

Complex Networks

- reprezentace reálného systému pomocí grafu
- sociální síť ↗ výhody = aktivity - uživatelé, sloužení, komunity
↳ hrany = interakce - friendship, collaboration
Facebook = (users, friendship) ... neorientovaný graf bez smyček

• Milgram experiment 1963 v Omaha

- ↳ struktura social networks
- ↳ kolik středů je potřeba aby rukopis poslal dopis koncovému?
- ⇒ six degrees of separation ... v průměru stačí 6

↳ small world phenomenon

- ↳ význam definice:
 1. vznikají clustery východů
 2. krátké vzdálenosti mezi východů
 3. ta síť je řidká

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

• Dynamic Systems

- systém se skládá z množiny menších podsystemů co spolu interagují
- vnitřní dynamika těch podsystemů většinou popsána diff. rovnicemi
- synchronizace = interakce mezi páry podsystemů

→ příklad: Snižování klima

- simulace celého světa naráží na memoře
- různí se menší regiony / fenomény a jak spolu interagují
- jednoduše vnitřní dynamika a synchronizace & máme dost dat
⇒ je to rezistenz, ale problém s výpočetní náročností a skobitou
- reálné systémy jsou nesty ↗ Elima - nemá dost dat
↳ mozek - neznáme vnitřní procesy

Strukturální × funkční konzervativita

- příklad: regiony mozků

• mozek je nejedna fyzická síť → měříme strukturální síť

• magnetická rezonance měří výšku jak silně spolu jednotlivé regiony komunikují
↳ funkční síť

! ale ty dvě sítě si někde v mozku nedoprovázají

→ rejstřík se, že mozek je small world

↳ regiony mozků = clusters, komunikace mezi nimi je velmi rychlá

→ Alzheimer: denser clusters, longer distances

↳ těžce formuje small world

↳ pustá ráz efektivní přenos informací mezi regiony

Protein interactions

vrcholy: proteiny

hrany: jde spolu ty proteiny interagují v rámci prostředí (urazitelné,...)

→ cíl: zjistit funkci proteinu podle jeho sousedů

↳ PPIN = Protein-Protein Interaction Network

• neighborhood methods: funkce proteinu ~ funkce jeho sousedů

• communities: vrcholy rozdělují na communities $V = C_1 \cup C_2 \cup \dots \cup C_m$

↳ jednotlivé communities by mohly být dosti dense

↳ každá komunita by snad měla mít nějakou klíčovou funkci

Internet - (IP adresy, konverza)

- percolation theory = rezistence síti vůči náhlému

↳ kolik vrcholů / bran. (DNS servery...) může smazat aby síť zůstala souvislá

WWW - (web pages, references)

- dělá se speed of fake news, communities, business analytics

Large networks ↑ některé sítě jsou opravdu obrovské

- WWW, internet nemá reálné možné analýzy, protože celé neto někam užít

→ dělá se sampling a modeluje se k řešení random grafy

→ tímto způsobem lze branism pustupují vše jaro a pravidelnost

→ jeden grafográfický limit: $P[\text{event}] \rightarrow 0$. as $|V(G)| \rightarrow \infty$

- Stupeň vrcholi reálných sítí

→ chceme aby se reprodukce výslednou grafu stereotypu se budeme modelovat

→ reálné sítě se chovají jinak

- few very high degree nodes
 - many low degree nodes
- } Scale-free property

random attacks jsou OK → množina vrcholi není důležitá

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

- Měření důležitosti vrcholů

→ různé spůsoby 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

↳ důležitostm vrcholům se říká hubs nebo centralities

- Super centralities

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

→ nejenže je s kohodl lepší, je to všechno nejblíže, ale
získaný je všechna spojení přes kohde lepší efektivně

Small-World Graphs

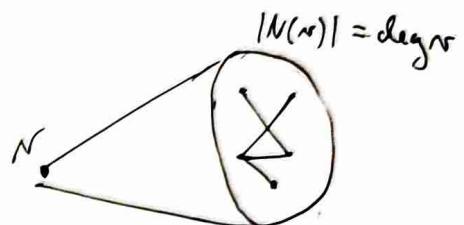
- locally dense subgraphs \rightarrow nodes create clusters ... $S(\text{cluster}) \in O(1)$
- globally the graph is sparse ... $S(G) \in O(\frac{1}{m})$, $O(\frac{\log n}{m})$... $\lim_{n \rightarrow \infty} S = 0$
- short average distance between nodes ... $d(u, v) \ll m$

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

Def: For $v \in V$, define

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

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



$$C(v) := S(G[N_G(v)]) = \frac{|E(G[N_G(v)])|}{\binom{\deg v}{2}} \dots \text{vertex clustering coefficient}$$

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

💡 $\forall v \in V: C(v) = 0 \iff \text{no } \Delta \text{ in } G$

Def: The average clustering coefficient is

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

💡 $C(G) = 0 \iff \text{no } \Delta \text{ in } G$

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



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

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

$$\begin{array}{c} \cdot \\ \checkmark \end{array} \longrightarrow \begin{array}{c} \cdot \\ \cdot \\ \cdot \end{array} \longrightarrow \begin{array}{c} \cdot \\ \cdot \\ \cdot \\ \cdot \end{array}$$

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

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

→ 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 partsizes having sizes as similar as possible.

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



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

Pf: Consider $n = q \cdot r$

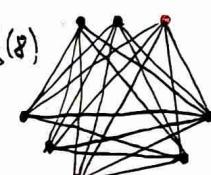
↪ r partsizes, each of size q

⇒ between any pair: complete bipartite graph $K_{\frac{m}{r}, \frac{m}{r}}$

$$\Rightarrow t_r(n) = \binom{r}{2} \cdot \left(\frac{m}{r}\right)^2 = \frac{r-1}{r} \cdot \frac{m^2}{2}$$

Consider $n = q \cdot r + s$

↪ $T(q \cdot r + s) \sim T(q \cdot r)$ and extra s vertices, each one in a different partsize 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 $\rightarrow \binom{s}{2}$

• connect them to partsize of $T(q, r)$ $\rightarrow s \cdot (r-1) \cdot q$:

⇒ in total:

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

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

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

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



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 & $|E| > \epsilon_n(n) \Rightarrow G$ contains K_{r+1} .

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

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

Average path length

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

Examples:

• Complete graphs: $L(K_m) = 1$

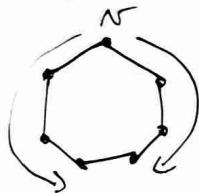
• Stars, S_m



$$L(S_m) = \frac{2}{(m+1)m} \left(m \cdot 1 + \binom{m}{2} \cdot 2 \right) = \frac{2}{m+1} \left(1 + \frac{m-1}{2} \cdot 2 \right) = \frac{2m}{m+1} \rightarrow 2$$

→ 1 center + m nodes

• Cycles, C_m



$$m = 2k+1 \Rightarrow L(C_m) = \frac{1}{m(m-1)} \cdot \left(m \cdot \left(2 \cdot \underbrace{(1+2+\dots+k)}_{m \text{ vertices}} \right) \right) = \frac{1}{m(m-1)} \cdot m \cdot \frac{m-1}{2} \cdot \frac{m+1}{2} = \frac{m+1}{4}$$

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

$$\underline{L(C_m) \rightarrow \frac{m}{4}}$$

Small world comparison of simple graphs

• Stars S_m : $L(S_m) \sim 2$, $C(S_m) = 0$, $S(S_m) = \frac{m}{\binom{m+1}{2}} = \frac{2m}{m(m+1)} = \frac{2}{m+1} \rightarrow 0$

↳ short distances, no clustering, low density

• Cycles C_m : $L(C_m) \sim \frac{m}{4}$, $C(C_m) = 0$, $S(C_m) = \frac{m}{\binom{m}{2}} = \frac{2}{m-1} \rightarrow 0$

↳ long distances, no clustering, low density

• Complete graphs K_m : $L(K_m) = 1$, $C(K_m) = 1$, $S(K_m) = 1$

↳ short distances, absolute clustering, full density

• Wheels W_m : $L(W_m) \sim 2$, $C(W_m) \sim \frac{2}{3}$, $S(W_m) = \frac{2m}{\binom{m+1}{2}} = \frac{4}{m+1} \rightarrow 0$

↳ short distances, high clustering, low density



$$C(W_m) = \frac{1}{m+1} \left(\frac{m}{\binom{m}{2}} + m \cdot \frac{2}{3} \right) = \frac{1}{m+1} \left(\frac{2}{m-1} + \frac{2m}{3} \right) \rightarrow \frac{2}{3}$$

$$L(W_m) = \frac{1}{m(m+1)} \left(m \cdot 1 + m \cdot (3 + (m-3) \cdot 2) \right) = \frac{1+3+2m-6}{m+1} = \frac{2(m-1)}{m+1} \rightarrow 2$$

$C_m + N$

- Complete bi-graph $K_{m,m}$: $L(K_{m,m}) \sim \frac{3}{2}$, $C(K_{m,m}) = 0$, $S(K_{m,m}) \sim \frac{1}{2}$

↳ short distances, no clustering, high density

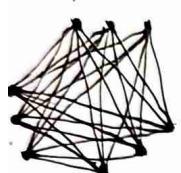
$$L(K_{m,m}) = \frac{2}{(2m)(2m-1)} \left(m \cdot m + 2 \cdot \binom{m}{2} \cdot 2 \right) = \frac{m^2 + 2m(m-1)}{m(2m-1)} = \frac{m+2(m-1)}{2m-1} = \frac{3m-2}{2m-1} \rightarrow \frac{3}{2}$$

$$C(K_{m,m}) = 0 \quad \because \Delta \notin K_{m,m}, \quad S(K_{m,m}) = \frac{m^2}{\binom{2m}{2}} = \frac{2m^2}{2m(2m-1)} = \frac{m}{2m-1} \rightarrow \frac{1}{2}$$

- Complete k-partite graphs K_{k+m} : $L(K_{k+m}) \sim \frac{k+1}{k}$, $C(K_{k+m}) \sim \frac{k-2}{k-1}$, $S(K_{k+m}) \sim \frac{k-1}{k}$

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

≥ 3 -partite



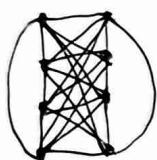
$$L(K_{k+m}) = \frac{1}{(km)(km-1)} \cdot (km) \left(1 \cdot (k-1)m + 2 \cdot (m-1) \right) = \frac{km - m + 2m - 2}{km-1} = \frac{m(k+1)-2}{m \cdot k - 1} \rightarrow \frac{k+1}{k}$$

$$C(K_{k+m}) = \frac{1}{km} \cdot (km) \cdot S(K_{(k-1)+m}) = \frac{m(k-2)}{m(k-1)-1} \rightarrow \frac{k-2}{k-1}$$

$$S(K_{k+m}) = \frac{1}{\binom{km}{2}} \cdot \frac{1}{2} (k \cdot m \cdot (k-1)m) = \frac{km(k-1)m}{km(km-1)} = \frac{m(k-1)}{m \cdot k - 1} \rightarrow \frac{k-1}{k}$$

- $B_{m,m}$ = 2 cycles + K_{m+m} : $L(B_{m,m}) \sim \frac{3}{2}$, $C(B_{m,m}) \sim 0$, $S(B_{m,m}) \sim \frac{1}{2}$

↳ short distances, no clustering, high density



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

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

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

- Ring Lattice $C_{m,\ell}$: $L(C_{m,\ell}) \sim \frac{m}{2\ell}$, $C(C_{m,\ell}) \sim \frac{3}{4}$, $S(C_{m,\ell}) = \frac{\ell}{m-1}$

↳ C_m + each vertex is connected to ℓ nearest neighbors

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

• clustering



$$\frac{(\ell/2-1)(\ell/2)}{2} = \binom{\ell/2}{2}$$



$$2 \cdot \binom{\ell/2}{2} + \underbrace{(1+2+\dots+(\ell/2-1))}_{= \binom{\ell}{2}} = 3 \cdot \binom{\ell/2}{2}$$

$$\Rightarrow C(C_{m,\ell}) = C(r) = \frac{3 \cdot \binom{\ell/2}{2}}{\binom{\ell}{2}} = \frac{3 \left(\frac{\ell}{2} \right) \left(\frac{\ell}{2} - 1 \right)}{\ell(\ell-1)} = \frac{3}{4} \cdot \frac{\ell-2}{\ell-1} \rightarrow \frac{3}{4}$$

• average path length $m = 2l+1$

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

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



$$l = q \cdot \frac{\ell}{2} + r \quad q \approx \frac{2\ell}{\pi} \approx \frac{m}{\ell}$$

$$\bullet \text{density}, \quad S = \frac{(m\ell)/2}{\binom{m}{2}} = \frac{\ell}{m-1}$$

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 $C_{n,\epsilon}$
2. For $v \in V(C_{n,\epsilon})$:
3. For $\#$ rightwards edge $\{v, u\}$: # $\frac{1}{2}$ of them in total
4. rewire $\{v, u\}$ with probability π

↳ change $\{v, u\}$ to $\{v, r\}$ where r is chosen uniformly from $\{r \in V \mid r \neq v \text{ & } \{v, r\} \notin E\}$... 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 $\# \{i, j\} \subset V$: insert an edge with probability $p \approx \text{density of final graph}$

WS($\pi \rightarrow 1$) $\rightsquigarrow ER(p = \frac{\ell}{m-1})$

Properties: $\bar{k} := \text{average degree} \Rightarrow p \approx \frac{\bar{k}}{m}$

$$L(G_{n,p}) \sim \frac{\log(m)}{\bar{k}} \sim \frac{1}{p} \cdot \frac{\log(m)}{\bar{k}} \Rightarrow \text{short distances}$$

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

Small World Coefficient

→ all small world notions are somewhat relative \Rightarrow 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 $\Leftrightarrow L(G) \gtrsim L(R_G)$ & $C(G) \gg C(R_G)$

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

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

Centralities

↳ important nodes in a network

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

• Excentricity: $e(v) := \max_{u \in V} d(v, u)$... 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 through v

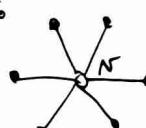
$C_B(v) = \frac{\# \text{ shortest paths in } G \text{ through } v}{\# \text{ shortest paths in } G}$

$v = \text{inner vertex}$

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

Examples

• Star S_m



leafs: $BC = 0$

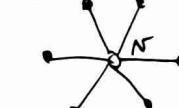
center: $C_B(v) = \sum_{s \neq v \neq t} \frac{1}{1} = \binom{m}{2}$

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

iterating pairs $\Rightarrow \binom{m}{2}$ summands

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

S_6



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

center: • adjacent vertices contribute 0

$\rightarrow m$ of shell

• nonadjacent vertices with a common neighbor: $\frac{1}{2}$

• more distant contribute 1 $\rightarrow \binom{m}{2} - m - m$ of shell

$$\begin{aligned} \Rightarrow C_B(v) &= m \cdot \frac{1}{2} + ((\binom{m}{2} - 2m) \cdot 1) = \frac{m}{2} + \frac{m(m-1)}{2} - 2m = \frac{m}{2}(1 + (m-1) - 4) \\ &= \underline{\underline{\frac{1}{2}m(m-4)}} \end{aligned}$$



• Complete graph K_m

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



\hookrightarrow all vertices are important \Rightarrow none of them is

Global betweenness centrality

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

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

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

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

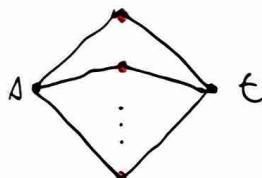
Proof: Define overall BC as $C_{BS} := m \overline{C}_B(G) = \sum_{v \in V} C_B(v)$

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

1) $d(s,t) = 1$: $s \rightarrow t$

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

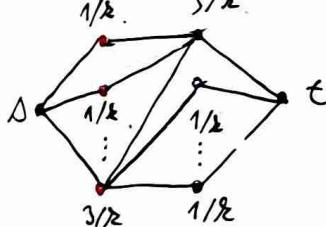
2) $d(s,t) = 2$:



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

→ exactly k shortest paths $\left\{ \begin{array}{l} \text{contribution} = k \cdot \frac{1}{k} = 1 \\ \text{so each vertex contributes } \frac{1}{k} \end{array} \right.$

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 l = 1

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

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

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

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



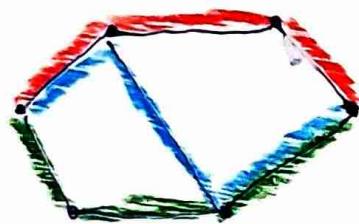
Note: We always assume G connected.

Alternative proof

connected

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

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 $(d(s,t)-1) \cdot \tilde{C}_{s,t}$



Theorem: $\overline{C}_B(G) = \frac{m-1}{2} (L(G)-1)$

$$\text{Pf: } \overline{C}_B(G) = \frac{1}{m} \sum_{v \in V} C_B(v) = \frac{1}{m} \sum_{v \in V} \sum_{\substack{s \neq v \neq t}} \frac{\tilde{C}_{st}(v)}{\tilde{C}_{st}} = \frac{1}{m} \sum_{s \neq t} \frac{1}{\tilde{C}_{st}} \sum_{v \in V \setminus \{s, t\}} \tilde{C}_{st}(v)$$

$$= \frac{1}{m} \sum_{s \neq t} \frac{1}{\tilde{C}_{st}} (d(s,t)-1) \cdot \tilde{C}_{st} = \frac{1}{m} \left(\sum_{s \neq t} d(s,t) - \sum_{s \neq t} 1 \right)$$

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



- Betweenss and adding edges

→ is BC monotone?

Lemma: Global BC decreases with adding edges

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

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

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

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

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

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

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

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

$$\frac{m-1}{2} L(G') \leq \frac{m-1}{2} L(G) - \frac{d-1}{m} \quad \& \text{ recall } \frac{m-1}{2} L(G) = \overline{C}_B(G) + \frac{m-1}{2}$$

- Global BS of spanning trees

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

$$\overline{C}_B(G) \leq \overline{C}_B(G') - \frac{r}{m}, \quad r = |E \setminus E'|$$

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

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

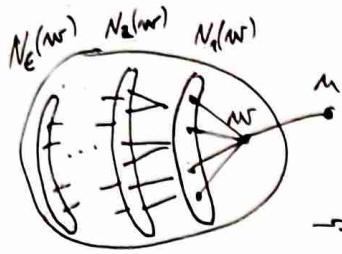
\Rightarrow BS decreased by $\frac{d-1}{m} \geq \frac{1}{m}$ each step (previous lemma) ■

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

$$\overline{C}_B(G) \leq \overline{C}_B(T) - \frac{m-m+1}{m} \quad \dots \because |E(T)| = m-1$$

Betweenness and adding vertices

→ let's just try adding an extra leaf to the graph



→ there are layers of ε -neighborhoods

$$N_\varepsilon(w) := \{v \in V \mid d(w, v) = \varepsilon\}$$

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

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

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

$$\overline{C}_B(G') = \frac{1}{m+1} \left(m \cdot \overline{C}_B(G) + \sum_{\varepsilon=1}^{e(w)} \varepsilon \cdot |N_\varepsilon(w)| \right), \quad e(w) = \text{eccentricity of } w \text{ in } G$$

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

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

$$= \sum d_G(x, y) + \sum_{\varepsilon} \varepsilon \cdot |N_\varepsilon(w)| + m = \sum d_G(x, y) + \underline{\Delta} + m$$

$$\Rightarrow \frac{1}{m+1} \sum d_{G'}(x, y) = L(G') = \frac{2}{m(m+1)} \sum d_G(x, y) + \frac{2}{m(m+1)} \Delta + \frac{2}{m+1} \\ = \frac{m-1}{m+1} \cdot \frac{1}{m} \cdot \sum + \dots = \frac{m-1}{m+1} L(G) + \frac{2}{m(m+1)} \Delta + \frac{2}{m+1}$$

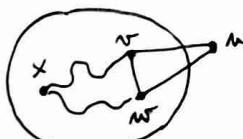
$$\bullet \quad \overline{C}_B(G') = \frac{m}{2} (L(G') - 1) = \frac{m}{2} \left(\frac{m-1}{m+1} L(G) + \frac{2}{m(m+1)} \Delta + \frac{2}{m+1} - 1 \right)$$

$$\Rightarrow \overline{C}_B(G') = \frac{m}{m+1} \cdot \frac{m-1}{2} (L(G) - 1 + 1) + \frac{1}{m+1} \Delta + \frac{m}{2} \cdot \frac{1-m}{m+1} \\ = \frac{m}{m+1} \cdot \overline{C}_B(G) + \frac{1}{m+1} \Delta + \underbrace{\frac{m(m-1)}{2(m+1)} + \frac{m(1-m)}{2(m+1)}}_{= \frac{1}{m+1} (m \overline{C}_B(G) + \Delta)} \quad \blacksquare$$

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

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

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



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

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

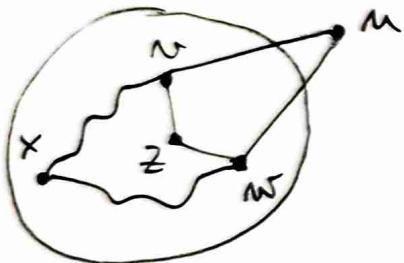
→ paths between x and u must use v or $w \Rightarrow +1$

$$\circlearrowleft C'_B(u) = 0$$

$$\circlearrowleft C'_B(v) + C'_B(w) = C_B(v) + C_B(w) + m - 2$$

→ other vertices (not only v, w) can also contribute $\Rightarrow \sum_{y \in V(G')} C'_B(y) \geq \sum_{x \in V(G)} C_B(x) + m - 2$ NOT END

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



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

↳ no: previously v and w contributed 1, now also 1

↳ see proof of relationship of $C_B(G)$ and $L(G)$

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



Calculating Betweenness

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

Algorithm:

Input: $G = (V, E)$

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

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

2. For $\forall s \in V$ do:

3. $S(v, t) \leftarrow 0 \quad \forall t, v \in V$

4. Run BFS to find $P_s(v)$ for $\forall v \neq s$

5. For $\forall t \neq s$ do:

6. Iterate back from s to t using $P_s(\cdot)$ and
increment $S(v, t)$ when we encounter v

7. Update $C_B(v) \leftarrow C_B(v) + \frac{S(v, t)}{S(s, t)}$ for $\forall s \neq v \neq t$

8. Return $C_B(v) / 2$

↳ $\tilde{O}(n^2)$

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

↳ but $C_B(v) = \sum_{\substack{\{s, t\} \subseteq V \\ s \neq v \neq t}} \frac{S(v, t)}{S(s, t)}$... 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 successors of v

③ We can accumulate $\delta(v)$ 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

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

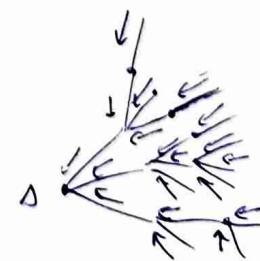
While Stack is not empty:

 top $v \leftarrow$ Stack

 for $\forall u \in P_s(v)$: $\delta(u) \leftarrow \delta(u) + \text{UPDATE}(\delta(v))$

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

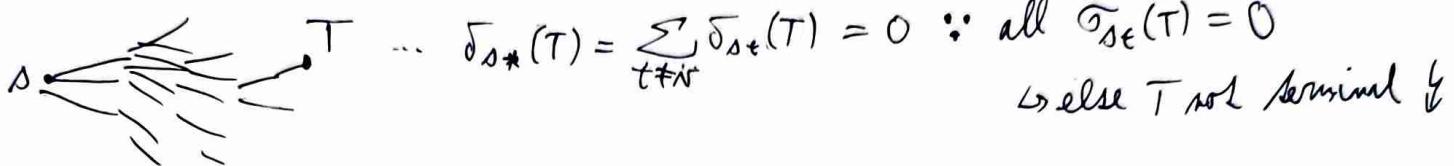
 ↓
 don't know



$$\text{Def: } \delta_{s,t}(v) = \frac{\tilde{\sigma}_{st}(v)}{\tilde{\sigma}_{st}} , \quad \delta_{s*}(v) = \sum_{t \in V} \delta_{st}(v) = \sum_{\substack{e \in V \\ t \neq s}} \delta_{st}(v)$$

$$\hookrightarrow \delta_{s,t}(s) = \delta_{st}(s) = 0 \quad \because \tilde{\sigma}_{st}(s) = \tilde{\sigma}_{st}(t) = 0 \quad \& \quad \tilde{\sigma}_{ss} = 1$$

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



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

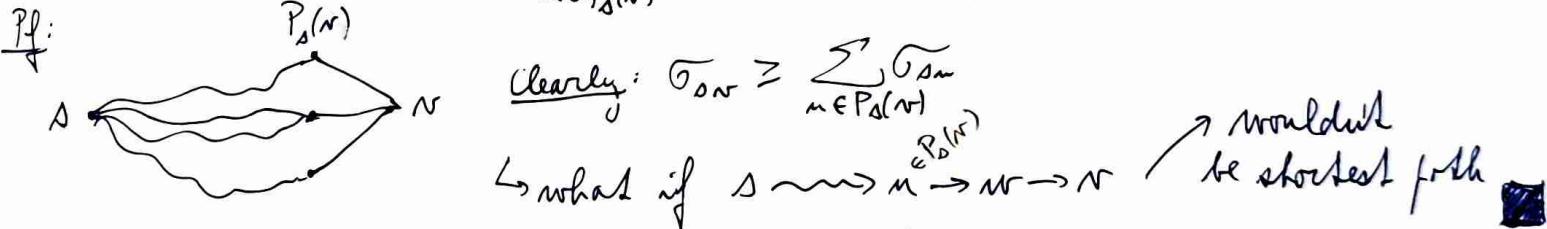
$$② \text{Want: } \delta_{s*}(v) = \sum_{w \in \text{succ}(v)} f(\delta_{sw}(w)) \quad , \quad v \neq s$$

Lemma 1: Let $s, v, t \in V$. Then the Δ -ineq ... two options

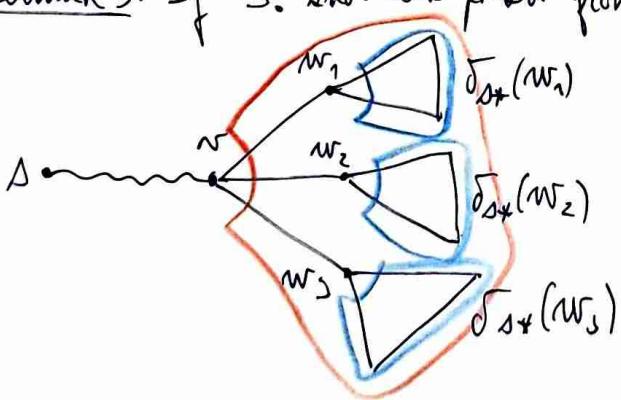
- $d(s, t) = d(s, v) + d(v, t) \Rightarrow \tilde{\sigma}_{st}(v) = \tilde{\sigma}_{sv} \cdot \tilde{\sigma}_{vt}$
- $d(s, t) < d(s, v) + d(v, t) \Rightarrow \tilde{\sigma}_{st}(v) = 0$

Pf: v lies on a shortest path $s \leftrightarrow t \Leftrightarrow d(s, t) = d(s, v) + d(v, t)$.

Lemma 2: For $s \neq v$: $\tilde{\sigma}_{sv} = \sum_{m \in P_s(v)} \tilde{\sigma}_{sm}$. $P_s(v)$... predecessors



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



$$\begin{aligned} \text{if } \tilde{\sigma}_{st} = 1 \quad \& \quad \tilde{\sigma}_{st}(v) \in \{0, 1\} \Rightarrow \delta_{st}(v) \in \{0, 1\} \\ \Rightarrow \delta_{s*}(v) &= \sum_{t \neq v} \delta_{st}(v) = \#\{t : d(s, t) > d(s, v)\} \\ \Rightarrow \boxed{\square} &= \sum_{w \in \text{succ}(v)} (1 + \delta_{sw}(w)) \end{aligned}$$

→ This is a special case of ②, we want this for general graphs.

Theorem: Let $s \in V$ and $r \neq s$. Then $\delta_{st}(r) = \sum_{w \in \text{succ}(r)} \frac{\overline{G_{sw}}}{\overline{G_{sr}}} (1 + \delta_{st}(w))$.

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

SP
δ_{st}(r) > 0 ⇔ r ∈ some shortest path P: s → t
for t such P ∃! edge irs in P s.t. irs ∈ succ(r)
⇒ extend G_{st}(r) to contain edge irs

$$\Rightarrow \delta_{st}(r) = \sum_{t \neq r} \delta_{st}(r) = \sum_{t \neq r} \sum_{w \in \text{succ}(r)} \delta_{st}(r, \{r, wr\}) = \sum_{w \in \text{succ}(r)} \sum_{t \neq r} \delta_{st}(r, wr)$$

→ How much is $\delta_{st}(r, wr) = \frac{\overline{G_{st}(r, wr)}}{\overline{G_{st}}}$? ↗ even for $t = wr \Leftrightarrow \overline{G_{tt}} = 1$

$$t = wr: \delta_{st}(r, wr) = \frac{\overline{G_{sr}} \cdot 1}{\overline{G_{st}}} = \frac{\overline{G_{sr}}}{\overline{G_{st}}}$$

$$t \neq wr: \delta_{st}(r, wr) = \frac{\overline{G_{sr}} \cdot \overline{G_{wt}}}{\overline{G_{st}}} = \frac{\overline{G_{sr}}}{\overline{G_{st}}} \cdot \frac{\overline{G_{st}(wr)}}{\overline{G_{sr}}} = \frac{\overline{G_{sr}}}{\overline{G_{sr}}} \cdot \delta_{st}(wr)$$

$$\Rightarrow \delta_{st}(r) = \sum_{w \in \text{succ}(r)} \left(\frac{\overline{G_{sr}}}{\overline{G_{sr}}} + \sum_{t \neq r, t \neq wr} \frac{\overline{G_{sr}}}{\overline{G_{sr}}} \cdot \delta_{st}(wr) \right) = \sum_{w \in \text{succ}(r)} \frac{\overline{G_{sr}}}{\overline{G_{sr}}} (1 + \delta_{st}(w))$$

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

Brandes Algorithm

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

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

1. $\forall v: C_B(v) \leftarrow 0$ $\delta_{\text{out}}(v)$ $\bar{\delta}_{\text{out}}$ $d(s, v)$ $P_s(v)$
2. For $\forall s \in V$ do:
3. $\forall v: \delta(v) \leftarrow 0$, $\bar{\sigma}(v) \leftarrow 0$, $d(v) \leftarrow \infty$, $P(v) \leftarrow \emptyset$
4. $\bar{\sigma}(s) \leftarrow 1$, $d(s) \leftarrow 0$
5. $Q \leftarrow \text{Queue}(s)$, $S \leftarrow \text{empty stack}$

6. While Q is not empty do:

7. $v \leftarrow Q.\text{dequeue}()$
8. $S.\text{push}(v)$

9. For $\forall w \in \text{Neighbors}(v)$ do:

10. if $d(w) = \infty$: $d(w) \leftarrow d(v) + 1$ & $Q.\text{enque}(w)$
11. if $d(w) = d(v) + 1$: $\bar{\sigma}(w) \leftarrow \bar{\sigma}(v) + \bar{\sigma}(v)$ & $P(w).\text{add}(v)$

12. While S is not empty do:

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

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

15. if $w \neq s$: $C_B(w) \leftarrow C_B(w) + \bar{\delta}(w)$ $\rightarrow C_B(w) = \sum_{\Delta \neq w} \bar{\delta}_{\Delta w}(w)$

16. end for

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

BFS is $\Theta(n \cdot m) = \Theta(n^2)$

$\hookrightarrow G$ connected

Total # predecessors is $\Theta(n^2)$

Time complexity

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

Generalizations:

• directed graphs: $C_B(v) := \sum_{s \neq v} \sum_{t \neq v} \frac{\bar{\delta}_{st}(v)}{\bar{\delta}_{st}} = 2 \cdot C_B$ for connected G

\hookrightarrow simply don't divide by 2 at the end

• weighted graphs:

\hookrightarrow use Dijkstra instead of BFS

Bounds on betweenness

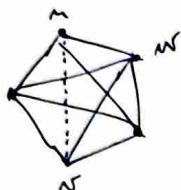
💡 $0 \leq C_B(n) \leq \binom{n-1}{2}$...  $\rightsquigarrow 0$

 $\rightsquigarrow \binom{n-1}{2}$

💡 $C_B(n) = 0 \Leftrightarrow G[N(n)] \cong K_{\deg(n)}$

Ex:

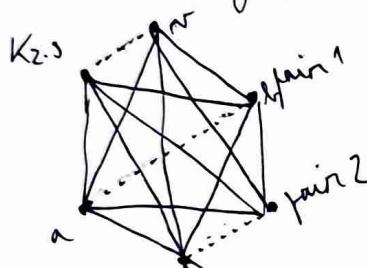
① Almost complete: $K_m \setminus \{u, v\}$



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

$$C_B(w) \dots \text{now } \tilde{\sigma}_{uw}(w) = 1, \quad \tilde{\sigma}_{vw} = m-2 \Rightarrow C_B(w) = \frac{1}{m-2}$$

② Cocktail graph: $K_{2m} - \{\text{Perfect matching of size } m\}$



$C_B(u) \dots \text{all pairs } \{a, b\} \text{ s.t. } a \dots b \text{ contribute } \frac{1}{2^{m-2}}$

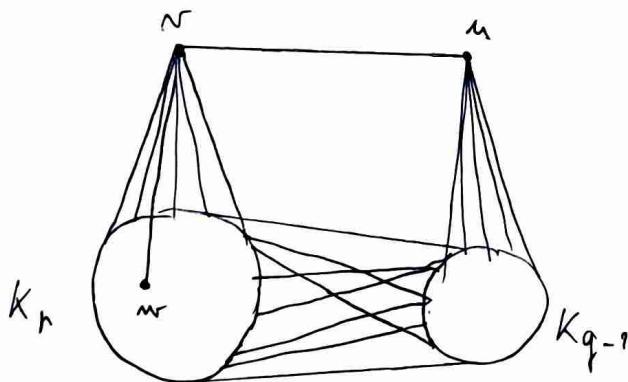
$$\hookrightarrow \tilde{\sigma}_{ab}(v) = 1, \quad \tilde{\sigma}_{ab} = 2m-2$$

↳ There are $m-1$ such pairs

$$\Rightarrow C_B(v) = \frac{m-1}{2^{m-2}} = \underline{\underline{\frac{1}{2}}}$$

③ Bicycle graph

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



$C_B(v) \dots \text{only pairs } \{u, w\}, \quad w \in K_q \text{ contribute}$

$$\hookrightarrow \tilde{\sigma}_{uvw}(v) = 1, \quad \tilde{\sigma}_{uvw} = 1 + (q-1) = q$$

$$\Rightarrow C_B(v) = \frac{k}{q} \quad \hookrightarrow |K_q| = k$$

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?

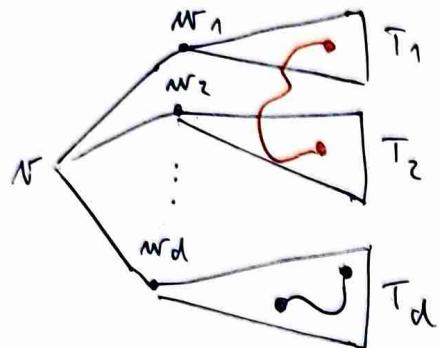
... can we create a model which generates random graphs with this score?

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

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

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 v_i of v we have trees T_i

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

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

$$\bullet x, y \in T_i: \overline{G}_{xy}^G(v) = \overline{G}_{xy}^T = 0 \quad \dots \text{BFS}$$

$$\bullet x \in T_i, y \in T_j: \frac{\overline{G}_{xy}^T(v)}{\overline{G}_{xy}^G} = 1 \geq \frac{\overline{G}_{xy}^G(v)}{\overline{G}_{xy}^G}$$

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

$$d \leq \Delta$$

*: 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 \blacksquare

Maximal betweenness

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

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

• What we know:

$$C_{B\max}(\mathcal{G}_m) = \binom{m-1}{2} \quad \dots \mathcal{G}_m = \text{graphs of size } m$$

$$C_{B\max}(\mathcal{G}_m^\Delta) = \frac{\Delta-1}{2\Delta} (n-1)^2 \quad \dots \mathcal{G}_m^\Delta = \text{graphs of size } m \text{ with max. deg } \Delta$$

• We will show:

$$C_{B\max}(\mathcal{G}_m^2) = \frac{(n-3)^2}{2} \quad \dots \mathcal{G}_m^2 = 2\text{-connected graphs of size } m$$

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

$$e_1(H) := \#\text{adjacent pairs in } H = |E(H)|$$

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

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

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

$$E_1 = E(H) \subseteq V_2$$

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

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

$$\text{(*) } \sigma_{xy}(v) = 0 \dots \leq \frac{1}{2} \checkmark$$

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

$$\Rightarrow \sigma_{xy} \geq 2$$

first fall

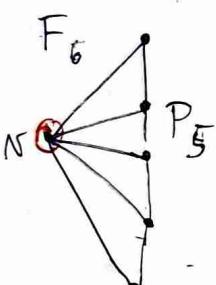
second fall

via v

$$\Rightarrow C_B(v) = \underbrace{\sum_{\{x, y\} \in E_1} \frac{\sigma_{xy}(v)}{\sigma_{xy}}} + \underbrace{\sum_{\{x, y\} \in E_2} \frac{\sigma_{xy}(v)}{\sigma_{xy}}} + \underbrace{\sum_{\{x, y\} \in R} \frac{\sigma_{xy}(v)}{\sigma_{xy}}}$$

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

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



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

• $d^P(x, y) = 1$... contributes 0

• $d^P(x, y) = 2$... contributes $\frac{1}{2}$, overall $\frac{1}{2} \cdot (|P|-2) = \frac{m-3}{2}$

• $d^P(x, y) \geq 3$... contributes 1, overall $\underbrace{\binom{m-1}{2}}_{\text{total}} - \underbrace{(m-2)}_{d=1} - \underbrace{(m-3)}_{d=2}$

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

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

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

$$\rightarrow \text{define } f(H) := e_1(H) + \frac{1}{2} e_2(H)$$

$$(*) \text{ it holds that } f(H) \geq \frac{3m-7}{2} \dots \text{ proof later}$$

\rightarrow using previous theorem

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

\rightarrow proof of $(*)$:

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

\rightarrow minimal such graph is P_{m-1} , then $f(P_{m-1}) = m-2 + \frac{1}{2}(m-3) = \frac{3m-7}{2}$

\rightarrow intuitively for "bigger" graphs, f should be larger

~~→ proof of (*)~~

min. day

$$\textcircled{1} \quad \underline{\delta(H) \geq 3}: \quad e(H) = |E(H)| = \frac{1}{2} \sum_{v \in V} \deg(v) \geq \frac{1}{2} \cdot 3 \cdot (m-1) = \frac{3m-3}{2} > \frac{3m-7}{2}$$

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

\rightarrow induction over $|H|$, base of induction = paths P_m

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

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

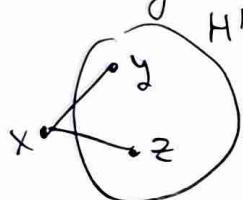
$$\text{I.H.: } f(H') \geq \frac{3(m-1)-7}{2} = \frac{3m-20}{2}, \text{ we add } x \text{ back for } H$$

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

$$\bullet \quad e_1(H) = e_1(H') + 1 \quad \dots \quad \Delta e_1 = 1$$

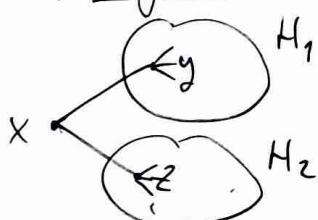
$$\bullet e_2(H) = e_2(H') + \deg_{H'}(y) \geq e_2(H') + 1 \quad \dots \quad e_2 \geq 1 \quad \left\{ \begin{array}{l} \Delta h = 1 + \frac{1}{2} \cdot 1 = \frac{3}{2} \\ \end{array} \right.$$

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



$$\bullet \quad \rho_1(H) = \rho_1(H^1) + 2 \dots + \rho_1 = 2 \quad \Rightarrow \quad \Delta f \geq 2 \quad \checkmark$$

c) $\deg(x)=2$ & H^1 disconnected = $H_1 \cup H_2$



$$\bullet \text{ If } H_1 \text{ size } m_1: f(H_1) \geq \frac{3(m_1+1)-7}{2} \quad \left\{ \begin{array}{l} \text{if } m_1 \leq 1 \\ \text{if } m_1 > 1 \end{array} \right. \quad \text{then } \frac{3(m_1+m_2+2)-14}{2} = \frac{3m-14}{2}$$

$$\text{• } \underline{H_2 \text{ size } m_2}: f(H_2) \geq \frac{s(m_2+1)-7}{2}$$

$$\frac{3(M_1+M_2+2)}{2} - 14 = \frac{3N}{2}$$

$$\cdot \mathcal{C}_1(\mu) = \mathcal{C}_1(\mu_1) + \mathcal{C}_1(\mu_2) + \dots + \Delta \mathcal{C}_1 = 2$$

$$\bullet \quad c_2(H) = c_2(H_1) + c_2(H_2) + 1 + \deg_{H_1}(y) + \deg_{H_2}(z) \dots \Delta c_2 \geq 3$$

$$\Rightarrow \Delta f = \Delta e_1 + \frac{1}{2} \Delta e_2 = 2 + \frac{3}{2} = \frac{7}{2} \quad \checkmark$$

Betweenness uniform graphs

BUGs

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

Intuition: All vertices are indistinguishable based on properties alone.

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

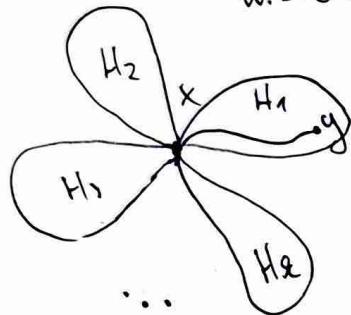
💡 Vertex transitive graphs are BUGs.

❗ \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 $k \geq 1$ components H_1, \dots, H_k with sizes h_1, \dots, h_k .
W.L.O.G let h_1 be minimal: $\forall i: h_1 \leq h_i$ (*)



💡 $C_B(x) \geq \sum_{\substack{i,j=1 \\ i < j}}^k h_i h_j$ $\rightarrow C_B(y) = ?$

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

\rightarrow pairs of m, n contributing to $C_B(y)$

- $m \in H_i, n \in H_j, i+1=j$: none
- $m \in H_i, n \in H_1, i \neq 1$: none
- $m, n \in (H_1 + x)$: some

$$\Rightarrow C_B(y) \leq \frac{h_1(h_1-1)}{2} < \frac{h_1^2}{2} \stackrel{(*)}{<} \sum_{\substack{i,j=1 \\ i < j}}^k h_i h_j \leq C_B(x) \Rightarrow G \text{ is not a BUG } \square$$

Theorem: Let $G = (V, E)$ be a BUG on n vertices with $n - \Delta(G) = k$. Then $d(G) \leq k$.

The equality holds for $k=1, 2$.

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

\rightarrow we will show:

$\bullet k=1 \dots$ equivalent to $G \cong K_n$

$\bullet k=2 \dots$ show something about BUGs with $\text{diam} = 2$

\rightarrow couldn't equality hold for $k \geq 3$?

\rightarrow for $k=3$ consider for $n \geq 5$ the complement of cycle C_n



💡 $G = \overline{C_5}$ is vertex transitive \Rightarrow it is a BUG

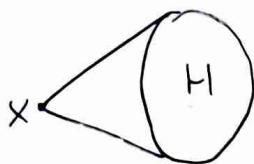
$$\Delta(G) = n-3 = n-k$$

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

\uparrow diameter

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

Pf: Let $x \in V$ have $\deg(x) = \Delta = m-1$ 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 $u \rightarrow v$ through x

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

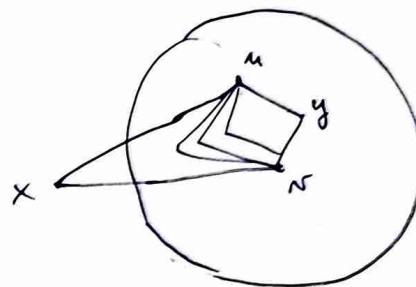
$\{u, v\} \in T \Rightarrow \tilde{\sigma}_{uv}(x) = 1$... only one SP of length 2 visits x

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

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

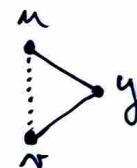
$$C_B(x) = \sum_{uv \in T} \frac{\tilde{\sigma}_{uv}(x)}{\tilde{\sigma}_{uv}} + \sum_{uv \in S} \frac{\tilde{\sigma}_{uv}(x)}{\tilde{\sigma}_{uv}} = \sum_{uv \in T} \frac{1}{1 + \tilde{\sigma}'_{u,v}} + \sum_{uv \in S} 1$$

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



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

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



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

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

$\circlearrowleft \tilde{\sigma}'_{u,v} = m-3 \dots$ for $\forall y \in H, y \notin \{u, v\}$: $uy, yv \in E$ using \star

$$C_B(x) = \sum_{uv \in T} \frac{1}{1 + \tilde{\sigma}'_{u,v}} = \sum_{uv \in T} \frac{1}{m-2} = \frac{1}{m-2} \cdot |E(H)| \quad \# \text{non-edges}$$

$$C_B(y) = \sum_{uv \in T} \frac{1}{m-2} < \sum_{uv \in T} \frac{1}{m-2} = C_B(x) \quad \nmid G \text{ is not a BUG}$$

\oplus Therefore $\forall y \in H: \deg(y) = m-1 \Rightarrow G \cong K_m$

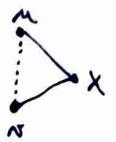
because $\deg(y) < m-1, \exists w \in V(H) \text{ s.t. } yw \notin E$

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

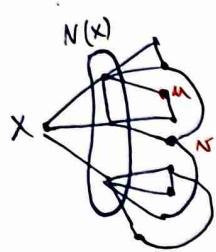
! but yw does not satisfy \star



Lemma: Let $G = (V, E)$ have diameter 2. Then $\forall x \in V$: $C_B(x) = \sum_{\substack{m, n \in N(x) \\ m \neq n}} \frac{1}{G_{mn}}$



Pf: $m, n \in E \Rightarrow G_{mn}(x) = 0$



$$C_B(x) = \sum_{\substack{m \neq x \neq n \\ mn \notin E}} \frac{G_{mn}(x)}{G_{mn}} = \sum_{\substack{m \neq x \neq n \\ mn \notin E}} \frac{G_{mn}(x)}{G_{mn}} = \sum_{\substack{m, n \in N(x) \\ mn \in E}} \frac{G_{mn}(x)}{G_{mn}} + \sum_{\substack{m, n \in N(x) \\ mn \notin E}} \frac{G_{mn}(x)}{G_{mn}} + \sum_{\substack{m \in N(x) \\ n \notin N(x)}} \frac{G_{mn}(x)}{G_{mn}}$$

② $P_{m,n}(x)$ has length 4 but $d(m, n) \leq \text{diam} = 2 \Rightarrow G_{mn}(x) = 0$

③ $P_{m,n}(x)$ has length 3 but $d(m, n) \leq \text{diam} = 2 \Rightarrow G_{mn}(x) = 0$

① $mn \notin E \Rightarrow d(m, n) = 2 \Rightarrow G_{mn}(x) = 1$



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

$$\begin{aligned} m &= |V| \\ m &= |E| \end{aligned}$$

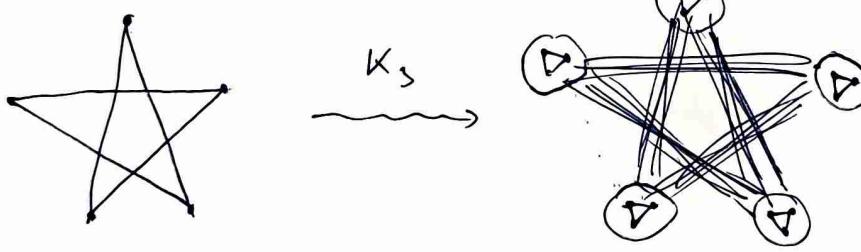
Pf Use previous lemma to calculate $\sum C_B(x)$ and divide by $m = |V|$.

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

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



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



$$n \mapsto \{N_1, \dots, N_n\}$$

$$nm \mapsto \{N_i M_j \mid i, j = 1, \dots, m\}$$

← blow-up graph

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

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