Random Walks in Graphs

Thomas Bonald
thomas.bonald@telecom-paristech.fr

April 2018

Consider a weighted, undirected graph G. The weights are non-negative and correspond to the strengths of the links between nodes. The graph is assumed to be connected and without self-loops. We denote by A the weighted adjacency matrix of the graph, i.e., A_{ij} is the weight of the edge between nodes i and j, if any, and is equal to 0 otherwise. We denote by 1 the vector of ones and by $d = A1$ the vector of node weights (i.e., sums of the weights of incident edges); for unit edge weights, d is the vector of node degrees.

We shall see that a number of fundamental tasks in graph analysis, like ranking and clustering nodes, can be derived from random walks in the graph. The analogy with electric networks, where electrons move at random through the transistors, will prove very useful. These notes are mainly based on [1, 2, 3, 4].

Contents

1 Random walk

Consider a random walk in the graph G with a probability of moving from node i to node j equal to A_{ij}/d_i. Let $X_0, X_1, X_2, \ldots$ be the sequence of nodes visited by the random walk. This defines an irreducible Markov chain on $\{1, \ldots, n\}$ with transition matrix $P = D^{-1}A$. We have for all $t \geq 0$:

$$\forall i = 1, \ldots, n, \quad \mathrm{P}(X_{t+1} = i) = \sum_{j=1}^{n} \mathrm{P}(X_t = j) P_{ji}.$$

The stationary distribution of the Markov chain is the unique vector π such that $\pi^T 1 = 1$ and:

$$\forall i = 1, \ldots, n, \quad \pi_i = \sum_{j=1}^{N} \pi_j P_{ji}, \tag{1}$$

that is

$$\pi^T = \pi^T P.$$

In particular, π is a left eigenvector of P for the eigenvalue 1. It can be easily verified that $\pi \propto d$, i.e., nodes are visited in proportion to their weights. That the stationary distribution π is unique, given by $\pi = d/|d|$ where $|d| = d^T 1$, follows from the fact that the eigenvalue 1 of P is simple, as proved in section 3.

Reversibility. A Markov chain is reversible if for any states i, j the frequency of transitions from state i to state j is equal to the frequency of transitions from state j to state i, that is

$$\forall i, j = 1, \ldots, n, \quad \pi_i P_{ij} = \pi_j P_{ji}.$$

These are the equations of *local balance*, that imply the equations of *global balance* (1). For a reversible Markov chain, the frequency of any sequence of states $i_1, \ldots, i_\ell$ is the same in both directions of time, that is

$$\pi_{i_1} P_{i_1 i_2} \ldots P_{i_{\ell-1} i_\ell} = \pi_{i_\ell} P_{i_\ell i_{\ell-1}} \ldots P_{i_2 i_1}. \tag{2}$$

That the above random walk in the graph G defines a reversible Markov chain with stationary distribution $\pi \propto d$ is a simple consequence of the equality:

$$d_i P_{ij} = A_{ij} = A_{ji} = d_j P_{ji}.$$

Return time. Let $\mathrm{P}_i = \mathrm{P}(\cdot | X_0 = i)$ and E_i the corresponding expectation. We denote by $\sigma_i = \mathrm{E}_i(\tau_i^+)$ the mean return time to node i, with $\tau_i^+ = \min\{t \geq 1 : X_t = i\}$. Since π_i is the frequency of visits to node i, we have

$$\sigma_i = \frac{1}{\pi_i}. \tag{3}$$

This will be proved in section 4.

Hitting time, commute time, escape probability. Let $H_{ij} = \mathrm{E}_i(\tau_j)$ be the mean hitting time of node j from node i, with $\tau_i = \min\{t \geq 0 : X_t = i\}$. Observe that $H_{ij} = 0$ for $j = i$. We denote by $\rho_{ij} = H_{ij} + H_{ji}$ the mean commute time between nodes i and j. The escape probability from node i to node j is $e_{ij} = \mathrm{P}_i(\tau_j < \tau_i^+)$, for any $i \neq j$. This is the probability of hitting node j before returning to node i.

Proposition 1 *We have:*

$$\rho_{ij} = \frac{1}{\pi_i e_{ij}}.$$

Proof. Let $\tau_{ij} = \min\{t > \tau_i : X_t = j\}$ be the hitting time of node j after having visited node i. We have:

$$\begin{aligned}
\rho_{ij} &= \mathrm{E}_i(\tau_{ji}), \\
&= \mathrm{E}_i(\tau_i^+) + \mathrm{E}_i(\tau_{ji} - \tau_i^+), \\
&= \mathrm{E}_i(\tau_i^+) + \mathrm{E}_i((\tau_{ji} - \tau_i^+) 1_{\{\tau_{ji} > \tau_i^+\}}), \\
&= \mathrm{E}_i(\tau_i^+) + \mathrm{P}_i(\tau_{ji} > \tau_i^+) \mathrm{E}_i(\tau_{ji} - \tau_i^+ | \tau_{ji} > \tau_i^+), \\
&= \mathrm{E}_i(\tau_i^+) + \mathrm{P}_i(\tau_j > \tau_i^+) \mathrm{E}_i(\tau_{ji}), \\
&= \mathrm{E}_i(\tau_i^+) + (1 - e_{ij}) \rho_{ij}.
\end{aligned}$$

The result then follows from (3). □

It follows from Proposition 1 that $\pi_i e_{ij} = \pi_j e_{ji}$ for all $i \neq j$. This is in fact a direct consequence of the reversibility of the Markov chain, on applying (2) to each direct path from node i to node j (that is, without return to node i).

2 Laplacian matrix

Let $D = \mathrm{diag}(d)$. The Laplacian matrix is defined by

$$L = D - A.$$

This is the discrete version of the usual Laplace operator.

Heat equation. Consider some strict subset S of $\{1, \dots, n\}$ and assume that the temperature of each node $i \in S$ is set at some fixed value T_i. We are interested in the evolution of the temperatures of the other nodes. Heat exchanges occur through each edge of the graph proportionally to the temperature difference between the corresponding nodes, with a coefficient equal to the weight of the edge, thus interpreted as *thermal conductivity*. Then,

$$\forall i \notin S, \quad \frac{dT_i}{dt} = \sum_{j=1}^{n} A_{ij}(T_j - T_i),$$

that is

$$\forall i \notin S, \quad \frac{dT_i}{dt} = -(LT)_i,$$

where T is the vector of temperatures. This is the heat equation in discrete space. At equilibrium, T satisfies Laplace's equation:

$$\forall i \notin S, \quad (LT)_i = 0, \tag{4}$$

We say that the vector T is *harmonic*. With the boundary conditions T_i for all $i \in S$, this defines a Dirichlet problem. Observing that $D^{-1}L = I - P$, Laplace's equation can be written equivalently

$$\forall i \notin S, \quad T_i = (PT)_i. \tag{5}$$

Proposition 2 (Uniqueness) *There is at most one solution to the Dirichlet problem.*

Proof. Since P is a stochastic matrix, it follows from (5) that the temperature of node i is the weighted average of the temperatures of its neighbors.

We first prove that the maximum and the minimum of the vector T are achieved on the boundary, that is for nodes in S. Let i be any node such that T_i is maximum. If $i \notin S$, it follows from (5) that T_j is maximum for all neighbors j of i. If no such node belongs to S, we apply again this argument until we reach a node in S. Such a node exists because the graph is connected. It achieves the maximum of the vector T. The proof is similar for the minimum.

Now consider two solutions T, T' to Laplace's equation. Then $\delta = T' - T$ is a solution of Laplace's equation with the boundary condition $\delta_i = 0$ for all $i \in S$. We deduce that $\delta_i = 0$ for all i (because both the maximum and the minimum are equal to 0), that is $T' = T$. □

Now let $\tau_S = \min\{t \geq 0 : X_t \in S\}$ be the hitting time of the set S. Define:

$$P^S_{ij} = \mathrm{P}_i(\tau_j = \tau_S)$$

This is the probability that the random walker first hits S in node j when starting from node i. Observe that P^S is a stochastic matrix. In particular, $P^S_{ij} = \delta_{ij}$ (Kronecker delta) for all $i \in S$. By first-step analysis, we have:

$$\forall i \notin S, \quad P^S_{ij} = \sum_{k=1}^{n} P_{ik} P^S_{kj}. \tag{6}$$

Proposition 3 (Existence) *The solution to the Dirichlet problem is*

$$\forall i \notin S, \quad T_i = \sum_{j \in S} P^S_{ij} T_j. \tag{7}$$

Proof. The vector T defined by (7) satisfies:

$$\forall i \notin S, \quad \sum_{j=1}^{n} P_{ij} T_j = \sum_{j=1}^{n} P_{ij} \sum_{k \in S} P^S_{jk} T_k = \sum_{k \in S} P^S_{ik} T_k = T_i,$$

where we have used (6). Thus T satisfies (5). The proof then follows from Proposition 2. □

3 Spectral analysis

The Laplacian matrix L is positive semi-definite:

Proposition 4 *We have:*

$$\forall v \in \mathbb{R}^n, \quad v^T L v = \sum_{i<j} A_{ij} (v_i - v_j)^2.$$

Proof. For all $v \in \mathbb{R}^n$,

$$\begin{aligned} v^T L v &= v^T (D - A) v, \\ &= \sum_{i,j=1}^{n} d_i v_i^2 - \sum_{i,j=1}^{n} v_j A_{ij} v_i, \\ &= \sum_{i,j=1}^{n} A_{ij} v_i (v_i - v_j), \\ &= \frac{1}{2} \sum_{i,j=1}^{n} A_{ij} (v_i - v_j)^2, \\ &= \sum_{i<j} A_{ij} (v_i - v_j)^2. \end{aligned}$$

□

The spectral theorem yields

$$L = V \Lambda V^T, \tag{8}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of eigenvalues of L, with $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, and $V = (v_1, \ldots, v_n)$ is the matrix of corresponding eigenvectors, with $V^T V = I$. In view of Proposition 4, $v^T L v = 0$ implies $v \propto 1$ (recall that the graph is connected) so that $\lambda_1 = 0 < \lambda_2$ and $v_1 = 1/\sqrt{n}$. This proves in turn that the eigenvalue 1 of P is simple, since $Pv = v$ if and only if $Lv = 0$.

A mechanical system. Consider n points of unit mass where points i and j are linked by a spring of stiffness A_{ij} following Hooke's law (i.e., force proportional to the distance). Now if the points are located according to some vector $v \in \mathbb{R}^n$ along a line, the potential energy accumulated in the springs is:

$$\frac{1}{2} \sum_{i<j} A_{ij} (v_i - v_j)^2,$$

that is $\frac{1}{2} v^T L v$ in view of Proposition 4.

We impose that the moment of inertia of the system (for a rotation around the origin) is equal to 1, that is $v^T v = 1$. Clearly, the vector v that minimizes the potential energy is $v = v_1$ (the corresponding potential energy is null). Now if we impose $1^T v = 0$, meaning that the centre of mass is at the origin, we obtain $v = v_2$ and $v^T L v = \lambda_2$, so that the eigenvalue λ_2 corresponds to twice the minimum value of potential energy. This is a consequence of the following characterization of the spectrum of the Laplacian.

Theorem 1 *For all $k = 1, \dots, n$,*

$$\lambda_k = \min_{\substack{v : v^T v = 1 \\ v_1^T v = 0, \dots, v_{k-1}^T v = 0}} v^T L v, \tag{9}$$

the minimum being attained for $v = v_k$.

Proof. Let $v \in \mathbb{R}^n$ such that $v^T v = 1$. The vector $x = V^T v$, giving the coordinates $x_1 = v_1^T v, \dots, x_n = v_n^T v$ of v in the basis of eigenvectors, satisfies:

$$x^T \Lambda x = v^T V \Lambda V^T v = v^T L v \quad \text{and} \qquad x^T x = v^T V V^T v = 1,$$

so that the optimization problem (9) is equivalent to:

$$\min_{\substack{x : x^T x = 1 \\ x_1 = 0, \dots, x_{k-1} = 0}} x^T \Lambda x.$$

The result then follows from the equality:

$$x^T \Lambda x = \sum_{i=1}^{n} \lambda_i x_i^2.$$

□

Assume the system has a uniform circular motion around its center of mass, taken as the origin, so that $1^T v = 0$, with $v \neq 0$. Let ω be the angular velocity of the system. By Newton's second law of motion, the system is in equilibrium if and only if

$$\forall i = 1, \dots, n, \quad \sum_{j=1}^{n} A_{ij}(v_j - v_i) = -v_i \omega^2,$$

that is

$$Lv = \omega^2 v.$$

This means that v is an eigenvector of L (different from v_1 since $1^T v = 0$) with eigenvalue ω^2. In particular, the only possible values of angular velocity are $\sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}$. Moreover,

$$v^T L v = v^T v \omega^2,$$

where $v^T v$ is the moment of inertia of the system. For a unit moment of inertia $v^T v = 1$, we obtain:

$$v^T L v = \omega^2.$$

Thus the eigenvalues $\lambda_2, \dots, \lambda_n$ are the squares of the possible values of angular velocities (the first eigenvalue $\lambda_1 = 0$ corresponding to the absence of rotation) and the eigenvectors $v_2, \dots, v_n$ are the corresponding equilibriums with unit moments of inertia.

4 Graph embedding

Let $L^+ = V \Lambda^+ V^T$ be the pseudo-inverse of L, with $\Lambda^+ = \operatorname{diag}\left(0, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_n}\right)$.

Proposition 5 *We have:*

$$L L^+ = L^+ L = I - \frac{1 1^T}{n}.$$

Proof. The proof follows from the fact that $v_1 = 1/\sqrt{n}$ on observing that

$$LL^+ = L^+L = V\Lambda^+\Lambda V^T = \sum_{k=2}^{n} v_k v_k^T = I - v_1 v_1^T.$$

□

Let:

$$X = \sqrt{|d|}Z(I - \pi 1^T),$$

where

$$Z = \sqrt{\Lambda^+}V^T.$$

Consider the embedding $X = (x_1, \ldots, x_n)$ of the graph, where node i is represented by the vector $x_i \in \mathbb{R}^n$. Observe that the first row of X is null so that only $n-1$ coordinates are informative. The embedding X is a shifted, rescaled version of Z so that:

$$X\pi = 0.$$

The Gram matrix of Z is the pseudo-inverse of the Laplacian L:

$$Z^T Z = V\Lambda^+ V^T = L^+.$$

We deduce the Gram matrix of X,

$$G = X^T X = |d|(I - 1\pi^T)L^+(I - \pi 1^T). \tag{10}$$

Observe that

$$G\pi = 0. \tag{11}$$

For any matrix M, we denote by $d(M)$ the diagonal matrix with the same diagonal as that of M.

Random walk. By one-step analysis, the mean hitting time of node j from node i satisfies:

$$H_{ij} = \begin{cases} 0 & \text{if } i = j, \\ 1 + \sum_{k=1}^{n} P_{ik} H_{kj} & \text{otherwise.} \end{cases} \tag{12}$$

We deduce that the matrix $(I - P)H - 11^T$ is diagonal. Equivalently, the matrix $LH - d1^T$ is diagonal.

Lemma 1 *There is at most one matrix H such that $d(H) = 0$ and the matrix $LH - d1^T$ is diagonal.*

Proof. Let H, H' be two such matrices and $\Delta = H - H'$. We have $L\Delta = 0$ so that each column of Δ is either null or proportional to 1. Since $d(\Delta) = 0$, we get $\Delta = 0$, that is $H' = H$. □

Theorem 2 *We have:*

$$H = 11^T d(G) - G, \tag{13}$$

where $G = X^T X$ is the Gram matrix of X.

Proof. Using the fact that $L1 = 0$, the matrix H defined by (13) satisfies:

$$\begin{aligned} LH &= -LG, \\ &= -|d|L(I - 1\pi^T)L^+(I - \pi 1^T), \\ &= -|d|LL^+(I - \pi 1^T), \\ &= -|d|(I - \frac{11^T}{n})(I - \pi 1^T), \\ &= -|d|(I - \pi 1^T), \\ &= -|d|I + d1^T, \end{aligned} \tag{14}$$

so that the matrix $LH - d1^T$ is diagonal. Since $d(H) = 0$, the proof follows from Lemma 1. □

Observe that the mean return time to node i satisfies:

$$\sigma_i = 1 + \sum_{j=1}^{n} P_{ij} H_{ji},$$

so that the corresponding vector σ forms the diagonal of the matrix $PH + 11^T$. The following result then proves (3).

Corollary 1 *We have:*

$$d(PH + 11^T) = \mathrm{diag}(\pi)^{-1}.$$

Proof. In view of (14), $(I - P)H = -\mathrm{diag}(\pi)^{-1} + 11^T$, and the result follows on observing that $d(H) = 0$. □

Let h_i be the mean hitting time of node i in stationary regime (that is, starting from a node chosen at random from the stationary distribution):

$$h_i = \sum_{j=1}^{n} \pi_j h_{ji}.$$

The corresponding vector h satisfies $h^T = \pi^T H$. In view of Theorem 2 and equation (11),

$$h^T = 1^T d(G),$$

In particular,

$$H = 1h^T - G,$$

that is

$$x_i^T x_j = h_j - H_{ij} = h_i - H_{ji},$$

and, since $h_{ii} = 0$,

$$||x_i||^2 = h_i.$$

In particular, the mean commute time between nodes i and j is given by:

$$\rho_{ij} = H_{ij} + H_{ji} = ||x_i - x_j||^2.$$

5 Electric networks

Consider the electric network induced by the graph, with a resistor of conductance A_{ij} between nodes i and j. We denote by 1_i the unit vector on component i.

Effective conductance, effective resistance. For any distinct nodes s, t, assume the electric potentials of s and t are set to 1 and 0, respectively. Let U_i be the electric potential of any node i. We have $U_s = 1$ and $U_t = 0$. By Ohm's law, the current that flows from i to j is

$$A_{ij}(U_i - U_j).$$

By Kirchoff's law, the net current at any node $i \neq s, t$ is null. We get

$$\sum_{j=1}^{n} A_{ij}(U_i - U_j) = 0,$$

that is $(LU)_i = 0$. Thus the vector of electric potentials U is harmonic. Moreover, $(LU)_s + (LU)_t = 0$, so that $LU = \alpha(1_s - 1_t)$ for some constant α, equal to the current flowing from s to t.

Proposition 6 *We have:*

$$U_i = \frac{(x_i - x_t)^T (x_s - x_t)}{||x_s - x_t||^2}.$$

Proof. In view of Proposition 5,

$$\left(I - \frac{11^T}{n}\right) U = L^+ L U = \alpha L^+ (1_s - 1_t),$$

that is

$$U = \alpha L^+ (1_s - 1_t) + \beta 1,$$

with $\beta = 1^T U/n$. We obtain:

$$\alpha z_s^T (z_s - z_t) + \beta = 1,$$
$$\alpha z_t^T (z_s - z_t) + \beta = 0,$$

so that

$$\alpha = \frac{1}{||z_s - z_t||^2}, \quad \beta = -\frac{z_t^T (z_s - z_t)}{||z_s - z_t||^2}$$

Finally,

$$U_i = \frac{(z_i - z_t)^T (z_s - z_t)}{||z_s - z_t||^2},$$

and the proof follows from the fact that $x_i - x_j = \sqrt{|d|}(z_i - z_j)$ for all $i, j = 1, \ldots, n$. □

The current α flowing from s to t, which is the current induced by a unit electric potential, is called the *effective conductance* between s and t. We have:

$$\alpha = \frac{1}{||z_s - z_t||^2} = \frac{|d|}{\rho_{st}},$$

so that the effective conductance between s and t is proportional to $1/\rho_{st}$, the inverse of the mean commute time of the random walk between s and t. Equivalently, the mean commute time ρ_{st} between nodes s and t can be interpreted as the *effective resistance* between s and t in the electric network, in some arbitrary unit.

Thompson's principle. The energy dissipation through any transistor is the product of voltage and current (both in absolute value), that is

$$A_{ij}(U_j - U_i)^2$$

between nodes i and j. We obtain the total energy dissipation:

$$E = \frac{1}{2} \sum_{i,j=1}^{n} A_{ij}(U_j - U_i)^2.$$

Thompson's principle states that the potential vector U minimizes energy dissipation. Taking the derivative in U_i, we obtain:

$$\sum_{j=1}^{n} A_{ij}(U_j - U_i) = 0,$$

that is $(LU)_i = 0$, which is Laplace's equation.

Interpretation of voltage and current. Observe that the electric potential is the solution to the heat equation with $T_s = 1$ and $T_t = 0$. In view of (7), we have $U_i = P_{is}^S$, i.e., the electric potential of any node is the probability that the random walk reaches node s before node t. Thus everything happens as if each electron were a random walker in the graph. We shall see that the current between two nodes can be interpreted as the net flow of electrons between these two nodes.

For convenience, we consider positive particles starting from node s and captured by node t (thus in the direction of the current) instead of electrons starting from node t and captured by node s, but the interpretation is exactly the same. Consider the path of a particule starting from node s before it is captured by node t. Let N_i be the mean number of times it visits node i before being captured by node t. We take the initial state into account in the number of visits so that, by Proposition 1,

$$N_s = \frac{1}{e_{st}} = \pi_s \rho_{st},$$

while $N_t = 0$. For any $i \neq s, t$, we have by one-step analysis,

$$N_i = \sum_{j=1}^{n} P_{ji} N_j.$$

Using the local balance equation $\pi_i P_{ij} = \pi_j P_{ji}$, we get

$$\frac{N_i}{\pi_i} = \sum_{j=1}^{n} P_{ij} \frac{N_j}{\pi_j}.$$

We deduce that the vector U defined by

$$U_i = \frac{N_i}{\pi_i \rho_{st}}$$

is harmonic, with $U_s = 1$ and $U_t = 0$. This is the electric potential. The net current from node i to node j is

$$A_{ij}(U_i - U_j) = \frac{1}{\rho_{st}} \left(\frac{N_i}{\pi_i} A_{ij} - \frac{N_j}{\pi_j} A_{ji} \right) = \frac{|d|}{\rho_{st}} (N_i P_{ij} - N_j P_{ji}),$$

This is the net frequency of particle moving from node i to node j, with a flow of particles entering the network at node s at rate

$$\alpha = \frac{|d|}{\rho_{st}},$$

which is the current flowing from node s to node t.

General solution. Now consider the case where the electric potential of node s is set to 1 while those of K other nodes, say $t_1, \ldots, t_K$, are set to 0. The following result extends Proposition 6.

Proposition 7 *We have:*

$$U_i = \sum_{k=1}^{K} \alpha_k (x_i - x_{t_l})^T (x_s - x_{t_k}),$$

where l is an arbitrary element of $\{1, \ldots, k\}$ and the vector $\alpha = (\alpha_1, \ldots, \alpha_K)^T$ is the unique solution to the equation $M\alpha = |d|1$, with M the Gram matrix of the vectors $(x_s - x_{t_1}, \ldots, x_s - x_{t_K})$.

Proof. Let α_k be the current going out of node t_k, for $k = 1, \ldots, K$. Then $\sum_{k=1}^{K} \alpha_k$ is the current entering node s and

$$LU = \sum_{k=1}^{K} \alpha_k (1_s - 1_{t_k}).$$

By Proposition 5,

$$\left(I - \frac{11^T}{n}\right) U = \sum_{k=1}^{K} \alpha_k L^+ (1_s - 1_{t_k}),$$

that is

$$U = \sum_{k=1}^{K} \alpha_k L^+ (1_s - 1_{t_k}) + \beta 1,$$

with $\beta = 1^T U/n$. We get:

$$\sum_{k=1}^{K} \alpha_k z_s^T (z_s - z_{t_k}) + \beta = 1,$$
$$\sum_{k=1}^{K} \alpha_k z_{t_l}^T (z_s - z_{t_k}) + \beta = 0, \quad l = 1, \ldots, K.$$

In particular,

$$\sum_{k=1}^{K} \alpha_k (z_s - z_{t_l})^T (z_s - z_{t_k}) = 1, \quad l = 1, \ldots, K,$$

so that α is the unique solution to the equation $M\alpha = |d|1$ (recall that $x_j - x_i = \sqrt{|d|}(z_j - z_i)$ for all i, j). The result then follows easily. □

Similarly, the electric potential U is the solution to the heat equation with $U_s = 1$ and $U_{t_1}, \ldots, U_{t_K} = 0$. It follows from (7) that $U_i = P^S_{is}$, the probability that a random walk starting from i hits the set $S = \{s, t_1, \ldots, t_K\}$ in s. Thus applying Proposition 7 to each $s \in S$ provides the full matrix P^S and thus the solution for any boundary condition. Specifically, setting the electric potential U_i of node i, for each $i \in S$, yields the solution:

$$\forall i \notin S, \quad U_i = \sum_{j \in S} P^S_{ij} U_j.$$

6 Applications

Finally, we show how to apply previous results to problems of node ranking and clustering. The first step consists in computing the embedding of the graph, $X = (x_1, \ldots, x_n)$:

Embedding

Parameter: k, dimension of the embedding

1. Check that the graph is connected
2. Form the Laplacian $L = D - A$
3. Compute $v_1, \ldots, v_k$, the k eigenvectors of L associated with the lowest eigenvalues, $\lambda_1 \leq \ldots \leq \lambda_k$
4. Compute $Z = \text{diag}\left(\frac{1}{\sqrt{\lambda_2}}, \ldots, \frac{1}{\sqrt{\lambda_k}}\right)(v_2, \ldots, v_k)^T$
5. Return $X = \sqrt{|d|} Z (I - \pi 1^T)$ where $d = D1$ and $\pi = d/|d|$

Ranking. A first way to rank nodes is to consider their *centrality*, in terms of mean hitting time: the more central the node, the shorter time it takes on average for a random walk to hit this node. By the results of section 4, we get the following ranking, the most central nodes appearing first:

Centrality

Output: nodes in increasing order of $||x_i||^2$

In practice, it is often interesting to rank nodes relative to another node. We then rank nodes with respect to their *local centrality*, defined as the mean hitting time from the node of interest. This approach easily extends to a set of nodes. By the results of section 4, we get:

Local centrality

Input: s, node of interest
Output: nodes in increasing order of $x_i^T(x_i - x_s)$

It may also be interesting to include, in addition to the node of interest, a repulsive node. We can then rank nodes with respect to their *directional centrality*, corresponding to the probability to hit the node of interest before the repulsive node (which can be interpreted as an electric potential in view of the results of section 5). Again, the approach easily extends to a set of repulsive nodes.

Directional centrality

Input: s, node of interest; t, repulsive node
Output: nodes in increasing order of $x_i^T(x_t - x_s)$

Clustering. For clustering a set of points in some Euclidian space, a classical algorithm consists in grouping these points into K clusters so that the sum of the square distances of each point to the centroid of its cluster is minimized. Specifically, we look for the partition $C_1, \ldots, C_K$ of $\{1, \ldots, n\}$ that minimizes:

$$J = \sum_{k=1}^{K} \sum_{i \in C_k} ||x_i - \mu_k||^2,$$

where

$$\mu_k = \frac{1}{|C_k|} \sum_{i \in C_k} x_i.$$

This is a combinatorial optimization problem that is known to be NP-hard. Only approximate solutions can be expected for a large number of points. Observe that, for some fixed values of $\mu_1, \ldots, \mu_K$, the cost function J is minimized by assigning each point to its closest cluster k, in terms of distance to the corresponding centroid μ_k. But this in turn changes the values of the centroids $\mu_1, \ldots, \mu_K$. This yields the following alternating optimization algorithm, known as K-means:

K-means algorithm

Input: K, number of clusters

Init $\mu_1, \ldots, \mu_K$ arbitrarily
Repeat until convergence:

- for each k, $C_k \leftarrow$ closest points of μ_k
- for each k, $\mu_k \leftarrow$ centroid of C_k

Output: Clusters $C_1, \ldots, C_K$

Observe that the cost function J decreases at each step, so that the algorithm converges in finite time (because there is a finite number of partitions). The output of the algorithm is a local minimum, which can be far from optimal, depending on the initial values of $\mu_1, \ldots, \mu_K$. In practice, these are chosen at random and the output is the best solution found after several independent runs of the algorithm.

Proposition 8 *Let μ be the centroid of n vectors $x_1, \ldots, x_n$. Then,*

$$\sum_{i=1}^{n} ||x_i - \mu||^2 = \frac{1}{2n} \sum_{i,j=1}^{n} ||x_i - x_j||^2.$$

Proof. We have

$$\sum_{i=1}^{n} ||x_i - \mu||^2 = \sum_{i=1}^{n} x_i^T (x_i - \mu) = \sum_{i=1}^{n} ||x_i||^2 - \frac{1}{n} \sum_{i,j=1}^{n} x_i^T x_j = \frac{1}{2n} \sum_{i,j=1}^{n} ||x_i - x_j||^2.$$

□

In view of Proposition 8, we have

$$J = \sum_{k=1}^{K} \frac{1}{2|C_k|} \sum_{i,j \in C_k} ||x_i - x_j||^2.$$

Thus the cost function J can be interpreted, up to a factor $n/2$, as the mean square distance of a random point to another random point of the same cluster. In view of the results of section 4, the best clustering for the cost function J is that minimizing the mean commute time of the random walk between a random node and another node taken uniformly at random in the same cluster.

References

[1] P. Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues.* Springer Science & Business Media, 2013.

[2] F. R. Chung. *Spectral graph theory.* American Mathematical Soc., 1997.

[3] L. Lovász. Random walks on graphs. *Combinatorics, Paul Erdos is eighty*, 1993.

[4] P. Snell and P. Doyle. Random walks and electric networks. *Free Software Foundation*, 2000.