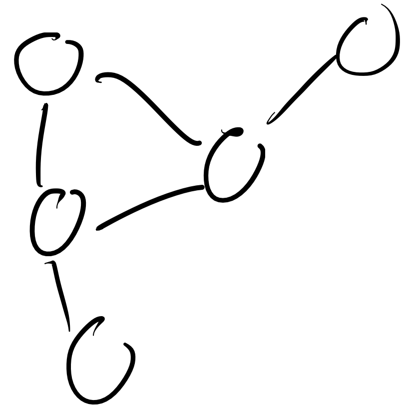
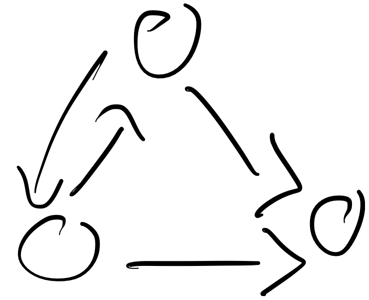
March 5, 2020

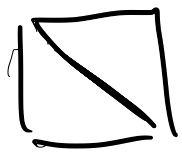
Last time: graphs





Fundamental matrices:

- A adjacency
- D



- $P = A^T D^{-1}$ random walk matrix $\quad D^{-1}A$
- $L = D - A$ Laplacian (undir)

social (FB)

information (Web)

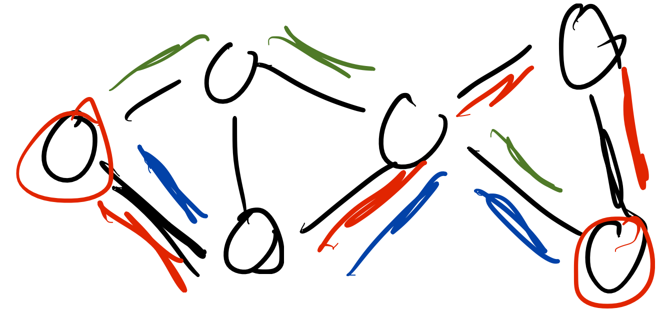
biological (PPI)

physical systems (Internet)

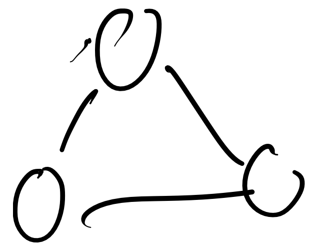
commerce (co-purchasing)

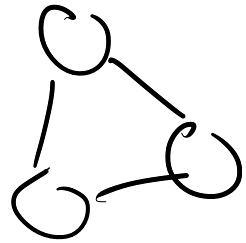
Common network properties

① Small "hop" distances

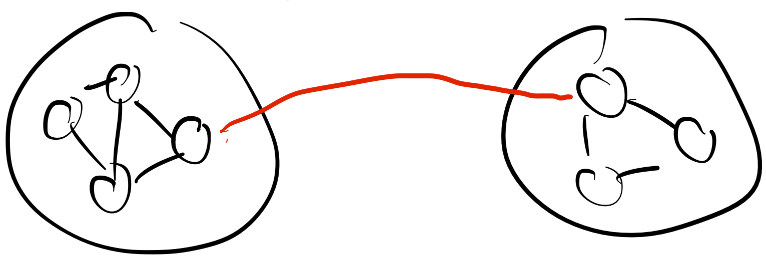


3





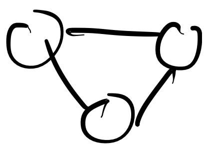
1a) giant component - most nodes in one component



more general than social networks (Watts / Strogatz 98)

---

② sparse (not too many edges)

③ clustered  ④ heavy-tailed degree dists

② sparsity $nnz(A) \ll n^2$

Example: FB (2011)

$n \simeq 721M$ users

$\sim 68.7B$ friendships

~ 95 friends/person

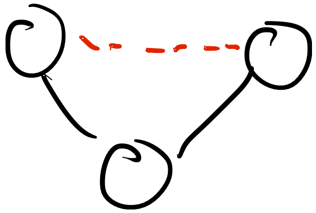
$Pr(edge) \simeq 3 \cdot 10^{-7}$

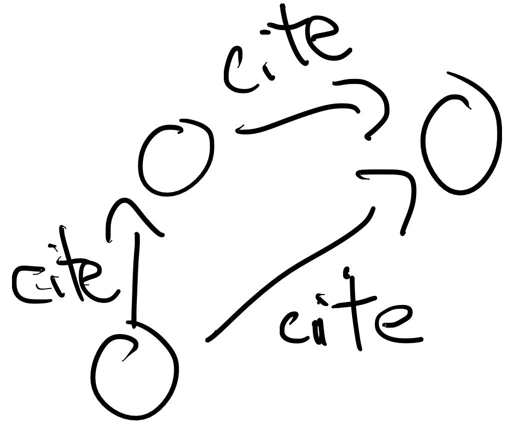
Matrix-vector products $O(\#edges)$

Linear systems

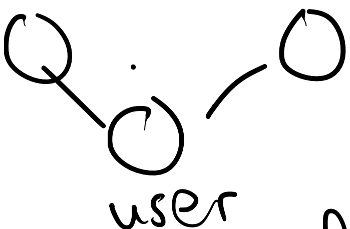
Eigenvectors

③ clustering





FB users with exactly 100 friends



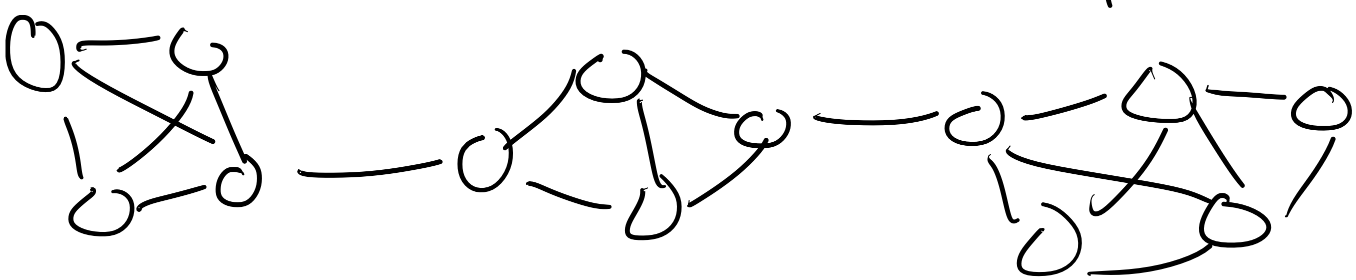
$\binom{100}{2} = 4950$ ≈ 700 connected (14%)

$Pr(edge) = 3 \cdot 10^{-7}$

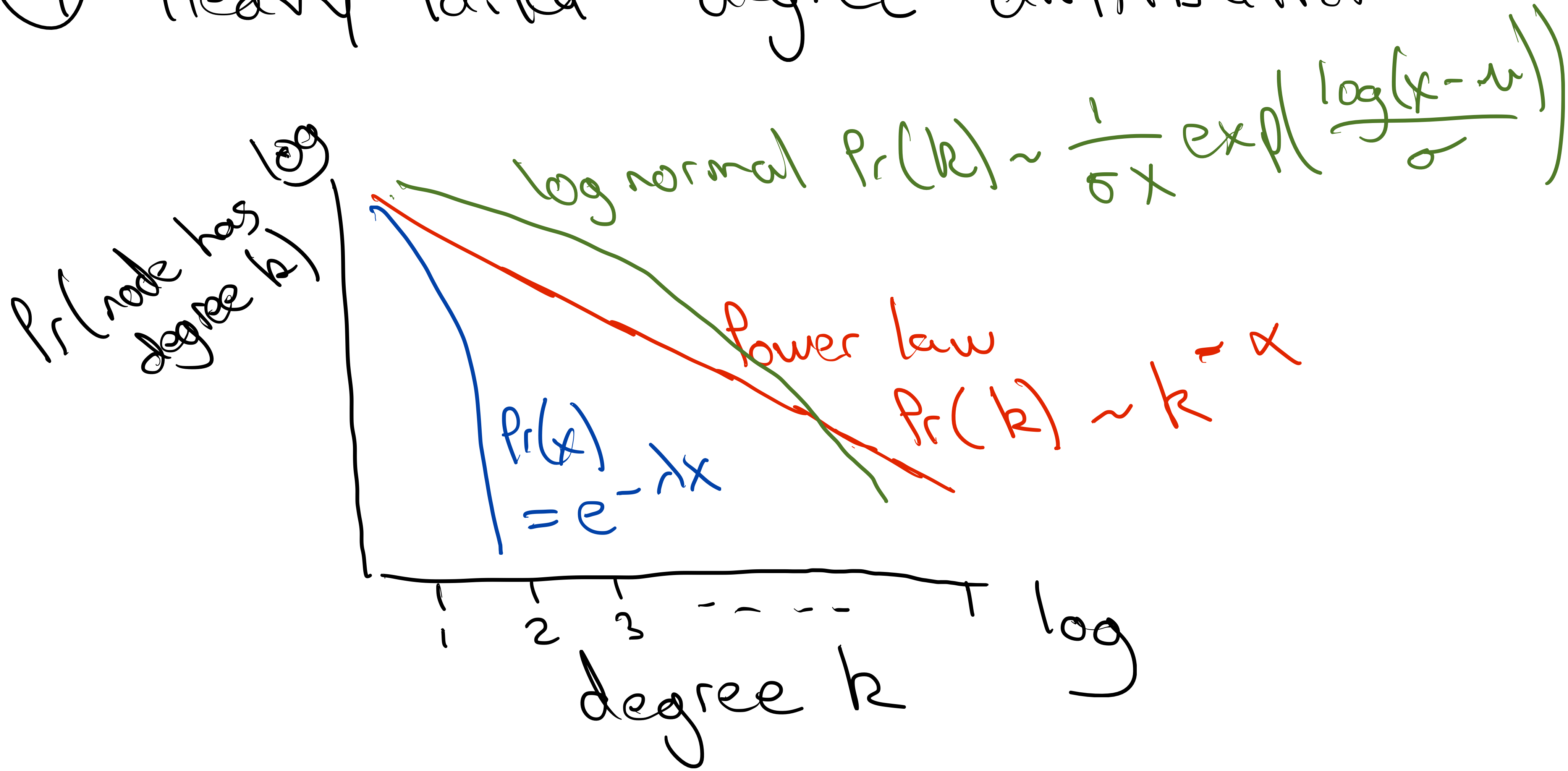
clustering coefficient

Local density + global sparsity

bio: functional groups



4 "Heavy-tailed" degree distribution



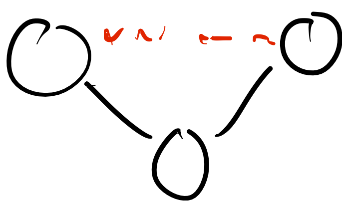
Random graph models

1. Baselines for what you could expect in data
   - is structure "real" or "random"
   - testbeds for algorithms
2. understanding phenomena / "explanatory"
   - fit model params data
   - why might you get a PL?

$G_{n,p}$: aka Erdős-Rényi

n nodes, add each edge w.p. p i.i.d

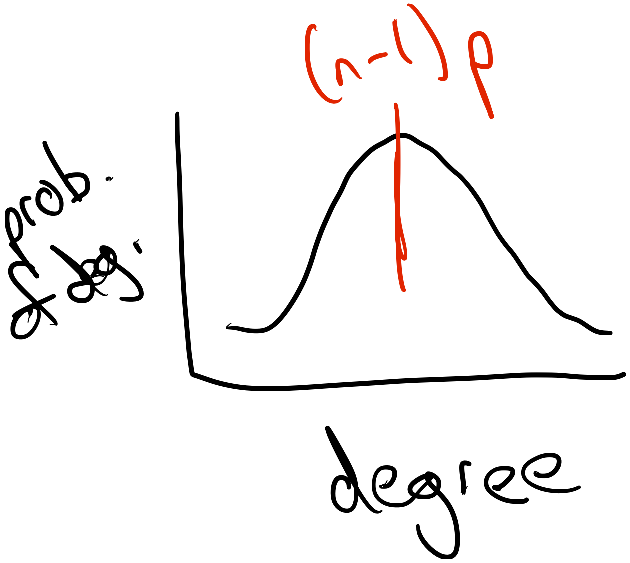
FB: $p \approx 3 \cdot 10^{-7}$ $\quad \dfrac{\#\text{edges}}{\binom{\#\text{nodes}}{2}}$



Hops? Giant component?

$np > 1+\varepsilon \Rightarrow$ giant component (constant frac)

others $O(\log n)$



$\dfrac{np}{\log n} > 2 \Rightarrow$ diameter $\approx \dfrac{\log n}{\log(np)}$

Stochastic block model (SBM)

2-block

$$\begin{array}{c|cc|} & n_1 & n_2 \\ \hline n_1 & p & q \\ n_2 & q & p \\ \hline \end{array}$$

$p > q \Rightarrow$ cluster structure

$p = \frac{a \log n}{n}$ $\quad q = \frac{b \log n}{n}$

$a > b$

$a > b \quad \sqrt{a} > \sqrt{b} + \sqrt{2}$

second eigenvector of A, take sign

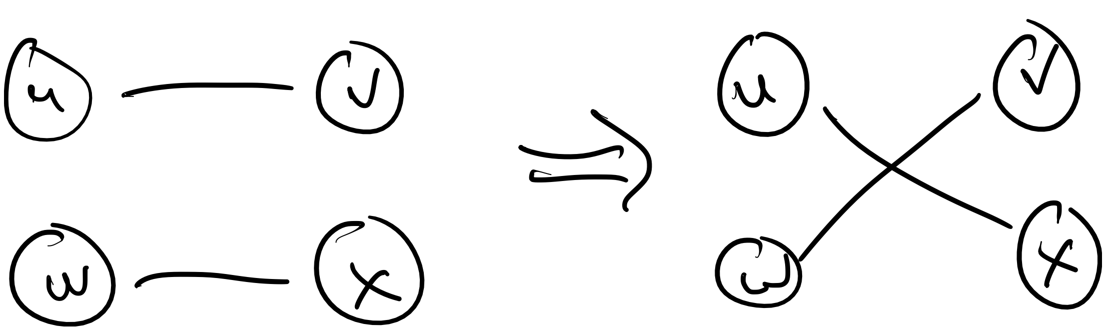
$$\max_{B, P} \Pr(A \mid B, P) \qquad \Pr((i,j) \in E) = P_{B(i), B(j)}$$

b blocks $\quad b \times b$ matrix P

Configuration model

1. Starts with degree sequence $d_1, \dots, d_n$ (D)
2. Samples uniformly at random from all graphs with this degree seq.

Swapping alg:



Random walk on the space of graphs

Markov chain: ergodic and aperiodic

stationary dist: uniform over graphs with $d_1, \dots, d_n$