March 5, 2019

HW1 due Thurs 11:59pm ET

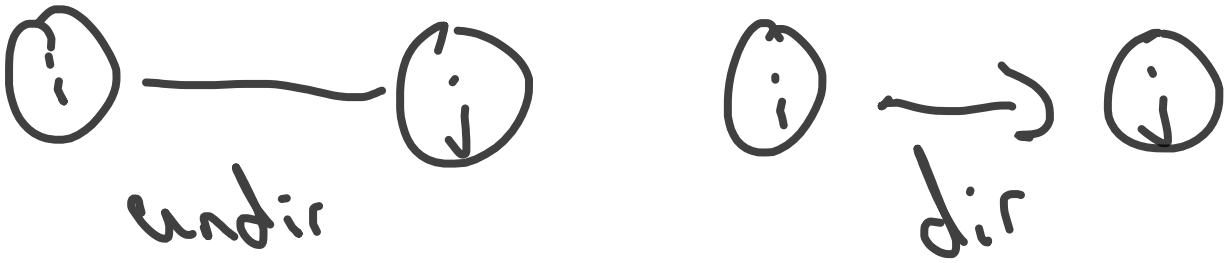
Class so far:

1. Least squares
   - ↳ objective function / subroutine
2. Dimensionality redux
   - Linear → spectral clustering
   - Multilinear (tensor) ↳ hypergraphs
   - Nonlinear ↳ representation learning

Next: network science methods for graph data

Today: ① matrices of graphs
② common structure

Graphs made up of nodes and edges



---

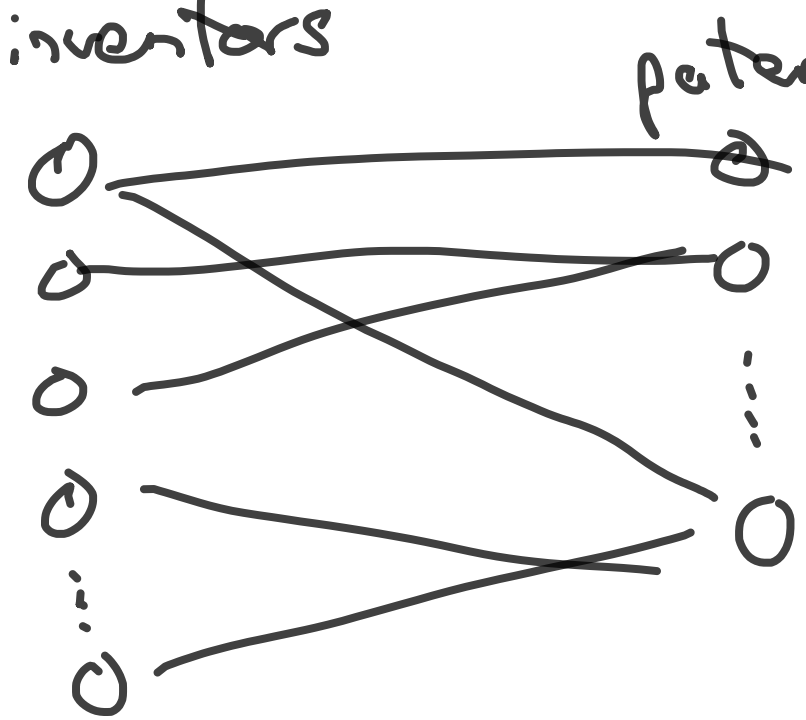
network	node	edge	dir?
Twitter	accounts	follow	yes
FB	"	friendships	no
gene regulation	genes/ proteins	regulation	yes
tectonic plates	plates	physical interaction	yes

multilayer networks

↳ diff edge labels

different physical interactions in tectonic plates

citation	papers	citation	yes
roads	destinations	roads	no/yes
co-invention	people	on a patent together	no



hypergraph?

Matrices associated with networks

Adjacency matrix A

$$A_{ij} = \begin{cases} 1 & i \to j \\ 0 & \text{o/w} \end{cases}$$

undirected $\Rightarrow A = A^T$

Diagonal degree matrix D vector d

$$d = A\mathbf{1}$$

$$d_i = \sum A_{ij} = \text{degree of } i$$

$$D_{ii} = d_i \quad D_{ij} = 0 \text{ if } i \neq j$$

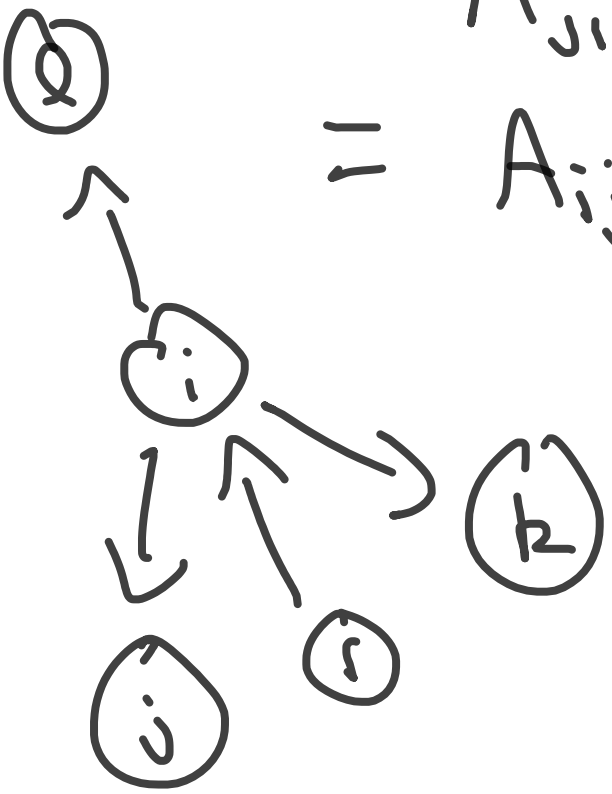
Random walk matrix P

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

$$P_{ji} = (A^T D^{-1})_{ji}$$

$$= A^T_{ji} D^{-1}_{ii}$$

$$= A_{ij} / d_i$$



$$P_{ji} = P_{ki}$$

$$= P_{qi}$$

$$= 1/3$$

Graph Laplacian L

$$L = D - A$$

(undir. graph)

E = set of edges

$(i,j) \in E$ is an undir. edge

Claim: $x^T L x = \sum_{(i,j) \in E} (x_i - x_j)^2$

Proof: $(x_i - x_j)^2 = x_i^2 + x_j^2 - 2x_i x_j$

$L = D - A$

$x^T L x = x^T D x - x^T A x$

$x^T D x = \sum_i d_i x_i^2$

$x^T A x = \sum_{i,j} A_{ij} x_i x_j$

$= 2 \sum_{(i,j) \in E} x_i x_j$

Classic spectral graph theory

Claim: number of connected comp.
= number of zero evals of L

conn. component C

$$c_i = \begin{cases} 1 & i \in C \\ 0 & \text{o/w} \end{cases}$$

$$c^T L c = \sum_{(i,j)\in E} (c_i - c_j)^2$$

$$= \sum_{\substack{(i,j)\in E \\ i,j \in C}} (1-1)^2 + \sum_{\substack{(i,j)\in E \\ i,j \notin C}} (0-0)^2$$

$$= 0$$

L p.s.d $\Rightarrow$ c eigenvector with eigenvalue 0

components disjoint

One vector c for each component

$\Rightarrow$ linearly independent

$\Rightarrow$ # of zero e-vals $\geq$ # of components

Suppose that $x^T L x = 0$

$$\Rightarrow \sum_{(i,j)} (x_i - x_j)^2 = 0$$

$x_i = \alpha_j \iff i \in j^{th}$ conn. comp.

$$x = \sum_{j=1}^{r} \alpha_j c_j$$ ← indicator vector

$$\in \text{span}\{c_1, \ldots, c_r\}$$

Matrices: A, D, P, L

Things we do with matrices:

(1) Solve systems of equations

$$(I - \alpha P)x = (1-\alpha)1$$

(PageRank)

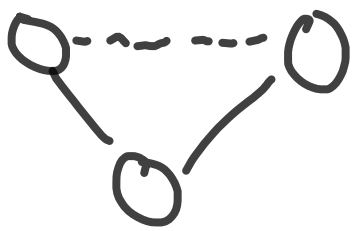
(2) Compute eigenpairs
(or singular pairs)

$$D^{-1/2} L D^{-1/2} x = \lambda x$$

(spectral clustering)

Properties common in graph datasets

1. sparse (not too many edges)
2. small hop distances ("6 degrees of separation")
3. clustered



lots of triangles

4. degree distributions heavy-tailed

① Sparsity

How sparse is sparse?

- sparse if you can benefit from it (computational)

$nnz(A) << n^2$ (total poss)

Facebook 2011
(Ugander et al.)

$n \approx 721$ million active users

~ 68.7 billion friendships

≈ 95 friends / person

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

Twitter 2009 (Kwak et al.)

$n \approx 41.7$ million

~ 1.47 billion following relationships

≈ 35 followers/followees per user

$Pr(\text{edge}) \approx 8 \cdot 10^{-7}$

Food Web (Florida Bay)

$n = 128$ species

2,106 edges

$Pr(\text{edge}) \approx 0.13$

② Small hop distances

"small world"

"six degrees of separation"

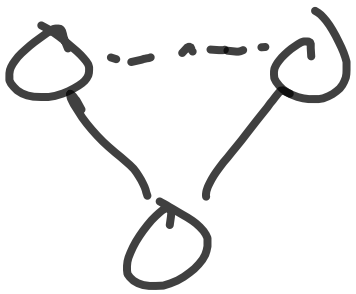
Milgram's experiment (1969)

296 "random" people in U.S.

Goal: forward chain letter to specific person in Boston

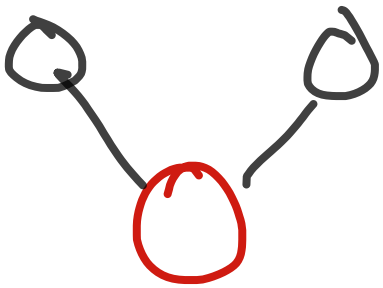
64 ~ 1/5 made it

3) Clustered



FB study:

users with exactly 100 friends



$\binom{100}{2} = 4950$

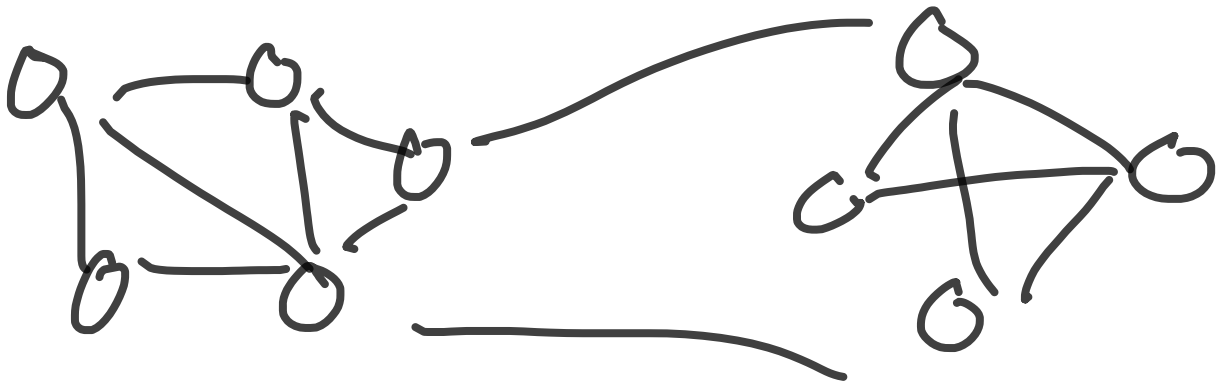
$\approx$ 700 / 4950 existed

(14%)

$P_c(\text{edge}) \approx 3 \cdot 10^{-7}$

in general, called clustering coefficient

Biology:
modular structure



(same with the Internet, power grids, ...)

---

(4) heavy-tailed degree distributions