March 7, 2019

Last time: network science

4 fundamental matrices

1. Adjacency matrix A
2. Diagonal degree matrix D

$$D = \mathrm{diag}(A\mathbb{1})$$

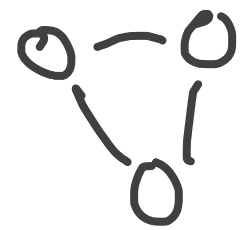
3. Random walk matrix P

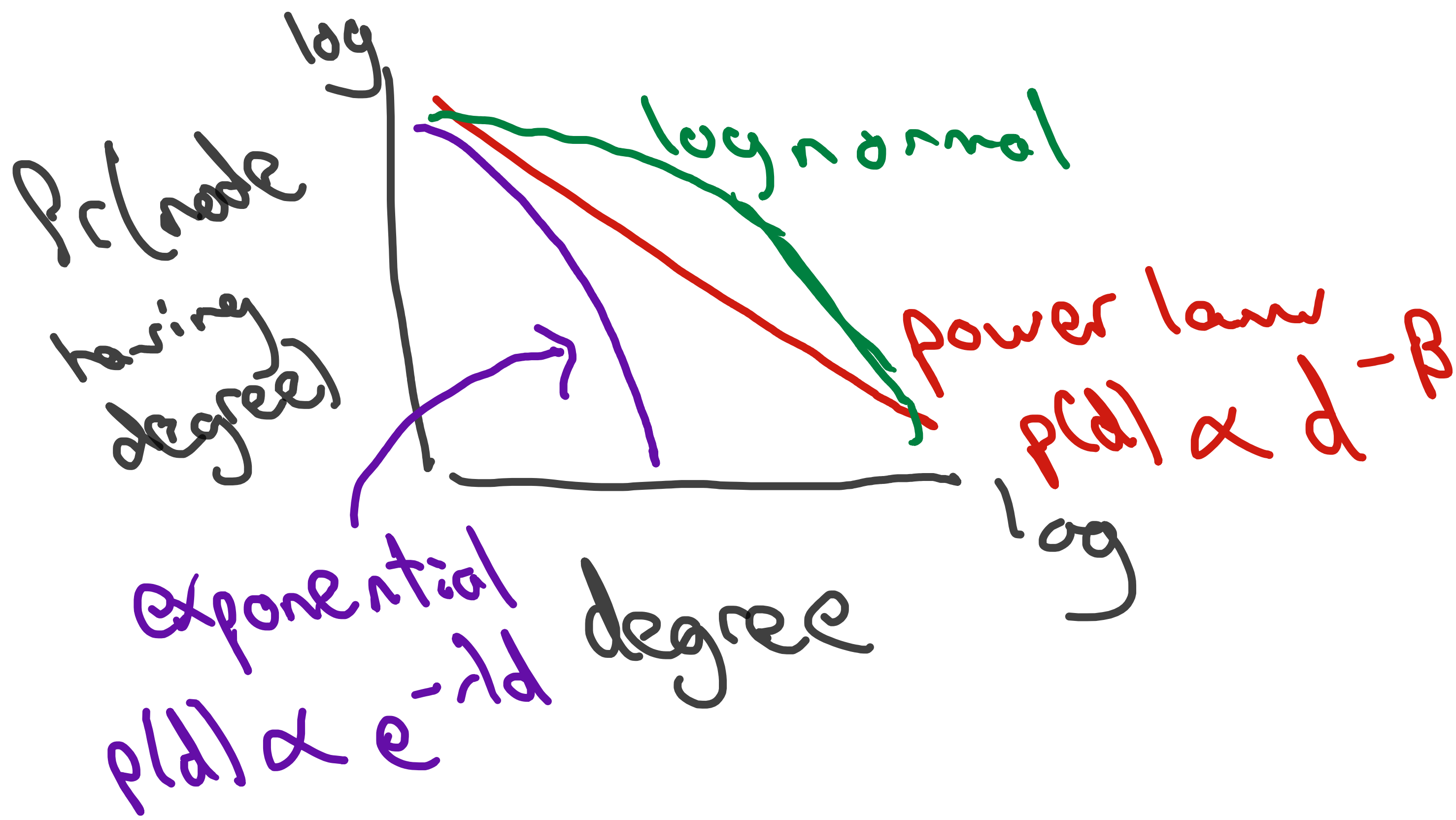
$$P = A^T D^{-1}$$

4. Graph Laplacian L

$$L = D - A \quad (A = A^T)$$

Common graph structure

1. sparse
2. Small hop distance
3. Clustered 
4. Heavy-tailed degree dist.



Why might we get heavy tails?

A model: preferential attachment

rich get richer

At each time step, a new node u gets added

w.p. α, add edge $u \to v$

v chosen uniformly at random

w.p. $1-\alpha$, add edge $u \to w$

w chosen proportional to in-degree

$X_j(t)$ = # nodes with in-degree j after t time steps

$X_j(t)$ $\uparrow$ or $\downarrow$

$$\Pr\left(X_j(t+1) > X_j(t) \right)$$

$$= \underbrace{\alpha X_{j-1}(t)/t}_{\text{uniform}} + \underbrace{(1-\alpha)(j-1) X_{j-1}(t)/t}_{\text{pref.}}$$

$$\Pr\left(X_j(t+1) < X_j(t) \right)$$

$$= \underbrace{\alpha X_j(t)/t}_{\text{uniform}} + \underbrace{(1-\alpha) j X_j / t}_{\text{pref.}}$$

$$\frac{dX_j}{dt} = \frac{\alpha (X_{j-1} - X_j) + (1-\alpha)\left[(j-1) X_{j-1} - j X_j \right]}{t} \qquad (j \geq 1)$$

$$j = 0 \quad \frac{dX_0}{dt} = 1 - \frac{\alpha X_0}{t}$$

$$x_j(t) = c_j \cdot t$$

$$\frac{dx_0}{dt} = c_0 = 1 - \frac{\alpha x_0}{t} = 1 - \alpha c_0$$

$$\Rightarrow c_0 = \frac{1}{1+\alpha}$$

$(*)$ $j \geq 1$

$$c_j(1+\alpha+j(1-\alpha))$$

$$= c_{j-1}(\alpha + (j-1)(1-\alpha))$$

$$\frac{c_j}{c_{j-1}} = \frac{\alpha + (j-1)(1-\alpha)}{1+\alpha+j(1-\alpha)}$$

$$= 1 - \frac{2-\alpha}{1+\alpha+j(1-\alpha)}$$

$$\sim 1 - \left(\frac{2-\alpha}{1-\alpha}\right)\left(\frac{1}{j}\right)$$

Guess: $c_j \sim c \cdot j^{-(2-\alpha)/(1-\alpha)}$

$$\frac{c_j}{c_{j-1}} = \left(\frac{j-1}{j}\right)^{(2-\alpha)/(1-\alpha)}$$

$$\sim 1 - \left(\frac{2-\alpha}{1-\alpha}\right)\left(\frac{1}{j}\right)$$

$$c_j = \left(\qquad \qquad \right) c_{j-1}$$

$\Rightarrow$ PL with exponent

$$B = \frac{2-\alpha}{1-\alpha}$$

(Mitzenmacher 2004)

$$p(d) \propto d^{-B}$$

Random graph models

Generate graph data via some model

Useful for

1. Understanding phenomena
   → heavy tails / rich get richer
2. Baseline for what you could expect
   → if "structure" is real
   → testing algorithms

---

3 Random graph models

1. $G_{n,p}$ / Erdős - Rényi
2. Configuration models
3. Stochastic block models

---

1. $G_{n,p}$

n nodes

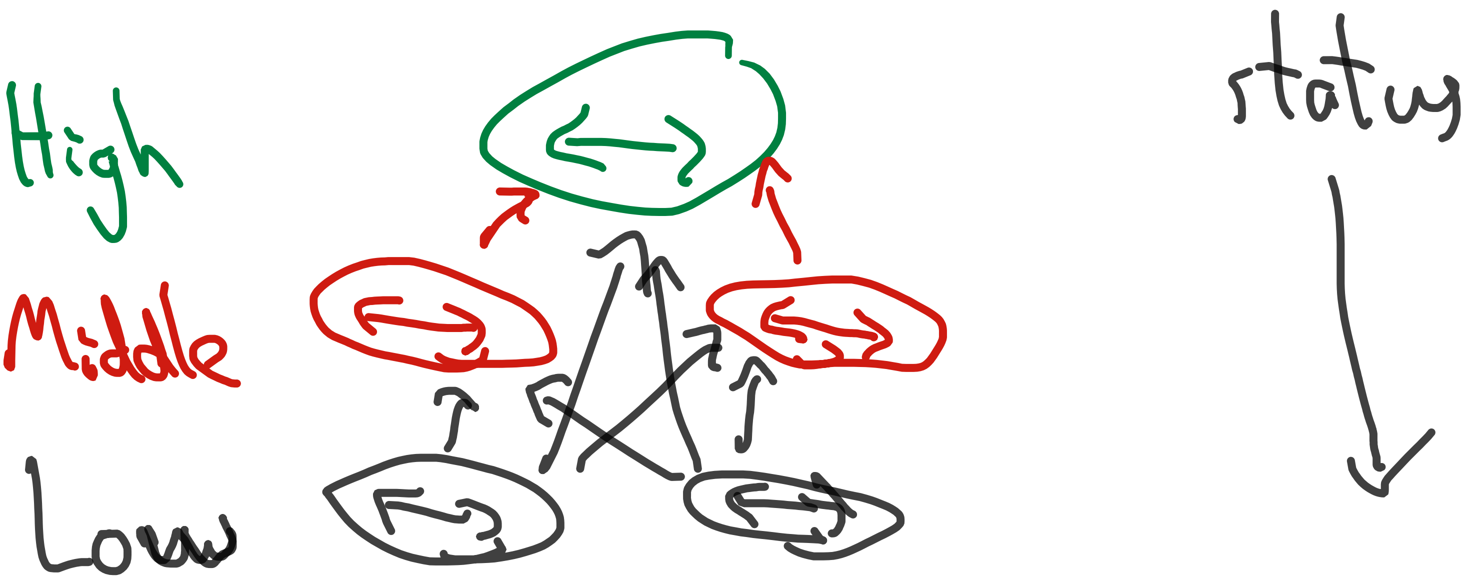
each edge appears w.p. p

Simple but often effective

Davis & Leinhardt

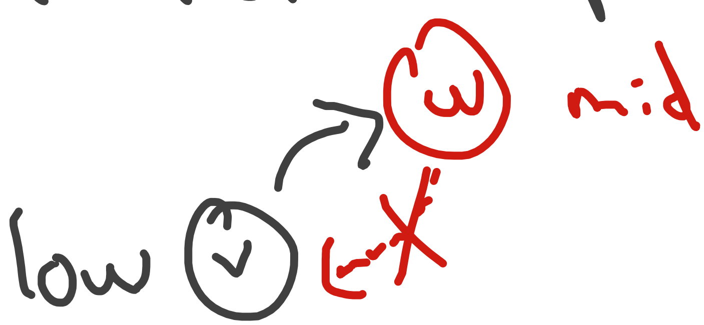
1967

Homans social theory



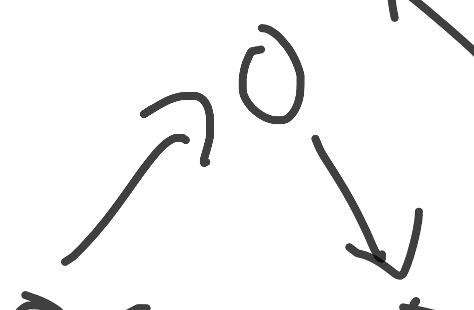
u -> v means that u thinks positively of v

Lower level thinks favorably of equal or higher levels, but not reciprocated



Newcomb survey data of 17 students in an experimental college dorm

$G_{n,p}$-type model



null: expected 50.9

in data: 29

match Homans theory

How do we sample $G_{n,p}$

when p is small?

Naively, $O(n^2)$

Ball-dropping

1. Sample # of edges m

   $m \sim \text{Binomial}\left(\binom{n}{2}, p\right)$

2. Repeat m times ....

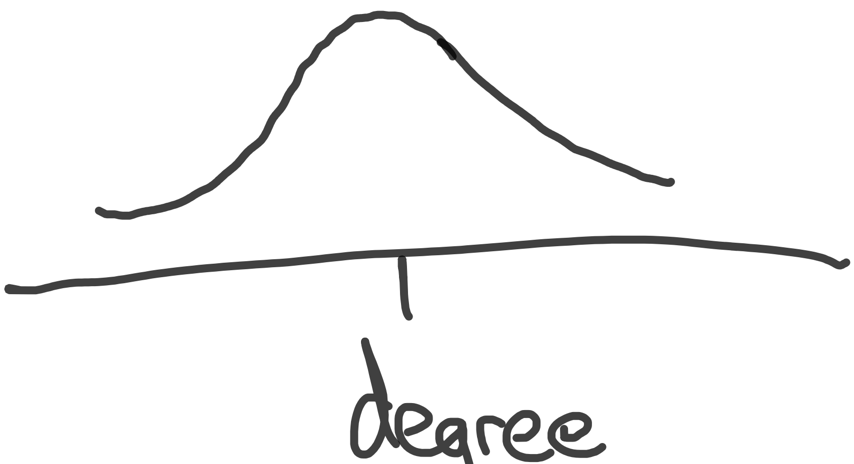
   Sample i, j from $\{1, \dots, n\}$

   Add (i, j) to graph

   (repeat if duplicates)

② Configuration models

$G_{n,p}$ a little simple



no heavy tails in the degree dist. of $G_{n,p}$

CM:

① Start with degree sequence $d_1, \dots, d_n$

② Samples uniformly at random

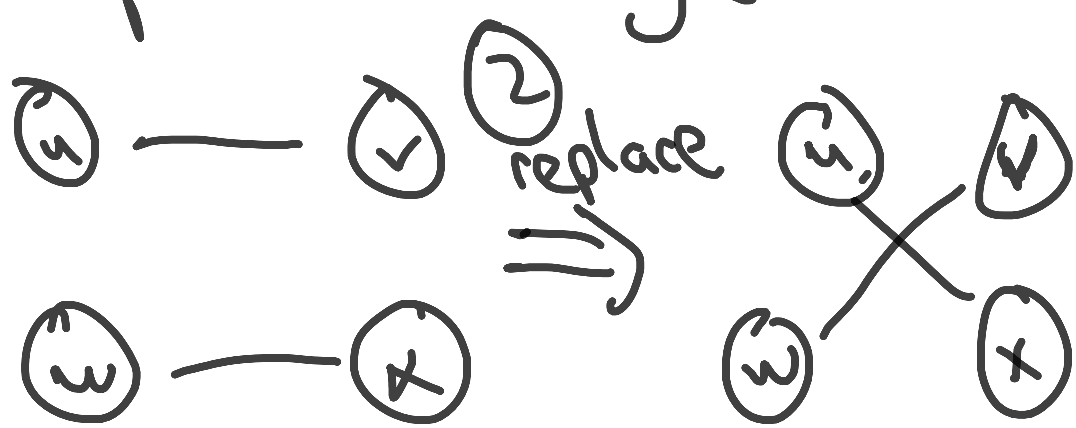
from all graphs with the same degree sequence

How to sample?

Swapping algorithm

for trial = 1, 2, 3, ----

1. Sample two edges



Random walk on the space of graphs with the same degree distribution

This RW is (as a Markov chain)

- ergodic
- aperiodic
- stationary distribution is uniform over all graphs with given deg. dist.

MCMC sampler

How fast does it mix

$\Leftrightarrow$ how many swaps do I need to do

rule of thumb: $10m$ swaps

3 Stochastic block model (SBM)

2-block SBM

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

usually assume $p \geq q$

General b-block model

$$\Pr((i,j) \in E) = P_{B(i),B(j)}$$

P $b \times b$ $\quad B(k) \in \{1, 2, \dots, b\}$