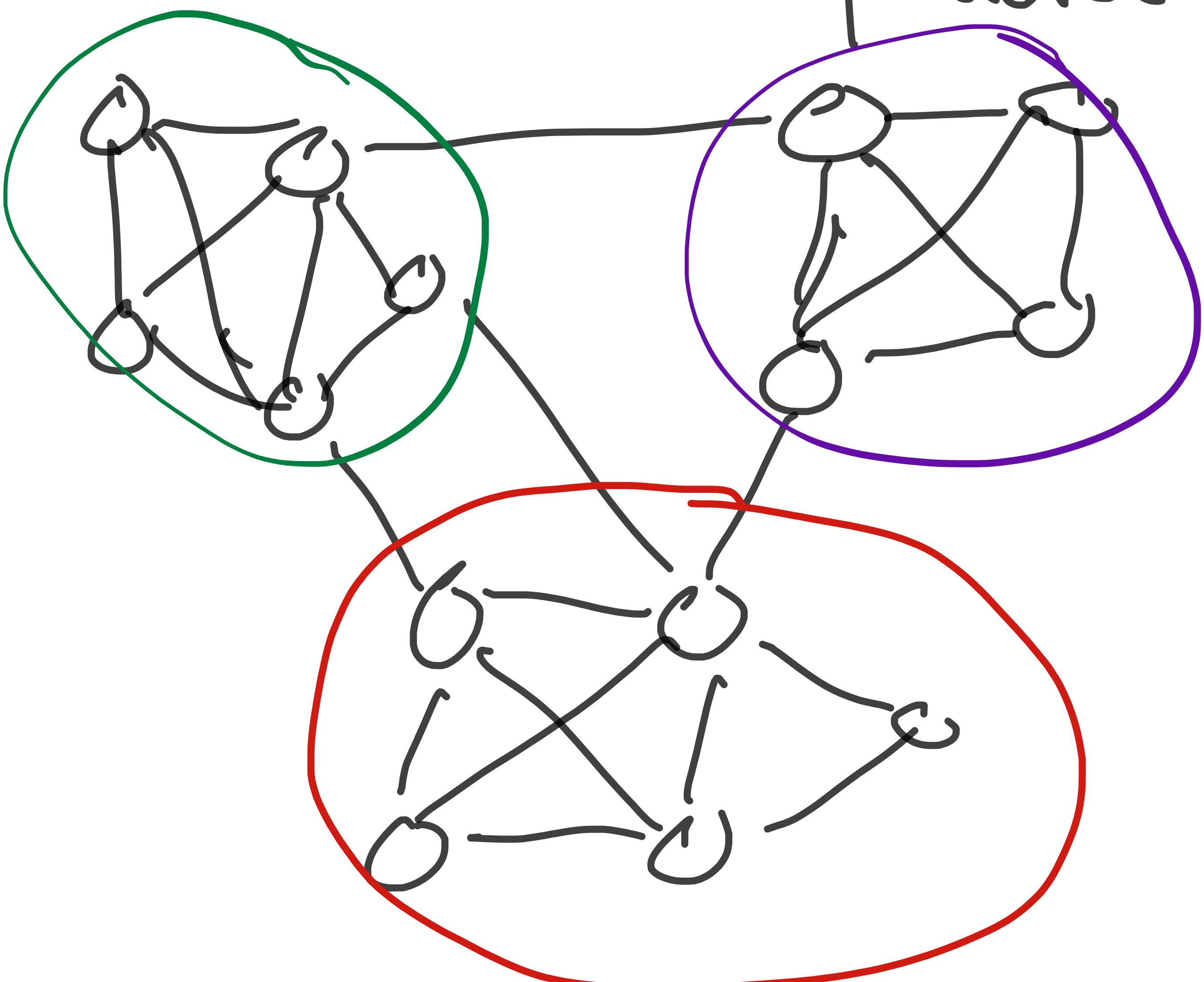
March 12, 2019

Graph clustering / partitioning / community detection



communities in social networks

Functional modules in bio

Food webs – different layers

Web graph – structure of the web

Lots of methods

Our focus: spectral methods

HW1 due

Reaction paper

End of lecture last time

Second eigenvector of L

$L = D - A$

$Lv = \lambda_2 v$

Math:

$$x_i = \begin{cases} +1 & i \in S \\ -1 & i \in T \end{cases}$$

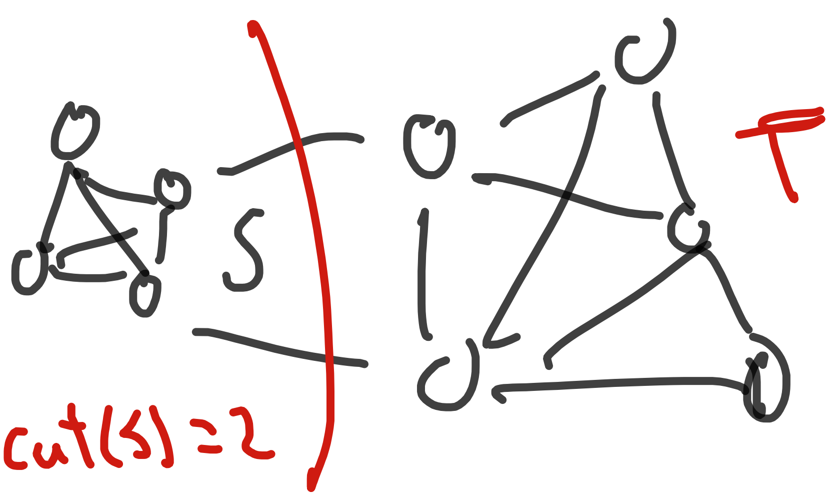
$$T = V \setminus S$$

$$S \cup T = V, \quad S \cap T = \emptyset$$

S = block 1

T = block 2

Claim: $x^T L x = c \cdot \text{cut}(S)$



$\text{cut}(S) = \text{cut}(T)$

$$\mathrm{cut}(S) = \#\text{edges leaving } S$$

$$= \sum_{\substack{i \in S \\ j \in T}} A_{ij}$$

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

$$= \sum_{\substack{i,j \in S \\ (i,j) \in E}} \cancel{(\overset{1}{x_i} - \overset{1}{x_j})^2} + \sum_{\substack{i,j \in T \\ (i,j) \in E}} \cancel{(\overset{-1}{x_i} - \overset{-1}{x_j})^2} + \sum_{\substack{i \in S \\ j \in T \\ (i,j) \in E}} (\overset{1}{x_i} - \overset{+1}{x_j})^2$$

$$\qquad\quad 0 \qquad\qquad\qquad\qquad\quad 0$$

$$= \sum_{\substack{i \in S \\ j \in T \\ (i,j) \in E}} 2^2 = 4 \cdot \mathrm{cut}(S)$$

$\square$

$$\min_{S} \ \mathrm{cut}(S)$$

$$\Longleftrightarrow \min_{x,S} \ x^T L x$$

$$x_i = \begin{cases} 1 & i \in S \\ -1 & i \in T \end{cases}$$

$S = V \Rightarrow$ objective takes 0

Constraint: bisection

$$|S| = |T| = |V|/2$$

$$\sum_{i \in S} x_i + \sum_{j \in T} x_j = 0$$

$$\mathbb{1}^T x = 0$$

$$x^T x = \sum x_i^2 = |V| = n$$

$$\min_{x, S} \; x^T L x \qquad \circledast$$

$$\text{s.t.} \quad \mathbb{1}^T x = 0 \qquad x^T x = n$$

tractable $\quad \cancel{x_i = \begin{cases} 1 & i \in S \\ -1 & i \in T \end{cases}}$

"graph bisection" NP-hard

Claim: solution to $\circledast$ is

$$L v_2 = \lambda_2 v_2 \qquad (v^T v = n)$$

(assume graph is conn)

Proof: $L = V \Lambda V^T$

$$z = \sum_{j=1}^{n} \alpha_j v_j \qquad \lambda_1 \le \lambda_2 \le \cdots$$

$$z^T L z = \sum_{j=1}^{n} \alpha_j^2 \lambda_j \qquad \lambda_j \ge 0$$

Last time: $\lambda_1 = 0$

Graph is connected:

$$v_1 = \mathbb{1}$$

$$\alpha_1 = \alpha_3 = \alpha_4 = \dots = \alpha_n = 0$$

$$\Rightarrow z = \alpha_2 v_2$$

$$\alpha v_2$$

$x^T x = n$

Thus far: v_2 is the solution to the relaxed objective function

Need to round back

to get a set S

$$S = \{ i : (v_2)_i > 0 \}$$

$$\mathrm{RatioCut}(S,T)$$

$$= \frac{\mathrm{cut}(S)}{|S|} + \frac{\mathrm{cut}(T)}{|T|}$$

$$x_i = \begin{cases} \sqrt{|T|/|S|} & i \in S \\ -\sqrt{|S|/|T|} & i \in T \end{cases} \quad (*)$$

Claim: $x^T L x = c \cdot \mathrm{RatioCut}(S,T)$
(under $(*)$)

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

$$= \sum_{\substack{i \in S \\ j \in T \\ (i,j) \in E}} \left(\sqrt{|T|/|S|} + \sqrt{|S|/|T|} \right)^2$$

$$\frac{|S|}{|S|} + \frac{|T|}{|T|}$$

$$= \sum \frac{|T|}{|S|} + \frac{|S|}{|T|} + 2$$

$$= \sum \frac{|T| + |S|}{|S|} + \frac{|T| + |S|}{|T|}$$

$$= n \sum \frac{1}{|S|} + \frac{1}{|T|}$$

$$= n \cdot \mathrm{RatioCut}(S, T)$$

$$\mathbf{1}^T x = \sum_{i \in S} \sqrt{|T|/|S|} - \sum_{j \in T} \sqrt{|S|/|T|}$$

$$= 0$$

$$x^T x = \sum_{i \in S} |T|/|S| + \sum_{j \in T} |S|/|T|$$

$$= |T| + |S| = n$$

$$\min_{x, S} \; x^T L x$$

s.t. $\mathbb{1}^T x = 0$ $\quad x^T x = n$

relax $\Rightarrow$ tractable ~~x as in (*)~~

NP-hard to optimize

Solution is v_2, $Lv_2 = \lambda_2 v_2$

Same relaxation as before!

Normalized cuts

$$NCUT(S,T) = \frac{cut(S)}{vol(S)} + \frac{cut(T)}{vol(T)}$$

$$vol(S) = \sum_{i \in S} d_i$$

$$x_i \overset{*}{=} \begin{cases} \sqrt{vol(T)/vol(S)} & i \in S \\ -\sqrt{vol(S)/vol(T)} & i \in T \end{cases}$$

Claim 1: $x^T L x = c \cdot NCUT(S,T)$

Claim 2: $\mathbb{1}^T (Dx) = 0$

Claim 3: $x^T D x = \text{const} \; (= 2 \cdot m)$

$$\min \; x^T L x$$

$$\text{s.t.} \; \mathbb{1}^T (Dx) = 0 \quad x^T D x = 2m$$

$\cancel{*}$ relax $\Rightarrow$ tractable

NP-hard to opt. NCUT

$z = D^{1/2} x$

Normalized Laplacian N

$$\min_{z} \; z^T \underbrace{D^{-1/2} L D^{-1/2}}_{\text{Normalized}} z$$

$$\text{s.t.} \quad \mathbb{1}^T D^{1/2} z = 0$$

$$z^T z = 2m$$

Claim: $\mathbb{1}^T D^{1/2}$ is an eigenvector with 0 eigenvalue of N

Proof: $L \mathbb{1} = 0$

$$L D^{-1/2} D^{1/2} \mathbb{1} = 0$$

$$\underbrace{D^{-1/2} L D^{-1/2}}_{N} (D^{1/2} \mathbb{1}) = 0$$

Claim: if graph is conn.,
N has exactly one 0 eval

$$\min \; z^T N z$$

s.t. 1 z orthog first evec of N

2 $z^T z = \text{const.}$

$$N v_2 = \lambda_2 v_2$$

$$z = D^{1/2} x$$

$x^* = D^{-1/2} v$ is the relaxed solution

$+$ entries $\Rightarrow S$

$-$ $\Rightarrow T$

① Order x^* by value

for $r = 1, 2, \dots, n-1$

① take first r nodes in ordering $= S$

② compute $NCUT(S, T)_r$

$$S_{opt} = \arg\min_r NCUT(S, T)_r$$

$$NCUT(S_{opt}, T_{opt}) \leq O(\sqrt{OPT})$$

(one side of Cheeger's inequality)