Graph Theory

Random walks on graphs

Instructor: *Nicolò Cesa-Bianchi* version of May 14, 2024

Most of the material in this handout is taken from: Daniel Spielman, Spectral and Algebraic Graph Theory. Manuscript, 2019.

Random walks on graphs have numerous applications in computer science and other disciplines. The well-known PageRank index, originally introduced as a way for ranking Web pages, is defined through a random walk on the Web graph. Random walks are also used to model information spreading in online social networks. Graph properties—such as size, diameter, degree distribution—can be efficiently approximated via random walks when the graph is so large that exact computations are not feasible.

Another important class of applications is the simulation of uniform draws from a finite combinatorial set. For example, all spanning trees of a graph, all permutations of a set that satisfy certain properties, all Hamiltonian cycles of a graph. Given the combinatorial set $\mathcal{S}$, one can define a graph with vertex set $\mathcal{S}$ and edges (u,v) whenever u can be obtained from v by a small change; for example, the substitution of an edge in a spanning tree. By designing a random walk on this graph that quickly converges to the uniform distribution on $\mathcal{S}$, one can efficiently simulate a uniform random draw from $\mathcal{S}$.

Let A be the adjacency matrix of a connected graph G and recall the normalized Laplacian matrix $L_{\mathrm{norm}} = I - D^{-1/2}AD^{-1/2}$ with entries

$$L_{\mathrm{norm}}(i,j) = \begin{cases} 1 & \text{if } i = j \\ -A(i,j)\big/\sqrt{d(i)d(j)} & \text{otherwise} \end{cases}$$

We use $\alpha_1 \ge \cdots \ge \alpha_n$ to denote the eigenvalues of A (note that they are ordered in the opposite direction with respect to the eigenvalues $\lambda_1 \le \cdots \le \lambda_n$ of L_{norm}).

If G is d-regular, then $L_{\mathrm{norm}} = I - \frac{1}{d}A$ and therefore $\lambda_i = 1 - \frac{1}{d}\alpha_i$. Since $\lambda_i \in [0,2]$ for any G (even not regular), we have that $\alpha_i \in [-d,d]$ for any d-regular graph.

Recall that $d(G)$ is the average degree of the nodes in G, whereas $\Delta(G)$ is the maximum degree of a node in G.

Fact 1 *For any graph $G = (V,E)$, $d(G) \le \alpha_1 \le \Delta(G)$.*

PROOF. Using the variational characterization of eigenvalues,

$$\alpha_1 = \max_{\boldsymbol{x} : \boldsymbol{x} \neq \boldsymbol{0}} \frac{\boldsymbol{x}^\top A \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{x}} \ge \frac{\boldsymbol{1}^\top A \boldsymbol{1}}{\boldsymbol{1}^\top \boldsymbol{1}} = \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n A(i,j) = \frac{1}{n}\sum_{i=1}^n d(i) = d(G)$$

For the other inequality, let $\boldsymbol{u}$ be an eigenvector for the eigenvalue α_1 and let $u_i > 0$ the largest component of $\boldsymbol{u}$ (if all components are negative, take $-\boldsymbol{u}$). Then

$$\alpha_1 = \frac{(A\boldsymbol{u})_i}{u_i} = \frac{1}{u_i}\sum_{j=1}^n A(i,j)u_j = \sum_{j=1}^n A(i,j)\frac{u_j}{u_i} \le \sum_{j=1}^n A(i,j) \le \Delta(G)$$

concluding the proof. □

The trace of a symmetric $n \times n$ matrix M is $M(1,1) + \cdots + M(n,n)$. One can show that the trace is equal to the sum of eigenvalues. Since $A_{i,i} = 0$, the trace of A is zero and so $\alpha_1 + \cdots + \alpha_n = 0$. Since we proved that $\alpha_1 \geq d(G) > 0$, this implies that $\alpha_n < 0$.

Lemma 2 *Let $G = (V, E)$ be a connected graph and let M be a nonnegative symmetric matrix such that $M(i,j) > 0$ if and only if $(i,j) \in E$. Assume that some nonnegative vector $\boldsymbol{u}$ is an eigenvector of M. Then $\boldsymbol{u}$ is strictly positive.*

PROOF. If $\boldsymbol{u}$ is nonnegative but not strictly positive, then there is some vertex r for which $u_r = 0$. As G is connected, there must be some edge (r, s) for which $u_r = 0$ but $u_s > 0$ (since $\boldsymbol{u}$ is an eigenvector, $\boldsymbol{u} \neq \boldsymbol{0}$). Let μ be the eigenvalue of $\boldsymbol{u}$. We obtain a contradiction from

$$0 = \mu u_r = (M\boldsymbol{u})_r = \sum_{i=1}^{n} M(r,i) u_i \geq M(r,s) u_s > 0$$

concluding the proof. □

The next result is the cornerstone for the analysis of random walks on graphs. It applies to many symmetric matrices defined on graphs, including the adjacency matrix and the normalized adjacency matrix.

Theorem 3 (Perron-Frobenius for symmetric matrices) *Let $G = (V, E)$ be a connected graph and let M be a nonnegative symmetric matrix such that, for all $i \neq j$, $M(i,j) > 0$ if and only if $(i,j) \in E$. Then the eigenvalues $\mu_1 \geq \cdots \geq \mu_n$ of M satisfy:*

1. *The largest eigenvalue μ_1 has a strictly positive eigenvector,*
2. *$\mu_1 \geq -\mu_n$,*
3. *$\mu_1 > \mu_2$, implying that μ_1 has multiplicity 1.*

PROOF. Note that M has trace zero and so $\mu_n < 0 < \mu_1$. In order to prove part 1, let $\boldsymbol{u}_1$ an eigenvector for μ_1 and define $x_i = |u_{1,i}|$. Then $\boldsymbol{x}^\top \boldsymbol{x} = \boldsymbol{u}^\top \boldsymbol{u} = 1$. Moreover, since M is nonnegative,

$$\mu_1 = \boldsymbol{u}_1^\top M \boldsymbol{u}_1 = \sum_{i=1}^{n} \sum_{j=1}^{n} M(i,j) u_i u_j \leq \sum_{i=1}^{n} \sum_{j=1}^{n} M(i,j) |u_i| |u_j| = \boldsymbol{x}^\top M \boldsymbol{x}$$

Therefore $\boldsymbol{x}$ satisfies $\boldsymbol{x}^\top \boldsymbol{x} = 1$ and

$$\boldsymbol{x}^\top M \boldsymbol{x} \geq \mu_i = \max_{\boldsymbol{v} \,:\, \boldsymbol{v} \neq \boldsymbol{0}} \frac{\boldsymbol{v}^\top M \boldsymbol{v}}{\boldsymbol{v}^\top \boldsymbol{v}}$$

So, according to the variational characterization of eigenvalues, $\boldsymbol{x}$ must be an eigenvector of μ_1. Since $\boldsymbol{x}$ is nonnegative, Lemma 2 implies that is strictly positive.

To prove part 2, let $\boldsymbol{u}_n$ be an eigenvector of μ_n and let $x_i = |u_{n,i}|$. Then, similarly to before and recalling that $\mu_n < 0$,

$$|\mu_n| = \left| \boldsymbol{u}_n^\top M \boldsymbol{u}_n \right| \leq \sum_{i=1}^{n} \sum_{j=1}^{n} M(i,j) |u_i| |u_j| = \boldsymbol{x}^\top M \boldsymbol{x} \leq \mu_1$$

To prove part 3, consider an eigenvector $\boldsymbol{u}_2$ of μ_2. Note that: $\boldsymbol{u}_2^\top \boldsymbol{u}_1 = 0$, $\boldsymbol{u}_1$ has strictly positive components, and $\boldsymbol{u}_2 \neq \boldsymbol{0}$. Hence $\boldsymbol{u}_2$ must contain positive and negative components. Now let $x_i = |u_{2,i}|$ and, once again, note that $\mu_2 = \boldsymbol{u}_2^\top M \boldsymbol{u}_2 \leq \boldsymbol{x}^\top M \boldsymbol{x} \leq \mu_1$. For the purpose of contradiction, assume $\mu_2 = \mu_1$. Then $\boldsymbol{x}$ is a nonnegative eigenvector of μ_1. Lemma 2 implies that $\boldsymbol{x}$ is strictly positive and so $\boldsymbol{u}_2$ has all components different from zero. Since $\boldsymbol{u}_2$ has positive and negative components and the graph is connected, there must be at least one edge $(i,j) \in E$ such that $u_{2,i} < 0 < u_{2,j}$. This edge gives a negative contribution to $\boldsymbol{u}_2^\top M \boldsymbol{u}_2$ and a positive contribution to $\boldsymbol{x}^\top M \boldsymbol{x} \leq \mu_1$ (recall that M is nonnegative). Hence the inequality $\boldsymbol{u}_2^\top M \boldsymbol{u}_2 \leq \boldsymbol{x}^\top M \boldsymbol{x}$ must be strict, implying $\mu_2 < \mu_1$. So we have a contradiction. □

The next observation (proof omitted) is important in the analysis of convergence of a random walk on a graph.

Fact 4 *G is bipartite if and only if $\mu_n = -\mu_1$.*

The random walk on a graph. Given a connected graph $G = (V, E)$ with $V = \{1, \dots, n\}$, we consider the random walk that starts from an arbitrary vertex $V_0 \in V$, and at each step $t = 0, 1, \dots$ moves from V_t to a random vertex V_{t+1} in the neighborhood of V_t. Therefore,

$$\mathbb{P}(V_{t+1} = i \mid V_t = j) = \frac{A(i,j)}{d(j)}$$

Let $\boldsymbol{e}_i$ be the canonical basis vector for the i-th coordinate (all zeros but a single 1 in position i). The state of the walk at time t is defined by a probability distribution $\boldsymbol{p}_t$ over V,

$$\mathbb{P}(V_t = i) = p_t(i)$$

Hence, if the walk starts at $V_0 = i$, then $\boldsymbol{p}_0 = \boldsymbol{e}_i$. At any time t we have

$$\boldsymbol{p}_t(i) = \mathbb{P}(V_t = i) = \sum_{j\,:\,(i,j)\in E} \mathbb{P}(V_t = i \mid V_{t-1} = j)\mathbb{P}(V_{t-1} = j) = \sum_{j=1}^{n} \frac{A(i,j)}{d(j)} p_{t-1}(j) \tag{1}$$

Let $D = \operatorname{diag}\big(d(1), \dots, d(n)\big)$ and note that $D^{-1}_{i,j} = \mathbb{I}\{i = j\}/d(j)$. Since

$$\big(AD^{-1}\big)(i,j) = \sum_{k=1}^{n} A_{i,k} \frac{\mathbb{I}\{k = j\}}{d(j)} = \frac{A(i,j)}{d(j)}$$

the right-hand side of (1) can be rewritten as $AD^{-1}\boldsymbol{p}_{t-1}$ Letting $W = AD^{-1}$, the evolution of our random walk is given by $\boldsymbol{p}_t = W\boldsymbol{p}_{t-1}$, or $\boldsymbol{p}_t = W^t \boldsymbol{p}_0$. W is a column-stochastic matrix, as it is a nonnegative matrix whose elements in each column sum to 1, that is $(\boldsymbol{1}^\top W)_i = 1$ for all i.

As $W_{i,j} = A(i,j)/d(j)$, the matrix W is not symmetric. However, it is related to the normalized adjacency matrix $A_{\text{norm}} = D^{-1/2} A D^{-1/2}$, which is symmetric with components

$$A_{\text{norm}}(i,j) = \begin{cases} 0 & \text{if } i = j \\ A(i,j)\big/\sqrt{d(i)d(j)} & \text{otherwise} \end{cases}$$

Indeed, $A_{\text{norm}} = D^{-1/2} W D^{1/2}$. The normalized adjacency matrix is in turn related to the normalized Laplacian as follows

$$L_{\text{norm}} = I - D^{-1/2} A D^{-1/2} = I - A_{\text{norm}}$$

Let $\alpha'_1 \geq \cdots \geq \alpha'_n$ be the eigenvalues of A_{norm}. Because $L_{\text{norm}} = I - A_{\text{norm}}$, we have $\alpha'_i = 1 - \lambda_i$ for all i. Moreover, because the eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ of L_{norm} belong to the interval $[0, 2]$, $\alpha'_i \in [-1, 1]$ for all i.

Fact 5 *The vector $\boldsymbol{\psi}$ is an eigenvector of A_{norm} of eigenvalue ω if and only if $D^{1/2}\boldsymbol{\psi}$ is an eigenvector of W of eigenvalue ω.*

PROOF. As $A_{\text{norm}} = D^{-1/2}WD^{1/2}$, we have that $D^{1/2}A_{\text{norm}} = WD^{1/2}$. Thus, if $A_{\text{norm}}\boldsymbol{\psi} = \omega\,\boldsymbol{\psi}$, then

$$WD^{1/2}\boldsymbol{\psi} = D^{1/2}A_{\text{norm}}\boldsymbol{\psi} = D^{1/2}\omega\,\boldsymbol{\psi} = \omega\big(D^{1/2}\boldsymbol{\psi}\big)$$

and, similarly, we can show that $W\boldsymbol{u} = \omega\,\boldsymbol{u}$ implies $A_{\text{norm}}\big(D^{-1/2}\boldsymbol{u}\big) = \omega\big(D^{-1/2}\boldsymbol{u}\big)$. □

This result implies that the eigenvalues $\omega_1 \geq \cdots \geq \omega_n$ of W are the same as the eigenvalues of A_{norm}.

An application of the Perron-Frobenius theorem to the normalized adjacency matrix A_{norm} gives:

1. $-1 \leq \omega_n < 0 < \omega_1 \leq 1$ and the unique eigenvector $\boldsymbol{\psi}_1$ of ω_1 has strictly positive components
2. $\omega_2 < \omega_1$
3. $\omega_n = -\omega_1$ if and only if G is bipartite.

Next, we show that $\omega_1 = 1$. However, ω_2 can be positive or negative. Moreover, if G is bipartite, then $\omega_1 = 1$ and $\omega_2 = -1$.

The stationary distribution. We say that a distribution $\boldsymbol{\pi}$ over V is the stationary distribution of W if $W\boldsymbol{\pi} = \boldsymbol{\pi}$. Hence, the stationary distribution is a (unnormalized) eigenvector of W with eigenvalue $1 = \omega_1$, as the eigenvalues of W range in $[-1, 1]$. Now let $\boldsymbol{d} = \big(d(1), \ldots, d(n)\big)$ be the vector of vertex degrees and consider the distribution

$$\boldsymbol{\pi} = \frac{\boldsymbol{d}}{\boldsymbol{1}^\top\boldsymbol{d}}$$

Since

$$(W\boldsymbol{\pi})_i = \sum_{j=1}^{n} W(i,j)\pi_j = \sum_{j=1}^{n} \frac{A(i,j)}{d(j)}\frac{d(j)}{\boldsymbol{1}^\top\boldsymbol{d}} = \frac{1}{\boldsymbol{1}^\top\boldsymbol{d}}\sum_{j=1}^{n} A(i,j) = \frac{d(i)}{\boldsymbol{1}^\top\boldsymbol{d}} = \pi(i)$$

this is the stationary distribution for W. Moreover, we also know that $\boldsymbol{\pi}$ is (up to normalization) the unique eigenvector of W for the eigenvalue 1. So Fact 5 implies $\boldsymbol{\pi} = D^{1/2}\boldsymbol{\psi}_1$.

Fact 6 *Let A be a $n \times n$ symmetric matrix with spectrum $\lambda_1, \ldots, \lambda_n, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$. Then, for any $t \in \mathbb{N}$,*

$$A^t = \sum_{i=1}^{n} \lambda_i^t \boldsymbol{u}_i \boldsymbol{u}_i^\top$$

PROOF. We use induction on t together with the spectral theorem and the orthonormality of the eigenvectors. For $t = 1$ the statement follows from the spectral theorem. Assume now the statement holds for $t-1$ and write

$$\begin{aligned}
A^t = AA^{t-1} &= \left(\sum_{i=1}^n \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top\right)\left(\sum_{j=1}^n \lambda_j^{t-1} \boldsymbol{u}_j \boldsymbol{u}_j^\top\right) \\
&= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j^{t-1} \boldsymbol{u}_i \boldsymbol{u}_i^\top \boldsymbol{u}_j \boldsymbol{u}_j^\top \\
&= \sum_{i=1}^n \lambda_i^t \boldsymbol{u}_i \boldsymbol{u}_i^\top && \text{(since } \boldsymbol{u}_i^\top \boldsymbol{u}_j = \mathbb{I}\{i=j\}\text{)}
\end{aligned}$$

concluding the proof. □

We are now ready to prove the convergence of the random walk to the stationary distribution.

Theorem 7 *For any connected graph G not bipartite,*

$$\lim_{t\to\infty} W^t \boldsymbol{p}_0 = \frac{\boldsymbol{d}}{\mathbf{1}^\top \boldsymbol{d}}$$

irrespective to the initial distribution $\boldsymbol{p}_0$.

Proof. To verify convergence to $\boldsymbol{\pi}$, we express $D^{-1/2}\boldsymbol{p}_0$ in the eigenbasis $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n$ of A_{norm},

$$D^{-1/2}\boldsymbol{p}_0 = \sum_{i=1}^n \left(\boldsymbol{\psi}_i^\top D^{-1/2}\boldsymbol{p}_0\right)\boldsymbol{\psi}_i = \sum_{i=1}^n c_i \boldsymbol{\psi}_i \tag{2}$$

Now we write

$$\begin{aligned}
\boldsymbol{p}_t = W^t \boldsymbol{p}_0 &= \left(D^{1/2} A_{\text{norm}} D^{-1/2}\right)^t \boldsymbol{p}_0 \\
&= D^{1/2} A_{\text{norm}} D^{-1/2} D^{1/2} A_{\text{norm}} D^{-1/2} \cdots D^{1/2} A_{\text{norm}} D^{-1/2} \boldsymbol{p}_0 \\
&= D^{1/2} A_{\text{norm}}^t D^{-1/2} \boldsymbol{p}_0 && \text{(using } D^{-1/2}D^{1/2} = I\text{)} \\
&= D^{1/2} A_{\text{norm}}^t \sum_{i=1}^n c_i \boldsymbol{\psi}_i && \text{(using (2))} \\
&= D^{1/2} \left(\sum_{j=1}^n \omega_j^t \boldsymbol{\psi}_j \boldsymbol{\psi}_j^\top\right) \sum_{i=1}^n c_i \boldsymbol{\psi}_i && \text{(using Fact 6)} \\
&= D^{1/2} \sum_{i=1}^n \sum_{j=1}^n c_i\, \omega_j^t \boldsymbol{\psi}_j \boldsymbol{\psi}_j^\top \boldsymbol{\psi}_i \\
&= D^{1/2} \sum_{i=1}^n c_i\, \omega_i^t \boldsymbol{\psi}_i && \text{(since } \boldsymbol{\psi}_i^\top \boldsymbol{\psi}_j = \mathbb{I}\{i=j\}\text{)}
\end{aligned}$$

Therefore, recalling that $\omega_1 = 1$,

$$\boldsymbol{p}_t = D^{1/2} c_1 \boldsymbol{\psi}_1 + D^{1/2} \sum_{i=2}^n c_i\, \omega_i^t \boldsymbol{\psi}_i \tag{3}$$

Now, if G is not bipartite, then $\omega_2, \ldots, \omega_n \in (-1, 1)$. Since $\lim_{x\to\infty} \omega^x = 0$ for all $\omega \in (-1, 1)$, we get

$$\lim_{t\to\infty} \boldsymbol{p}_t = D^{1/2} c_1 \boldsymbol{\psi}_1$$

Now recall that $\boldsymbol{\pi}$ is a unnormalized eigenvector of W. Therefore, using Fact 5, $\boldsymbol{\pi} = \frac{\boldsymbol{d}}{\mathbf{1}^\top \boldsymbol{d}} \propto D^{1/2}\boldsymbol{\psi}_1$ implying $\boldsymbol{\psi}_1 = D^{-1/2}\boldsymbol{d}/\|D^{-1/2}\boldsymbol{d}\|$. Therefore

$$c_1 = \boldsymbol{\psi}_1^\top D^{-1/2}\boldsymbol{p}_0 = \frac{\left(D^{-1/2}\boldsymbol{d}\right)^\top}{\|D^{-1/2}\boldsymbol{d}\|} D^{-1/2}\boldsymbol{p}_0 = \frac{\boldsymbol{d}^\top D^{-1}\boldsymbol{p}_0}{\|D^{-1/2}\boldsymbol{d}\|} = \frac{\mathbf{1}^\top \boldsymbol{p}_0}{\|D^{-1/2}\boldsymbol{d}\|} = \frac{1}{\|D^{-1/2}\boldsymbol{d}\|}$$

because $\boldsymbol{p}_0$ is a probability vector. So,

$$D^{1/2}c_1\boldsymbol{\psi}_1 = \frac{1}{\|D^{-1/2}\boldsymbol{d}\|} D^{1/2}D^{-1/2}\frac{\boldsymbol{d}}{\|D^{-1/2}\boldsymbol{d}\|} = \frac{\boldsymbol{d}}{\|D^{-1/2}\boldsymbol{d}\|^2} = \frac{\boldsymbol{d}}{\sum_{j=1}^n d(j)^2/d(j)} = \boldsymbol{\pi}$$

concluding the proof. □

Speed of convergence of the random walk. Assume that the random walk starts at some vertex $u \in V$. For every vertex $v \in V$, we will bound how far $p_t(v)$ can be from $\pi(v)$.

Theorem 8 *For all $u, v \in V$ and $t \in \mathbb{N}$, if $\boldsymbol{p}_0 = \boldsymbol{e}_u$, then*

$$\left|p_t(v) - \pi(v)\right| \leq \left(\sqrt{\frac{d(v)}{d(u)}}\right)\kappa^t$$

where $\kappa = \max\left\{|\omega_n|, |\omega_2|\right\}$.

Proof. We start by writing $p_t(v) = \boldsymbol{e}_v^\top \boldsymbol{p}_t$. Recalling (3),

$$p_t(v) = \boldsymbol{e}_v^\top \boldsymbol{p}_t = \pi(v) + \boldsymbol{e}_v^\top D^{1/2}\sum_{i=2}^n \omega_i^t c_i \boldsymbol{\psi}_i \tag{4}$$

Using (2), we know that

$$c_i = \boldsymbol{\psi}_i^\top D^{-1/2}\boldsymbol{e}_u = \frac{\boldsymbol{\psi}_i^\top \boldsymbol{e}_u}{\sqrt{d(u)}}$$

So, from (4) and $\boldsymbol{e}_v^\top D^{1/2} = \sqrt{d(v)}\boldsymbol{e}_v^\top$,

$$\boldsymbol{e}_v^\top D^{1/2}\sum_{i=2}^n \omega_i^t c_i\boldsymbol{\psi}_i = \left(\sqrt{\frac{d(v)}{d(u)}}\right)\boldsymbol{e}_v^\top \sum_{i=2}^n \omega_i^t \boldsymbol{\psi}_i\boldsymbol{\psi}_i^\top \boldsymbol{e}_u$$

Now we look at the last part of the above expression. We can write

$$
\begin{aligned}
\boldsymbol{e}_v^\top \sum_{i=2}^n \omega_i^t \boldsymbol{\psi}_i \boldsymbol{\psi}_i^\top \boldsymbol{e}_u &= \sum_{i=2}^n \omega_i^t \big(\boldsymbol{e}_v^\top \boldsymbol{\psi}_i\big)\big(\boldsymbol{\psi}_i^\top \boldsymbol{e}_u\big) \\
&\le \sum_{i=2}^n |\omega_i|^t \big|\boldsymbol{e}_v^\top \boldsymbol{\psi}_i\big| \, \big|\boldsymbol{\psi}_i^\top \boldsymbol{e}_u\big| \\
&\le \kappa^t \sum_{i=1}^n \big|\boldsymbol{e}_v^\top \boldsymbol{\psi}_i\big| \, \big|\boldsymbol{\psi}_i^\top \boldsymbol{e}_u\big| \\
&\le \kappa^t \sqrt{\sum_{i=1}^n \big(\boldsymbol{e}_v^\top \boldsymbol{\psi}_i\big)^2} \sqrt{\sum_{i=1}^n \big(\boldsymbol{\psi}_i^\top \boldsymbol{e}_u\big)^2} && \text{(using the Cauchy-Schwartz inequality)} \\
&= \kappa^t \, \|\boldsymbol{e}_v\| \, \|\boldsymbol{e}_u\| && \text{(because } \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_n \text{ is an orthonormal basis)} \\
&= \kappa^t
\end{aligned}
$$

This concludes the proof. □

The lazy random walk. In order to directly relate the speed of convergence to the spectrum of the Laplacian of G, we replace W by

$$W' = \frac{1}{2}\big(I + W\big)$$

With this new matrix, with equal probabilities we have that $V_{t+1} = V_t$ or V_{t+1} is a random neighbor of V_t . Also the eigenvalues $\omega'_1 \ge \cdots \ge \omega'_n$ of W' satisfy $\omega'_i = \frac{1}{2}(1 + \omega_i) \in [0, 1]$. It is easy to check that $\boldsymbol{\pi} = \frac{\boldsymbol{d}}{\boldsymbol{1}^\top \boldsymbol{d}}$ is the stable distribution also for W'. Indeed, $W'\boldsymbol{\pi} = \frac{1}{2}\big(I + W\big)\boldsymbol{\pi} = \frac{1}{2}\boldsymbol{\pi} + \frac{1}{2}\boldsymbol{\pi} = \boldsymbol{\pi}$. The relation between W' and L_{norm} is now

$$W' = \frac{1}{2}\Big(I + D^{1/2} A_{\text{norm}} D^{-1/2}\Big) = \frac{1}{2}\Big(I + D^{1/2}(I - L_{\text{norm}}) D^{-1/2}\Big) = I - \frac{1}{2} D^{1/2} L_{\text{norm}} D^{-1/2}$$

If $(\lambda, \boldsymbol{u})$ is an eigenpair for L_{norm}, then let $\boldsymbol{v} = D^{1/2}\boldsymbol{u}$ and note that

$$D^{1/2} L_{\text{norm}} D^{-1/2} \boldsymbol{v} = D^{1/2} L_{\text{norm}} \boldsymbol{u} = \lambda \boldsymbol{v}$$

Hence, $D^{1/2} L_{\text{norm}} D^{-1/2}$ has the same eigenvalues as L_{norm} and thus $\omega'_i = 1 - \frac{\lambda_i}{2}$ for $i = 1, \dots, n$. Therefore, $\omega'_2 \ge \omega'_n \ge 0$ which implies that in Theorem 8 we have $\kappa = \max\big\{|\omega'_n|, |\omega'_2|\big\} = \omega'_2 = 1 - \frac{\lambda_2}{2}$. Doing again the proof of Theorem 8 we obtain

$$\boldsymbol{p}_t = \boldsymbol{\pi} + D^{1/2} \sum_{i=2}^n c_i \left(1 - \frac{\lambda_i}{2}\right)^t \boldsymbol{\psi}_i$$

and so, for $\boldsymbol{p}_0 = \boldsymbol{e}_u$,

$$p_t(v) \le \pi(v) + \left(\sqrt{\frac{d(v)}{d(u)}}\right) \left(1 - \frac{\lambda_i}{2}\right)^t$$

Mixing time. For any $u \in V$, let $\boldsymbol{p}_0(u, \cdot) = \boldsymbol{e}_u$ and $\boldsymbol{p}_t(u, \cdot) = \big(W'\big)^t \boldsymbol{p}_0(u, \cdot)$. Define

$$d(t) = \max_{u,v \in V} \big|p_t(u, v) - \pi(v)\big|$$

The mixing time of W' is defined by $t_{\text{mix}}(\varepsilon) = \min\{t \geq 0 : d(t) \leq \varepsilon\}$ for any $\varepsilon > 0$. For any sufficiently small $\varepsilon > 0$ and for any $t \geq t_{\text{mix}}(\varepsilon)$, V_t approximates an independent random draw from π. This is how we can use the random walk to approximately draw from the stationary distribution.

For simplicity, let $t_{\text{mix}} = t_{\text{mix}}(1/4)$. Then $t \geq t_{\text{mix}}$ is implied by

$$\max_{u,v\in V}\left(\sqrt{\frac{d(v)}{d(u)}}\right)\left(1-\frac{\lambda_2}{2}\right)^t \leq \frac{1}{4}$$

Using $d(u) \leq 1$, $d(v) \leq n-1$, and $1+x \leq e^x$ for all $x \in \mathbb{R}$ the above is implied by

$$t \geq \frac{2}{\lambda_2}\ln\left(4\sqrt{n-1}\right) \qquad \text{which gives} \qquad t_{\text{mix}} = \mathcal{O}\left(\frac{\ln n}{\lambda_2}\right).$$

This reveals the important connection between mixing time, clusterability and the role of λ_2.

The mixing time is a lower bound on the hitting time. Let $t_{\text{hit}}(u,v)$ the smallest t such that $V_t = v$ given that $V_0 = u$, and let the hitting time be defined as

$$t_{\text{hit}} = \max_{u,v\in V}\mathbb{E}\big[t_{\text{hit}}(u,v)\big]$$

Then one can show that $t_{\text{mix}} \leq 2t_{\text{hit}} + 1$. The hitting time is also a lower bound on the cover time. Let $t_{\text{cov}}(u)$ be the smallest time t such that the random walk visited all $v \in V$ starting from u. The cover time is

$$t_{\text{cov}} = \max_{u\in V}\mathbb{E}\big[t_{\text{cov}}(u)\big]$$

Then one can show that $t_{\text{hit}} \leq t_{\text{cov}} \leq \big(1 + \ln|V|\big)t_{\text{hit}}$.

Some examples. We now bound the mixing time of some graphs. Let L be the Laplacian matrix of some graph G of order n. Then, for any $\boldsymbol{x} \in \mathbb{R}^n$ and any $i = 1,\dots,n$ we have

$$\big(L\boldsymbol{x}\big)_i = d(i)x_i - \sum_{j=1}^{n} A(i,j)x_j = \sum_{j\,:\,(i,j)\in E}(x_i - x_j) \tag{5}$$

In the following, we write $\lambda'_1 \leq \cdots \leq \lambda'_n$ to denote the eigenvalues of L and $\boldsymbol{u}_1,\dots,\boldsymbol{u}_n$ the associated eigenvectors. We know that $\boldsymbol{u}_1$ is a multiple of $\mathbf{1}$, and so $\mathbf{1}^\top\boldsymbol{u}_i = 0$ for any $i > 1$.

Since computing the eigenvalues of L is easier than computing the eigenvalues of L_{norm}, we need the following result relating the eigenvalues of L_{norm} with those of L.

Theorem 9 *Let L be the Laplacian matrix of a graph with eigenvalues $\lambda'_1 \leq \cdots \leq \lambda'_n$ and let L_{norm} be its normalized Laplacian with eigevalues $\lambda_1 \leq \cdots \leq \lambda_n$. Then, for all $i = 1,\dots,n$ we have*

$$\frac{\lambda'_i}{\Delta} \leq \lambda_i \leq \frac{\lambda'_i}{\delta}$$

Proof. By the Courant-Fischer theorem, for all $i = 1,\dots,n$ we have

$$\lambda_i = \min_{S\,:\,\dim(S)=i}\ \max_{\boldsymbol{x}\in S\setminus\{\mathbf{0}\}} \frac{\boldsymbol{x}^\top L_{\text{norm}}\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}} = \min_{T\,:\,\dim(T)=i}\ \max_{\boldsymbol{y}\in T\setminus\{\mathbf{0}\}} \frac{\boldsymbol{y}^\top L\boldsymbol{y}}{\boldsymbol{y}^\top D\boldsymbol{y}}$$

beause the change of variables $\boldsymbol{y} = D^{-1/2}\boldsymbol{x}$ is a bijection (and so the dimensionality of the subspace is preserved). Using

$$\boldsymbol{y}^\top D\boldsymbol{y} = \sum_i d(i) y_i^2 \le \Delta \boldsymbol{y}^\top \boldsymbol{y}$$

we obtain

$$\min_{T\,:\,\dim(T)=i} \max_{\boldsymbol{y}\in T\setminus\{\boldsymbol{0}\}} \frac{\boldsymbol{y}^\top L\boldsymbol{y}}{\boldsymbol{y}^\top D\boldsymbol{y}} \ge \frac{1}{\Delta}\left(\min_{T\,:\,\dim(T)=i} \max_{\boldsymbol{y}\in T\setminus\{\boldsymbol{0}\}} \frac{\boldsymbol{y}^\top L\boldsymbol{y}}{\boldsymbol{y}^\top \boldsymbol{y}}\right) = \frac{\lambda_i'}{\Delta}$$

The other inequality is proved similarly. □

Consider now K_n, the complete graph of order n and consider any eigenvector $\boldsymbol{u}$ orthogonal to $\boldsymbol{1}$. Using (5), for any $i = 1, \ldots, n$ we have that

$$\big(L\boldsymbol{u}\big)_i = \sum_{j\neq i}(u_i - u_j) = (n-1)u_i - \sum_{j\neq i} u_j = nu_i - \sum_j u_j = nu_i$$

because $\boldsymbol{1}^\top\boldsymbol{u} = 0$. Therefore, any $\boldsymbol{u}$ such that $\boldsymbol{1}^\top\boldsymbol{u} = 0$ satisfies $L\boldsymbol{u} = n\boldsymbol{u}$. This implies that the eigenvalue n has multiplicity $n-1$. So $\lambda_2' = n$ and, because K_n is regular, Theorem 9 gives $\lambda_2 = \frac{n}{n-1}$ implying that the mixing time is $\mathcal{O}(\ln n)$.

For n evem, the dumbell graph D_n consists of two copies of $K_{n/2}$, joined by one edge (called the bridge). So all vertices have degree $n/2-1$ or n and we get

$$\lambda_2' = \min_{\substack{\boldsymbol{x}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\}\\ \boldsymbol{x}^\top\boldsymbol{1}=0}} \frac{\sum_{(i,j)\in E}(x_i - x_j)^2}{\boldsymbol{x}^\top\boldsymbol{x}} \le \frac{4}{n}$$

where we chose the vector $\boldsymbol{x}$ such that $x_i = 1$ if i belongs to the first clique and $x_i = -1$ otherwise. Using Theorem 9 we get that $\lambda_2 \le \frac{4}{n(n-1)}$. This implies that the mixing time is $\mathcal{O}(n^2 \ln n)$.

The two bounds for the mixing time of the clique and the dumbell are tight. This implies an exponential gap between the mixing time $\Theta(\ln n)$ of K_n and the mixing time $\Theta(n^2 \ln n)$ of D_n.

Distributed consensus. Given numbers $\boldsymbol{x}_0 = \big(x_0(1), \ldots, x_0(n)\big)$ at each vertex of a connected graph $G = (V, E)$, we want each node in V to compute the average

$$\mu = \frac{1}{n}\sum_{v\in V} x_0(v)$$

by communicating only with its neighbors in G.

We run at each node $v \in V$ an algorithm that, at each time step $t = 0, 1, \ldots$, updates the node's state $x_t(v)$ according to

$$x_{t+1}(v) = \sum_{u\,:\,(v,u)\in E} W(v,u)\, x_t(u) \tag{6}$$

We can write the update as $\boldsymbol{x}_{t+1} = W\boldsymbol{x}_t$. We let the matrix W to be a **gossip matrix**. This is any nonnegative symmetric matrix, doubly stochastic ($W\boldsymbol{1} = \boldsymbol{1}$ and $\boldsymbol{1}^\top W = \boldsymbol{1}^\top$), and such that $W(i,j) > 0$ if and only if $(i,j) \in E$.

Fact 10 *The largest eigenvalue of a row-stochastic matrix is* 1.

PROOF. Let W be a row-stochastic matrix. Then $W\mathbf{1} = \mathbf{1}$ and so 1 is an eigenvalue of W. Now suppose there exists $\mu > 1$ and $\boldsymbol{x} \neq \mathbf{0}$ such that $W\boldsymbol{x} = \mu\,\boldsymbol{x}$. Let x_k be a largest element of $\boldsymbol{x}$. Since $-\boldsymbol{x}$ also satisfies this equation we can assume, without loss of generality, that $x_k > 0$. Since the elements $W(i,1),\ldots,W(i,n)$ on each row of W are nonnegative and sum to 1, for any $i = 1\ldots,n$ we have

$$(W\boldsymbol{x})_i = \sum_{j=1}^{n} W(i,j)x_j \leq \max_{j=1,\ldots,n} x_j = x_k$$

Thus, no entry in $\mu\,\boldsymbol{x} = W\boldsymbol{x}$ can be larger than x_k. But since $\mu > 1$, $\mu\,x_k > x_k$ and we have a contradiction. Therefore, the largest eigenvalue of W is 1. $\square$

Let $\omega_n \leq \cdots \leq \omega_{n-1} < \omega_1 = 1$ be the eigenvalues of W. Let

$$\boldsymbol{x} = \mu\mathbf{1} = \frac{1}{n}\mathbf{1}\mathbf{1}^\top\boldsymbol{x}_0$$

Then $\boldsymbol{x}$ is the stationary distribution for W. Indeed, because W is row-stochastic,

$$W\boldsymbol{x} = \frac{1}{n}W\mathbf{1}\mathbf{1}^\top\boldsymbol{x}_0 = \frac{1}{n}\mathbf{1}\mathbf{1}^\top\boldsymbol{x}_0 = \boldsymbol{x}$$

Note also that

$$\frac{1}{n}\mathbf{1}\mathbf{1}^\top\boldsymbol{x}_t = \frac{1}{n}\mathbf{1}\mathbf{1}^\top W\boldsymbol{x}_{t-1} = \frac{1}{n}\mathbf{1}\mathbf{1}^\top\boldsymbol{x}_{t-1} = \cdots = \frac{1}{n}\mathbf{1}\mathbf{1}^\top\boldsymbol{x}_0 = \frac{1}{n}\mathbf{1}\mathbf{1}^\top\boldsymbol{x} \tag{7}$$

In order to prove convergence of $\boldsymbol{x}_t$ to $\boldsymbol{x} = \mu\mathbf{1}$, we first observe that

$$\max_v \left|x_t(v) - \mu\right| = \|\boldsymbol{x}_t - \boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}_t - \boldsymbol{x}\|$$

because the infinity norm is never larger than the Euclidean norm. Hence, it is enough to measure how fast $\|\boldsymbol{x}_t - \boldsymbol{x}\|$ vanishes as $t \to \infty$. We use the operator norm $\|W\|$ of a symmetric matrix W, which is the largest absolute value of an eigenvalue of W. For any vector $\boldsymbol{z}$, we have the following inequality: $\|W\boldsymbol{z}\| \leq \|W\|\,\|\boldsymbol{z}\|$. Using that inequality and (7) we can write

$$\|\boldsymbol{x}_{t+1} - \boldsymbol{x}\| = \left\|W\big(\boldsymbol{x}_t - \boldsymbol{x}\big)\right\| = \left\|\left(W - \frac{\mathbf{1}\mathbf{1}^\top}{n}\right)\big(\boldsymbol{x}_t - \boldsymbol{x}\big)\right\| \leq \left\|W - \frac{\mathbf{1}\mathbf{1}^\top}{n}\right\|\,\|\boldsymbol{x}_t - \boldsymbol{x}\|$$

Fact 11 *If G is not bipartite, then*

$$\left\|W - \frac{\mathbf{1}\mathbf{1}^\top}{n}\right\| = \max\big\{|\omega_2|, |\omega_n|\big\}$$

PROOF. By Fact 10, $\omega_1 = 1$ with eigenvalue $\boldsymbol{u}_1 = \frac{1}{\sqrt{n}}\mathbf{1}$. Let $W = U\Lambda U^\top$ be the spectral decomposition of W, where $\Lambda = \mathrm{diag}(1,\omega_2,\ldots,\omega_n)$ and $U = \big[\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\big]$. Let M be the $n \times n$ diagonal matrix $\mathrm{diag}(1,0,\ldots,0)$. Then, $UMU^\top = \boldsymbol{u}_1\boldsymbol{u}_1^\top = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ and

$$\left\|W - \frac{\mathbf{1}\mathbf{1}^\top}{n}\right\| = \left\|U(\Lambda - M)U^\top\right\| = \left\|U\,\mathrm{diag}(0,\omega_2,\ldots,\omega_n)U^\top\right\| = \max\big\{|\omega_2|, |\omega_n|\big\}$$

because G is not bipartite and so $|\omega_n| < 1$. $\square$

Now let $\kappa = \max\{|\omega_2|, |\omega_n|\}$. We have that $\max_v |x_t(v) - \mu| \le \kappa^t \|\boldsymbol{x}_0 - \boldsymbol{x}\|$, implying that the speed of convergence is dictated by the spectrum of the gossip matrix W if the underlying graph is not bipartite.

A reasonable choice for the gossip matrix is $W = I - \alpha L$, where $0 < \alpha < 1/\big(2\Delta(G)\big)$ and $L = D - A$ is the unnormalized Laplacian of G (check that this choice of W is indeed a gossip matrix). The eigenvalues of W are thus $\omega_i = 1 - \alpha\lambda_i$, where $0 = \lambda_1 < \lambda_2 \le \cdots \le \lambda_n$ are the eigenvalues of L. Recall (Theorem 9) that $\lambda_n < 2\Delta(g)$ because the largest eingenvalue of L_{norm} for a non-bipartite graph is smaller than 2. Hence, $\omega_n = 1 - \alpha\lambda_n > 0$, implying $\kappa = \omega_2 = 1 - \alpha\lambda_2 < 1 - \lambda_2/\big(2\Delta(G)\big)$. The update (6) in this case can be written as

$$\boldsymbol{x}_{t+1}(v) = \boldsymbol{x}_t(v) + \alpha \sum_{u\,:\,(u,v)\in E} \big(\boldsymbol{x}_t(u) - \boldsymbol{x}_t(v)\big)$$

and the speed of convergence is dictated by $\lambda_2/\Delta(G)$,

$$\omega_2^t \le \left(1 - \frac{\lambda_2}{\Delta(G)}\right)^t \le \exp\left(-\frac{\lambda_2 t}{\Delta(G)}\right)$$

Exercises

1. Show that if G is connected and $\mu_1 = \Delta(G)$, then G is $\Delta(G)$-regular.
2. Prove Fact 4.