LECTURE 30

Last class: matrix concentration inequalities. We upgraded

(scalar) Hoeffding inequality: let a_i be numbers $\Rightarrow$

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n} \underset{\pm 1 \text{ iid}}{\varepsilon_i} a_i\right| \geq t\right\} \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right), \quad \text{where } \sigma^2 = \sum_{i=1}^{n} a_i$$

to Matrix Hoeffding inequality (MHI): let A_i be $d \times d$ symmetric matrices $\Rightarrow$

$$\mathbb{P}\Big\{ \underbrace{\Big\| \sum_{i=1}^{n} \varepsilon_i A_i \Big\|}_{S_n} \geq t \Big\} \leq 2d \cdot \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad \text{where } \sigma^2 = \Big\| \sum_{i=1}^{n} A_i^2 \Big\|$$

- The cost of the upgrade $= d$. Is it big? Curse of H.D.?

$$\mathbb{P}\{\|S_n\| \geq t\} \leq 2\exp\left(\log d - \frac{t^2}{2\sigma^2}\right) = \frac{2}{d} \ll 1 \quad \text{in H.D.}$$

choose $t = 2\sigma\sqrt{\log d}$

$\Rightarrow$ $\|S_n\| \leq 2\sigma\sqrt{\log d}$ with high probability

the cost is logarithmic ☺

- Expectation = ? Recall the integral identity for $\forall$ r.variable $Z \geq 0$:

$$\mathbb{E}X = \int_0^\infty \mathbb{P}\{X \geq s\}\, ds$$

$$\mathbb{E}\|S_n\| \leq \left(\mathbb{E}\|S_n\|^2\right)^{1/2} \quad (L_1 \text{ norm} \leq L_2 \text{ norm}). \qquad (*)$$

$$\mathbb{E}\|S_n\|^2 = \int_0^\infty \mathbb{P}\{\|S_n\|^2 \geq s\}\, ds \overset{t=\sqrt{s}}{\leq} \int_0^\infty 2d \cdot \exp\left(-\frac{s}{2\sigma^2}\right) ds = 4\sigma^2 d$$

too large! not logarithmic ☹

- The loss occurs for small t.

For example, for $t=0$, M.H.I gives $\mathbb{P}\{\cdots\} \le \underbrace{2d}$

Useless! The trivial bound $\mathbb{P}\{\cdots\} \le 1$ is better

- $\Rightarrow$ Use the trivial bound (1) for small t, M.H.I's bound for large t. | For simplicity, let $\sigma=1$.

$$\mathbb{E}\|S_n\|^2 \le \int_0^{s_0} \underbrace{\mathbb{P}\{\|S_n\|^2 \ge s\}}_{\le 1} ds + \int_{s_0}^{\infty} \underbrace{\mathbb{P}\{\|S_n\|^2 \ge s\}}_{\le 2d\exp(-s/2)} ds$$

$$\le s_0 + 4d\exp(-s_0/2). \quad \text{Choose } s_0 = 2\log d \Rightarrow$$

$$\le 2\log d + 4 \le 8\log d \qquad (d \ge 2)$$

For general σ:

$$\mathbb{E}\|S_n\|^2 \le 8\sigma^2 \log d \qquad \text{(DIY)}$$

def. of σ in M.H.I $\Downarrow$

$$\boxed{\begin{array}{l}\text{Cor } \forall\ d\times d \text{ symmetric matrices } A_i, \\ \mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i A_i\right\|^2 \le 8\log(d)\left\|\sum_{i=1}^n A_i^2\right\|\end{array}}$$

- This is a matrix version of the scalar identity

$$\mathbb{E}\left(\sum_{i=1}^n \varepsilon_i a_i\right)^2 \underset{\substack{\uparrow \\ \text{var(sum)} = \text{sum(var)}}}{=} \sum_{i=1}^n a_i^2 \qquad \forall a_i \in \mathbb{R}$$

- (*) p.1 $\Rightarrow$ $\mathbb{E}\|\sum \varepsilon_i A_i\|^2 \le 3\sqrt{\log d}\,\left\|\sum_{i=1}^n A_i^2\right\|^{1/2}.$

- Similarly, we can upgrade Chernoff inequality **(HW)** and Bernstein's ineq. from Oct. 3, lec 14:

Matrix Bernstein Inequality (MBI)

> Let X_i be independent, mean zero, symmetric $d \times d$ random matrices. Assume $\|X_i\| \le 1$ with prob. 1. Then $\forall t \ge 0$:
>
> $$\mathbb{P}\left\{ \left\| \sum_{i=1}^n X_i \right\| \ge t \right\} \le 2d \cdot \exp\left[-c \left(\frac{t^2}{\sigma^2} \wedge \frac{t}{K} \right) \right]$$
>
> minimum
>
> where $\sigma^2 = \left\| \sum_{i=1}^n \mathbb{E} X_i^2 \right\|$.

"matrix variance" absolute constant >0 (see Thm 5.4.1 in Book)

- If we choose $t = C\sigma\sqrt{\log d} \vee CK\log d$ (maximum) with a sufficiently large absolute const C, we get $\mathbb{P}\{\cdots\} \le 1/d \ll 1$.

  $\Rightarrow$ $\left\| \sum_1^n X_i \right\| \le C\sigma\sqrt{\log d} + CK\log d$ with high probability ☺

- Also, in expectation (argue similarly to p. 1):

Cor $$\boxed{\mathbb{E} \left\| \sum_{i=1}^n X_i \right\| \le C\sigma\sqrt{\log d} + K\log d.}$$ (DIY)

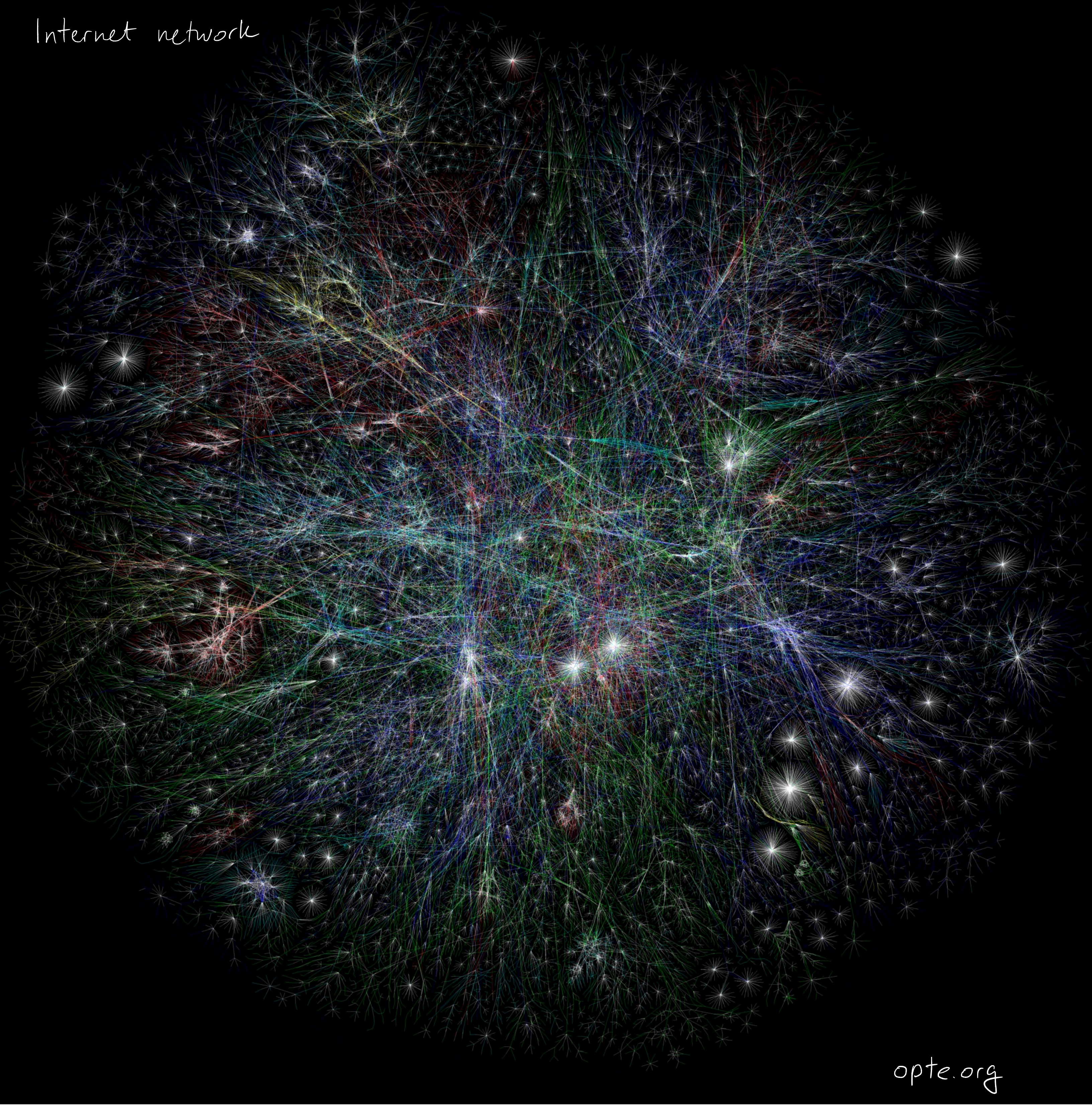
A VERY USEFUL TOOL. No assumptions on random matrices X_i — no independence of entries!

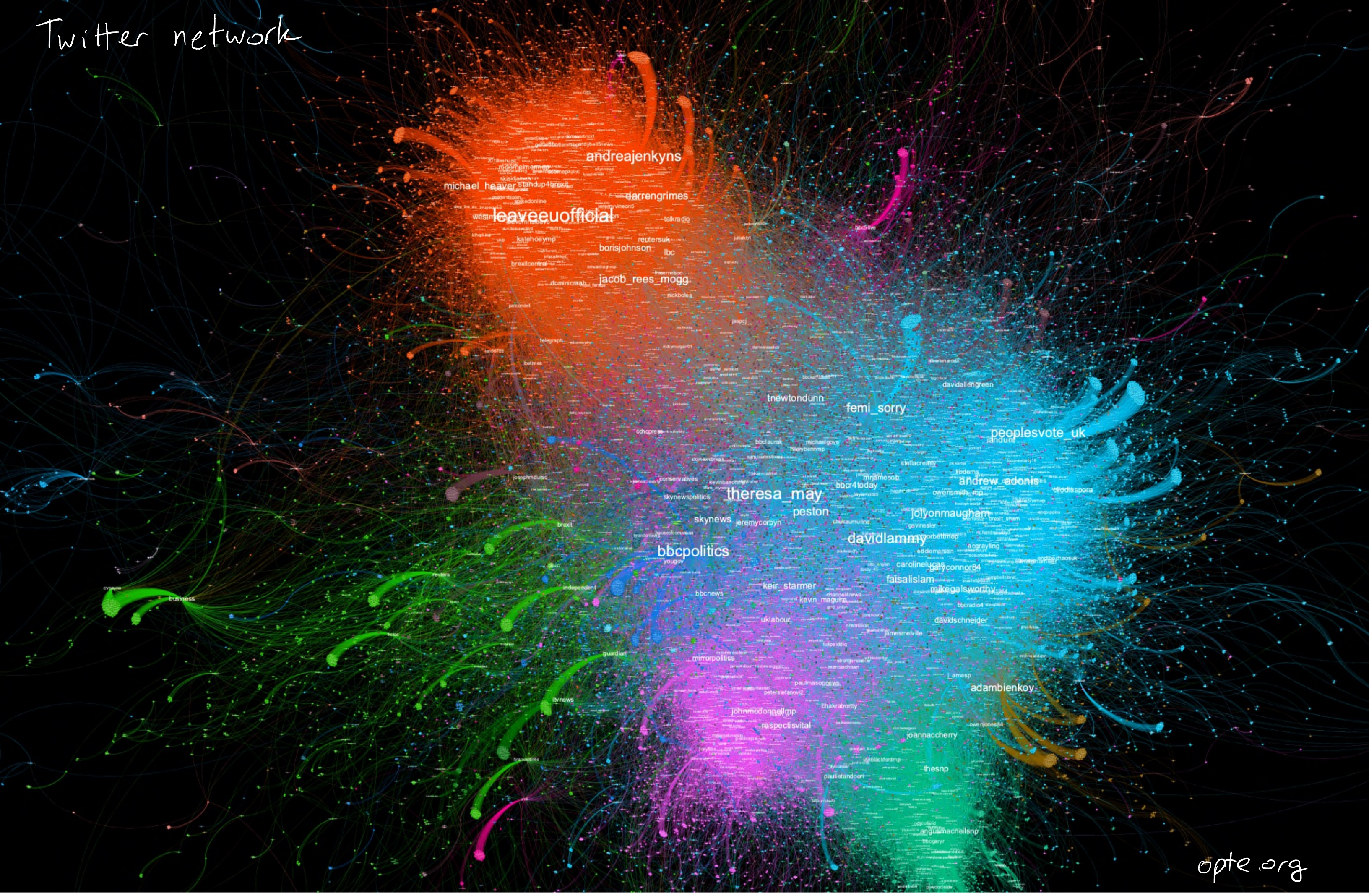
APPLICATION: COMMUNITY DETECTION IN NETWORKS

lec. 8 (Sep. 19): a network is an (undirected) graph

$$G = (\underset{\text{vertices}}{V}, \underset{\text{edges}}{E})$$

- Examples:


Internet network
opte.org

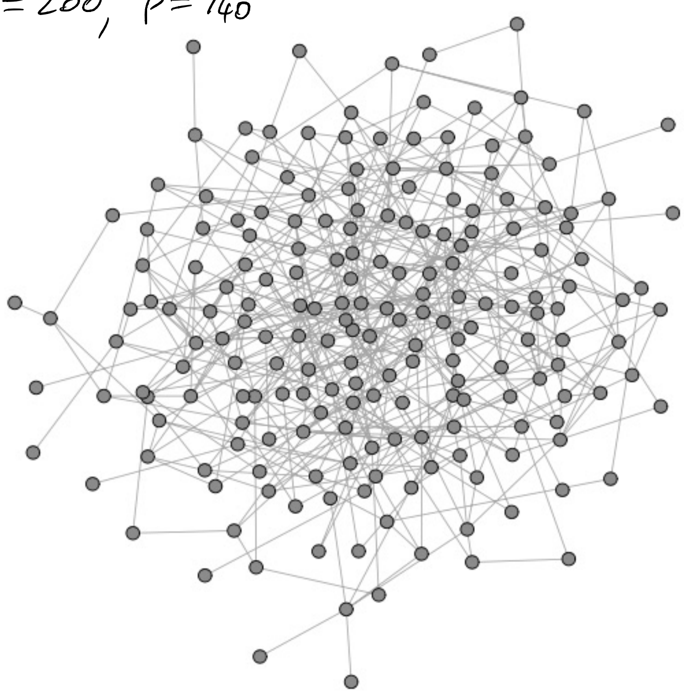

Twitter network
andreajenkyns
michael_heaver
darrengrimes_
leaveeuofficial
borisjohnson
jacob_rees_mogg
femi_sorry
peoplesvote_uk
theresa_may
peston
andrew_adonis
jolyonmaugham
davidlammy
bbcpolitics
skynews
faisalislam
keir_starmer
adambienkov
opte.org

- Problem: find communities (clusters).

- lec 8 (Sep 19): a mathematical model of a complex network
  = random graph.

- Erdös-Rényi model $G(n,p)$:
  n vertices; connect each pair $i \neq j$ prob. p independently

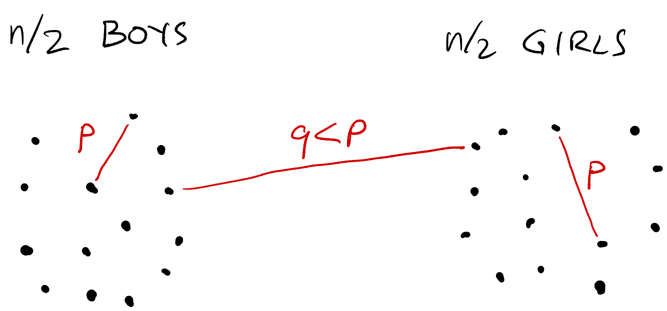
$n = 200$, $p = 1/40$



Pro of $G(n,p)$: elegant, simple model
Con: too simple. No communities.

How to model communities?

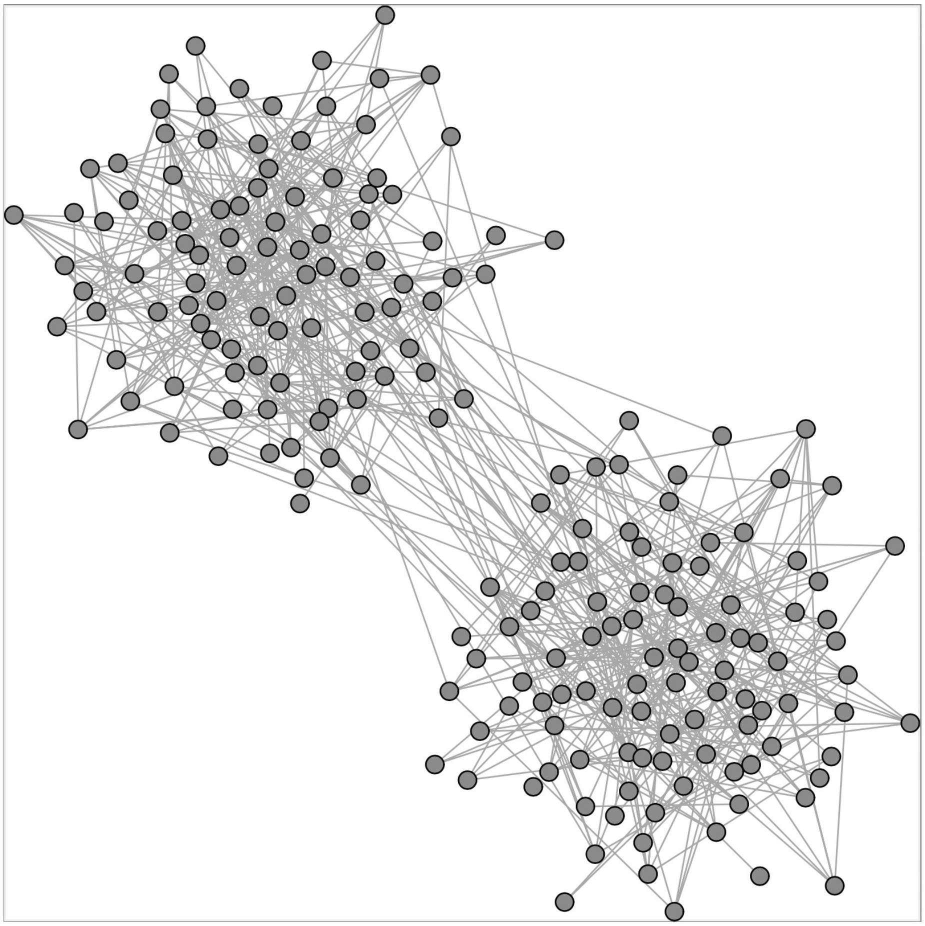
- Answer: allow different probabilities of edges



def "stochastic block model"

$G(n,p,q)$

Example: $G(n,p,q)$ with $n = 200$, $p = 1/20$, $q = 1/200$.



COMMUNITY DETECTION PROBLEM:

$\boxed{\text{Given a random graph drawn from a SBM, find the 2 communities.}}$

- Unsupervised learning
- For what p, q is this possible? and how?

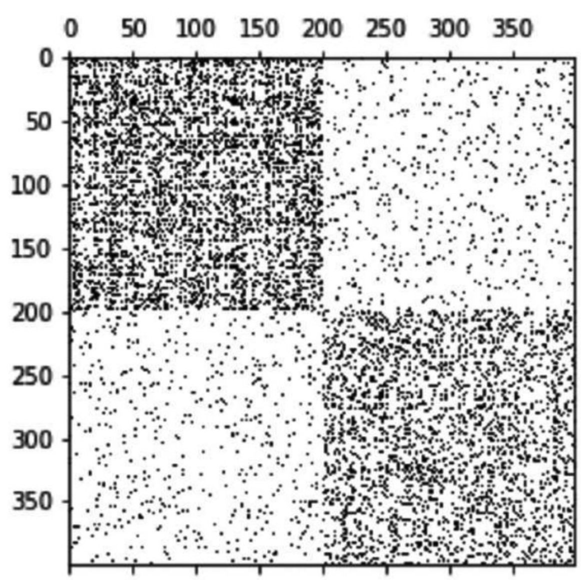
(e.g. for $p = q$: impossible.)

- Key to success: random matrix theory

Def The adjacency matrix A of a graph $G = (V, E)$, $V = \{1, 2, \dots, n\}$, is the $n \times n$ matrix with entries

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{if not} \end{cases}$$

Example: Adj. matrix of a SBM $\longrightarrow$



Brzoska, Fischer, Lentz