University of Edinburgh
INFR11156: Algorithmic Foundations of Data Science (2019)
Homework 7

Problem 1: Prove that G has exactly k connected components if and only if $\lambda_k = 0$ and $\lambda_{k+1} > 0$.

Solution:

1. First, we will show that if G has exactly k connected components, then $\lambda_k = 0$ and $\lambda_{k+1} > 0$. Let $S_1, S_2, \ldots, S_k$ be the connected components of G, and we define the indicator vector $\mathbf{1}_{S_i} \in \{0,1\}^n$ corresponding to each S_i by
$$(\mathbf{1}_{S_i})_v = \begin{cases} 1 & \text{if } v \in S_i, \\ 0 & \text{otherwise.} \end{cases}$$
Since
$$\mathcal{L} D^{1/2} \mathbf{1}_{S_i} = D^{-1/2}(D-A)D^{-1/2}D^{1/2}\mathbf{1}_{S_i} = D^{-1/2}(D-A)\cdot \mathbf{1}_{S_i} = 0,$$
and $\langle \mathbf{1}_{S_i}, \mathbf{1}_{S_j} \rangle = 0$ for any $1 \le i \ne j \le k$, $D^{1/2}\mathbf{1}_{S_i}$ is an eigenvector of $\mathcal{L}$ corresponding to the eigenvalue $\lambda = 0$. Hence, $\lambda_1 = 0$ has multiplicity at least k, i.e., $\lambda_k = 0$.

   Now suppose for contradiction that $\lambda_{k+1} = 0$ and let x be an eigenvector of λ_{k+1}, linearly independent from the set $\{D^{1/2}\mathbf{1}_{S_i}\}_i$. We have that
$$0 = x^\intercal \mathcal{L} x = \sum_{u \sim v} \left(\frac{x_u}{\sqrt{d_u}} - \frac{x_v}{\sqrt{d_v}} \right)^2.$$
Therefore, we see that $(D^{-1/2}x)_u = (D^{-1/2}x)_v$ holds for every $u \sim v$. Let $y = D^{-1/2}x$ and notice that y is constant on every connected component of G. Therefore $y \in \operatorname{span}\{\mathbf{1}_{S_i}\}_i$ and hence $x \in \operatorname{span}\{D^{1/2}\mathbf{1}_{S_i}\}_i$. This contradicts the choice of x being linearly independent from $\{D^{1/2}\mathbf{1}_{S_i}\}_i$. Therefore, we have that $\lambda_{k+1} > 0$.

2. Now we prove that, if $\lambda_k = 0$ and $\lambda_{k+1} > 0$, then G has exactly k connected components. Let $S_1, S_2, \ldots, S_r$ be the connected components of G and $x_1, x_2, \ldots, x_k$ be linearly independent eigenvectors corresponding to $\lambda = 0$. By the previous part we see that for every $j \le k$, $x_j \in \operatorname{span}\{D^{1/2}\mathbf{1}_{S_i}\}_i$. Since $\{x_j\}$'s are chosen to be linearly independent, we must have $k \le r$. If $r > k$, again by the previous part we can conclude that $0 = \lambda_1 = \cdots = \lambda_k = \lambda_{k+1} = \cdots = \lambda_r$, which contradicts the fact that $\lambda_{k+1} > 0$. Therefore, we have that $k = r$, i.e., G has exactly k connected components.

Problem 2: Prove that, for any connected graph G with diameter α, it holds that
$$\lambda_2 \ge \frac{1}{\alpha \cdot \operatorname{vol}(G)}.$$

Solution: Let $f \in \mathbb{R}^n$, $f \perp D\mathbf{1}$ be such that
$$\lambda_2 = \frac{\sum_{u \sim v}(f_u - f_v)^2}{\sum_u d_u \cdot f_u^2}.$$
Let v_0 be such that $|f_{v_0}| = \max_v |f_v|$. Since $\sum_v f_v d_v = 0$, there exists a vertex u_0 such that $f_{v_0} \cdot f_{u_0} < 0$. Let $\mathcal{P}$ be the shortest path between u_0 and v_0. Since it holds by the Cauchy-Schwarz inequality that
$$|f_{u_0} - f_{v_0}| = \left| \sum_{\{u,v\} \in \mathcal{P}} (f_u - f_v) \right| \le \sum_{\{u,v\} \in \mathcal{P}} |f_u - f_v| \le \sqrt{\sum_{\{u,v\} \in \mathcal{P}} (f_u - f_v)^2} \cdot \sqrt{|\mathcal{P}|},$$

where $|\mathcal{P}|$ is the length of path $\mathcal{P}$, we have

$$\lambda_2 = \frac{\sum_{u\sim v}(f_u - f_v)^2}{\sum_u d_u f_u^2} \geq \frac{\sum_{\{u,v\}\in\mathcal{P}}(f_u - f_v)^2}{\mathrm{vol}(G)\cdot f_{v_0}^2} \geq \frac{(1/|\mathcal{P}|)\cdot(f_{u_0} - f_{v_0})^2}{\mathrm{vol}(G)\cdot f_{v_0}^2} \geq \frac{1}{|\mathcal{P}|\cdot\mathrm{vol}(G)} \geq \frac{1}{\alpha\cdot\mathrm{vol}(G)},$$

Problem 3: Prove that for any graph G, $\lambda_2 \leq n/(n-1)$ and $\lambda_2 = n/(n-1)$ if and only if G is the complete graph.

Solution: We first show that, if G is the complete graph, then $\lambda_2 = n/(n-1)$. Let $v_1,\ldots,v_n$ be the vertices of G. We will construct $n-1$ linearly independent vectors $f^1,\ldots,f^{n-1}$ as follows: for all $1 \leq i \leq n-1$ let

$$\left(f^i\right)_u = \begin{cases} 1 & \text{if } u = v_i, \\ -1 & \text{if } u = v_n, \\ 0 & \text{otherwise.} \end{cases}$$

Since G is the complete graph, it is $(n-1)$-regular and $\mathcal{L} = I - \frac{1}{n-1}A$. We also have that for every $i \leq n-1$, it holds that

$$\mathcal{L}f^i = f^i - \frac{1}{n-1}Af^i = f^i + \frac{1}{n-1}f^i = \frac{n}{n-1}f^i.$$

Hence, each of the f^is is an eigenvector of $\mathcal{L}$, and the eigenvalue $\lambda = n/(n-1)$ has multiplicity $n-1$. Along with $\lambda_1 = 0$ they form the entire spectrum of $\mathcal{L}$. We conclude that $\lambda_2 = \lambda_3 = \cdots = \lambda_n = n/(n-1)$.

For the other implication we can use the result from the lecture notes which says that if G is not the complete graph, then $\lambda_2 \leq 1$. We see that if $\lambda_2 = n/(n-1) > 1$ then G must be the complete graph.

Problem 4: [1] Let G be an undirected and connected d-regular graph, and let $\mathbf{A}$ be its adjacency matrix with the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$.

1. Prove that $\lambda_n = -d$ if and only if G is bipartite.
2. Assume that $G = (V, W, E)$ is bipartite and $\max_{2\leq i\leq n-1}|\lambda_i| = \mu$. For any $S \subseteq V$ and $T \subseteq W$, let $e(S,T)$ be the number of edges between S and T. Prove that

$$e(S,T) \leq \frac{2d|S||T|}{|S|+|T|} + \mu n.$$

Solution:

1. We first prove that $\lambda_n \geq -d$, which is equivalent to prove $d + \lambda_n \geq 0$. Since

$$\begin{aligned} d + \lambda_n &= \min_{x\in\mathbb{R}^n, x\neq 0} \frac{dx^\intercal x + x^\intercal Ax}{x^\intercal x} = \min_{x\in\mathbb{R}^n, x\neq 0} \frac{\sum_u d_u \cdot x_u^2 + \sum_{u\sim v} 2x_u x_v}{x^\intercal x} \\ &= \min_{x\in\mathbb{R}^n, x\neq 0} \frac{\sum_{u\sim v}(x_u + x_v)^2}{x^\intercal x} \geq 0, \end{aligned}$$

we have $\lambda_n \geq -d$ for any graph. Moreover, our calculation shows that $\lambda_n = -d$ if and only if $x_u + x_v = 0$ for any edge $u \sim v$, which is the case if and only if the set $L = \{u : x_u > 0\}$ and $R = \{u : x_u < 0\}$ forms a bipartition of G.

[1] This problem appeared in the last year's AFDS final exam. You should be able to answer questions of a similar level of difficulty in your final exam.

2. Let $S \subseteq V$ and $T \subseteq W$ be two arbitrary subsets. Let $\mathbf{1}_S$ and $\mathbf{1}_T$ be indicator vectors of sets S and T. Assume that $\lambda_1 \geq \ldots \geq \lambda_n$ are the eigenvalues of matrix $\mathbf{A}$ with the corresponding orthonormal eigenvectors $v_1, \ldots v_n$. Then, we can write $\mathbf{1}_S = \sum_{i=1}^n \alpha_i v_i$ and $\mathbf{1}_T = \sum_{i=1}^n \beta_i v_i$, and

$$e(S,T) = \left(\sum_{i=1}^n \alpha_i v_i\right)^\intercal A \left(\sum_{i=1}^n \beta_i v_i\right) = \sum_{i=1}^n \lambda_i \alpha_i \beta_i.$$

Since G is connected, we know that

$$v_1 = \frac{\mathbf{1}_V + \mathbf{1}_W}{\sqrt{n}} = \frac{\mathbf{1}}{\sqrt{n}},$$

and also because G is bipartite, we can write without loss of generality that

$$v_n = \frac{\mathbf{1}_V - \mathbf{1}_W}{\sqrt{n}}.$$

Moreover, it holds that

$$\alpha_1 = \langle \mathbf{1}_S, v_1 \rangle = \frac{|S|}{\sqrt{n}}, \qquad \beta_1 = \langle \mathbf{1}_T, v_1 \rangle = \frac{|T|}{\sqrt{n}},$$

as well as

$$\alpha_n = \langle \mathbf{1}_S, v_n \rangle = \frac{|S|}{\sqrt{n}}, \qquad \beta_n = \langle \mathbf{1}_T, v_n \rangle = -\frac{|T|}{\sqrt{n}}.$$

Hence, we have that

$$\begin{aligned} e(S,T) &= \frac{d|S||T|}{n} + \frac{d|S||T|}{n} + \sum_{i=2}^{n-1} \lambda_i \alpha_i \beta_i \\ &\leq \frac{2d|S||T|}{n} + \mu \sqrt{\sum_{i=2}^{n-1} \alpha_i^2} \sqrt{\sum_{i=2}^{n-1} \beta_i^2} \\ &\leq \frac{2d|S||T|}{n} + \mu \sqrt{|S| \cdot |T|} \\ &\leq \frac{2d|S||T|}{|S|+|T|} + \mu n, \end{aligned}$$

which proves the statement.

Problem 5 (challenging): Let $G = (V,E)$ be an undirected graph with m edges. Prove that, for any fixed k, the number of cycles of length k in G is at most $O(m^{k/2})$.

***Solution*:** Let A be the adjacency matrix of graph G, and $\lambda_1(A), \ldots, \lambda_n(A)$ are the eigenvalues of A. We recall the fact that the trace of A satisfies the following property:

$$\mathrm{tr}(A) = \sum_{i=1}^n \lambda_i(A) = \sum_{i=1}^n A_{i,i}.$$

Hence, the number of cycles of length k in G is at most $\sum_{i=1}^n A_{i,i}^k = \sum_{i=1}^n \lambda_i^k(A)$. Since it holds for any $k \geq 3$ that

$$\left(\sum_{i=1}^n \lambda_i^k(A)\right)^{1/k} \leq \left(\sum_{i=1}^n \lambda_i^2(A)\right)^{1/2} = \left(\sum_{i=1}^n A_{i,i}^2\right)^{1/2} = (2m)^{1/2},$$

we have that the number of cycles in G is at most $O\left(m^{k/2}\right)$. Here, the first inequality holds by the fact that for any $\{x_i\}_{i=1}^n$ and $p \geq q$, it holds that

$$\left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \leq \left(\sum_{i=1}^n |x_i|^q\right)^{1/q}.$$