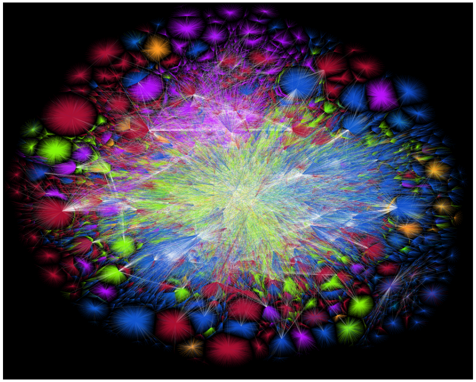


Bringing order to centrality measures

Remco van der Hofstad

CNI Seminar Series, IISc Bengaluru, April 20, 2026

Joint work with:

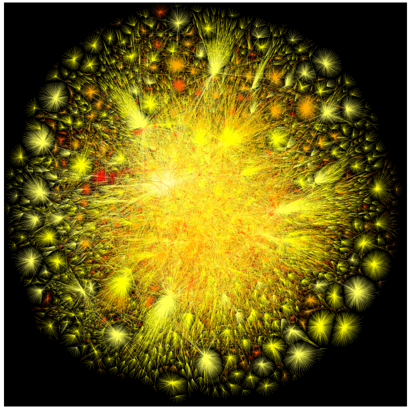
▷ Alessandro Garavaglia (Gucci)
▷ George Exarchakos (TU/e)
▷ Florian Henning (TU/e)
▷ Nelly Litvak (TU/e)
▷ Manish Pandey (Aarhus)
▷ Oliver Nagy (Leiden)

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY
EURANDOM
NETWORKS

Part 1:

Real-world networks and network science

Complex networks



Burst of activity in past 25 years.
See books Newman (2010) or Barabási (web book) for examples and theory.

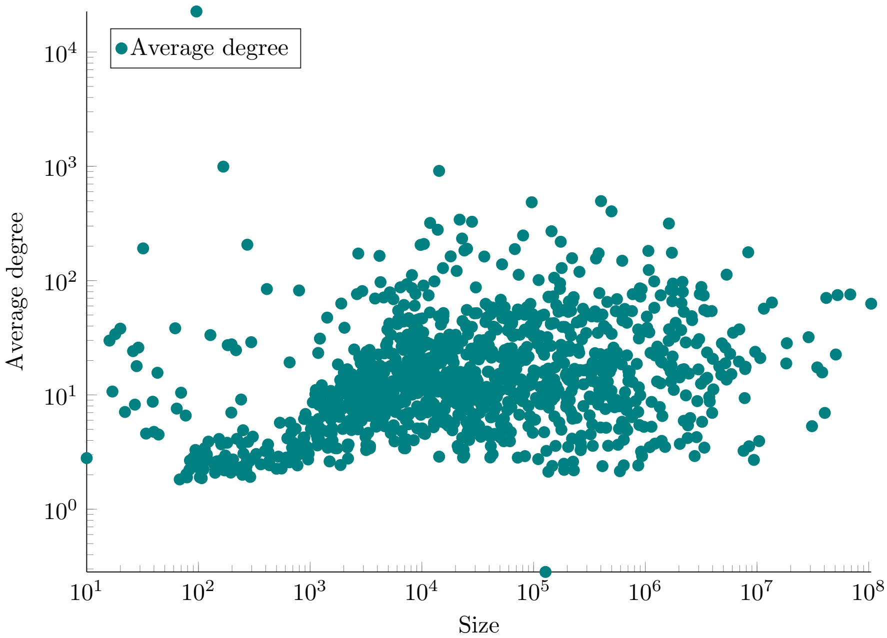
Opte project, Barrett Lyon, 2010]

Networks come in different flavours:

- ▷ Social networks: Acquaintances, sexual relations,...
- ▷ Information networks: Collaboration graphs, WWW,...
- ▷ Technological networks: Internet, power/telephone grids,...
- ▷ Biological networks: Food webs, neurons, protein interactions,...

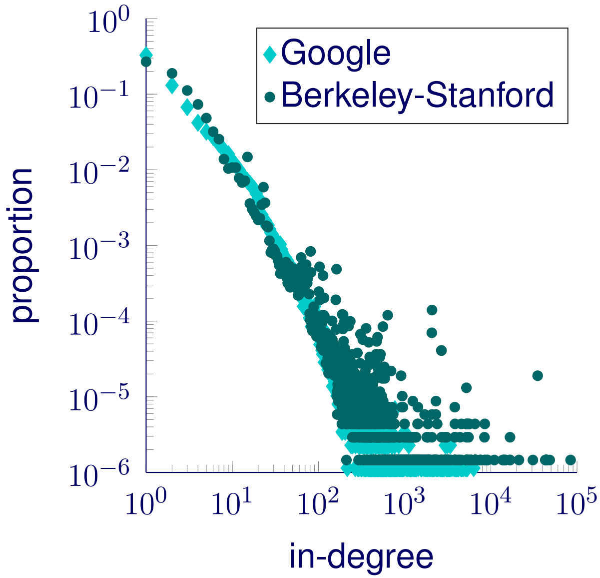
Attention focussing on unexpected commonality.

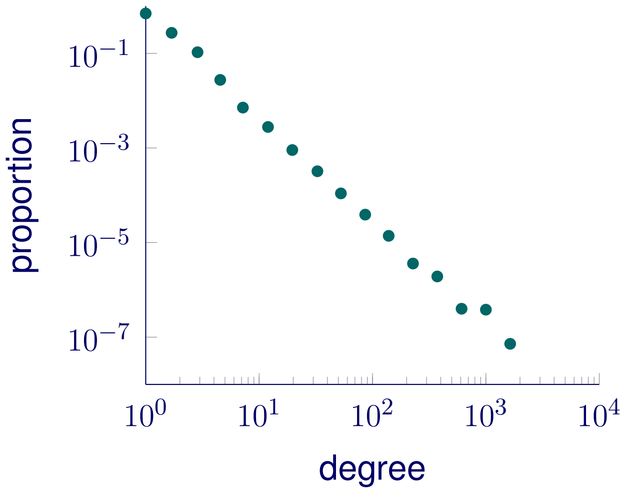
Networks are sparse



Average degrees of 1,203 networks in KONECT

Scale-free paradigm





Loglog plot (in-)degree sequences WWW and Internet (courtesy Krioukov)

▷ Straight line: proportion p_k vertices of degree k satisfies $p_k = ck^{-\tau}$.

▷ Empirical data finds $\tau \in (2, 3)$: highly-variable number of neighbours

Network science

▷ Complex networks modelled using

random graphs.

▷ Network functionality modelled by stochastic processes on, or algorithms for, them.

▷ A plethora of examples:

Disease spread	Synchronization
Information diffusion	Robustness to failures
Consensus reaching	Information retrieval

▷ Network algorithms: PageRank, community detection,...

▷ Prominent part of applied math for decades to come.

Model jungle

Plethora of sparse random graph models have been invented:

⋆ Static models:
Erdős-Rényi random graph, inhomogeneous random graph (IRG), configuration model (CM), exponential random graphs,...

⋆ Dynamic models:
Growing models such as preferential attachment model (PAM), copying models, as well as dynamic versions of above models of fixed size...

Need techniques to deal with many models at once:
graph limits.

Part 2:

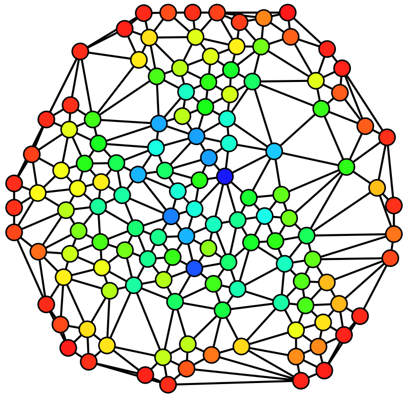
Centrality measures:
Who is important in network?

Centrality measures

▷ Closeness centrality:
Measures to what extent vertex can reach others using few hops.
Vertices with low closeness centrality are central in network.

▷ Betweenness centrality:
Measures extent to which vertex connects various parts of network.

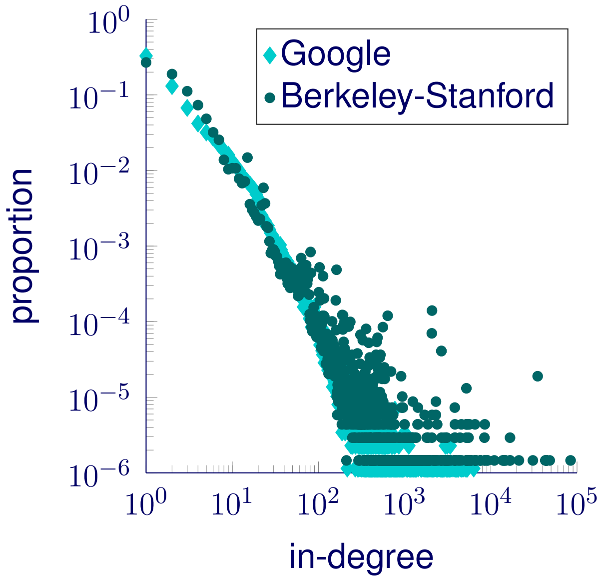
Betweenness large for bottlenecks.

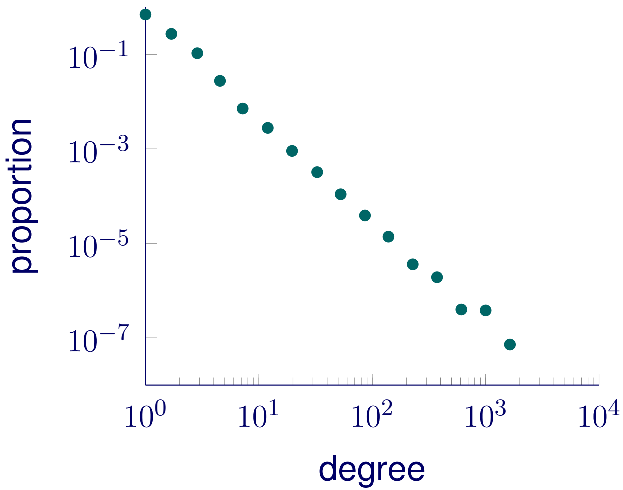


▷ PageRank:
Measures extent to which vertex is visited by random walk.
Used in Google to rank importance in web pages.

Degree centrality

⋆ **Simplest centrality measure, and basis of graph connectivity.**





Loglog plot (in-)degree sequences WWW and Internet (courtesy Krioukov)

▷ Straight line: proportion p_k vertices of degree k satisfies $p_k = ck^{-\tau}$.

PageRank

$\star$ Fix damping factor c (or restart probability $1-c$,) and let P denote random walk transition probability on graph G, i.e.,

$$P_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i) = \frac{1}{d_i^{(\mathrm{out})}} \mathbb{1}_{\{(i,j) \in E(G)\}}.$$

Then, PageRank $\pi^{(G)}$ satisfies

$$\pi^{(G)} = c\pi^{(G)} P + \frac{1-c}{n}\mathbf{1}.$$

$\star$ Often more convenient to deal with graph-normalized PageRank, which is just $R^{(G)} = n\pi^{(G)}$, and which satisfies

$$R^{(G)}(v) = c \sum_{u \to v} \frac{1}{d_v^{(\mathrm{out})}} R^{(G)}(u) + 1 - c.$$

$\star$ Then, denoting V_n vertex chosen uniformly at random from $[n]$,

$$\mathbb{E}[R^{(G)}(V_n)] = 1.$$

Closeness centrality

⋆ Closeness centrality function is

$$c(v) = \frac{|V(G)| - 1}{\sum_{u \in V(G)} d_G(u, v)},$$

where $d_G(u, v)$ is length of shortest path between u and v in G.

$\boxed{\text{Closely related to Kevin Bacon game, or Erdős number.}}$

⋆ Related is harmonic centrality, given by

$$h(v) = \sum_{u \neq v} \frac{1}{d_G(u, v)}.$$

Betweenness centrality

⋆ Betweenness centrality function is

$$b(v) = \sum_{i \in V(G)} \sum_{j \in V(G)} \frac{\sigma_{i,j}(v)}{\sigma_{i,j}},$$

where $\sigma_{i,j}(v)$ counts number of shortest paths from i to j containing v and $\sigma_{i,j}$ is number of shortest paths from i to j.

Computationally heavy for large graphs!

⋆ Related is load centrality, given by

$$l(v) = \frac{\sum_{i \in V(G)} \sum_{j \in V(G)} \sigma_{i,j}(v)}{\sum_{i \in V(G)} \sum_{j \in V(G)} \sigma_{i,j}},$$

Related centralities

⋆ Katz centrality function is

$$K_\alpha(i) = \sum_{k\geq 1} \sum_{j\in V(G)} \alpha^k A_{i,j}^k,$$

where A is adjacency matrix of G, $\alpha \in (0,1)$ is attenuation factor, which must be smaller than $1/\lambda$ with λ largest eigenvalue of A.

[Also called Bonacich centrality, particularly in economics.]

⋆ Eigenvector centrality satisfies

$$x^t A = \lambda x^t,$$

where again A is adjacency matrix of G and λ its largest eigenvalue, and x its (left-) eigenvector normalised to sum to 1.

[Problematic when network is not (strongly) connected.]

Properties centralities

⋆ Topic of centrality measures is very broad, and I could speak hours about this alone.

⋆ From now on, let $G = (V(G), E(G))$ be connected undirected graph, or strongly connected directed graph.

First focus on PageRank as prominent network measure.
Focus is to derive its analytical properties.

Final topic is how to compare centrality measures,
which is topic of recent research.

Part 3:

Large graph limits of centrality measures

Recall: PageRank

$\star$ Fix damping factor c (or restart probability $1-c$,) and let P denote random walk transition probability on graph G, i.e.,

$$P_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i) = \frac{1}{d_i^{(\mathrm{out})}} \mathbb{1}_{\{(i,j)\in E(G)\}}.$$

Then, PageRank $\pi^{(G)}$ satisfies

$$\pi^{(G)} = c\pi^{(G)} P + \frac{1-c}{n}\mathbf{1}.$$

$\star$ Often more convenient to deal with graph-normalized PageRank, which is just $R^{(G)} = n\pi^{(G)}$, and which satisfies

$$R^{(G)}(v) = c\sum_{u\to v} \frac{1}{d_v^{(\mathrm{out})}} R^{(G)}(u) + 1 - c.$$

$\star$ Then, denoting V_n vertex chosen uniformly at random from $[n]$,

$$\mathbb{E}[R^{(G)}(V_n)] = 1.$$

PageRank Power Law

PageRank is large for vertices that have
many important vertices pointing to them,
not just those that have many vertices pointing to them.

⋆ Thus, seems highly appropriate way to

classify importance vertices in networks.

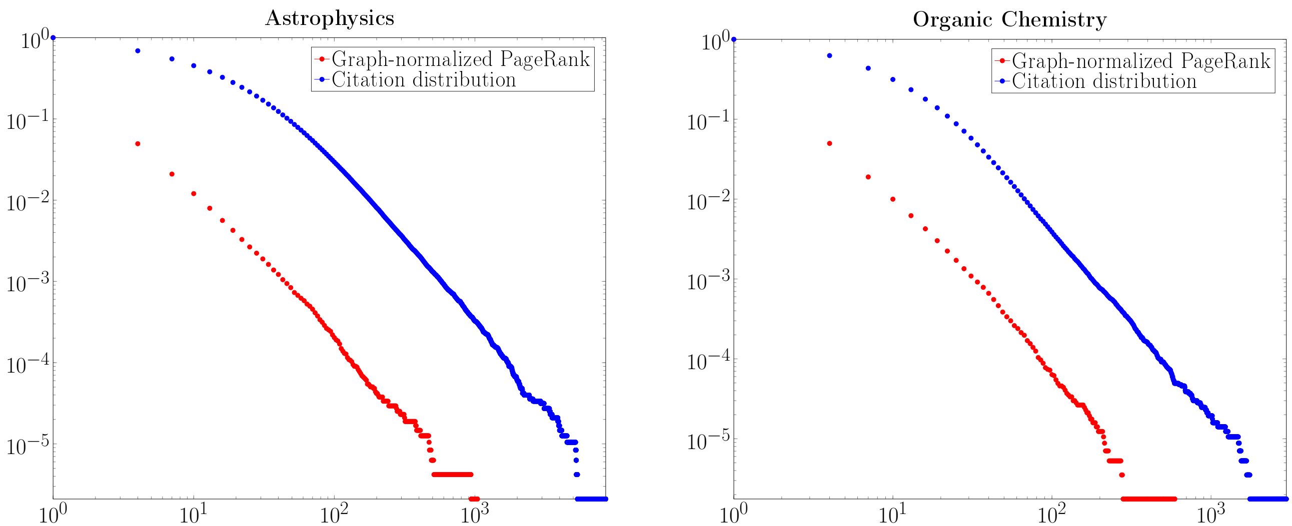
⋆ Further, empirical data suggests that

$$\frac{1}{n}\sum_{v\in[n]} \mathbb{1}_{\{R^{(G)}(v)>x\}} = \mathbb{P}(R^{(G)}(V_n) > x) \approx x^{-(\tau-1)} :$$

Power-law tails for PageRank distribution.

Reason excellent performance PageRank?

Power-law hypothesis



PageRank and In-degree in citation networks from
Web of Science in Astrophysics and Organic chemistry

Same power-law exponents!

Power-law hypothesis

Theorem 1. [vdH-L-H 25, B-OC23,...]

$\star$ Power-law hypothesis is true for

(a) Directed configuration model;

(b) Directed generalized random graph;

(c) Most undirected random graphs (e.g., ERRG, CM, GRG, PAM.)

$\star$ Power-law hypothesis is not true for directed preferential attachment model, with edges directed from young to old.

In latter case, tails PageRank are heavier than in-degree.

$\star$ Result for undirected graphs follows from bound, for all $v \in V(G)$,

$$R_v^{(G)} \leq d_v.$$

PageRank is local

Theorem 2. [GarHofLit18]

Consider sequence of directed random graphs $(G_n)_{n\in\mathbb{N}}$.

(a) If G_n converges locally weakly, then there exists limiting distribution $R_o^{(G)}$, with $\mathbb{E}[R_o^{(G)}] \leq 1$, such that

$$R_{o_n}^{(G_n)} \xrightarrow{d} R_o^{(G)};$$

(b) If G_n converges locally in probability, then there exists limiting distribution $R_o^{(G)}$, with $\mathbb{E}[R_o^{(G)}] \leq 1$, such that, for every continuity point $r > 0$ of $r \mapsto \mu\left(R_o^{(G)} > r\right)$,

$$\frac{1}{n}\sum_{v\in[n]} \mathbb{1}_{\{R_v^{(G_n)}>r\}} \xrightarrow{\mathbb{P}} \mu\left(R_o^{(G)} > r\right).$$

PageRank determined by large-graph limit!

But what IS local convergence??

Local convergence

Has been proved for many sparse random graph models.

Local convergence implies

- ▷ one-sided law of large numbers $|\mathscr{C}_{\max}|/n$;
- ▷ convergence proportion neighborhoods of specific shape;
- ▷ convergence various other functionals:

Examples include PageRank distribution, and under more restrictions, log partition function Ising model, while through a lot more work, densest subgraph.

Local convergence gives good starting point analysis
for many more graph properties.

Preliminaries

⋆ Graph: $G = (V(G), E(G))$ with $V(G)$ vertex set, $E(G)$ edge set.

⋆ Rooted graph: (G, v) with G graph and $v \in V(G)$ vertex or root.

⋆ Neighbourhood: r-neighbourhood $B_r^{(G)}(v)$ of $v \in V(G)$ is rooted subgraph induced by all vertices at distance at most r from root v.

⋆ Isomorphisms: Two rooted graphs $(G_1, v_1), (G_2, v_2)$ are isomorphic when there is bijection $\phi \colon V(G_1) \to V(G_2)$ mapping edges to edges and root to root. Denoted $(G_1, v_1) \simeq (G_2, v_2)$.

⋆ Metric: Distance of rooted connected graphs $(G_1, v_1), (G_2, v_2)$ is

$$\mathrm{d}_{\mathscr{G}_\star}\big((G_1, v_1), (G_2, v_2)\big) = \frac{1}{1 + R^\star},$$

where $R^\star$ is largest value for which $B_r^{(G_1)}(v_1) \simeq B_r^{(G_2)}(v_2)$, and $\mathscr{G}_\star$ is space of rooted graphs modulo isomorphisms.

[Metric space well defined: $\mathscr{G}_\star$ Polish space under this metric [RGCNII, Chapter 2].]

Local convergence

$\star$ Random graph sequence is $(G_n)_{n\geq 1}$ satisfying $|V(G_n)| \to \infty$.

[Will often take $V(G_n) = [n] \equiv \{1, \ldots, n\}$.]

$\star$ Local weak convergence holds when, with o_n chosen uniformly from $V(G_n)$,

$$\mathbb{E}[h(G_n, o_n)] = \frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} \mathbb{E}[h(G_n, v)] \to \mathbb{E}_{\bar{\mu}}\big[h(G, o)\big],$$

for any bounded and continuous functions $h \colon \mathscr{G}_\star \to \mathbb{R}$ and $\bar{\mu}$ some probability measure on $\mathscr{G}_\star$. [Benjamini-Schramm (2001), Aldous-Steele (2004).]

$\star$ Local convergence in probability holds when, instead,

$$\mathbb{E}[h(G_n, o_n) \mid G_n] = \frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} h(G_n, v) \xrightarrow{\mathbb{P}} \mathbb{E}_{\mu}\big[h(G, o)\big],$$

for any bounded and continuous function $h \colon \mathscr{G}_\star \to \mathbb{R}$ and μ some probability measure on $\mathscr{G}_\star$.

Proof PageRank

$$
\begin{tikzcd}
R_{o_n}(n) \arrow[r, dashed, "\text{(A) } n \to \infty"'] \arrow[d, "\text{(a) finite } N"'] & R_{\varnothing} \\
R^{(N)}_{o_n}(n) \arrow[r, "\text{(b) } n \to \infty"] & R^{(N)}_{\varnothing} \arrow[u, "\text{(c) } N \to \infty"']
\end{tikzcd}
$$

Part 4:

Bringing order to centralites: Centrality comparison curve

Comparing centralities

⋆ As mentioned before, many different centrality measures exist.

⋆ Some of these have parameters, of which it is unclear how they should be chosen, or whether choice matters at all.

Need principled way to compare centralities.

⋆ Correlation is sometimes used, but that is sensitive to monotone, non-linear transformations. Since often only ranks produced by centrality measures matter, we need a more robust way:

Centrality Comparison Curve (CCC)

CCC

Definition 1 (CCC). Fix graph $G = (V(G), E(G))$ with $|V(G)| = n$.

$\star$ Define induced total ordering $\prec_{(R,S)}$ by $a \prec_{(R,S)} b$ for $a, b \in V(G)$ iff (a) $R(a) < R(b)$; or (b) $R(a) = R(b)$ and $S(a) < S(b)$; or (c) $R(a) = R(b)$ and $S(a) = S(b)$ and $u_a < u_b$.

$\star$ Let $\mathsf{Top}^G_{\prec}(k)$ denote set of first k elements in set $(V(G), \prec)$.

$\star$ For two centrality measures R, S, let $\mathsf{CCC}^G_{R,S} \colon (0,1] \to [n]/n$ be centrality comparison curve (CCC) defined by

$$\mathsf{CCC}^G_{R,S}(x) = \frac{|\mathsf{Top}^G_{\prec_{(R,S)}}(\lceil xn \rceil) \cap \mathsf{Top}^G_{\prec_{(S,R)}}(\lceil xn \rceil)|}{n}.$$

$\mathsf{CCC}^G_{R,S}(k/n)$ represents overlap of k most central vertices according to R and S.

Properties CCC

Centrality Comparison Curve has many desirable properties, e.g., it is insensitive to any monotone transformation.

How to read CCC in practice?

$\star \mathrm{CCC}^G_{R,S}(x) \leq x$, and $\mathrm{CCC}^G_{R,S}(x) \approx x$ for similar centralities;

$\star \mathrm{CCC}^G_{R,S}(x) \approx x$ for small x for similar highest ranked vertices;

$\star \mathrm{CCC}^G_{R,S}(x) \approx x^2$ when two centralities are almost independent;

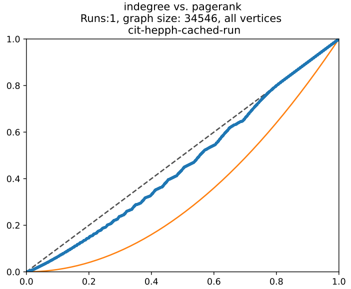
$\star \mathrm{CCC}^G_{R,S}(x) \approx \min\{0, 2x-1\}$ for maximally different centralities;

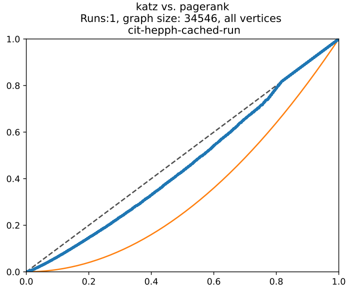
CCC in practice

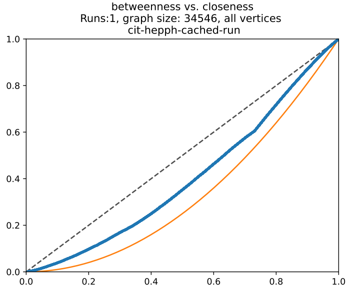
Have many CCC plots, both for real-world networks, as well as for (un)directed CMs and certain graphon models.

Will show some, and then conclusions.

CCC “Hep” arXiv

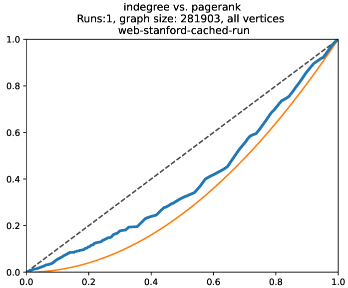


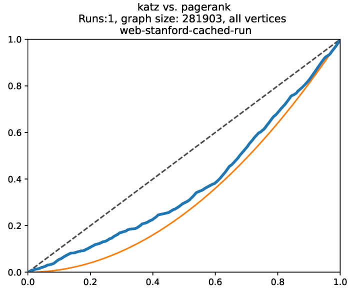


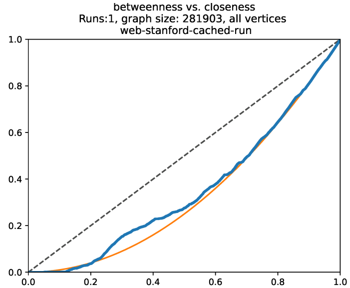


In-degree vs. Pagerank, Katz vs. PageRank, and Betweenness vs. Closeness for Hep arXiv

CCC Stanford Web graph



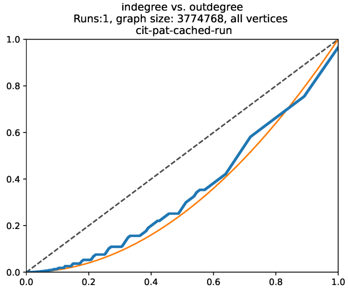


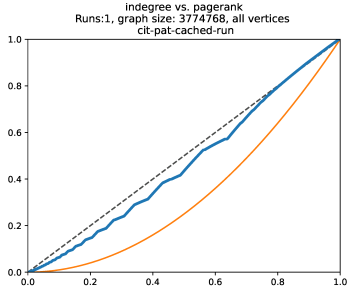


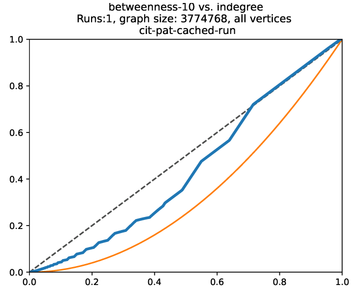
In-degree vs. Pagerank, Katz vs. PageRank, and

Betweenness vs. Closeness for Stanford Web graph

CCC Patent graph

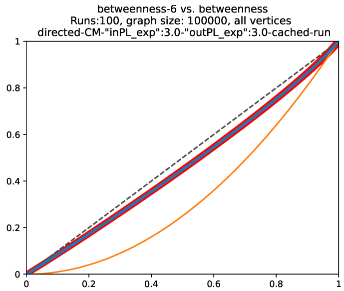


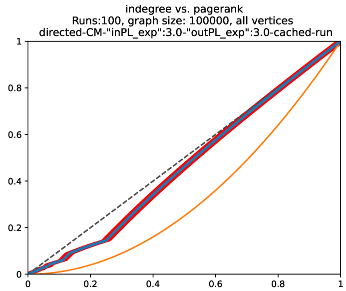


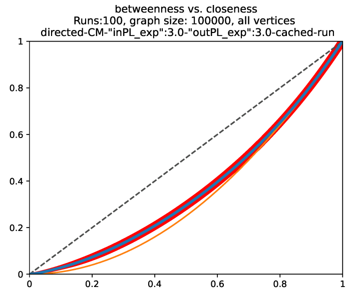


In-degree vs. out-degree, in-degree vs. PageRank, and betweenness10 vs. in-degree for Patent graph

CCC DCM







Betweenness6 vs. betweenness, in-degree vs. PageRank, and betweenness vs. closeness for Directed Configuration Model

Conclusions CCC

⋆ PageRank insensitive to restart probability=damping factor;

⋆ Closeness centrality virtually identical to harmonic centrality;

⋆ In-degree quite similar to PageRank;

⋆ Out-degree rather different from PageRank;

⋆ Betweenness, closeness well approximated by local versions.

CCC useful addition to network science toolbox:
It brings order to centrality measures.

Open problems CCC

Show that CCC behaves well under graph limits [with Manish Pandey.]

Do large-scale clustering analysis of centralities using CCC.

Conclusions

▷ Networks useful to interpret real-world phenomena:
centrality measures.

▷ Unexpected commonality networks:
scale free and small worlds.

▷ Random graph models explain properties networks.
Universality?

Example: Local limits often branching processes