About efficient randomized algorithms for PageRank problem

Alexander Gasnikov (PreMoLab MIPT)

avgasnikov@gmail.com

V Traditional Mathematical School, 21 June 2013

PageRank problem formulation

Assume that we have directed graph $G = \langle V_G, E_G \rangle$ of the Internet network. In this graph vertexes denote web-pages and directed edges denote references from web-page to web-page. Denote by N the number of users of the Internet (for simplicity we assume that this quantity doesn't change in time and $N \gg |V_G| \gg 1$). Note that rather often the stochastic matrix corresponding to this Internet graph chooses in form $\boxed{P = (1-\delta)I + \delta\tilde{P}}$ ($\delta \in (0,1]$), where $I = \operatorname{diag}\{1,\ldots,1\}$ – the identity matrix or matrix with all element equals $1/|V_G|$ and

$$\boxed{\tilde{p}_{ij} = \left|\left\{k :\ (i,k) \in E\right\}\right|^{-1}, i \neq j, \text{else} = 0}.$$

Let's introduce $n_i(t)$ as a number of users at the web-page i at the moment of time t. For one step (unit of time) each of the users (independently of everything) with probability $\delta\tilde{p}_{ij}$ move to the web-page j and with probability $1-\delta$ remains in place.

We assume matrix P to be irreducible (it can be generalized). So we are ready to formulate main result.

Why it is necessary to solve: $\vec{p}^T = \vec{p}^T P$?

PageRank theorem:

$$\exists \ \lambda_{0.99} > 0; \ T_N > 0: \ \ \forall \ k = 0, \ldots, n = |V_G|; \ t \geq T_N$$

$$P\left(\left|\frac{n_k(t)}{N} - p_k\right| \leq \frac{\lambda_{0.99}}{\sqrt{N}}\right) \geq 0.99,$$

where $\vec{p}^T = \vec{p}^T P$ (solution of this equation is unique in class of probability distributions because of irreducibility of P).

Axillaries facts about stochastic matrix
(Elements of Perron–Frobenius theory)

Let’s reformulate our problem of finding left *eigenvector* of matrix P corresponds to the Frobenius–Perron *eigenvalue* = 1, as matrix game

$$\boxed{f\left(\vec{p}\right)=\left\|A\vec{p}\right\|_{\infty}=\max_{\vec{u}\in S_n(1)}\left\langle \vec{u},A\vec{p}\right\rangle \to \min_{\vec{p}\in S_n(1)},} \qquad \text{(MG)}$$

where $A=P^T-I$, $S_n(1)=\left\{\vec{p}\geq\vec{0}:\ \sum_{k=1}^{n}p_k=1\right\}$. Note that $f\left(\vec{p}\right)\geq 0$ for all $\vec{p}\in S_n(1)$, and iff $\vec{p}$ = PageRank , than $f\left(\vec{p}\right)=0$.

Antisymmetric form of the game

$$\max_{\vec{u}\in S_{2n+1}(1)} \langle \vec{u}, \mathrm{A}\vec{x} \rangle \to \min_{\vec{x}:=(\vec{y},\vec{p}',u)\in S_{2n+1}(1)},$$

$$\mathrm{A} := \begin{bmatrix} 0 & A & -\vec{e} \\ A^T & 0 & \vec{e} \\ \vec{e}^T & -\vec{e}^T & 0 \end{bmatrix},$$

where $\vec{e} = (1,...,1)^T$, $\mathrm{A} = \|a_{ik}\|$. Then

$$f(\vec{p}) \le 2\varepsilon, \text{ where } \vec{p} = \vec{p}' \big/ \left(\vec{e}^T \vec{p}'\right), \ \vec{e}^T \vec{p}' \ge 1/2 - \varepsilon \text{ if } \mathrm{A}\vec{x} \le \varepsilon\vec{e}.$$

So we reduce PageRank problem to the following: $\boxed{\mathrm{A}\vec{x} \le \varepsilon\vec{e}}$.

Possible conditions on matrix P

- *G-condition* – there exists such a vertex (for example vertex corresponds to **G**oogle, without restriction of generality let's assume that this vertex is 1), that from any vertex we can come to 1 for one step with the probability greater than $\alpha > cn^{-1}$;
- *S-condition* – **S**parseness condition of P (in rows) means that each of the vertexes leave no more than $s \ll n$ edges (references on the other vertexes), we will write *Src* for the sparseness of P both in rows and columns;
- *SG-condition* – **S**pectral **G**ap of matrix P equals $\alpha \gg n^{-2}$.

Method	Cond.	Total complexity	Goal (min)
Grigoriadis–Khachyan, 95; 13 *Stochastic mirror descent*	Src	$O\left(\frac{s\ln n\left(\ln n+\ln\sigma^{-1}\right)}{\varepsilon^2}\right)$	$\left\Vert P^T\vec{p}-\vec{p}\right\Vert_\infty$
Nazin–Polyak, 2009 *Stochastic mirror descent*	no	$O\left(\frac{n\left(\ln n+\ln\sigma^{-1}\right)}{\varepsilon^2}\right)$	$\left\Vert P^T\vec{p}-\vec{p}\right\Vert_2^2$
Nemirovski et al., 2009 *Stochastic mirror descent*	no	$O\left(\frac{n\left(\ln n+\ln\left(\sigma^{-1}\right)\right)}{\varepsilon^2}\right)$	$\left\Vert P^T\vec{p}-\vec{p}\right\Vert_\infty$

Spielman, 2009	G, S	$\frac{3s}{\alpha\varepsilon^2}$	$\left\Vert \vec{p} - \vec{p}_* \right\Vert_\infty$
Polyak–Tremba, 2012	S	$\frac{4sn}{\varepsilon}$	$\left\Vert P^T \vec{p} - \vec{p} \right\Vert_1$
Nesterov–Nemirovski, 2012	G, S	$\frac{sn}{\alpha} \ln\left(\frac{2}{\varepsilon}\right)$	$\left\Vert \vec{p} - \vec{p}_* \right\Vert_1$
MCMC, 2013	SG	$O\left(\frac{\ln n}{\alpha\varepsilon^2} \ln\left(\frac{1}{\sigma}\right)\right)$	$\left\Vert \vec{p} - \vec{p}_* \right\Vert_1$

We assume that: $n^{-2} \ll \varepsilon^2 \ll n^{-1}$**.**

Markov chain Monte Carlo algorithm

$$E\left[\left\|\frac{1}{T}\sum_{t=T_0}^{T+T_0}\vec{x}_t-\vec{\pi}\right\|_1^2\right]=(\mathrm{Bias})^2+\mathrm{Var}=\left(\mathrm{O}\left(\frac{(1-\kappa)^{T_0}}{\kappa T}\right)\right)^2+\mathrm{O}\left(\frac{1}{\kappa T}\right),$$

$$\exists\ c>0:\ P\left(\left\|\frac{1}{T}\sum_{t=T_0}^{T+T_0}\vec{x}_t-\vec{\pi}\right\|_1>\mathrm{Bias}+\Omega\right)\le\exp\left(-c\Omega^2/\mathrm{Var}\right),$$

where $\vec{\pi}$ is invariant measure of considered Markov chain (MC), and random vector $\vec{x}_t$ have all components equal to 0 expect one component (corresponding to the position of MC at the moment of time t) equals to 1, κ is coarse Ricci curvature of MC ($\kappa\sim\alpha$) [J–O].

MCMC algorithm

1. **Initialization:** $\vec{X} = \vec{0}$, vertex $= 1$, $t = 0$.
2. **Counter of iterations:** $t := t + 1$.
3. **Modification $\vec{X}$:** if $t \geq T^0_{\varepsilon,\alpha,n}$, than $X_k := X_k + 1$, where k is a number of current vertex.
4. **Vertex modification:** according to stochastic matrix P choose the next vertex at random.
5. **Stopping rule:** if $t < T_{\varepsilon,\sigma,\alpha}$, than go to step 2, else go to step 6.
6. **Answer:** $\vec{p} = \vec{X} \Big/ \left(T_{\varepsilon,\sigma,\alpha} - T^0_{\varepsilon,\alpha,n} \right)$.

We choose $T^0_{\varepsilon,\alpha,n} = \mathrm{O}\left(\dfrac{\ln n + \ln(1/\varepsilon)}{\alpha}\right)$, $T_{\varepsilon,\sigma,\alpha} = \mathrm{O}\left(\dfrac{1}{\alpha\varepsilon^2}\ln\left(\dfrac{1}{\sigma}\right)\right)$.

Problems:

1. We don't know, as a rule, a spectral gap of P. This problem has in most of the cases an efficient solution ($\tau \sim 10^2$):

$$\alpha \simeq 1 - \frac{1}{\tau}\sum_{t=T}^{T+\tau} \frac{\langle \vec{p}_t - \vec{p}_{t-1}, \vec{p}_t - \vec{p}_{t-1} \rangle}{\langle \vec{p}_{t-1} - \vec{p}_{t-2}, \vec{p}_t - \vec{p}_{t-1} \rangle}.$$

2. We don't know, as a rule, explicit values of $T^0_{\varepsilon,\alpha,n}$ and $T_{\varepsilon,\sigma,\alpha}$.

The problem of finding $T_{\varepsilon,\sigma,\alpha}$ can be solved by controlling $\left\| \vec{p}_{t+\tau} - \vec{p}_t \right\|_1 / \tau$. It will costs $\mathrm{O}(n)$. As for $T^0_{\varepsilon,\alpha,n}$, we can start with $T^0_{\varepsilon,\alpha,n} = 0$ at first, then we correct the obtained answer to put $T^0_{\varepsilon,\alpha,n} = T_{\varepsilon,\sigma,\alpha}/5$. This procedure doesn't change the total number of computation in sense $\mathrm{O}(\cdot)$.

Grigoriadis–Khachyan algorithm for $A\vec{x} \le \varepsilon\vec{e}$

Grigoriadis–Khachyan alghorithm

1. **Initialization:**

$$\vec{X} = \vec{0},\ \vec{p}^T = \frac{1}{2n+1}\underbrace{(1,...,1)}_{2n+1}^T,\ t = 0.$$

2. **Counter of iterations:** $t := t+1$.

3. **Generation of random variables:** choose $k \in \{1,...,2n+1\}$

with probability p_k.

Grigoriadis–Khachyan algorithm for $A\vec{x} \le \varepsilon\vec{e}$

4. Modification $\vec{X}$**:** $X_k := X_k + 1.$

5. Modification $\vec{p}$**:** $i = 1,\ldots,2n+1$

$$p_i := p_i \exp\left(\frac{\varepsilon a_{ik}}{2}\right)\left(\sum_{j=1}^{2n+1} p_j \exp\left(\frac{\varepsilon a_{jk}}{2}\right)\right)^{-1}.$$

6. Stopping rule: if $t < 12\left(\ln\left(2n+1\right) + \ln\sigma^{-1}\right)\varepsilon^{-2}$ then go to step 2, else to step 7.

7. Answer: $\vec{x} = \vec{X}/t$.

For simplicity let's change $2n+1$ to n. Introduce $r>0$ by the equation: $\boxed{\sigma \simeq n^{-r}}$. Let's show that Grigoriadis–Khachyan algorithm after $\boxed{t^* = 3\left(\ln n + \ln \sigma^{-1}\right)\varepsilon^{-2} = 3(1+r)\varepsilon^{-2}\ln n}$ iteration return such a random vector $\vec{x}$, that with probability greater than $1-\sigma$ the following inequality $\boxed{\mathrm{A}\vec{x} \le \varepsilon\vec{e}}$ is true.

First of all, we introduce Freund–Schapire potentials

$$p_i(t) = P_i(t)\left(\sum_{j=1}^{n} P_j(t)\right)^{-1},\ P_i(t) = \exp\left(\varepsilon U_i(t)/2\right) \text{ and}$$

$$\Phi(t) = \sum_{i=1}^{n} P_i(t), \text{ where } \vec{U}(t) = \mathrm{A}\vec{X}(t).$$

After that we have

$$\Phi(t+1)=\sum_{i=1}^{n} P_i(t)\exp(\varepsilon a_{ik}/2)=\Phi(t)\sum_{i=1}^{n} p_i(t)\exp(\varepsilon a_{ik}/2),$$

$$E\left[\Phi(t+1)\middle|\vec{P}(t)\right]=\Phi(t)\sum_{i,k=1}^{n} p_i(t)p_k(t)\exp(\varepsilon a_{ik}/2),$$

$$\exp(\varepsilon a_{ik}/2)\le 1+\varepsilon a_{ik}/2+\varepsilon^2/6,\ |a_{ik}|\le 1.$$

Since $\sum_{i,k=1}^{n} p_i(t)\,p_k(t)=\left(\sum_{i=1}^{n} p_i(t)\right)^2=1,\ \sum_{i,k=1}^{n} p_i(t)\,p_k(t)\frac{\varepsilon^2}{6}=\frac{\varepsilon^2}{6},$

$$\sum_{i,k=1}^{n} p_i(t)\,p_k(t)a_{ik}=\left\langle \vec{p}(t),\mathrm{A}\vec{p}(t)\right\rangle=0,\ \text{than}$$

$$E\left[\Phi(t+1) \middle| \vec{P}(t)\right] \le \Phi(t)\left(1+\varepsilon^2/6\right),$$

$$E\left[\Phi(t+1)\right] \le E\left[\Phi(t)\right]\left(1+\varepsilon^2/6\right).$$

According to this inequality and

$$E\left[\Phi(0)\right] = \Phi(0) = n$$

we have

$$E\left[\Phi(t)\right] \le n\left(1+\varepsilon^2/6\right)^t.$$

Thus,

$$E\left[\Phi(t)\right] \le n\exp\left(t\varepsilon^2/6\right) \text{ and } E\left[\Phi\left(t^*\right)\right] \le n^{3/2+r/2}.$$

From the Markov inequality

$$\forall \;\; r.\text{v. } \xi \ge 0, t > 0 \rightarrow P\left(\xi \ge t\right) \le E\xi/t,$$

we have

$$P\left(\Phi\left(t^*\right) \le n^{3/2(1+r)}\right) = 1 - P\left(\Phi\left(t^*\right) \ge n^{3/2(1+r)}\right) \ge$$
$$\ge 1 - E\left[\Phi\left(t^*\right)\right]\Big/ n^{3/2(1+r)} \ge 1 - \sigma.$$

Thus, with the probability greater than $1-\sigma$

$$\exp\left(\varepsilon U_i\left(t^*\right)/2\right)=P_i\left(t^*\right)\le\sum_{i=1}^{n}P_i\left(t^*\right)=\Phi\left(t^*\right)\le n^{3/2(1+r)},$$

hence

$$P\left(\varepsilon U_i\left(t^*\right)/2\le 3/2\left(1+r\right)\ln n,\ \ i=1,\dots,n\right)\ge 1-\sigma,$$

$$P\left(U_i\left(t^*\right)/\left(3\left(1+r\right)\varepsilon^{-2}\ln n\right)\le\varepsilon,\ \ i=1,\dots,n\right)\ge 1-\sigma.$$

So we have just shown that

$$\boxed{P\left(\mathrm{A}\vec{x}\left(t^*\right)\le\varepsilon\vec{e}\right)\ge 1-\sigma}\text{, QED.}$$

Literature

[1]. *Назин А.В., Поляк Б.Т.* Рандомизированный алгоритм нахождения собственного вектора стохастической матрицы с применением к задаче PageRank // Автоматика и телемеханика, № 2. 2011. С. 131–141.

[2]. *Nesterov Y.E.* Subgradient methods for huge-scale optimization problems // CORE Discussion Paper; 2012/2, 2012.
http://dial.academielouvain.be/handle/boreal:107876

[3]. *Nesterov Y.E.* Efficiency of coordinate descent methods on large scale optimization problem // SIAM Journal on Optimization. 2012. V. 22. № 2. P. 341–362.
http://dial.academielouvain.be/handle/boreal:121612

[4]. *Nesterov Y.*, *Nemirovski A.* Finding the stationary states of Markov chains by iterative methods // CORE Discussion Paper; 2012/58, 2012.
http://dial.academielouvain.be/handle/boreal:122163

[5]. *Spielman D.* Lecture № 7, 2009 http://www.cse.cuhk.edu.hk/~chi/csc5160/notes/L07.pdf

[6]. *Juditsky A., Lan G., Nemirovski A., Shapiro A.* Stochastic approximation approach to stochastic programming // SIAM Journal on Optimization. 2009. V. 19. № 4. P. 1574–1609. http://www2.isye.gatech.edu/~nemirovs/

[7]. *Хачиян Л.Г.* Избранные труды / сост. С. П. Тарасов. М.: МЦНМО, 2009. С. 38–48.

[8]. *Поляк Б.Т., Тремба А.А.* Решение задачи PageRank для больших матриц с помощью регуляризации // Автоматика и телемеханика. 2012. № 11. С. 144–166.

[9]. *Joulin A., Ollivier Y.* Curvature, concentration and error estimates for Markov chain Monte Carlo // Ann. Prob. 2010. V. 38. № 6. P. 2418–2442. http://www.yann-ollivier.org/rech/publs/surveycurvmarkov.pdf

[10]. *Fan Chung* Laplacians and the Cheeger inequality for directed graphs // Annals of Combinatorics. 2005. V. **9**. P. 1–19.