Chapter 6 PageRank

Angsheng Li

Institute of Software
Chinese Academy of Sciences

Advanced Algorithms, U CAS
11th, April, 2016

Outline

1. Backgrounds
2. Web graph
3. Google's matrix
4. Teleportation
5. Personalised vector
6. Sensitivity
7. Proofs
8. Local algorithms

The new phenomena

Brin and Page, 1995 - 1998

1. The current-generation search engine
2. Billions of queries everyday
3. What is the principle behind?
4. How good is the current-generation search engine?

The graph

- Massive directed graph
- Nodes: webpages
- Directed edges, hyperlines, including inlinks and outlinks
- The question: Rank the web pages by importance.

The PageRank thesis

A page is important, if it is pointed to by many important pages.

Brin and Page, 1998

Established the equation of the PageRank thesis.
The PageRank of a page P_i, written $r(P_i)$, is the sum of the PageRanks of all the pages pointing to P_i, that is,

$$r(P_i) = \sum_{P_j \in B_i} \frac{r(P_j)}{|P_j|}, \tag{1}$$

- B_i: the set of pages pointing to P_i,
- $|P_j|$: the number of outlinks from page P_j.

Recurrence of the PageRank

$$\begin{cases} r_{k+1}(P_i) = \sum_{P_j \in B_i} \frac{r_k(P_j)}{|P_j|} \\ r_0(P_i) = \frac{1}{n} \end{cases} \tag{2}$$

The stationary solution of the recursive equation in Equation (2) gives rise to the PageRank of a graph G.

Matrix representation

$$H_{ij} = \begin{cases} \frac{1}{|P_i|} & \text{if there is an edge from node } i \text{ to node } j, \\ 0 & \text{o.w.} \end{cases} \tag{3}$$

$|P_i|$: The number of outlinks from node i.
$H = (H_{ij})$ is the PageRank matrix of G.

PageRank solution

Let π^{T} be a $1 \times n$ vector.
Set

$$\begin{cases} \pi^{(k+1)\mathrm{T}} = \pi^{(k)\mathrm{T}} H, \\ \pi^{(0)\mathrm{T}} = \frac{1}{n} e^{\mathrm{T}}, \end{cases} \tag{4}$$

where $e^{\mathrm{T}} = (1, 1, \cdots, 1)$.
For the equation (4), we require:

- convergence and the interpretation of the solution
- Uniqueness of the solution
- Invariance of $\pi^{(0)}$
- The number of iterations of the convergent solution

Rank sinks

$$\begin{tikzcd}
1 \arrow[rr, shift left] \arrow[dr] & & 2 \arrow[ll, shift left] \arrow[dl] \\
& 3 &
\end{tikzcd}$$

All the PageRanks go to node 3.

Matrix S

To solve the sink problem, define a vector **a**,

$$a_i = \begin{cases} 1 & \text{if node } i \text{ has no outgoing links,} \\ 0 & \text{o.w.} \end{cases} \tag{5}$$

Definition

Define

$$S = H + \frac{1}{n}\mathbf{a}\mathbf{e}^{\mathrm{T}},$$

where $\mathbf{e}^{\mathrm{T}} = (1, 1, \cdots, 1)$.

Intuition: If node i has no outgoing link, then from node i, the randomly walks to any other nodes uniformly.

S is the transition probability matrix of a Markov chain.

Google's matrix G

Definition

Define the Google's matrix by

$$G = \alpha S + (1-\alpha)J,$$

where $J_{ij} = \frac{1}{n}$.

- J is called *teleportation matrix*
- $1-\alpha$ is called the *teleportation parameter*.

Expander

Recall: If G is a graph with $\lambda = \lambda(G) < 1$, then for $A = A_G$,

$$A = (1-\lambda)J + \lambda C,$$

for some C with $\|C\| \leq 1$.
We thus know that Google's matrix is an expander. However, the parameter α is chosen arbitrarily. Of course, α determines the spectral gap of the graph.

Properties of G - I

(1) G is stochastic
It is a convex combination of two stochastic matrices S and J.

(2) G is irreducible.
Every page is directly connected to every other page.

(3) G is aperiodic.
$G_{ii} > 0$. Every node has a self-loop.

(4) G is primitive.
There exists a k such that $G^k > 0$
Because: G is an expander. There is a unique π^{T} such that

$$\|pG^l - \pi^{\mathrm{T}}\| \approx 0$$

for a small l. – Power method works

Properties of G - II

(5) G is rank-one updated

$$\begin{aligned} G &= \alpha S + (1-\alpha)\frac{1}{n}ee^{\mathrm{T}} \\ &= \alpha(H + \frac{1}{n}ae^{\mathrm{T}}) + (1-\alpha)\frac{1}{n}ee^{\mathrm{T}} \\ &= \alpha H + (\alpha\frac{1}{n}a + (1-\alpha)\frac{1}{n}e)e^{\mathrm{T}}. \end{aligned} \tag{6}$$

- H is sparse
- $\alpha\frac{1}{n}a + (1-\alpha)\frac{1}{n}e$ is dense, but only one-dimensional vector.

(6) G is artificial due to the choice of α.
G may not well reflect the real world H.

Computation of π^{T}

Power method

$$\begin{aligned}
\pi^{(k+1)\mathrm{T}} &= \pi^{(k)\mathrm{T}} G \\
&= \alpha \pi^{(k)\mathrm{T}} S + \frac{1-\alpha}{n} \pi^{(k)\mathrm{T}} e e^{\mathrm{T}} \\
&= \alpha \pi^{(k)\mathrm{T}} H + (\alpha \pi^{(k)\mathrm{T}} a + (1-\alpha) e) e^{\mathrm{T}} / n.
\end{aligned} \tag{7}$$

Suppose that $1, \lambda_2, \cdots, \lambda_n$ are the eigenvalues of G with $1 > |\lambda_2| \geq \cdots \geq |\lambda_n|$.
Then:

$$G = G_1 + \lambda_2 G_2 + \cdots + \lambda_n G_n,$$

– $G_i^2 = G_i$,
– For $i \neq j$, $G_i G_j = 0$.
Then

$$G^l = G_1 + \lambda_2^l G_2 + \cdots \lambda_n^l G_n$$

Since $\lambda_2 < 1$, G^l quickly converges to G_1.

$\lambda(G)$

Lemma

For the Google matrix $G = \alpha S + (1-\alpha)J$,

$$|\lambda_2(G)| \leq \alpha.$$

$\lambda(G)$ again

Lemma

If the spectrum of the stochastic matrix S is $\{1, \lambda_2, \cdots, \lambda_n\}$, then the spectrum of the Google matrix $G = \alpha S + (1-\alpha)ev^{\mathrm{T}}$ is

$$\{1, \alpha\lambda_2, \cdots, \alpha\lambda_n\},$$

where v^{T} is the personalised vector.

Proofs - I

Since S is stochastic, $(1, \mathbf{e})$ is an eigenpair of S. Let $Q = (eX)$ be a nonsingular matrix that has the eigenvector $\mathbf{e}$ as its first column.
Set

$$Q^{-1} = \begin{pmatrix} y^{\mathrm{T}} \\ Y^{\mathrm{T}} \end{pmatrix} \tag{8}$$

Then:

$$Q^{-1}Q = \begin{pmatrix} y^{\mathrm{T}}\mathbf{e} & y^{\mathrm{T}}X \\ Y^{\mathrm{T}}\mathbf{e} & Y^{\mathrm{T}}X \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & I \end{pmatrix} \tag{9}$$

Proofs - II

Similarly,

$$Q^{-1}SQ = \begin{pmatrix} y^{\mathrm{T}}e & Y^{\mathrm{T}}SX \\ Y^{\mathrm{T}}\mathbf{e} & Y^{\mathrm{T}}SX \end{pmatrix} = \begin{pmatrix} 1 & y^{\mathrm{T}}SX \\ 0 & Y^{\mathrm{T}}SX \end{pmatrix} \tag{10}$$

This implies that $Y^{\mathrm{T}}SX$ contains the remaining eigenvalues of S, i.e., $\lambda_2, \cdots, \lambda_n$.
In addition,

$$Q^{-1}GQ = \begin{pmatrix} 1 & \alpha y^{\mathrm{T}}SX + (1-\alpha)v^{\mathrm{T}}X \\ 0 & \alpha Y^{\mathrm{T}}SX \end{pmatrix} \tag{11}$$

The eigenvalues of G are

$$\{1, \alpha\lambda_2, \cdots, \alpha\lambda_n\}.$$

Since $\lambda_2 \leq 1$, $\alpha\lambda_2 \leq \alpha$.

The role α

$$G = (1-\alpha)J + \alpha S.$$

If α is small, then $1-\alpha$ is large, G is basically an artificial random graph, failing to reflect the real world matrix S.
If α is large, then

- there is no unique stationary distribution
- even if there is a stationary distribution, it is hard to compute
- the power method fails

Google's choice: $\alpha = 0.85$.

Personalised PageRank

For a personalised probability vector v^{T},

$$G = \alpha S + (1-\alpha) e v^{\mathrm{T}}.$$

The power method works as before.
The stationary distribution is a personalised PageRank.
Significance: Real applications.

The stationary distribution

Theorem

The Pagerank $\pi^{\mathrm{T}}(\alpha)$ of G_α is

$$\pi^{\mathrm{T}}(\alpha) = \frac{1}{\sum_{i=1}^{n} D_i(\alpha)} (D_1(\alpha), D_2(\alpha), \cdots, D_n(\alpha))$$

where $D_i(\alpha)$ is the i-th principal minor determinant of order $n-1$ in $I - G_\alpha$.
Furthermore, every $D_i(\alpha)$ is differentiable for α.

Proof.

By definition. $\square$

Differential

Theorem

If $\pi^{\mathrm{T}}(\alpha) = (\pi_1(\alpha), \pi_2(\alpha), \cdots, \pi_n(\alpha))$*, then*

1. *For each* j,
$$\left|\frac{d\pi_j(\alpha)}{d\alpha}\right| \leq \frac{1}{1-\alpha}.$$
2. $$\left\|\frac{d\pi^{\mathrm{T}}(\alpha)}{d\alpha}\right\|_1 \leq \frac{2}{1-\alpha}.$$

- If α is small, then the PageRank $\pi^{\mathrm{T}}(\alpha)$ is not sensitive.
- If α is large, then the upper bounds $\frac{1}{1-\alpha}$ and $\frac{2}{1-\alpha}$ are both approaching to infinity.

Representation

Theorem

$$\frac{d\pi^{\mathrm{T}}(\alpha)}{d\alpha} = -v^{\mathrm{T}}(I-S)(I-\alpha S)^{-2}.$$

Sensitive to H

1. $$\frac{d\pi^{\mathrm{T}}(h_{ij})}{dh_{ij}} = \alpha\pi_i(e_j^T - v^T)(I - \alpha S)^{-1}$$

2. $$(I - \alpha S)^{-1} \to \infty,$$

   as α goes to 1.

π^{T} is sensitive to perturbations in H is $\alpha \approx 1$.
Therefore, if $\alpha \approx 1$, then π^{T} is sensitive to small changes of the matrix H.

Sensitive to v^T

$$\frac{d\pi^{\mathrm{T}}(v^T)}{dv^T} = (1-\alpha+\alpha\sum_{i\in D}\pi_i)(I-\alpha S)^{-1},$$

D is the set of nodes that have no outgoing links.
The same as before, as α goes to 1, $(I-\alpha S)^{-1}$ goes to ∞.

Summary of sensitivity

If $\alpha \approx 1$, then

1. Computing $\pi^{\mathrm{T}}(\alpha)$ is hard, since the power method fails
2. $\pi^{\mathrm{T}}(\alpha)$ is sensitive to the perturbation of *H*
3. $\pi^{\mathrm{T}}(\alpha)$ is sensitive to the personalised vector v^{T}

Google's tradeoff:

$$\alpha = 0.85$$

Proof of upper bounds - I

Theorem

If $\pi^{\mathrm{T}}(\alpha) = (\pi_1(\alpha), \pi_2(\alpha), \cdots, \pi_n(\alpha))$, *then*

1. *For each j,*

$$|\frac{d\pi_j(\alpha)}{d\alpha}| \leq \frac{1}{1-\alpha}.$$

2.

$$\|\frac{d\pi^{\mathrm{T}}(\alpha)}{d\alpha}\|_1 \leq \frac{2}{1-\alpha}.$$

$\pi^T(\alpha)$ is a probability vector, so

$$\sum_{i=1}^{n} \pi_i(\alpha) = 1$$

giving
$\pi^{\mathrm{T}}(\alpha)e = 1$, $e^T = (1, 1, \cdots, 1)$.

Proof of upper bounds- II

By definition,
$\pi^T(\alpha) = \pi^T(\alpha)G(\alpha) = \pi^T(\alpha)(\alpha S + (1-\alpha)ev^T)$.
By differential,

$$\frac{d\pi^T(\alpha)}{d\alpha} = \pi^T(\alpha)(S - ev^T)(I - \alpha S)^{-1}. \tag{12}$$

For (1). For every real x, $xT \perp e$, i.e., $\sum x_i = 0$, and for all real vector y, column vector,

$$\begin{aligned} |x^{\mathrm{T}} y| &= |\sum_{i=1}^{n} x_i y_i| \\ &\leq \|x^{\mathrm{T}}\|_1 \cdot \frac{y_{\max} - y_{\min}}{2}. \end{aligned} \tag{13}$$

By Equation (12),

$$\frac{d\pi_j(\alpha)}{d\alpha} = \pi^{\mathrm{T}}(\alpha)(S - ev^{\mathrm{T}})(I - \alpha S)^{-1} e_j.$$

Prof of upper bounds - III

Since $\pi^{\mathrm{T}}(\alpha)(S-ev^{\mathrm{T}})e=0$, set $x^{\mathrm{T}}=\pi^{\mathrm{T}}(\alpha)(S-ev^{\mathrm{T}})$ and $y=(I-\alpha S)^{-1}e_j$.
By Inequality (13),

$$|\frac{d\pi_j(\alpha)}{d\alpha}| \le \|\pi^{\mathrm{T}}(\alpha)(S-ev^{\mathrm{T}})\|_1 \cdot \frac{y_{\max}-y_{\min}}{2}.$$

Since $\|\pi^{\mathrm{T}}(\alpha)(S-ev^{\mathrm{T}})\|_1 \le 2$,
$|\frac{d\pi_j(\alpha)}{d\alpha}| \le y_{\max}-y_{\min}$.
Since $(I-\alpha S)^{-1} \ge 0$ and $(I-\alpha S)e=(1-\alpha)e$, and hence $(I-\alpha S)^{-1}=(1-\alpha)^{-1}e$.
This shows that $y_{\min} \ge 0$.
For $y_{\max}$, we have
$y_{\max} \le \max_{i,j}[(I-\alpha S)^{-1}]_{ij} \le \frac{1}{1-\alpha}$.
(1) follows.

Proof of upper bounds - IV

For (2).

$$\begin{aligned}\left\|\frac{d\pi^{\mathrm{T}}(\alpha)}{d\alpha}\right\|_1 &= \|\pi^{\mathrm{T}}(\alpha)(S-ev^{\mathrm{T}})(I-\alpha S)^{-1}\|_1 \\ &\le \|\pi^{\mathrm{T}}(\alpha)(S-ev^{\mathrm{T}})\|_1 \cdot \|(I-\alpha S)^{-1}\|_\infty \\ &\le 2\frac{1}{1-\alpha} = \frac{2}{1-\alpha}. \end{aligned} \tag{14}$$

Conductance

Given a graph $G = (V, E)$ and $S \subset V$, the conductance of S in G is:

$$\Phi(S) = \frac{|E(S, \bar{S})|}{\min\{\mathrm{vol}(S), \mathrm{vol}(\bar{S})\}}.$$

The conductance of G is
$\Phi = \min\{\Phi(S) \mid |S| \leq \frac{n}{2}\}$.

$\mathrm{Push}(u)$

Andersen, Chung and Lang, FOCS, 2006.
Define an operator
$\mathrm{Push}(u)$:

1. $p(u) \leftarrow p(u) + \alpha r(u)$
2. $r(u) \leftarrow (1-\alpha) r(u)/2$
3. For each v with $v \sim u$,
   set

$$r(v) \leftarrow r(v) + (1-\alpha) r(u)/(2d(u)).$$

Approximate PageRank

Given a node v,

1. set $p = 0$, $r(v) = 1$, and $r(u) = 0$ for all $u \neq v$.
2. For every u, if $r(u) \geq \epsilon d(u)$, then:
   – Apply $\mathrm{push}(u)$.
3. Otherwise, Then output p and r.

ACL local algorithm

1. To find the RageRank from a given input vertex v,
2. To rank the pages by decreasing of the normalised PageRank, i.e., $\frac{p_v}{d(v)}$. Suppose that $v_1, v_2, \cdots, v_l$ is listed such that

$$\frac{p_{v_1}}{d(v_1)} \geq \frac{p_{v_2}}{d(v_2)} \geq \cdots \geq \frac{p_{v_l}}{d(v_l)}.$$

3. (Pruning) To take an initial segment of the list as a community associated with the given input v.
   Let j be such that

$$\Phi(X_j) = \min\{\Phi(X_i) \mid 1 \leq i \leq l\},$$

   where $\phi(X)$ is the conductance of X in G, and $X_k = \{v_1, \cdots, v_k\}$.
   Output X_j.

Question for local algorithm

For every query Q, we rank the set of answers for the query by PageRank, however, the list is a too long list.
The question is to determine a short list of ranks as the output of the query.
Still open.

The great idea

- The PageRank thesis
- The teleportation parameter $1-\alpha$.
  This is a great idea, which may be used in many other areas, such as learning, data processing.
  The essence of the idea here is to make sure that the Ranking matrix is a well-defined stochastic procedure so that PageRank exists and can be computed.
  We may also regard the introduction of $1-\alpha$ as amplifying noises, playing a role similar to that in the error correcting codes.
- Google's success: Making big money by randomness

A grand challenge

- What is the principle for determining α? Is there a metric of networks which determines the optimum α?
- What are principles for structuring the unstructured and noisy data?
- Making money by connection and interaction???

Reference

1. Amy N. Langville and Carl D. Meyer, Google's PageRank and Beyond: The Science of Search Engine Ranking, Princeton University Press, 2006.
2. Andersen, Chung and Lang, Local graph partitioning using PageRank vectors, FOCS, 2006.