Lecture 10: Page Ranking

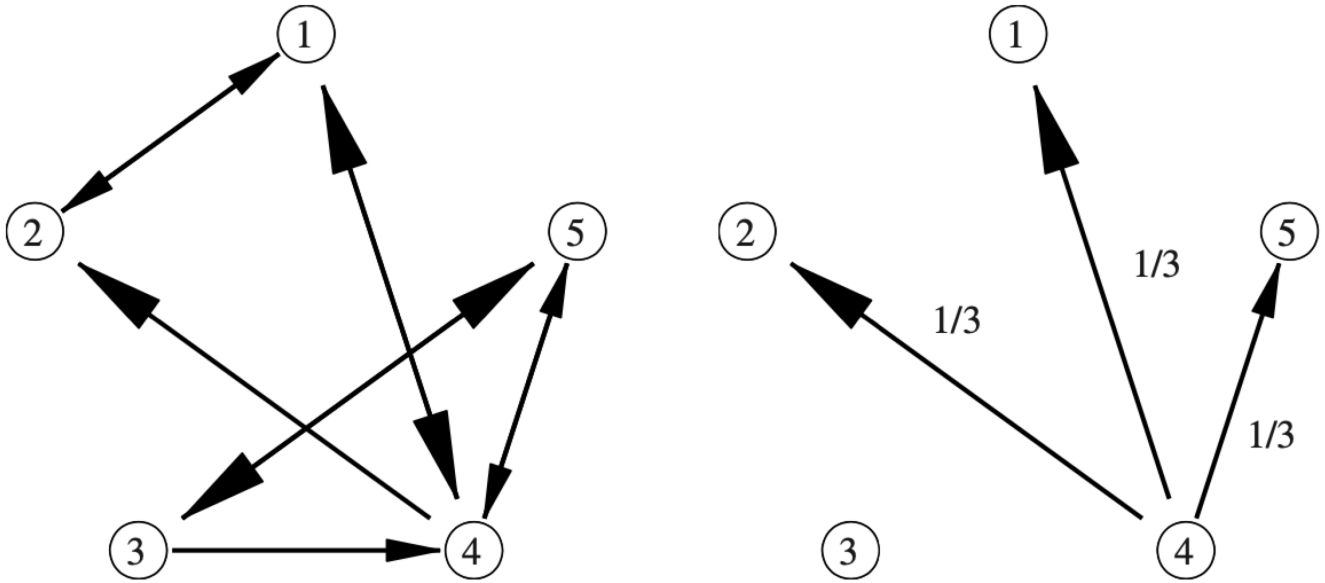
School of Mathematical Sciences, Xiamen University

1. A simple model problem

- Consider a small network consisting of *nodes* and *directed links*.

  The nodes of the network $\leftrightarrow$ sites of an internet.

  Site A contains a link to Site B $\leftrightarrow$ a directed link from A to B.



The network (left) and the transition probabilities for moving from node 4 (right).

- Random walk: when surfing the network randomly, always going from one node to another by selecting one of the directed links randomly, the most visited node would seem to be the most important one.
- Transition matrix: defining the probabilities of transition between the nodes. For example:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 1/3 & 0 \\ 1/2 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/3 & 0 \end{bmatrix}$$

The kth column of $\mathbf{P}$ contains the transition probabilities from the kth node. Hence, if we start from node 1, the probability density of the next state is

$$\pi_2 = \mathbf{P}\pi_1 = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}^\top, \quad \pi_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}^\top.$$

- Observe that the choice of π_1 reflects the fact that initially we are at node 1 with certainty. Similarly, the probability density after n steps is

$$\pi_n = \mathbf{P}\pi_{n-1} = \mathbf{P}^2\pi_{n-2} = \cdots = \mathbf{P}^n\pi_1.$$

- Define a sequence of random variables $\mathbf{x}^n \in \mathbb{R}^5$ as follows: the components of $\mathbf{x}^n$ all vanish except one, and the index of the non-vanishing component is drawn from the density π_n. Hence, in the previous example, the possible realizations of the random variable $\mathbf{x}^2$ are $\mathbf{e}_2$ and $\mathbf{e}_4$, and

$$\mathbb{P}\{\mathbf{x}^2 = \mathbf{e}_2\} = \mathbb{P}\{\mathbf{x}^2 = \mathbf{e}_4\} = 1/2.$$

- If we want to move one step further, we have

$$\pi_3 = \mathbf{P}^2\pi_1 = \begin{bmatrix} \frac{2}{3} & \frac{1}{6} & 0 & 0 & \frac{1}{6} \end{bmatrix}.$$

- These are the probabilities of the third state, assuming that we know that the sequence starts at node 1, i.e.,

$$\pi(\mathbf{x}^3 \mid \mathbf{x}^1) = \pi_3, \quad \mathbf{x}^1 = \mathbf{e}_1.$$

But what if we know also the second state? Suppose that $\mathbf{x}^2 = \mathbf{e}_2$. Then, evidently, the probability distribution of $\mathbf{x}^3$ becomes

$$\pi(\mathbf{x}^3 \mid \mathbf{x}^1, \mathbf{x}^2) = \mathbf{P}\mathbf{e}_2 = \mathbf{e}_1, \quad \mathbf{x}^1 = \mathbf{e}_1, \ \mathbf{x}^2 = \mathbf{e}_2.$$

The relevant point here is that this result does not depend on the value of $\mathbf{x}^1$: whichever way we came to the second node $\mathbf{x}^2 = \mathbf{e}_2$, the probability density of $\mathbf{x}^3$ is always $\mathbf{P}\mathbf{e}_2$, that is,

$$\pi(\mathbf{x}^3 \mid \mathbf{x}^1, \mathbf{x}^2) = \pi(\mathbf{x}^3 \mid \mathbf{x}^2).$$

- More generally, if we know the state $\mathbf{x}^n$, the next state has density $\mathbf{P}\mathbf{x}^n$ regardless of the more remote past states. We may write

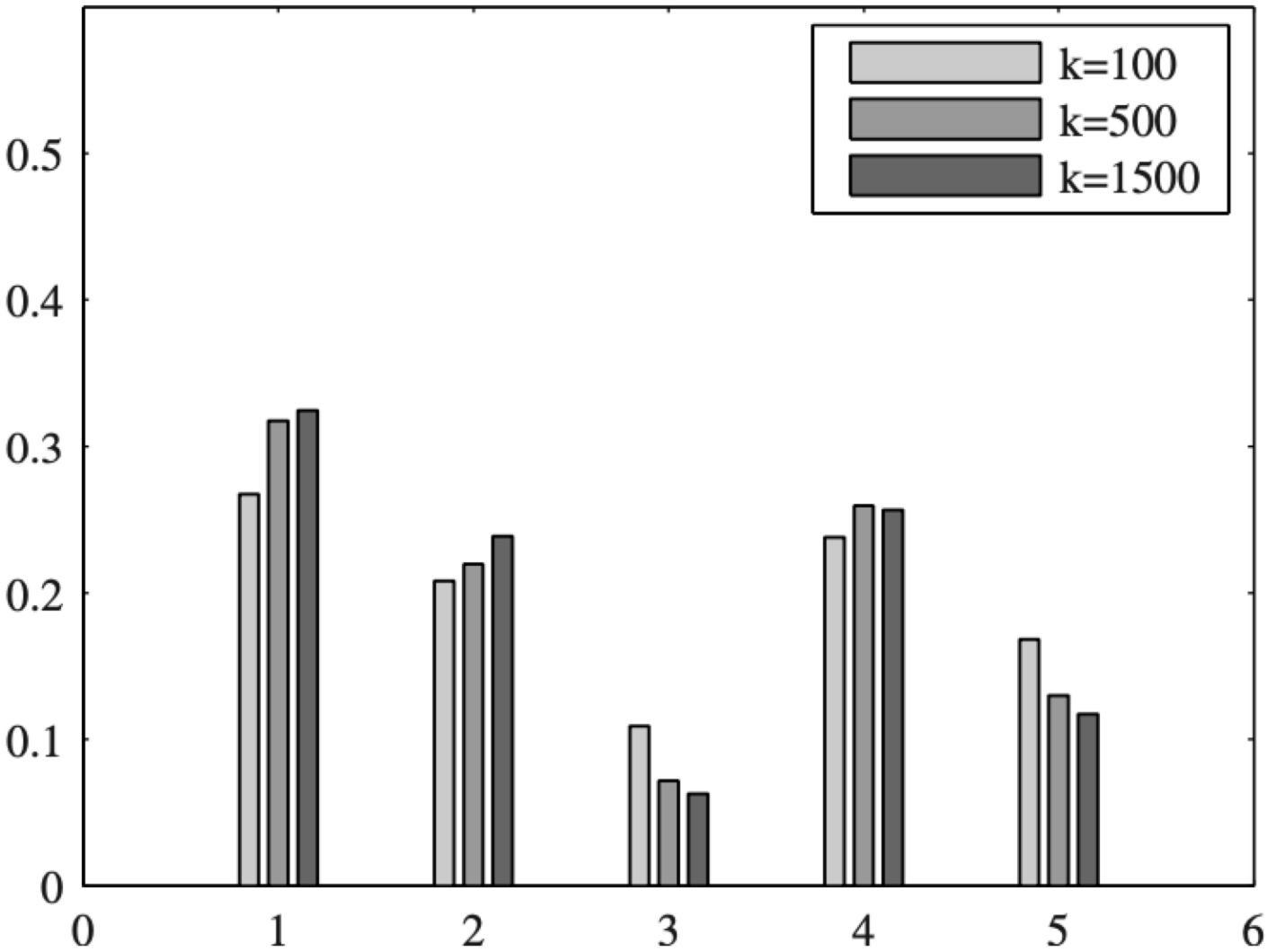
$$\pi(\mathbf{x}^{n+1} \mid \mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^n) = \pi(\mathbf{x}^{n+1} \mid \mathbf{x}^n),$$

  and say that the discrete time stochastic process $\{\mathbf{x}^1, \mathbf{x}^2, \cdots\}$ is a Markov process.
- Consider now the question of how to assess the importance of the nodes. One way is to start a realization of the Markov chain at some node, say, $\mathbf{x}^1 = \mathbf{e}_1$, generate a sample,

$$S = \{\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^k\},$$

  by drawing $\mathbf{x}^j$ from the probability density $\mathbf{P}\mathbf{x}^{j-1}$ and compute the visiting frequency of the different nodes.

- The relative visiting frequencies at different nodes with three different chain lengths $k = 100, 500, 1500$



- Finally, let us search for a more efficient way of reaching the result that we obtained above by random walk. Assume that, as the number of iterations grows, the probability distribution converges towards a limit distribution,

$$\lim_{n\to\infty} \mathbf{P}^n \pi_1 = \pi_\infty.$$

Is there a simple way to find the limit distribution? By writing

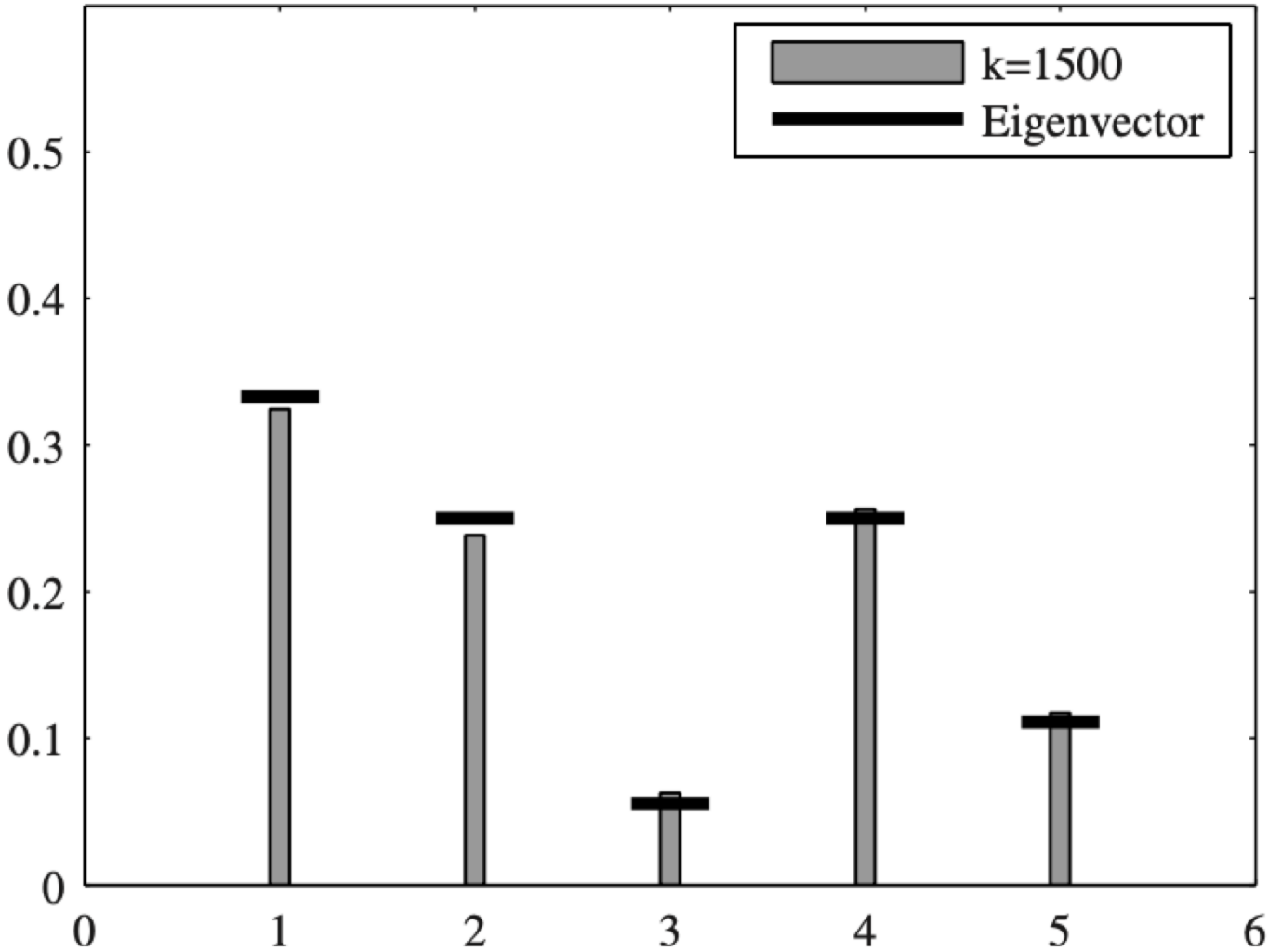
$$\pi_{n+1} = \mathbf{P}\pi_n,$$

and taking the limit on both sides, we obtain

$$\pi_\infty = \mathbf{P}\pi_\infty.$$

In other words, the limit distribution, if it exists, is the normalized eigenvector (scale to have unit 1-norm) of the transition matrix $\mathbf{P}$ corresponding to the unit eigenvalue.

- The relative visiting frequencies and the components of the eigenvector corresponding to eigenvalue one.



2. Eigenvalue analysis

- Let λ be an eigenvalue of the transition matrix $\mathbf{P}$. We have $|\lambda| \leq 1$.

  Proof. Consider eigenpair $(\lambda, \mathbf{v})$, i.e., $\mathbf{P}\mathbf{v} = \lambda\mathbf{v}$. We have

$$
\begin{aligned}
|\lambda| \|\mathbf{v}\|_1 = \|\mathbf{P}\mathbf{v}\|_1 &= \sum_i \left| \sum_j \mathbf{P}_{ij} \mathbf{v}_j \right| \\
&\leq \sum_i \sum_j \mathbf{P}_{ij} \left| \mathbf{v}_j \right| \\
&= \sum_j \sum_i \mathbf{P}_{ij} \left| \mathbf{v}_j \right| \\
&= \sum_j \left| \mathbf{v}_j \right| = \|\mathbf{v}\|_1 .
\end{aligned}
$$

- If $\mathbf{P}$ is a transition matrix with all entries positive, i.e., $\mathbf{P}_{ij} > 0$, then $\lambda = 1$ is an eigenvalue, and the corresponding eigenvector can be chosen so that all its entries are positive. (Frobenius–Perron.)

- Given a transition matrix $\mathbf{P} \in \mathbb{R}^{N\times N}$, we replace $\mathbf{P}$ with a nearby approximation. For $0 < \varepsilon < 1/N$, define

$$\mathbf{P}_\varepsilon = (1 - N\varepsilon)\mathbf{P} + \varepsilon \mathbf{1}\mathbf{1}^\top.$$

  We observe that since $N\varepsilon < 1$, all the entries of $\mathbf{P}_\varepsilon$ are positive. Moreover, $\mathbf{P}_\varepsilon$ is also a transition matrix, because its column sums are equal to 1, i.e., $\mathbf{P}_\varepsilon^\top \mathbf{1} = \mathbf{1}$..

- Assume that $\mathbf{P}$ has eigenvalues $\{1, \lambda_2, \cdots, \lambda_N\}$, $|\lambda_i| \leq 1$. Then $\mathbf{P}_\varepsilon$ has eigenvalues $\{1, \alpha\lambda_2, \cdots, \alpha\lambda_N\}$, where $\alpha = (1 - N\varepsilon) < 1$. In particular,

$$|\alpha\lambda_i| \leq \alpha < 1, \quad i = 2, 3, \cdots, N.$$

  Therefore, there is a spectral gap of size at least $1 - \alpha = N\varepsilon$ between the largest and second largest eigenvalues of $\mathbf{P}_\varepsilon$.

  Proof. Exercise.

PageRank algorithm

1. Construct the network and the transition matrix $\mathbf{P} \in \mathbb{R}^{N \times N}$.
2. Choose a small parameter ε, $0 < \varepsilon < 1/N$, and let
$$\mathbf{P}_\varepsilon = (1 - N\varepsilon)\mathbf{P} + \varepsilon \mathbf{1}\mathbf{1}^\top.$$
3. Use the power method to find the eigenvector $\mathbf{v}$ of $\mathbf{P}_\varepsilon$ corresponding to the largest eigenvalue, and scale it so that $\mathbf{v}_j > 0$ for all j, $1 \le j \le N$.
4. Sort the pages in the descending order of the entries of $\mathbf{v}$.

- The matrix-vector product $\mathbf{P}_\varepsilon \mathbf{z}$ can be computed by
$$\mathbf{P}_\varepsilon \mathbf{z} = (1 - N\varepsilon)\mathbf{P}\mathbf{z} + \varepsilon \mathbf{1}^\top \mathbf{z} \mathbf{1},$$
which takes advantage of the sparsity of $\mathbf{P}$.