Graph mining
SD212
Graph embedding

Thomas Bonald

2018 – 2019

These lecture notes introduce graph embedding, a technique consisting in transforming graph data into vector data. Specifically, each node is represented by a vector of low dimension. This embedding can in turn be used to apply classical learning techniques, either semi-supervised (with labels attached to some nodes) or unsupervised (like ranking or clustering nodes). We here present a classical appraoch based on the spectral decomposition of the Laplacian matrix. The analogy with various fields of physics, like thermodynamics and mechanics, will emerge naturally. For the interpretation in terms of electrical networks, we refer the reader to [5]. These notes are mainly based on [1, 2, 3]. See [4] for an overview on spectral clustering.

1 Notion of embedding

Consider a weighted, undirected graph $G = (V, E)$ of n nodes and m edges. Without loss of generality, we assume that $V = \{1, \dots, n\}$. 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 $w = A1$ the vector of node weights (i.e., sums of the weights of incident edges); for unit edge weights, w is the vector of node degrees.

We aim at representing the graph in some Euclidian space of low dimension, say $\mathbb{R}^k$ with $k << n$. Specifically, each node $i \in V$ is represented by some vector $x_i \in \mathbb{R}^k$. The structure of the graph must be encoded in its representation $x_1, \dots, x_n$ in the sense that two "close" nodes i, j in the graph should correspond to two "close" vectors x_i, x_j in the embedding space.

2 Discrete-time 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}/w_i. This defines a Markov chain on $\{1, \dots, n\}$ with transition matrix $P = D^{-1}A$, where $D = \text{diag}(w)$. Denoting the distribution of the random walk at time t as a row vector π_t, we get:

$$\pi_t = \pi_{t-1} P, \tag{1}$$

so that $\pi_t = \pi_0 P^t$, where π_0 is the initial distribution. If the graph is connected and not bipartite, the following limit exists and is unique:

$$\pi = \lim_{t \to +\infty} \pi_t. \tag{2}$$

This is the stationary distribution, which satisfies the balance equations:

$$\pi = \pi P. \tag{3}$$

In particular, π is the unique left eigenvector of P for the eigenvalue 1 such that $\pi 1 = 1$. Since the graph is undirected, we have $\pi \propto w^T$, i.e., nodes are visited in proportion to their weights. The convergence to the stationary distribution π is geometric at rate equal to the second largest eigenvalue of P (in modulus).

Theorem 1 (Spectral decomposition of the transition matrix) *We have:*

$$PV = V\Lambda, \quad V^T DV = I,$$

where $V = (v_1, \ldots, v_n)$ is the matrix of right eigenvectors and $\Lambda = (\lambda_1, \ldots, \lambda_n)$ the diagonal matrix of eigenvalues, with $\lambda_1 = 1 > \lambda_2 \geq \ldots \geq \lambda_n > -1$.

Proof. Consider the normalized adjacency matrix $\bar{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. This is a symmetric matrix, with spectral decomposition $\bar{A} = U\Lambda U^T, \quad U^T U = I$. Letting $V = D^{-\frac{1}{2}} U$ yields the result. The fact that $\lambda_1 = 1 > \lambda_2 \geq \ldots \geq \lambda_n > -1$ is a consequence of Perron-Frobenius theorem. □

In particular, we have $P = V\Lambda V^T D$ so that for all $t \geq 0$:

$$P^t = V\Lambda^t V^T D = \sum_{k=1}^{n} \lambda_k^t v_k v_k^T D \to v_1 v_1^T D = 1\pi,$$

with geometric convergence at rate $\max_{k>1} |\lambda_k|$.

3 Continuous-time random walk

Consider a random walk in the graph G where the move from node i to node j occurs at rate A_{ij}, i.e., after an exponential time with parameter A_{ij}. This defines a continuous-time Markov chain on $\{1, \ldots, n\}$ with infinitesimal generator $Q = A - D$. The time evolution is governed by the differential equation:

$$\frac{\mathrm{d}\pi_t}{\mathrm{d}t} = \pi_t Q,$$

so that

$$\forall t \geq 0, \quad \pi_t = \pi_0 e^{Qt},$$

with

$$e^{Qt} = \sum_{p=0}^{+\infty} \frac{Q^p t^p}{p!}.$$

Observe that this matrix is stochastic since $Q1 = 0$. Since the graph is connected, the following limit exists and is unique:

$$\pi = \lim_{t \to +\infty} \pi_t. \tag{4}$$

This is the stationary distribution, which satisfies the balance equations:

$$\pi Q = 0. \tag{5}$$

In particular, π is the unique eigenvector of Q for the eigenvalue 0 such that $\pi 1 = 1$. We have $\pi = 1^T/n$, i.e., the stationary distribution is uniform. The convergence to the stationary distribution π is exponential at rate equal to the second smallest eigenvalue of the Laplacian $L = -Q$. For convenience, we use the same notation of the eigenvectors and eigenvalues of P and L but these are different.

Theorem 2 (Spectral decomposition of the Laplacian) *Let $L = D - A$. We have:*

$$LV = V\Lambda, \quad V^T V = I,$$

where $V = (v_1, \ldots, v_n)$ is the matrix of right eigenvectors and $\Lambda = (\lambda_1, \ldots, \lambda_n)$ the diagonal matrix of eigenvalues, with $\lambda_1 = 0 < \lambda_2 \leq \ldots \leq \lambda_n$.

Proof. The matrix L is symmetric and semi-definite positive (see Proposition 1 below). The result follows from the spectral theorem. □

In particular, we have $L^p = V\Lambda^p V^T$ for all integers $p \geq 0$ so that for all $t \geq 0$:

$$e^{Qt} = e^{-Lt} = Ve^{-\Lambda t}V^T = \sum_{k=1}^{n} e^{-\lambda_k t} v_k v_k^T \to v_1 v_1^T = 1\pi,$$

with convergence at the exponential rate λ_2.

Proposition 1 *We have:*

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

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

$$\begin{aligned} u^T L u &= u^T(D - A)u, \\ &= \sum_{i,j=1}^{n} w_i u_i^2 - \sum_{i,j=1}^{n} u_j A_{ij} u_i, \\ &= \sum_{i,j=1}^{n} A_{ij} u_i (u_i - u_j), \\ &= \frac{1}{2} \sum_{i,j=1}^{n} A_{ij}(u_i - u_j)^2, \\ &= \sum_{i<j} A_{ij}(u_i - u_j)^2. \end{aligned}$$

□

Proposition 1 shows that the Laplacian matrix L is positive semi-definite. Moreover, $u^T L u = 0$ implies $u \propto 1$ (recall that the graph is connected) so that the eigenvalue $\lambda_1 = 0$ is simple and $v_1 \propto 1$. This proves in turn that the eigenvalue 1 of P is simple, since $Pu = u$ if and only if $Lu = 0$.

4 Heat equation

The Laplacian matrix $L = D - A$ is the discrete version of the usual Laplace operator. We shall see that it is related to the heat equation.

Consider some strict subset S of $\{1, \ldots, n\}$ 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:

$$\forall i \notin S, \quad (LT)_i = 0, \tag{6}$$

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

$$\forall i \notin S, \quad T_i = (PT)_i. \tag{7}$$

The temperature of each node is equal to the weighted average of those of its neighbors.

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

Proof. 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 (7) 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 P^S_{ij} be 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:

$$\forall i \notin S, \quad P^S_{ij} = \sum_{k=1}^{n} P_{ik} P^S_{kj}. \tag{8}$$

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

$$\forall i \notin S, \quad T_i = \sum_{j \in S} P^S_{ij} T_j. \tag{9}$$

Proof. The vector T defined by (9) 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 (8). Thus T satisfies (7). The proof then follows from Proposition 2. □

A practically interesting case is that where $S = \{s, t\}$ with $T_s = 0$ (source) and $T_t = 1$ (target). In this case, all temperatures at equilibrium are in $[0, 1]$ and can be interpreted as probabilities that the random walk hits the target t before the source s. The solution to the Dirichlet problem is then explicit and involves the pseudo-inverse of the Laplacian, defined by:

$$L^+ = V \Lambda^+ V^T, \quad \Lambda^+ = (0, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_n}).$$

Proposition 4 *We have:*

$$LL^+ = L^+ L = I - \frac{11^T}{n}.$$

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

$$LL^+ = L^+ L = V \Lambda^+ \Lambda V^T = \sum_{k=2}^{n} v_k v_k^T = V V^T - v_1 v_1^T = I - \frac{11^T}{n}.$$

□

Proposition 5 *For $S = \{s, t\}$ with $T_s = 0$ and $T_t = 1$, the solution to the Dirichlet problem is given by:*

$$T_i = \frac{(e_i - e_s)^T L^+ (e_t - e_s)}{(e_t - e_s)^T L^+ (e_t - e_s)}.$$

Proof. It is sufficient to prove that the solution to the Dirichlet problem is of the form $T = \alpha L^+(e_t - e_s) + \beta 1$ for some constants α, β (determined by the conditions $T_s = 0$, $T_t = 1$ on the boundary). Now by Proposition 4, $LT = \alpha L L^+(e_t - e_s) = \alpha(e_s - e_t)$, which proves that T is the unique solution to the Dirichlet problem. □

5 Mechanics

A useful interpretation of the spectrum of the Laplacian is through the following 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., attractive force proportional to the distance). Now if the points are located according to some vector $u \in \mathbb{R}^n$ along a line, the potential energy accumulated in the springs is:

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

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

We impose that the moment of inertia of the system (for a rotation around the origin) is equal to 1, that is $u^T u = 1$. Clearly, the vector u that minimizes the potential energy is $u = v_1$ (the corresponding potential energy is null). Now if we impose $1^T u = 0$, meaning that the centre of mass is at the origin, we obtain $u = v_2$ and $u^T L u = \lambda_2$, so that the eigenvalue λ_2 corresponds to twice the minimum value of potential energy. More generally, the spectrum of the Laplacian can be interpreted as levels of energy of the mechanical system, as shown by the following result.

Theorem 3 *For all $k = 1, \ldots, n$,*

$$\lambda_k = \min_{\substack{u : u^T u = 1 \\ v_1^T u = 0, \ldots, v_{k-1}^T u = 0}} u^T L u, \tag{10}$$

the minimum being attained for $u = v_k$.

Proof. Let $u \in \mathbb{R}^n$ such that $u^T u = 1$. The vector $x = V^T u$, corresponding to the coordinates of u in the basis of eigenvectors, satisfies:

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

so that the optimization problem (10) is equivalent to:

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

The result then follows from the equality:

$$x^T \Lambda x = \sum_{k=1}^{n} \lambda_k x_k^2.$$

□

Now consider the dynamical system, with nodes of unit mass. Let $s \in \mathbb{R}^n$ be the (time-dependent) state of the system (location of each node on the line). The force exerted on node i is:

$$\sum_j A_{ij}(s_j - s_i) = -(Ls)_i.$$

By Newton's law, we get:

$$-Ls = \ddot{s}.$$

Letting $s = ue^{-\omega t}$, we deduce that the eigenmodes u of the system are the eigenvectors of the Laplacian:

$$Lu = \omega^2 u.$$

Specifically, the eigenvectors of L correspond to the eigenmodes u with unit moments of inertia $u^T u = 1$.

Letting the mass of node i be w_i instead of 1, the eigenmodes become:

$$Lu = \omega^2 Du \quad \Longleftrightarrow \quad Pu = (1 - \omega^2)u.$$

We deduce that the eigenvectors of P correspond to the eigenmodes u with unit moments of inertia $u^T Du = 1$.

6 Spectral embedding

Let V be the matrix of eigenvectors of the Laplacian and

$$X = V\sqrt{\Lambda^+}.$$

Consider the embedding $X = (x_1, \ldots, x_n)^T$ of the graph, where node i is represented by the i-th row of the matrix X. Observe that the first column of X is null so that only $n - 1$ coordinates are informative. We have $1^T X = 0$, so that the center of mass of the embedding is at the origin:

$$\sum_{i \in V} x_i = 0.$$

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

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

In view of Proposition 5, the solution to the Dirichlet problem with one source node s and one target node t is given by:

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

Letting $x_s = 0$, which amounts to not specifying the source (as the origin is the center of mass of the embedding) shows that the proximity of node i to some node $t = j$ is given by $x_i^T x_j / ||x_j||^2$. Similarly, the proximity of node j to node i is given by $x_i^T x_j / ||x_i||^2$. The geometric mean of both provides a symmetric, normalized proximity measure between nodes i and j: it is equal to $\cos(x_i, x_j)$ and referred to as the *cosine similarity*.

Another embedding is based on the spectral decomposition of the transition matrix P. Let V be the matrix of eigenvectors of P and

$$Y = V\sqrt{(I - \Lambda)^+}.$$

Consider the embedding $Y = (y_1, \ldots, y_n)^T$ of the graph, where node i is represented by the i-th row of the matrix Y. Again, the first column of Y is null so that only $n - 1$ coordinates are informative. We have $1^T DY = 0$, so that the center of mass of the embedding is at the origin:

$$\sum_{i \in V} w_i y_i = 0.$$

It can be proved that the geometry of the embedding is the same as that of the previous embedding in the sense that $y_i^T y_j = (x_i - \bar{x})^T (x_j - \bar{x})$ for all pairs of nodes i, j, with $\bar{x} = \sum_{i \in V} w_i x_i / \sum_{i \in V} w_i$. In low dimension (that is, based on the first k eigenvectors with $k << n$), the embedding Y based on the spectral decomposition of the transition matrix P proves to better represent the structure of the graph.

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] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 2007.

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