Graph Theory and Applications

Session 4 — Scores in a social network and on the web

Master 1 APE

1 Problem 1 — Scores in a social network and on the web

Let $G = (S, A)$ be an undirected graph. To motivate this exercise, we will assume that each vertex $i \in S$ represents an agent, while each edge $(i, j) \in A$ represents a connection. When a graph is interpreted in this way, it is referred to as a social network. A significant portion of the literature on social networks seeks to study ways of measuring an agent's prestige, influence, or power based on the agents they are connected to or, more generally, their position in the network. As we will discuss in the final question, this exercise focuses on a measure very similar to that used by Google to rank websites in search results. The idea is as follows: we want to assign a score $x_i \geq 0$ to each vertex $i \in S$ based on the premise that a vertex's score increases with its connection to vertices with high scores. More precisely, the score $x_i \in [0, +\infty[$ of a vertex $i \in S$ is proportional to the scores of its neighbors $V(i) := \{j \in S, (i, j) \in A\}$. Formally, a graph's score is a vector $(x_i)_{i \in S}$ such that $x_i \in [0, +\infty[$ for each $i \in S$ and there exists $\lambda > 0$ such that for each $i \in S$,

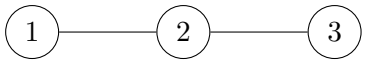
$$x_i = \frac{1}{\lambda} \sum_{j \in V(i)} x_j. \tag{1}$$

The issue with this definition is that it is self-referential. The score of an agent depends on the scores of their neighbors, which in turn depend on the scores of their neighbors, and so forth. When such a problem arises, a common approach is to try to find a "fixed point." This is what we will do in this exercise.

1.1. Show that if $(x_i)_{i \in S}$ is a score of the graph G, then for any $k \geq 0$, $(kx_i)_{i \in S}$ is also a score of the same graph G. In particular, observe that the zero vector is always a possible score.

To select a non-zero score, we will impose that $\sum_{i \in S} x_i = 1$ in the following. We will refer to this as a normalized score, meaning a nonnegative vector whose coordinates satisfy equation (1) and sum to 1.

1.2. Compute the unique (strictly positive) value of λ and the unique possible normalized score for the following graph.



1.3. Write the adjacency matrix of the above graph. We will denote this matrix by M.

1.4. Let $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ be a column vector in $[0, +\infty[^3$. Compute the product $M \cdot x$.

1.5. Deduce that if the coordinates of $x = (x_1, x_2, x_3)^\mathsf{T}$ represent the sought-after score, then $M \cdot x = \lambda x$.

1.6. Using your knowledge of linear algebra, determine the nature of x and λ for x to be non-zero.

1.7. Using your knowledge of linear algebra and the calculations already performed in question 2, verify that the three eigenvalues of M are $\sqrt{2}$, $-\sqrt{2}$, and 0. *Comment: This question can help you verify the accuracy of your results from question 2.*

1.8. Which method (using linear algebra) would you propose to calculate the normalized score of a more complex network?

1.9. Google ranks web pages in a similar way. With a few details aside, it views the web as a strongly connected directed graph (S, A) where the vertices $S = \{1, \ldots, s\}$ are the pages, and the arcs are the links going from one page to another. Each arc is assigned a weight, corresponding to the probability of moving from one page to the next. For simplicity, assume that every page has at least one outgoing link and that if there are n links from a page, each link has a weight of $\frac{1}{n}$. The score $x = (x_1, \ldots, x_s)^\mathsf{T}$ assigned to each vertex is a positive vector whose coordinates sum to 1 and satisfies the following matrix equation:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_s \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \vdots & \vdots \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{pmatrix}^\mathsf{T} \begin{pmatrix} x_1 \\ \vdots \\ x_s \end{pmatrix},$$

where for each $i, j \in S$, p_{ij} represents the probability of moving from site i to site j. We will assume that at least one site refers to itself, so there exists $i \in S$ such that $p_{ii} > 0$. Explain why such a score x exists and is unique. What additional role does the self-link play if we compute the score by repeatedly applying the transition matrix?