tr(Q\Lambda Q^{-1}) = tr(QQ^{-1}\Lambda) = tr(\Lambda) = \sum \lambda_i$

① ... if $\sum \lambda_i = 0$ then $\exists$ eigenvalue ≥ 0 and λ_1 is the biggest

③ $\det(A) = \det(Q)\det(\Lambda)\det(Q^{-1}) = \det(Q)\det(Q^{-1})\det(\Lambda) = \underbrace{\det(QQ^{-1})}_{1} \cdot \prod \lambda_i$ ■

Theorem: Let G be an undirected graph and let A_G have (real) eigenvalues $\lambda_1 \geq \dots \geq 0 \geq \dots \geq \lambda_n$ with orthonormal eigenvectors $v_1 \perp v_2 \perp \dots \perp v_n$.

① $\rho(A) = \lambda_1$ $(\lambda_1 \geq |\lambda_n|)$

② suppose λ_1 has multiplicity k, that is $\lambda_1 = \lambda_2 = \dots = \lambda_k > \dots$, then $\exists j \in [k]$ s.t. $v_j \geq 0$ (or $v_j \leq 0$)

③ if G is connected, then λ_1 is simple $(\mathrm{Geom}(\lambda_1) = 1)$, therefore it is strictly dominant $(\lambda_1 > \lambda_2)$ and moreover $v_1 > 0$ (or $v_1 < 0$)

Proof: We first introduce Rayleigh quotient: $A \in \mathbb{R}^{n\times n}$ symmetric, $x \in \mathbb{R}^n \to R(x) := \frac{x^T A x}{x^T x}$

$\to$ define $|x| := (|x_1|, \dots |x_n|)$ and note: $x^T x = |x|^T |x| = \sum x_i^2$

① denote $x := v_n$, $A := A_G$

- $|\lambda_n| x^T x = |\lambda_n x^T x| = |x^T A x| = \left|\sum_{ij} A_{ij} x_i x_j\right| \leq \sum_{ij} |A_{ij} x_i x_j| = \sum_{ij} A_{ij} |x_i| \cdot |x_j| = |x|^T A |x|$ ($A \geq 0$)

$\Rightarrow |\lambda_n| \leq \frac{|x|^T A |x|}{x^T x} = \frac{|x|^T A |x|}{|x|^T |x|} = R(|x|)$

- eigenbasis $\{v_i\} \Rightarrow$ arbitrary $w = \sum \alpha_i v_i$ where $v_i^T v_i = 1$, $v_i^T v_j = 0$

$\Rightarrow R(w) = \frac{(\sum \alpha_i v_i^T) A (\sum \alpha_i v_i)}{(\sum \alpha_i v_i^T)(\sum \alpha_i v_i)} = \frac{(\sum)\cdot(\sum \alpha_i A v_i)}{\sum \alpha_i^2} = \frac{(\sum)(\sum \alpha_i \lambda_i v_i)}{\sum \alpha_i^2} = \frac{\sum \lambda_i \alpha_i^2}{\sum \alpha_i^2} = \sum \beta_i \lambda_i$ $\quad \beta_i := \frac{\alpha_i^2}{\sum \alpha_j^2}$

👁 $\sum \beta_i = 1$ & $0 \leq \beta_i \leq 1 \Rightarrow$ convex combination

- $\lambda_n \leq R(w) \leq \lambda_1 \Rightarrow |\lambda_n| = R(|x|) \leq \lambda_1 \Rightarrow$ ① ✓

- ②: note: $R(w) = \lambda_1 \Leftrightarrow \sum_{i=1}^{Alg(\lambda_1)} \beta_i = 1 \Rightarrow w = \sum_{i=1}^{Alg(\lambda_1)} v_i \Rightarrow$ w is eigenvector of λ_1

② proof: we have shown $R(w) = \lambda_1 \Rightarrow w$ is eigenvector of λ_1

→ because $\lambda_1 \geq 0$ we can redo the argument for λ_n to get

$$\lambda_1 v_1^T v_1 = |\lambda_1 v_1^T v_1| = |v_1^T A v_1| = \dots \leq |v_1|^T A |v_1| \Rightarrow \lambda_1 \leq R(|v_1|)$$

→ but always $R(w) \leq \lambda_1$ therefore $R(|v_1|) = \lambda_1$ and $|v_1|$ is eigenvector of λ_1

→ let $Alg(\lambda_1) = k$ and the eigenspace of λ_1 have orthonormal basis $v_1, \dots v_k$

→ because $\exists w \geq 0$, eigenvector of λ_1 we have $w = \sum \alpha_i v_i$

→ intuitively: there is at least 1 eigenvector in the orthant $\mathbb{R}^n_+$ (or $\mathbb{R}^n_-$) ⇒ ②

③ proof: let G be connected and let v be an eigenvector of λ_1

→ we know that $|v|$ is also an eigenvector. We will show $|v| > 0$

→ suppose $v_i = 0$, then $Av = \lambda_1 v \Rightarrow \sum_j A_{ij} v_j = \lambda_1 v_i = 0$ ← i-th row

$$\sum_j A_{ij} v_j = \sum_{j \sim i} v_j = 0$$ where $j \sim i \equiv$ vertex j is connected to vertex i

→ meaning: if $v_i = 0$, then all neighbours j of i also have $v_j = 0$

⇒ because G is connected, we recursively get $v = (0, 0, \dots 0) = \vec{0}$ ↯

→ we have shown that when v is eigenvector of λ_1, then all $v_i \neq 0$

→ want: λ_1 has multiplicity 1

↳ suppose λ_1 has another eigenvector w, not a scalar multiple of $|v| > 0$

→ we will choose w to be orthonormal to $|v|$

$$\Rightarrow w^T |v| = 0 \Rightarrow \sum w_i |v_i| = 0 \xRightarrow{\forall v_i \neq 0} w_i \text{ must have mixed signs}$$

→ claim: w having mixed signs contradicts connectedness

↳ recall: $|w|$ is also an eigenvector & $|w| > 0 \Rightarrow \forall w_i \neq 0$

→ w has mixed signs $\Rightarrow \exists\, w_i > 0$

$$\sum_j A_{ij} (|w_j| - w_j) = \lambda_1 (|w_i| - w_i) = 0$$ ← i-th row of $A(|w| - w)$

→ because always $|w_j| - w_j \geq 0$ and $|w_j| - w_j = 0 \iff w_j > 0$,

it must be that $A_{ij} \neq 0 \Rightarrow w_j > 0$

⇒ if $w_i > 0$ then for all neighbours j of vertex i we have $w_j > 0$

⇒ because G is connected, we recursively get $\forall w_j > 0$

⇒ therefore w doesn't have mixed signs ↯

Spectral characteristics of directed graphs   ⚠ standard network science definition ⚠

→ adjacency matrix is not symmetric $A_{ij} = [v_j \to v_i]$ not $v_i \to v_j$

↳ this makes things more complicated but the theorem can be generalized

Def: Let $A \in \mathbb{R}^{n \times n}$. The vector $x \in \mathbb{R}^n$ is a

- right eigenvector of λ $\equiv$ $Ax = \lambda x$
- left eigenvector of λ $\equiv$ $x^T A = \lambda x^T$

👁 if G is directed with adjacency matrix A, then

- right eigenvectors ~ incoming edges

$$Ax = \lambda x \Rightarrow \lambda x_i = \sum_j A_{ij} x_j = \sum_{j: j \to i} x_j \quad \Rightarrow \lambda x_i \text{ considers } v_j \to v_i$$

- left eigenvectors ~ outgoing edges

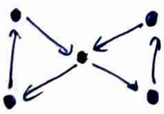
$$x^T A = \lambda x^T \Rightarrow \lambda x_i = \sum_j x_j A_{ji} = \sum_{j: i \to j} x_j \quad \Rightarrow \lambda x_i \text{ considers } v_i \to v_j$$

Theorem (Perron-Frobenius): Let G be a strongly connected digraph. Then

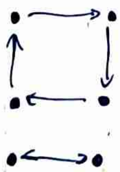
1. $\rho(A_G) \in \mathbb{R}^+$ is a simple eigenvalue of A_G → $\rho = 0 \iff G$ no edges
2. ρ has a unique (up to scaling)
   - positive right eigenvector $x > 0$
   - positive left eigenvector $y > 0$
3. If G is aperiodic, then ρ is strictly dominant : $\lambda \neq \rho \Rightarrow |\lambda| < \rho$ (complex)
4. If G has period p, then there are exactly p eigenvalues with $|\lambda| = \rho$ and they are exactly the p-th roots of unity times ρ

Def: The period of a digraph G is $p(G) := \gcd\{|C| \mid C \text{ is a cycle in } G\}$

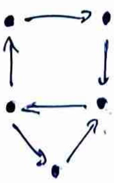
Intuition: $p(G)$ is the largest number p s.t. when you start at any vertex v and take any $v \to v$ walk P, then always $|P| = p \cdot n$, $n \in \mathbb{N}$



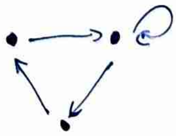
$p = 3$



$p = 2$



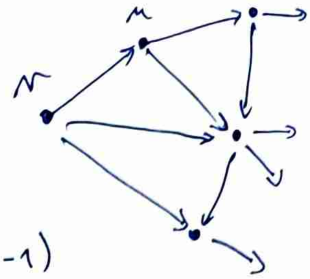
$p = 1$



$p = 1$

👁 The graph "ticks" in periods of length p, it doesn't make sense to check my position when time $\neq p \cdot n$, I will not have returned

Eigenvector centrality



- we want an eigenvector centrality for digraphs
- recall the approach:

$$\tilde{x}(0) = (1, \dots 1), \quad \tilde{x}(1) = A\tilde{x}(0), \quad \tilde{x}(t) = A\,\tilde{x}(t-1)$$

$$\Rightarrow \tilde{x}(1)_i = \sum_j A_{ij} = \sum_{j: j \to i} 1 = \text{indeg}(v_i)$$

$$\Rightarrow \tilde{x}(2)_i = \sum_j A_{ij}\,\tilde{x}(1)_j = \sum_{j: j \to i} \tilde{x}(1)_j = \sum_{j: i \to j} \text{indeg}(v_j)$$

this has <u>issues if G is not strongly connected</u>

- $k \geq 1 \Rightarrow v$ has score $\tilde{x}(k)_v = 0$
- $k \geq 2 \Rightarrow u$ has score $\tilde{x}(k)_u = 0$

if G is acyclic, then $\tilde{x}(k) \to 0$ as $k \to \infty$

$\Rightarrow$ only vertices in strongly connected components and their out-components have nonzero centrality in the limit

<u>Def</u>: Let G be a <u>strongly connected digraph</u> with vertices $V = \{1, \dots n\}$

Perron-Frobenius: $\rho(A)$ is a simple e.v. of A_G with right-eigenvector $v > 0$

The <u>eigenvector centrality</u> of G is normalized v:

$$\Rightarrow \underline{A\tilde{c} = \rho(A) \cdot \tilde{c}} \quad \text{s.t.} \quad \tilde{c} > 0 \;\;\&\;\; \|\tilde{c}\| = 1$$

<u>Intuition</u>: The process limit is (as we have analyzed earlier)

- suppose A_G diagonalizable: $\lambda_1 \geq \dots \geq \lambda_n$ & $v_1, \dots v_n$ eigenbasis

$\Rightarrow$ write $\tilde{x}(0) = \sum \alpha_i v_i$

$$\Rightarrow \tilde{x}(k) = A^k \tilde{x}(0) = A^k \sum \alpha_i v_i = \sum \alpha_i A^k v_i = \sum \alpha_i \lambda_i^k v_i = \rho^k \cdot \sum \alpha_i \left(\frac{\lambda_i}{\rho}\right)^k v_i$$

- if G is aperiodic then ρ is strictly dominant and we get

$\circledast$ $\tilde{x}(k) \longrightarrow \rho^k \alpha_1 v_1$ as $k \to \infty$

$\Rightarrow \tilde{x}(k)$ is determined by $v_1 > 0 \Rightarrow$ we just normalize it

<u>Note</u>: A_G is not guaranteed to be diagonalizable

$\Rightarrow$ use Jordan normal form, argument still work $\because \rho$ is a simple eigenvalue

$\circledast$ not quite true if G has period > 1, but the other dominant eigenvalues have complex eigenvectors, so we use only the real one

Katz centrality

→ improved eigenvector centrality, works even for not strongly connected graphs

→ imagine social network → instagram, twitter, ...

→ user is important if other important users follow him

⇒ better: if Trump → A → B, then Trump is also close to B

Def: Let G be a digraph with vertices $V = \{1, \dots n\}$, adjacency matrix A.
The Katz centrality of G is the vector $\tilde{c} \in \mathbb{R}^n$

$$\tilde{c} := \sum_{k=1}^{\infty} \alpha^k A^k \beta \quad , \; 0 < \alpha < \frac{1}{\rho(A)} \text{ decay factor}, \; \beta \in \mathbb{R}^n = \text{source weight}$$

Intuition: $\beta = \tilde{x}(0)$ serves as the strenght of the node when acting as a source

$$\to c_i = \sum_{k=1}^{\infty} \alpha^k \sum_j A_{ij}^k \beta_j$$

↳ # $j \to i$ walks of lenght k weighted by β_j

⇒ $\alpha < 1$ ⇒ longer paths are penalized

⇒ high score for vertices which can be reached by many other important nodes, with shorter paths contributing more

Theorem: The Katz centrality converges to

$$\tilde{c} = (I_n - \alpha A)^{-1} \beta - \beta \qquad \to \text{provided } 0 < \alpha < \frac{1}{\rho(A)}$$

Proof: Consider the series

$$\tilde{c} = \alpha A \beta + \alpha^2 A^2 \beta + \alpha^3 A^3 \beta + \dots = (\alpha A + \alpha^2 A^2 + \dots) \beta$$

$$= (I_n + \alpha A + \alpha^2 A^2 + \dots) \beta - \beta$$

$$= (I_n - \alpha A)^{-1} \beta - \beta \qquad \dots \; 1 + a + a^2 + \dots = \frac{1}{1-a}$$

But we need $I_n - \alpha A$ to be invertible ≡ regular ≡ 0 is not an eigenvalue

→ if A has eigenvalues $\lambda_1, \dots \lambda_n$, then $I_n - \alpha A$ has eigenvalues $1 - \alpha \lambda_i$

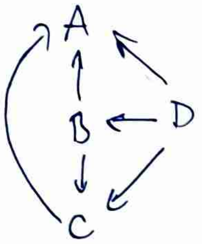
claim: $0 < \alpha < \frac{1}{\rho(A)} \Rightarrow 1 - \alpha \lambda_i \neq 0 \equiv \alpha \lambda_i \neq 1$

- $\mathrm{Im}(\lambda_i) \neq 0 \Rightarrow \mathrm{Im}(\alpha \lambda_i) \neq 0 \Rightarrow \alpha \lambda_i \neq 1$
- $\lambda_i \leq 0 \Rightarrow \alpha \lambda_i \leq 0 < 1 \Rightarrow \alpha \lambda_i \neq 1$
- $\lambda_i > 0 \Rightarrow \alpha \lambda_i < \frac{1}{\rho(A)} \lambda_i \leq 1 \Rightarrow \alpha \lambda_i \neq 1$

Page rank

→ used by Google to rank importance of web pages

→ idea: consider that there exist only 4 web pages A, B, C, D



Init: $p(A) = p(B) = p(C) = p(D) = \frac{1}{n} = 0.25$

Iter 1: B transfers $p(B)/\text{outdeg}(B) = 0.125$ to A and C
C transfers $p(C)/\text{outdeg}(C) = 0.25$ to A
D transfers $p(D)/\text{outdeg}(D) = 0.083$ to A, B, C

$\Rightarrow p(A) = 0.458,\ p(B) = 0.083,\ p(C) = 0.208,\ p(D) = 0$

Iter 2: $p(A) = \frac{p(B)}{2} + \frac{p(C)}{1} + \frac{p(D)}{3} = 0.25$
$p(B) = \frac{p(D)}{3} = 0$
$p(C) = \frac{p(B)}{2} + \frac{p(D)}{3} = 0.042$
$p(D) = 0$

Iter 3: $p(A) = 0.042,\ p(B) = 0,\ p(C) = 0,\ p(D) = 0$

Iter 4: $p(A) = p(B) = p(C) = p(D) = 0$

→ we assign each source a weight $= \frac{1}{\text{outdegree}}$ like β in Katz centrality

→ but the rank is calculated iteratively like in eigenvector centrality

this will converge to zero if the graph is not strongly connected

$\Rightarrow$ fix: add small chance to teleport to a random page.

↳ damping factor $d \in (0,1)$ → $\mathbb{P}[\text{teleport}] = 1-d$

↳ suppose $d = 0.9$, then $\mathbb{P}[\text{teleport to } D] = (1-d) \cdot \frac{1}{n} = 0.1 \cdot 0.25 = 0.025$

Setup

- $A \in \mathbb{R}^{n \times n}$ adjacency matrix
- $D \in \mathbb{R}^{n \times n}$ diagonal with $D_{jj} = \text{outdeg}(v_j)$ $\Rightarrow D^{-1}_{jj} = \frac{1}{\text{outdeg}(v_j)}$
- $M := AD^{-1}$ column-stochastic matrix of neighbour jump probabilities

$\Rightarrow M_{ij} = A_{ij} \cdot \frac{1}{\text{outdeg}(v_j)} = \mathbb{P}[\text{going } j \to i] \geq 0$

$\Rightarrow$ the elements of each column sum up to 1

Fact: The spectral radius of a column stochastic matrix is $\rho(M) = 1$

$M, N \in \mathbb{R}^{n \times n}$ col. stoch $\Rightarrow M \cdot N$ col. stoch

Def: Let G be a directed graph with neighbour jump probability matrix $M \in \mathbb{R}^{n\times n}$.
The page rank of G is initialized and iteratively computed as

$$p^{(0)} = \frac{\tilde{1}}{n} = \left(\frac{1}{n}, \frac{1}{n}, \dots \frac{1}{n}\right) \in \mathbb{R}^n$$

usually $d = 0.85$

$$p^{(k+1)} = d M p^{(k)} + (1-d)\cdot p^{(0)} \quad , \quad d \in (0,1) = \text{damping factor}$$

Intuition: At the start, the user randomly chooses a page $\Rightarrow p^{(0)}$

→ then, with probability d, he follows a random link on this page transferring a portion of its pagerank (equal to $\frac{1}{\text{outdeg}}$) to the new page

→ with probability $1-d$, he teleports to a completely random page $\Rightarrow p^{(0)}$

$$p_i^{(k+1)} = d \sum_j M_{ij}\, p_j^{(k)} + (1-d)\cdot\frac{1}{n}$$

$$= d \sum_{j:\, j\to i} \frac{1}{\text{outdeg}(v_j)}\, p_j^{(k)} + \frac{1-d}{n}$$ → randomly teleported here with $\mathbb{P} = \frac{1-d}{n}$

↳ $p(A) = \frac{p(B)}{2} + \frac{p(C)}{1} + \frac{p(D)}{3}$

Properties: Damping ensures that the process is strongly connected and doesn't get stuck in cycles (the resulting Markov chain has a unique stationary dist.)

$\Rightarrow$ page rank always exists, is unique and strictly positive

Theorem: Page rank converges to $p = (I_n - dM)^{-1}\cdot(1-d)\, p^{(0)}$

Proof: Denote $A := dM$, $B := (1-d)\, p^{(0)}$, then

$$p^{(1)} = Ap^{(0)} + B \quad , \quad p^{(2)} = A(Ap^{(0)} + B) + B = A^2 p^{(0)} + AB + B$$

$$p^{(3)} = A^3 p^{(0)} + A^2 B + AB + B$$

$1 + a + a^2 + \dots = \frac{1}{1-a}$

$$p^{(k)} \to A^k p^{(0)} + (I_n + A + A^2 + \dots)B = A^k p^{(0)} + (I_n - A)^{-1} B \to (I_n - A)^{-1} B$$

! for the series to converge we need

- $\rho(A) < 1$ ... $\rho(A) = \rho(dM) = d\cdot\rho(M) = d\cdot 1 < 1$ ✓
- from this we can show that $I_n - A$ is invertible, as with Katz

👁 $A^k = d^k M^k \to O_{n\times n}$ ∵ $d^k \to 0$ and M^k is col. stochastic (small)

Remark: In practice, page rank is calculated iteratively, because matrix inversion is costly.