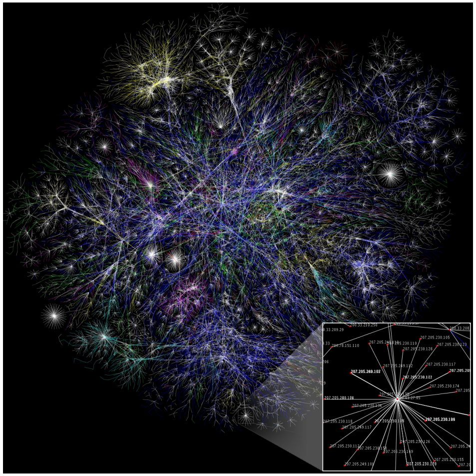


Information Diffusion on Random Graphs: Small Worlds, Percolation and Competition

Remco van der Hofstad

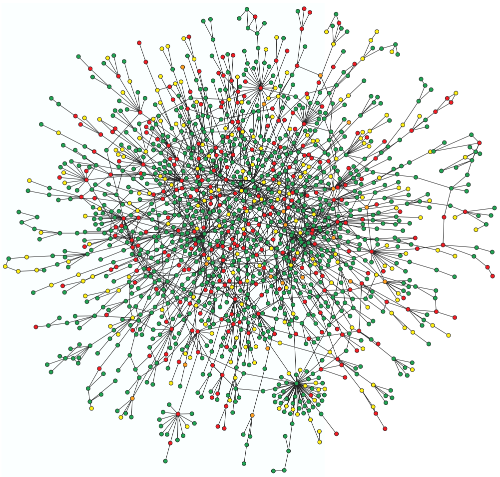
Simons Conference on Random Graph Processes,
May 9–12, 2016, UT Austin

TU/e Technische Universiteit Eindhoven University of Technology

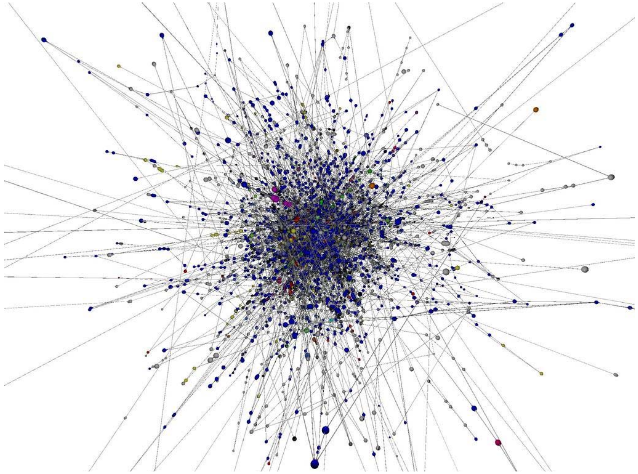
EURANDOM

NETWORKS

Complex networks



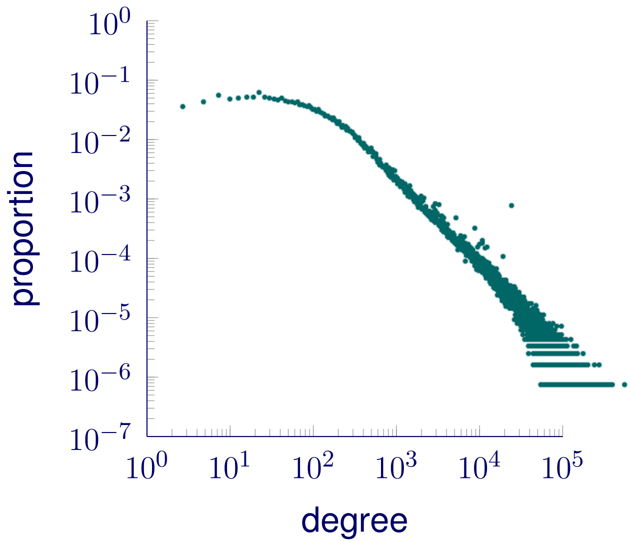
Yeast protein interaction network

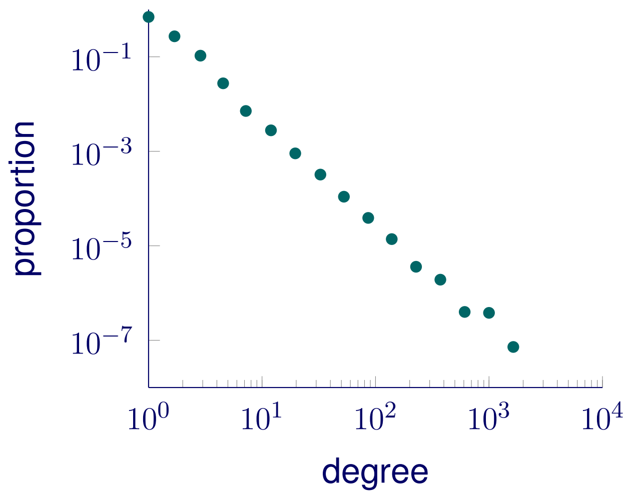


Internet topology in 2001

Attention focussing on unexpected commonality.

Scale-free paradigm



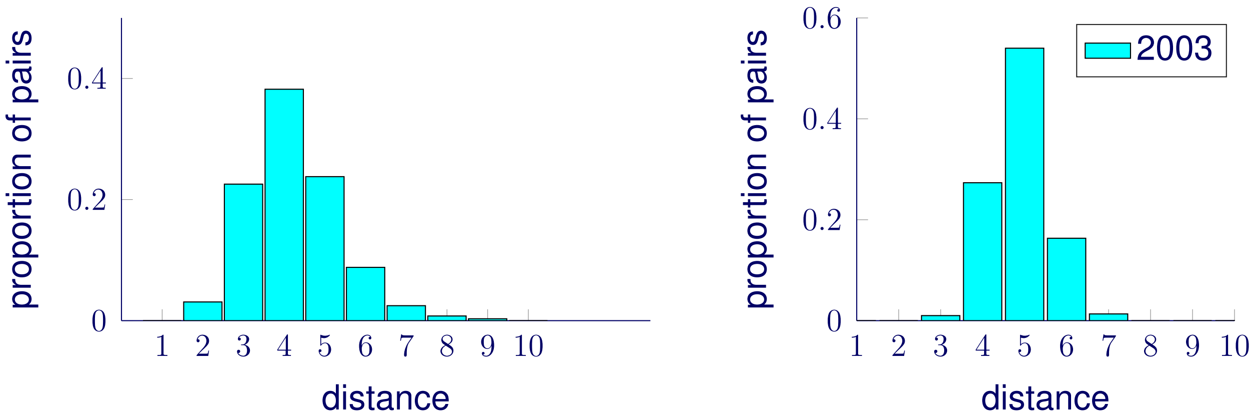


Loglog plot degree sequences Internet Movie Database and Internet

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

▷ Empirically: often $\tau \in (2, 3)$ found.

Small-world paradigm



Distances in Strongly Connected Component WWW and IMDb in 2003.

Facebook

Largest virtual friendship network:
721 million active users,
69 billion friendship links.

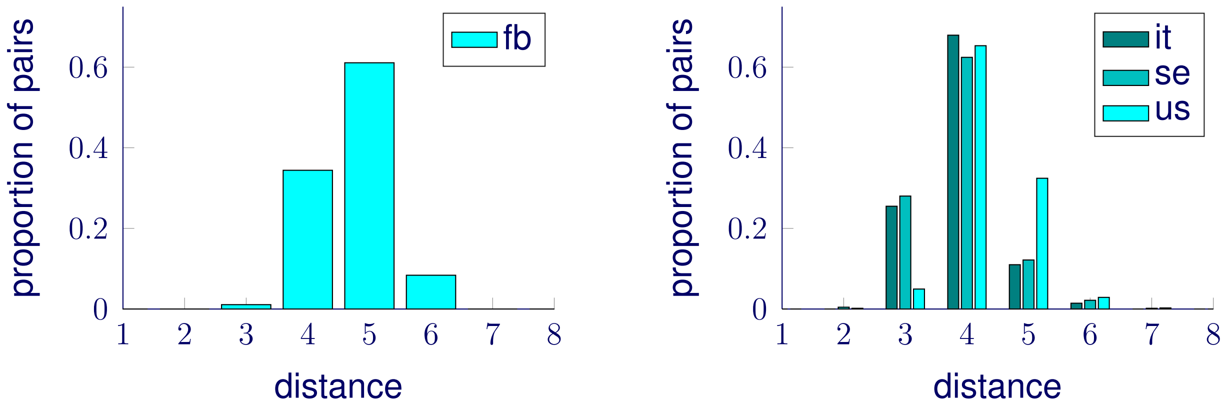
Typical distances on average four:

Four degrees of separation!

Fairly homogeneous (within countries, distances similar).

Recent studies:
Ugander, Karrer, Backstrom, Marlow (2011): topology
Backstrom, Boldi, Rosa, Ugander, Vigna (2011): graph distances.

Four degrees of separation



Distances in FaceBook in different subgraphs

Backstrom, Boldi, Rosa, Ugander, Vigna (2011)

Modeling networks

Use random graphs to model uncertainty in formation connections between elements.

▷ Static models:
Graph has fixed number of elements:

Configuration model and **Inhomogeneous random graphs**

▷ Dynamic models:
Graph has evolving number of elements:

Preferential attachment model

Universality??

Configuration model

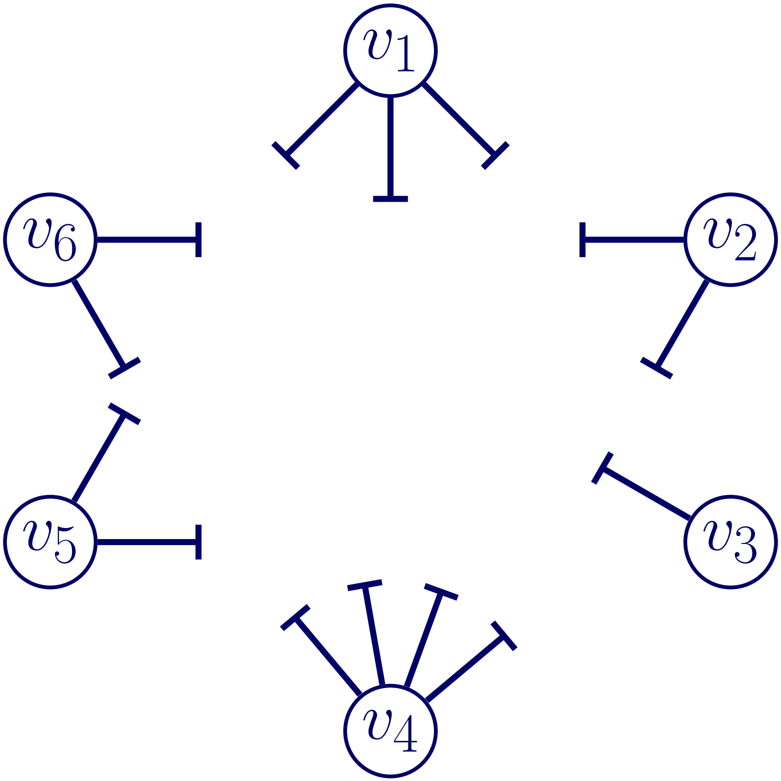
$\triangleright$ n number of vertices;
$\triangleright$ $\boldsymbol{d} = (d_1, d_2, \ldots, d_n)$ sequence of degrees.

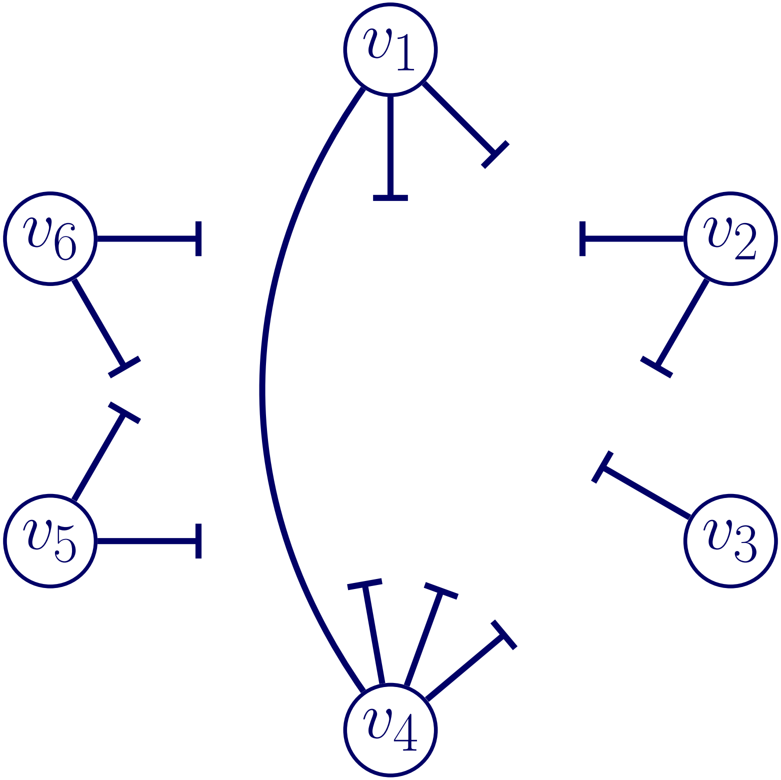
$\triangleright$ Assign d_j half-edges to vertex j. Assume total degree even, i.e.,

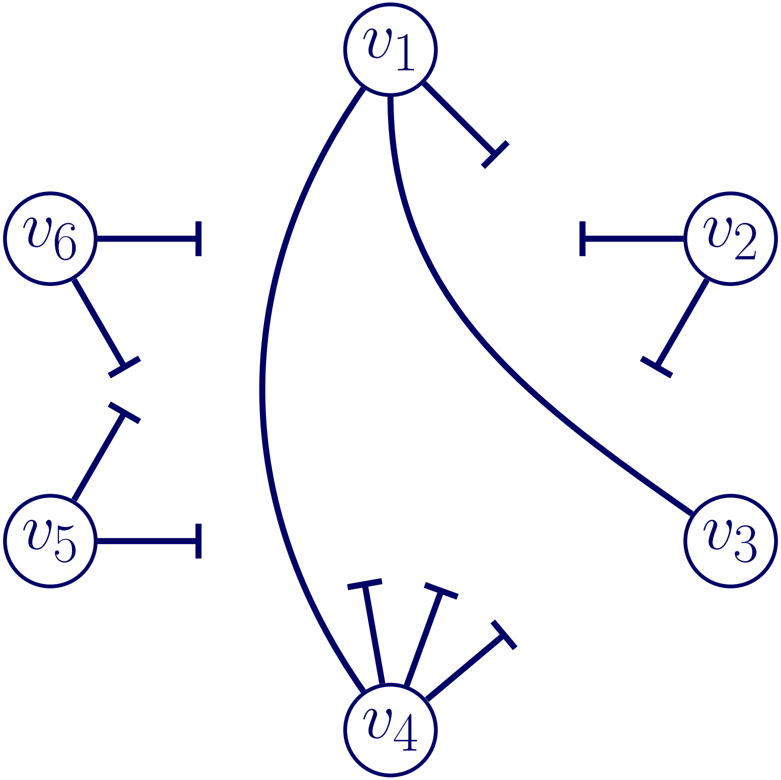
$$\ell_n = \sum_{i \in [n]} d_i \qquad \text{even.}$$

$\triangleright$ Pair half-edges to create edges as follows:
Number half-edges from 1 to ℓ_n in any order.
First pair first half-edge at random to one of other $\ell_n - 1$ half-edges.

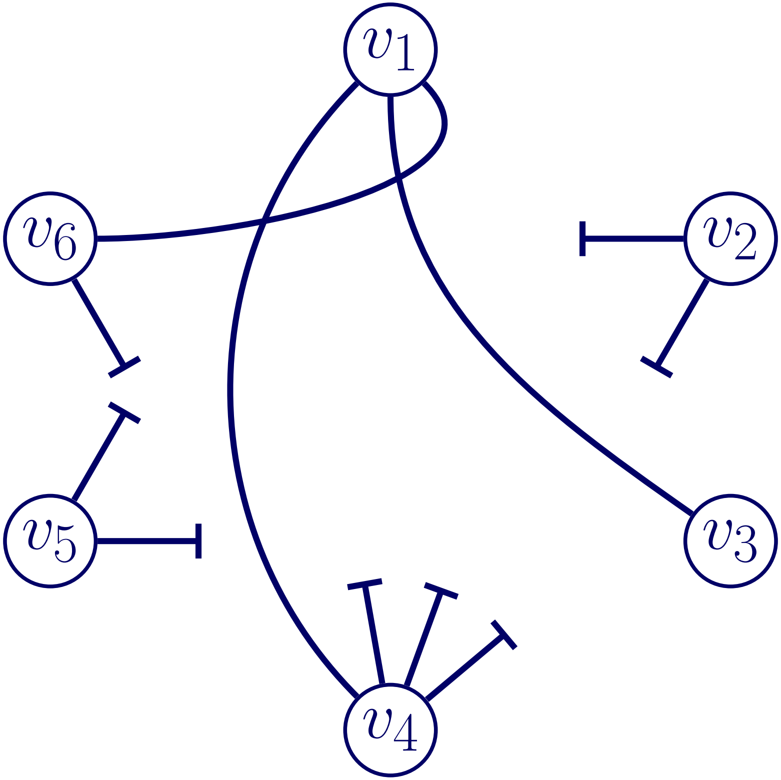
$\triangleright$ Continue with second half-edge (when not connected to first) and so on, until all half-edges are connected.


v_1
v_2
v_3
v_4
v_5
v_6

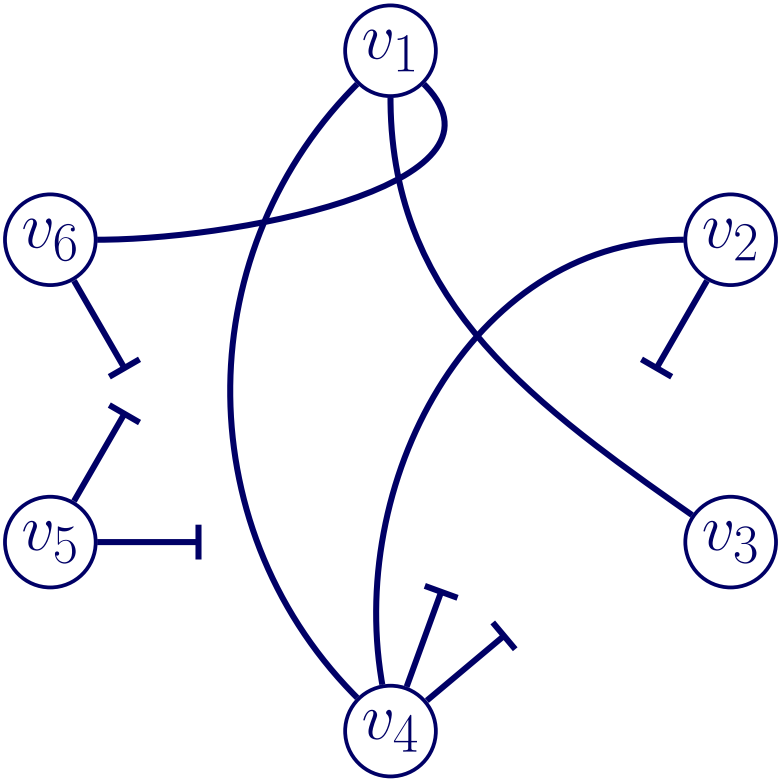

v_1
v_2
v_3
v_4
v_5
v_6

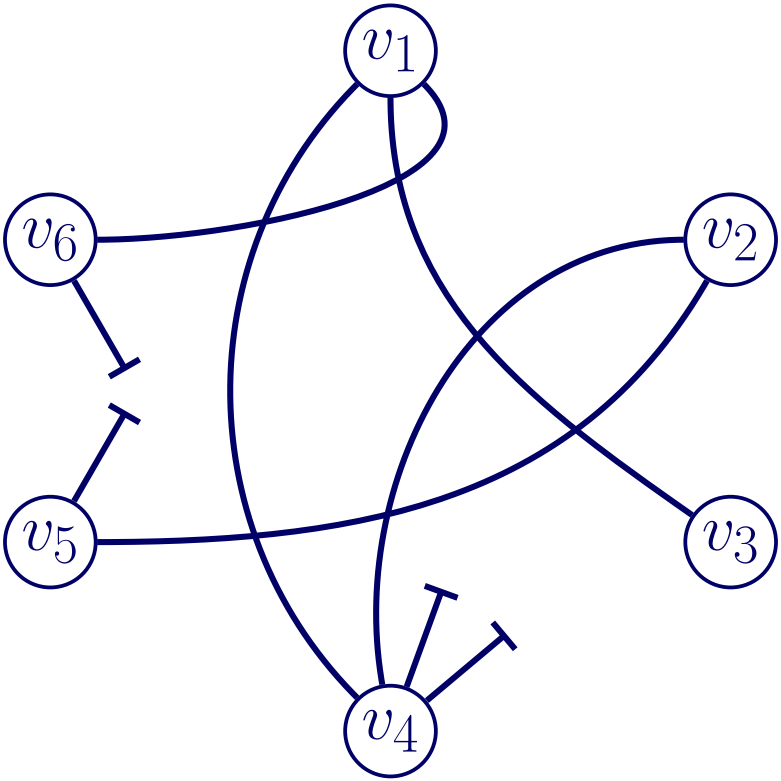


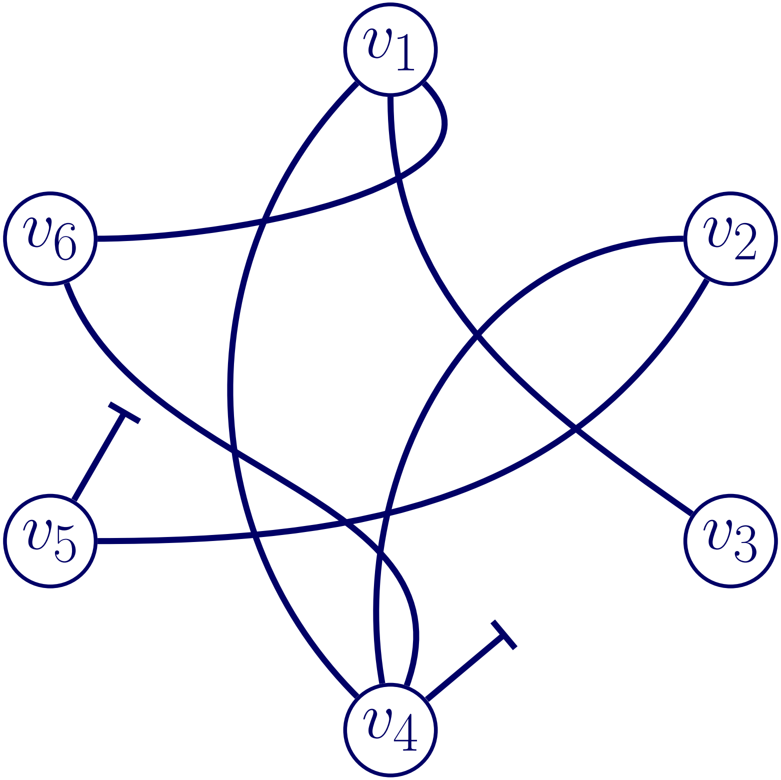
v_1
v_2
v_3
v_4
v_5
v_6



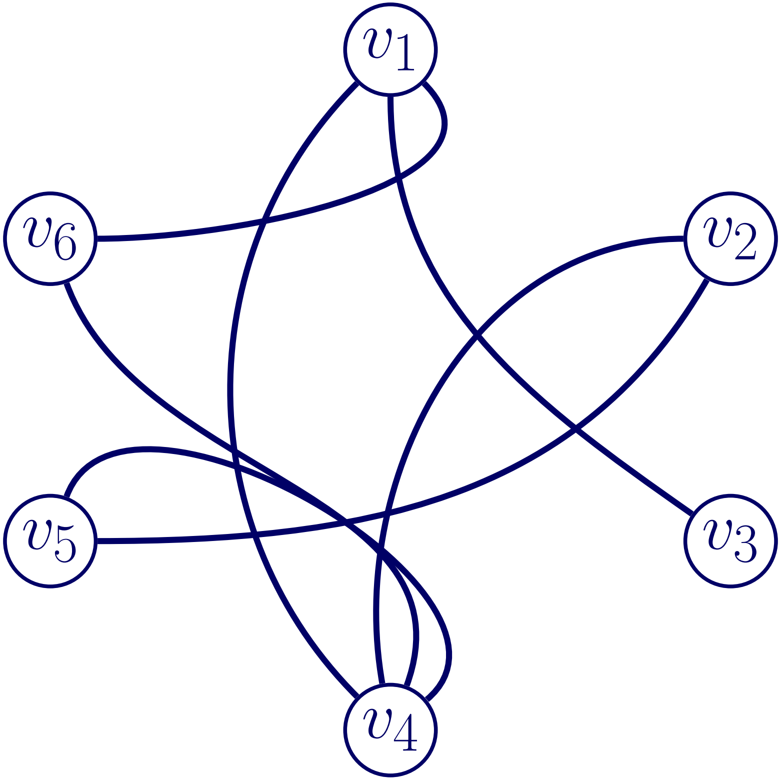
v_1
v_2
v_3
v_4
v_5
v_6


v_1
v_2
v_3
v_4
v_5
v_6

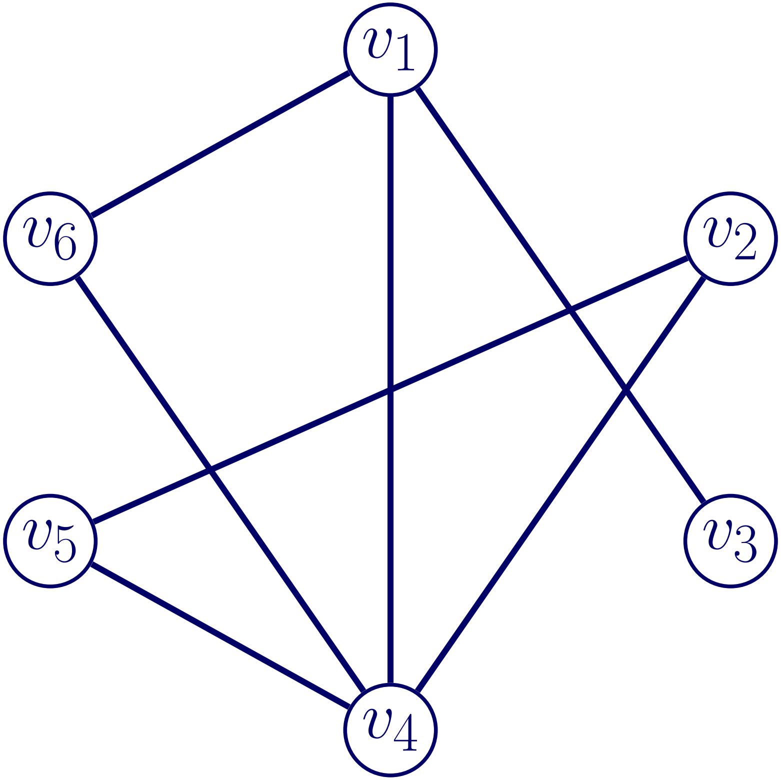

v_1
v_2
v_3
v_4
v_5
v_6



v_1
v_6
v_2
v_5
v_3
v_4



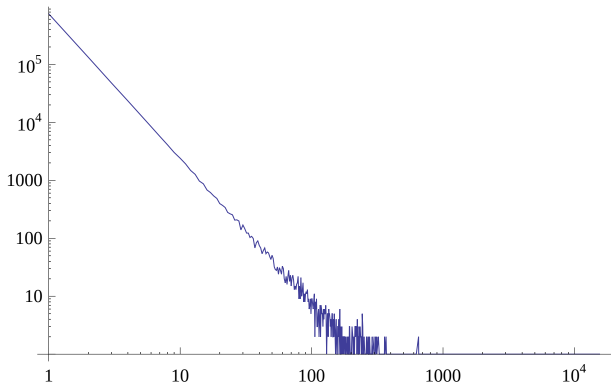
v_1
v_2
v_3
v_4
v_5
v_6

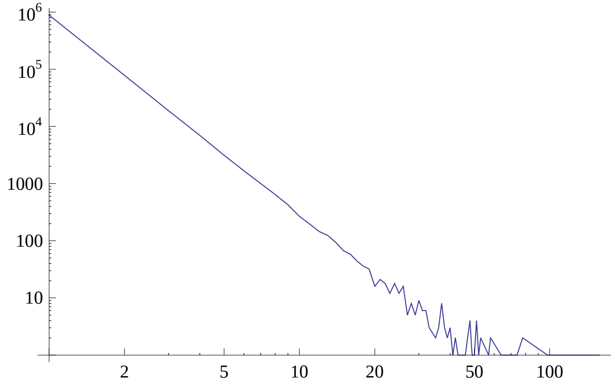

v_1
v_6
v_2
v_5
v_3
v_4

Power-laws CM

▷ Special attention to power-law degrees, i.e., for $\tau > 1$ and c_τ

$$\mathbb{P}(d_1 \geq k) = c_\tau k^{-\tau+1}(1 + o(1)).$$





Loglog plot of degree sequence CM with i.i.d. degrees $n = 1,000,000$ and $\tau = 2.5$ and $\tau = 3.5$, respectively.

Graph distances CM

H_n is graph distance between uniform pair of vertices in graph.

Theorem 1. [vdH-Hooghiemstra-Van Mieghem RSA05] When $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] \in (1,\infty)$, conditionally on $H_n < \infty$,

$$\frac{H_n}{\log_\nu n} \xrightarrow{\mathbb{P}} 1.$$

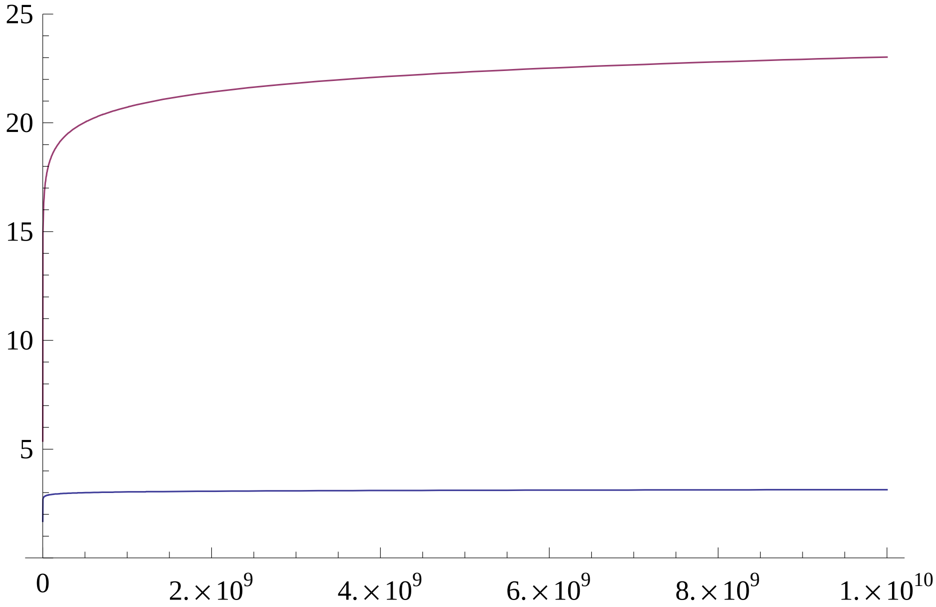
For i.i.d. degrees having power-law tails, fluctuations are bounded.

Theorem 2. [vdH-Hooghiemstra-Znamenski EJP07, Norros+Reittu 04] When degrees have power-law distribution with $\tau \in (2,3)$, conditionally on $H_n < \infty$,

$$\frac{H_n}{\log\log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau-2)|}.$$

For i.i.d. degrees having power-law tails, fluctuations are bounded.

$x \mapsto \log\log x$ grows extremely slowly



Plot of $x \mapsto \log x$ and $x \mapsto \log\log x$.

Preferential attachment

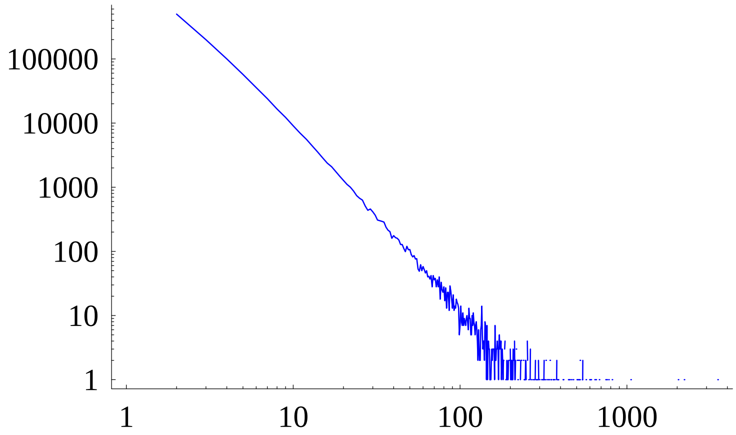
At time n, single vertex is added with m edges emanating from it. Probability that edge connects to ith vertex is proportional to

$$D_i(n-1)+\delta,$$

where $D_i(n)$ is degree vertex i at time n, $\delta > -m$ is parameter.

Yields power-law degree sequence with exponent $\tau = 3 + \delta/m > 2$.

Rich get richer!



$m = 2, \delta = 0, \tau = 3, n = 10^6$

Competition

$\triangleright$ Viral marketing aims to use social networks so as to excellerate adoption of novel products.

$\triangleright$ Observation: Often one product takes almost complete market. Not always product of best quality:

Why?

$\triangleright$ Setting:
– Model social network as random graph;
– Model dynamics as competing rumors spreading through network, where vertices, once occupied by certain type, try to occupy their neighbors at (possibly) random and i.i.d. time:

$\triangleright$ Fastest type might correspond to best product.

Competition and rumors

$\triangleright$ In absence competition, dynamics is rumor spread on graph.

$\triangleright$ Central role for spreading dynamics of such rumors=

first-passage percolation on graph with i.i.d. random weights.

$\triangleright$ Main object of study: $\mathcal{C}_n$ is weight of smallest-weight path two uniform connected vertices:

$$\mathcal{C}_n = \min_{\pi\colon U_1 \to U_2} \sum_{e\in\pi} Y_e,$$

where π is path in G, while $(Y_e)_{e\in E(G)}$ are i.i.d. collection of weights.

$\triangleright$ Focus here on exponential or deterministic weights.

Deterministic spreading

Theorem 3. [Baroni-vdH-Komjáthy (2014)] Fix $\tau \in (2,3)$. Consider competition model, where types compete for territory with deterministic traversal times. Without loss of generality, assume that traversal time type 1 is 1, and of type 2 is $\lambda \geq 1$.

Fastest types wins majority vertices, i.e., for $\lambda > 1$,

$$\frac{N_1(n)}{n} \xrightarrow{\mathbb{P}} 1.$$

Number of vertices for losing type 2 satisfies that there exists random variable Z s.t.

$$\frac{\log(N_2(n))}{(\log n)^{2/(\lambda+1)} C_n} \xrightarrow{d} Z.$$

$\triangleright$ Here, C_n is some random oscillatory sequence.

Deterministic spreading

Theorem 4. [vdH-Komjáthy (2014)] Fix $\tau \in (2,3)$.
Consider competition model, where types compete for territory with deterministic equal traversal times.

$\triangleright$ When starting locations of types are sufficiently different,

$$\frac{N_1(n)}{n} \xrightarrow{d} I \in \{0,1\},$$

and number of vertices for losing type satisfies that exists C_n s.t. whp

$$\frac{\log(N_{\mathrm{los}}(n))}{C_n \log n} \xrightarrow{d} Z.$$

$\triangleright$ When starting locations are sufficiently similar, **coexistence** occurs, i.e., there exist $0 < c_1, c_2 < 1$ s.t. whp

$$\frac{N_1(n)}{n}, \frac{N_2(n)}{n} \in (c_1, c_2).$$

Markovian spreading

Theorem 5. [Deijfen-vdH (2013)] Fix $\tau \in (2,3)$.
Consider competition model, where types compete for territory at fixed, but possibly unequal rates. Then, each of types wins majority vertices with positive probability:

$$\frac{N_1}{n} \xrightarrow{d} I \in \{0,1\}.$$

Number of vertices for losing type converges in distribution:

$$N_{\text{los}}(n) \xrightarrow{d} N_{\text{los}} \in \mathbb{N}.$$

The winner takes it all, the loser's standing small...

▷ Who wins is determined by location of starting point types:

Location, location, location!

Neighborhoods CM

$\triangleright$ Important ingredient in proof is description local neighborhood of uniform vertex $U_1 \in [n]$. Its degree has distribution $D_{U_1} \stackrel{d}{=} D$.

$\triangleright$ Take any of D_{U_1} neighbors a of U_1. Law of number of forward neighbors of a, i.e., $B_a = D_a - 1$, is approximately

$$\mathbb{P}(B_a = k) \approx \frac{(k+1)}{\sum_{i\in[n]} d_i} \sum_{i\in[n]} \mathbb{1}_{\{d_i=k+1\}} \xrightarrow{\mathbb{P}} \frac{(k+1)}{\mathbb{E}[D]} \mathbb{P}(D = k+1).$$

Equals size-biased version of D minus 1. Denote this by $D^\star - 1$.

Local tree-structure CM

$\triangleright$ Forward neighbors of neighbors of U_1 are close to i.i.d. Also forward neighbors of forward neighbors have asymptotically same distribution...

$\triangleright$ Conclusion: Neighborhood looks like branching process with offspring distribution $D^\star - 1$ (except for root, which has offspring D.)

$\triangleright$ $\tau \in (2,3):$ Infinite-mean BP, which has double exponential growth of generation sizes:

$$(\tau - 2)^k \log(Z_k \vee 1) \xrightarrow{a.s.} Y \in (0, \infty).$$

$\triangleright$ In absence of competition, it takes each of types about $\frac{\log \log n}{|\log(\tau-2)|}$ steps to reach vertex of maximal degree.

$\triangleright$ Type that reaches vertices of highest degrees (=hubs) first wins. When $\lambda > 1$, fastest type wins whp.

Proof Winner takes it all

Theorem 6. [Bhamidi-vdH-Hooghiemstra AoAP10]. Fix $\tau \in (2,3)$. Then,

$$\mathcal{C}_n \xrightarrow{d} \mathcal{C}_\infty,$$

for some limiting random variable $\mathcal{C}_\infty$:

Super efficient rumor spreading.

▷ $\mathcal{C}_\infty \stackrel{d}{=} V_1 + V_2$, where V_1, V_2 are i.i.d. explosion times of CTBP starting from vertices U_1, U_2. Then,

$$I = \mathbb{1}_{\{V_1 < \lambda V_2\}}.$$

Law of $N_{\rm los}$ much more involved, as competition changes dynamics after winning type has found hubs.

Conclusions

▷ Networks useful to interpret real-world phenomena: competition.

▷ Unexpected commonality networks: scale free and small worlds.

▷ Random graph models: Explain properties real-world networks:

Universality?

Example: Distances in preferential attachment model similar to those in configuration model with same degrees.
Poster Alessandro Garavaglia diameters in scale-free CM & PAM.

Poster Clara Stegehuis on more realistic model for real-world networks on mesoscopic scale.

▷ Book: Random Graphs and Complex Networks
`http://www.win.tue.nl/~rhofstad/NotesRGCN.html`