Link Analysis in Graphs: PageRank The Math Behind PageRank

Matrices

A matrix $A = [a_{ij}]$ is called positive iff for all i and j, $a_{ij} > 0$.

Nonnegative matrices. A matrix $A = [a_{ij}]$ is called non-negative iff for all i and j, $a_{ij} \geq 0$.

Eigenvalues and eigenvectors. Let A be a matrix, $\bar{x}$ be a vector and λ be a scalar (number). If

$$Ax = \lambda x,$$

then λ is called an **eigenvalue** and x is called an **eigenvector** of A.

The set $\sigma(A)$ of **eigenvalues** of A is called a **spectrum** of A. The **spectral radius** of A, $\rho(A)$ is

$$\rho(A) = \max_{\lambda \in \sigma(A)} (|\lambda|).$$

The circle with the radius $\rho(A)$ centered at the origin is called the **spectral circle** of A.

The **eigenvectors** of A are all the roots of the **characteristic polynomial** $p(\lambda)$ of A:

$$p(\lambda) = det(A - \lambda I),$$

where I is the unit matrix (diagonal matrix with 1 on the diagonal and 0 everywhere else).

The **algebraic multiplicity** of an **eigenvalue** λ, denoted $algmult_A(\lambda)$ is the number of times it is repeated as the root of $p(\lambda)$. λ is a **simple eigenvalue** if $algmult(\lambda) = 1$.

Perron's Theorem for Positive Matrices. Let $A = [a_{ij}]$ be a **positive** matrix. Let $r = \rho(A)$ be its *spectral radius*. The following statements hold:

1. $r > 0$. r is called the *Perron root*.
2. $r \in \sigma(A)$. r is an **eigenvalue** of r.
3. $algmult_A(r) = 1$. The Perron root (r) is simple.
4. There exists a positive vector $\hat{x}$ such that $A\hat{x} = r\hat{x}$.
5. A vector $\hat{p}$ such that $A\hat{p} = r\hat{p}$, $p > 0$, $||p||_1 = \sum |p_i| = 1$, is **unique**. (it is called the *Perron vector*).
6. r is the only value on the spectral circle of A.

Irreducible matrices. A square matrix M is said to be **irreducible** iff the graph G_M it induces is **strongly connected**, i.e., *if there is a path from every node in the graph to every other node in the graph.*

(alternatively, a square matrix M is **reducible** if there exists such a symmetric permutation of rows and columns P that transforms M into a matrix of the form

$$P^T MP = \begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix},$$

where X and Z are square. M is *irreducible* iff such a transformation is impossible.)

(A permutation matrix P is any square matrix that has exactly one 1 in each row and in each column of the matrix; all other elements of the matrix are 0s).

Question: Why are the two definitions above equivalent?

Perron-Frobenius Theorem for irreducible matrices. Let A be a *non-negative irreducible matrix*. The following statements are true:

1. $r = \rho(A) > 0$. The spectral radius of A is non-zero.
2. $r \in \sigma(A)$. r is the Perron root.
3. $algmult_A(r) = 1$. r is simple.
4. There exists $\hat{x} > 0$, such that $A\hat{x} = r\hat{x}$. ($\hat{x}$ is an eigenvector of A for the Perron root).
5. The Perron vector of A, defined as $\hat{p} > 0$, $||\hat{p}||_1 = 1$, $A\hat{p} = r\hat{p}$ is *unique*.

   There are no non-negative eigenvectors for A **except for positive multiples of $\hat{p}$, regardless of eigenvalue**.
6. r need not be the only eigenvalue on the spectral circle of A.

Primitive matrices. A matrix A is **primitive** iff A has only one eigenvalue $r = \rho(A)$ on its spectral circle.

That is, in **primitive matrices**, the ”largest” eigenvalue is unique.

A *nonnegative irreducible* matrix that has $h > 1$ eigenvalues on its spectral circle[1] is said to be **imprimitive** and h is called its *index of imprimitivity*.

Markov Chains

Stochastic Process. A set of random variables $\{X_t\}$, $t = 1 \ldots \infty$, such that $dom(X_t) = \{S_1, \ldots, S_n\}$ for all t is called a **discrete finite-state stochastic process**.

Elements of the set $S = \{S_1, \ldots S_n\}$ are called **states** and set S is called the **state space**.

Markov Chains. **A Markov chain** is a **stochastic process** that satisfies the following property (called *Markov property*):

$$Pr(X_{t+1} = S_i | X_t = S_{i_t} \wedge \ldots \wedge S_0 = S_{i_0}) = Pr(X_{t+1} = S_i | X_t = S_{i_t}.$$

The *Markov property* reads:

> The value of the random variable X at each time is conditionally dependent **only** on the value of X at the previous moment of time.

Transition probabilities. The probability

$$p(t)_{ij} = Pr(X_{t+1} = S_i | X_t = S_j)$$

is called **transitional probability** from state S_j to state S_i.

Stationary Markov Chains. A Markov Chain $\{X_t\}$ is **stationary** if its transitional probabilities do not vary over time, i.e., if for each $t, t' > 0$, and for each pair i, j

$$p(t)_{ij} = p(t')_{ij} = p_{ij}.$$

Transition probability matrix. Transition probabilities p_{ij} of a stationary Markov chain form an $n \times n$ matrix of transitional probabilities:

$$\mathbf{P} = [p_{ij}] = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}$$

[1] Some eigenvalues my be complex numbers.

Irreducible Markov Chain. A *Markov chain* is **irreducible** iff **each state** S_i **of the state space is reachable from each state** S_j.

In the terms of linear algebra, a Markov chain is **irreducible** iff its transition probability matrix is **irreducible**.

Periodic Markov Chain. A state S_i in a *Markov chain* $\{X_t\}$ is **periodic** with the period $k > 1$ iff k is the smallest number such that all paths leading from state S_i back to S_i have a length hat is a multiple of k.

A *Markov chain* is **periodic**, iff at least one its state S_i is **periodic**.

A *Markov chain* is **aperiodic**, iff all its states are **aperiodic**.

Periodic Markov Chain (revisited). A *Markov chain* is **periodic** iff it is *irriducible* and its transition matrix is **imprimitive**.

In a periodic Markov chain, each state can occur only every h steps, where h is the index of imprimitivity for the transition probability matrix.

A *Markov chain* is **aperiodic** iff its transition probability matrix is *primitive*.

Probability distribution vector. A **probability distribution vector** or **probability vector** $p^T = (p_1, \ldots, p_n)$ is a *non-negative* ($p_i \geq 0$) row vector such that

$$\sum_{i=1}^{n} p_i = 1.$$

Stationary probability distribution vector. Let $\{X_t\}$ be a *Markov chain* with transition probability matrix $\mathbf{P}$. A **stationary probability distribution vector** π^T for $\{X_t\}$ is a probability vector π^T such that

$$\pi^T \mathbf{P} = \pi^T.$$

(i.e., the vector π^T is the stationary point of the transformation $\mathbf{P}$.)

Why stationary probability distribution vectors? Stationary probability distribution vectors, **when they exist**, represent the proportion of time X_t spends in each of the states in its state space.

Question: When do stationary probability distributions vectors **exist**?

Why stationary probability distribution vectors? Part 2. Consider the following iterative schema:

$$p_0^T = (p_{01}, \ldots, p_{0n}).$$
$$p_{t+1}^T = p_t^T \mathbf{P}.$$

The sequence $p_0^T, p_1^T, \ldots, p_t^T, \ldots$ **converges** iff Markov chain $\{X_t\}$ has a **stationary probability distribution vector** π^T.

If it is the case,

$$\lim_{t\to\infty} p_t^T = \lim_{t\to\infty} p_0^T \mathbf{P}^t = \pi^T.$$

Why this is all important (PageRank Revisited)

PageRank is based on a traversal of the World Wide Web graph $G_{WWW} = \langle V, L \rangle$, where $V = \{v_1, \ldots v_N\}$ is the set of all web pages, and L is the set of all `<a href="...">` hyperlinks connecting two pages.

Consider a user observing in his/her browser some web page $v_i \in V$. Suppose this web page has hyperlinks $(v_i, v_{j_1}), \ldots, (v_i, v_{j_s})$ on it[2]

Consider for a moment that user makes his/her decisions about further web traversal as follows:

1. After viewing page v_i, the users selects as his/her next page one of the pages $v_{j_1}, \ldots, v_{j_s}$ with probability $p_{si} > 0$. ($\sum_{k=1} s p_{si} = 1$.)
2. The probabilities p_{si} do not change over time, i.e., the user makes his/her choices with the same probability **each time** the user visits page v_i.

If the two properites of the user traversal hold than the web surfing of the user **is described by a Markov chain** $\{X_t\}$ as follows:

1. $S = V$: the state space of the Markov chain is the set of all web pages.
2. Transitional probability matrix has $p_{ij} > 0$ for $(v_i, v_j) \in L$ (there is a link from v_i to v_j) and $p_{ij} = 0$ otherwise.

Let $\mathbf{P} = [p_{ij}]$ be the transition probability matrix for the Markov chain described above.

Suppose $q_0^T = (q_{01}, \ldots, q_{0N})$ is the probability distribution specifying the probability of the user to select a starting web page, i.e.,

$$Pr(X_0 = v_i) = q_{0i}.$$

Then, the probability q_{1j}^T, of the user electing to visit page v_j on the first step of the web traversal can be specified as

$$q_{1j} = Pr(X_1 = v_j) = \sum_{k=1} N Pr(X_1 = v_j | X_0 = v_k) \cdot Pr(X_0 = v_k) = \sum_{k=1} N p_{jk} \cdot q_{0k}.$$

Or, in vector notation:

$$q_1^T = q_0^T \mathbf{P}.$$

Similarly, we obtain:

$$q_{t+1}^T = q_t^T \mathbf{P} = q_{t-1}^T \mathbf{P}\mathbf{P} = \ldots = q_0^T \mathbf{P}^{t+1}.$$

A **stationary probability distribution** for this Markov chain specifies *for each web page v_i the percentage of time a user spends visiting it.* This is the **core idea** behind **PageRank** — the more often the page is visited, the more important it is.

[2] Multiple hyperlinks pointing to one page are counted as one hyperlink.

Question: Does the sequence of vectors $q_0^T, q_1^T \ldots, q_t^T, \ldots$ converge? I.e., is there a **stationary probability distribution** for the web traversal process? I.e., *can we compute the eventual probability of visiting page v on a step of the traversal*?

Finding the Right Computation For PageRank

Existence of stationary probability distribution (the Ergodic Theorem). Let $\{X_t\}$ be a Markov chain with the state space S and the probability transition matrix $\mathbf{P}$. $\{X_t\}$ has a **unique stationary probability distribution** iff

1. $\{X_t\}$ is **irreducible** (i.e., each state is reachable from each state).
2. $\{X_t\}$ is **aperiodic**.

(such Markov chains are called **ergodic**, hence the name of the theorem.)

Idea of PageRank. **PageRank** of a web page is the percentage of time the user spends observing this web page over a large number of traversal steps. As such, **PageRank** is **stationary probability distribution** of a *Markov chain describing the traversal of the World Wide Web.*

Fixing the World Wide Web Traversal

The World Wide Web traversal procedure described above has the following issues:

1. **Non-stochastic transition matrix.** If a web page v_i has **no out-links**, $p_{ji} = 0$ for any $v_j \in V$. This means that $\sum_{j=1} N p_{ji} = 0 \neq 1$.
2. **Reducible transition matrix.** $\mathbf{P}$ is irreducible if there is a path from every web page to every other web page. This, generally speaking, need not be the case.
3. **Periodic transition matrix.** $\mathbf{P}$ is periodic, if for at least one web page $v \in V$, there is a periodicity $k > 1$ with which the user can return to the page (i.e., if the user, starting at v, needs to visit a multiplicative of k pages prior to returning back to v).

Example. Consider the three cases illustrated on Figure **??**.

(A) Graph (A) illustrates the situation when two web pages (A and C) do not have *out-links*. If a user starts on page B, (s)he will traverse to either A or C on step 1, but will not be able to apply the *select an out-link and follow it* rule any further.

(B) Graph (B) illustrates the situation when the web graph has disconnected components. Pages C and C are unreachable from pages A and B and vice

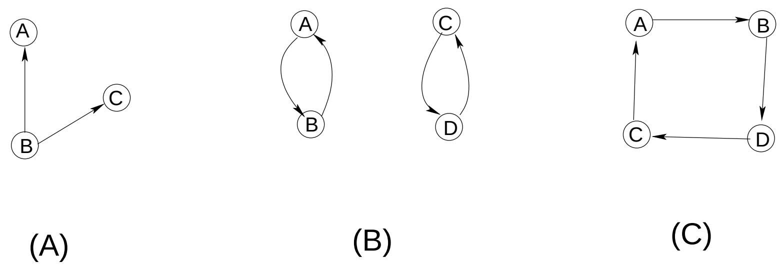

Figure 1: Problems with web structure. (A) ”Sinks”: web pages without links cause transition probability matrix to be non-stochastic. (B) Disconnected web graph causes transition probability matrix to be reducible. (C) Periodic web pages (each web page is visited with the period of 4) cause transition probability matrix to be periodic.

versa. The transition probability matrix for this case will look as follows:

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

This matrix is **reducible**.

(C) Graph (C) illustrates a fully-connected set of pages, which yields an **irreducible** transition probability matrix. However, in order to get from page A back to page A, the user **must** visit pages B, D and C in succession, causing page A to be periodic with period 4 (four steps from A to A). Note, that all other pages in this graph are also periodic. The transition probability matrix is

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

This matrix is *irreducible* but **periodic**. $\mathbf{P} = \mathbf{P}^4$.

To be able to obtain rank of a page we must construct a Markov chain for web traversal that alleviates each of the problems above.

Fixing the Markov Chain for Web Traversal

Stochastic Matrix. To make the **transaction probability matrix** stochastic we can do one of two things:

1. **Exclude** all ”sinks”, i.e., web pages with no links from consideration (at least for now). That is, define graph $G_{WWW} = \langle V, L\rangle$ in which V is the set of all web pages **with at least one outgoing link**.

   What it does. The transition probability matrix now will have **at least one** non-zero entry in each row. All such rows will obey the stochastisity condition (values in the row add to 1).

2. **Change** the traversal. Add to the traversal the following condition:

If the user is observing a page v which has no out-links, then on the next step the user randomly selects a web page from V to visit.

What it does. This approach replaces the row of zeroes ($p_{ij} = 0$) for a page v_i with no out-links with the row $p_{ij} = \frac{1}{N}$. We use random selection with a uniform distribution to ensure that each page has equal chance of being visited on the next step.

Irreducible Matrix. We need to ensure that regardless of the "real" structure of the Web, the graph for the transition probability matrix is strongly connected. We can do it by implementing the following change to the main traversal rule:

On each step, the user either follows one of the links on the page, or, the user gets bored with link-by-link traversal of the web, and instead, visits a randomly selected web page from V.

We formalize this rule, by assuming that the user will do the former with probability d and the latter with probability $1 - d$.

What it does. For each $(v_i, v_j) \notin L$, then the original probability $p_{ij} = 0$ is replaced with the non-zero probability $p_{ij} = (1 - d) \cdot \frac{1}{N}$.

For each $v_i \in V$, if $(v_i, v_j) \in L$, the probability p_{ij} is multipled by d — the probability of choosing to follow a link. Additionally, it is increased by $(1-d) \cdot \frac{1}{N}$: the chance of getting to page v_j from page v_i via the *"I am bored"* procedure.

This **guarantees** to make the matrix $\mathbf{P}$ **irreducible**, as the graph induced by $\mathbf{P}$ will now be *strongly connected*.

Aperiodic Matrix. In addition to making $\mathbf{P}$ irreducible, we must ensure that it becomes **aperiodic**.

Turns out, that the change in the main traversal rule, proposed above, takes care of that. This is because, for each $v_i \in V$, $p_{ii} \neq 0$ in the new matrix. Therefore, there is a non-zero probability of going from any page v_i to itself in one step, which is counter to the definition of a periodic state. Each v_i is **aperiodic** and therefore $\mathbf{P}$ is primitive.

Creating Transition Probability Matrix. Our final goal is to actually instantiate the transition probability matrix $\mathbf{P}$.

The original web traversal process did not specify the values of p_{ij}s. However **we do need to specify them**. This can be done, by augmenting the traversal rule as follows:

On each step, the user either follows one of the links on the page, **choosing each link with equal probability**, or, the user gets bored with link-by-link traversal of the web, and instead, visits a randomly selected web page from V.

What it does. This rule allows us to define transition probabilities as follows:

$$p_{ij} = \begin{cases} (1-d)\cdot\frac{1}{N}: & \text{if } (v_i,v_j)\notin L; \\ d\cdot\frac{1}{N_i}+(1-d)\cdot\frac{1}{N}: & \text{if } (v_i,v_j)\in L. \end{cases}$$

Here N_i is the number of out-links from page v_i.

Payoff

The combined traversal rules:

1. User starts at some page $v \in V$ selected randomly with equal probability.
2. On each step, the user either follows one of the links on the page, **choosing each link with equal probability**, or, the user gets bored with link-by-link traversal of the web, and instead, visits a randomly selected web page from V.

These rules define a Markov chain with the transition probability matrix $\mathbf{P} = [p_{ij}]$ defined as follows:

$$p_{ij} = \begin{cases} (1-d)\cdot\frac{1}{N}: & \text{if } (v_i,v_j)\notin L; \\ d\cdot\frac{1}{N_i}+(1-d)\cdot\frac{1}{N}: & \text{if } (v_i,v_j)\in L. \end{cases}$$

The **PageRank** of a page $v_j \in V$ is computed as

$$PageRank(v_j) = (\lim_{t\to\infty} q_0^T\mathbf{P}^t)[j],$$

where $q_0^T = \left(\frac{1}{N},\ldots,\frac{1}{N}\right)$[3].

Because $\mathbf{P}$ is **stochastic**, **irreducible** and **primitive**, the process above converges to a unique **stationary probability distribution** π^T:

$$\pi^T = \pi^T\mathbf{P}.$$

$$PageRank(v_j) = \pi_j^T.$$

Multiplying π^T by $\mathbf{P}$ yields:

$$\pi_j = \sum_{(v_i,v_j)\notin L} p_{ij}\cdot\pi_i + \sum_{(v_i,v_j)\in L} p_{ij}\cdot\pi_i = \sum_{(v_i,v_j)\notin L}(1-d)\cdot\frac{1}{N}\cdot\pi_i + \sum_{(v_i,v_j)\in L}\left(d\cdot\frac{1}{N_i}+(1-d)\cdot\frac{1}{N}\cdot\pi_i\right) =$$

$$\cdot\sum_{i=1}^{N}(1-d)\cdot\frac{1}{N}\cdot\pi_i + \sum_{(v_i,v_j)\in L} d\cdot\frac{1}{N_i}\cdot\pi_i = \frac{(1-d)}{N}\cdot\sum_{i=1}^{N}\pi_i + \sum_{(v_i,v_j)\in L} d\cdot\frac{1}{N_i}\cdot\pi_i =$$

[3]Note, that for ergodic Markov chains, the sequence converges **regardless of the starting probability distribution**.

$$\frac{(1-d)}{N} + d \cdot \sum_{(v_i,v_j)\in L} \frac{1}{N_i} \cdot \pi_i.$$

This forms the set of (linear) equations, which can be solved in any known manner (e.g., Gaussian reduction). However, because N is **very large**, direct solution methods turn out to be inefficient, and we resort to the approximation scheme:

$$\pi_0^T = \left(\frac{1}{N}, \ldots, \frac{1}{N}\right)$$

$$\pi_{k+1}[j] = \frac{(1-d)}{N} + d \cdot \sum_{(v_i,v_j)\in L} \frac{1}{N_i} \cdot \pi_k[i].$$

Because $\mathbf{P}$ is **stochastic**, **irreducible** and **primitive**, the approximation process converges.

References

[1] Page, Lawrence; Brin, Sergey; Motwani, Rajeev and Winograd, Terry (1998). The PageRank citation ranking: Bringing order to the Web. *Technical Report*, *Department of Computer Science*, *Stanford University*. http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf

[2] Amy N. Langville, Carl D. Meyer. *Google's PageRank and Beyond: the Science of Search Engine Rankings*, Princeton University Press. 2006.