Applied spectral graph analysis—Solutions

Dario Fasino

1 What is an irreducible matrix?

Theorem 1.1. *A matrix A is irreducible if and only if $\mathcal{G}(A)$ is strongly connected.*

PROOF. Suppose that A is reducible. Apart of a permutation, we can assume that A is already in reduced block form:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ O & A_{22} \end{pmatrix}, \qquad A_{11} \in \mathbb{R}^{n_1 \times n_1}, \quad A_{22} \in \mathbb{R}^{n_2 \times n_2}, \tag{1}$$

with $n_1 + n_2 = n$. Hence, in $\mathcal{G}(A)$ there are no edges connecting the nodes $n_1 + 1, \ldots, n$ with the nodes $1, \ldots, n_1$. As a consequence, there are no walks going from nodes $n_1 + 1, \ldots, n$ to nodes $1, \ldots, n_1$, and the graph is not strongly connected.

Conversely, if $\mathcal{G}(A)$ is not strongly connected then there are two distinct nodes, say i and j, such that there is no walk from i to j. Let $\mathcal{I}$ be the set of all nodes that can be reached by a walk starting from i, and let $\mathcal{J}$ be its complementary set. Note that $i \in \mathcal{I}$ and $j \in \mathcal{J}$. By renumbering nodes, we can suppose that $\mathcal{I} = \{n_1 + 1, \ldots, n\}$ and $\mathcal{J} = \{1, \ldots, n_1\}$. Observe that there are no edges from nodes in $\mathcal{I}$ to nodes in $\mathcal{J}$. Hence A has the form (1), that is, A is reducible. ■

2 A closeness centrality index

Let $\mathcal{G}$ be a graph and let L be its Laplacian matrix. The *resistance distance* between nodes i and j is given by

$$R(i,j) = (e_i - e_j)^T L^+ (e_i - e_j).$$

Define the *current-flow closeness centrality* of node i as

$$C(i) = \frac{1}{n} \sum_{j=1}^{n} R(i,j).$$

1. Prove that the resistance distance is really a distance.

   Answer: The identity $R(i,j) = R(j,i)$ is trivial. The inequality $R(i,j) \geq 0$ follows from positive semidefiniteness of L^+. Moreover,

   $$R(i,j) = 0 \iff e_i - e_j \in \mathrm{Ker}(L^+) \iff e_i = e_j \iff i = j.$$

   In order to prove the triangle inequality, consider the following preliminary result:

   Lemma 2.1. *For all i, j, k we have*

   $$e_j^T L^+ (e_i - e_j) \leq e_k^T L^+ (e_i - e_j) \leq e_i^T L^+ (e_i - e_j).$$

   PROOF. *Suppose $i \neq j$, otherwise the claim is trivial. For notational simplicity, let $v = L^+(e_i - e_j)$. From the relation $Lv = e_i - e_j$ we deduce that v is not constant, otherwise we have $e_i - e_j = Lv = 0$. Let M be the maximum value of v. Then, there exist two distinct nodes p and q such that $v_p = M$, $v_q < M$ and moreover $p \sim q$ (otherwise either v is constant or $\mathcal{G}$ is not connected). If $p \neq i$ then, by looking at the p-th equation of $Lv = e_i - e_j$ we have*

   $$\deg(p)\, v_p - \sum_{\ell : \ell \sim p} v_\ell \leq 0 \quad \rightsquigarrow \quad M \leq \frac{1}{\deg(p)} \sum_{\ell : \ell \sim p} v_\ell < M.$$

Indeed, all terms in the last sum are not larger than M and at least one of them (that is v_q) is strictly smaller than M. Then $p = i$. From $v_k \le v_i$ we have $e_k^T L^+(e_i - e_j) = v_k \le v_i = e_i^T L^+(e_i - e_j)$ for $k = 1, \dots, n$. The other inequality is shown analogously, by arguing on the minimum of v. ■

With the help of the previous lemma we obtain, for any i, j, k,

$$\begin{aligned} R(i,j) + R(j,k) &= (e_i - e_j)^T L^+(e_i - e_j) + (e_j - e_k)^T L^+(e_j - e_k) \\ &\ge (e_i - e_k)^T L^+(e_i - e_j) + (e_i - e_k)^T L^+(e_j - e_k) \\ &= (e_i - e_k)^T L^+(e_i - e_k) \;=\; R(i,k), \end{aligned}$$

and the proof is complete.

2. Prove that
$$C(i) = (L^+)_{ii} + \frac{1}{n}\,\mathrm{trace}(L^+).$$

Answer: by denoting $e = (1, \dots, 1)^T$ we have

$$\begin{aligned} C(i) = \frac{1}{n}\sum_{j=1}^{n}(e_i - e_j)^T L^+(e_i - e_j) &= \frac{1}{n}\sum_{j=1}^{n}\left[(L^+)_{ii} - 2e_i^T L^+ e_j + (L^+)_{jj}\right] \\ &= (L^+)_{ii} - \frac{2}{n}\underbrace{e_i^T L^+ e}_{=0} + \frac{1}{n}\sum_{j=1}^{n}(L^+)_{jj} \\ &= (L^+)_{ii} + \frac{1}{n}\,\mathrm{trace}(L^+). \end{aligned}$$

Indeed $\mathrm{Ker}(L) = \mathrm{Ker}(L^+)$, *whence* $L^+ e = 0$.

Remark 2.2. *It is not difficult to see that the function $\|v\|_* = \sqrt{v^T L^+ v}$ is a norm on the orthogonal complement of* $\mathrm{Ker}(L^+)$*, that is, the set $\{v : e^T v = 0\}$. Hence, if we could define $R(i,j) = \|e_i - e_j\|_*$ then the first part of the exercise becomes immediate. Unfortunately this is not the case, because of the square root. The resistance distance has been introduced in the paper*

D. J. Klein, M. Randić. *Resistance distance.* J. of Mathematical Chemistry, *12 (1993), 81–95.*

In that paper (which is freely available on the Internet) the authors prove that $R(i,j)$ is a distance function by using a very brief argument which exploits the electric circuit analogy.

3 HITS scores on a modified network

Suppose that in the digraph $\mathcal{G}$ node j has no inlinks ($\deg^-(j) = 0$). Let i ($i \ne j$) be a node having positive hub score ($h_i > 0$). We modify $\mathcal{G}$ by adding the oriented edge (i,j). Prove the strict inequality

$$\forall \ell \ne i, \quad \frac{\bar{h}_\ell}{h_\ell} < \frac{\bar{h}_i}{h_i},$$

where $\bar{h}_i$ is the updated hub score of node i, which shows that, for any fixed normalization, the hub score of node i gets the largest relative increase.

Answer: Let H and $\bar{H}$ denote the hub matrix before and after the addition of the new edge. Owing to the hypotheses, simple computations show that

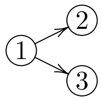
$$\bar{H} = H + e_i e_i^T.$$

Recall that a hub matrix is nonnegative, symmetric, and positive semidefinite. Hence $\rho(H) = \lambda_{\max}(H)$, the hub vector being an associated eigenvector. As a consequence,

$$\rho(\bar{H}) = \sup_{v \ne 0}\frac{v^T \bar{H} v}{v^T v} \ge \frac{h^T \bar{H} h}{h^T h} = \frac{h^T(H + e_i e_i^T)h}{h^T h} = \frac{h^T H h + h_i^2}{h^T h} > \frac{h^T H h}{h^T h} = \rho(H).$$

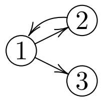
Hence $\rho(\bar{H}) > \rho(H)$. The claim follows immediately from Dietzenbacher's theorem.

Remark 3.1. *The following example shows that the hypothesis $h_i > 0$ is necessary to conclude that $\rho(\bar{H}) > \rho(H)$:*



$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad H = AA^T = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence $\rho(H) = 2$ and $h = (1,0,0)^T$. Note that HITS behaves fairly on this graph. Adding the edge $(2,1)$ we have



$$\bar{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \bar{H} = \bar{A}\bar{A}^T = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence $\rho(\bar{H}) = 2$ and $\bar{h} = (1,0,0)^T$. Thus, any solution of this exercise that does not exploit the assumption $h_i > 0$ is not entirely correct.

4 A workload balancing process

To every node of a graph we allocate an initial *workload* $x_i(0) \in \mathbb{R}$, $i = 1, \dots, n$. The workload is then updated at discrete time instants acording to the following rule: At every time step, every node sends a fixed fraction $0 < \alpha < 1$ of its current workload to its neighbors (and receives their workload fractions):

$$x_i(t+1) = x_i(t) + \alpha \sum_{j : i \sim j} (x_j(t) - x_i(t)).$$

1. Prove that $\sum_{i=1}^n x_i(t)$ is constant over time.

   Answer: Firstly, observe that the vector $x(t) = (x_1(t), \dots, x_n(t))^T$ fulfills

   $$x(t+1) = x(t) - \alpha L x(t). \tag{2}$$

   Now, recalling that L is symmetric and $Le = 0$ we obtain

   $$\sum_{i=1}^n x_i(t+1) = e^T x(t+1) = e^T x(t) - \alpha \underbrace{e^T L}_{=0} x(t) = e^T x(t) = \sum_{i=1}^n x_i(t).$$

2. Describe the stationary workload distributions.

   Answer: we have $x(t+1) = x(t) \;\Leftrightarrow\; Lx(t) = 0 \;\Leftrightarrow\; x(t) \in \mathrm{Ker}(L)$. The latter holds if and only if $x(t)$ is a multiple of $e = (1, \dots, 1)^T$, the uniform workload.

3. Find a necessary and sufficient condition (in terms of the Laplacian spectrum) so that any initial workload distribution evolves toward a stationary distribution.

   Answer: The recurrence (2) is exactly the power method applied to the matrix $I - \alpha L$. The sought condition is then equivalent to the requirement that $1 = \lambda_1(I - \alpha L)$ is a dominant eigenvalue, that is, $-1 < \lambda_i(I - \alpha L) < 1$ for $i = 2, \dots, n$. The latter is equivalent to the condition $0 < \lambda_i(L) < 2/\alpha$ for $i = 2, \dots, n$, that is, the graph is connected and $\rho(L) < 2/\alpha$.

Remark 4.1. *In the third part of this exercise, the inequality $\rho(I - \alpha L) < 1$ implies $x(t) \to 0$ as $t \to \infty$, which is not the desired behavior.*