Project proposal
Team “Untitled”
Numerical Linear Algebra, Fall 2016

1. Project Name

RepRank implementation

2. Team

Alexandr Demidov, Oleg Khomenko, Polina Polunina, Ruslan Yusupov

3. Background

Due to popularity surge social networks became lucrative targets for spammers and guerilla marketers, who are trying to game ranking systems and broadcast their messages at little to none cost. Ranking systems, for example Twitter’s Trends, can be gamed by scripted users also called bots, who are automatically or semi-automatically twitting essentially the same message. Judging by the prices and abundance of supply from PR firms this is an easy to implement and widely used tactic, at least in Russian blogosphere. Aggregative analysis of social networks should at best mark those messages as spam or at least correctly downplay their importance as they represent opinions only of a few, if dedicated, users. Hence bot detection plays a crucial role in social network mining and analysis.

Original method is just iterative method with matrix-by-vector product.

4. Problem formulation

Implement “RepRank” algorithm, described in [1], and improve it using Anderson acceleration.

$$t = \alpha_1 F_{t+} + \alpha_2 B t_- + \alpha_3 d$$

$$
\begin{aligned}
&\texttt{anderson}(u_0, G, m)\\
&\quad u_1 = G(u_0);\ F_0 = G(u_0) - u_0\\
&\quad \textbf{for } k = 1, \ldots \textbf{ do}\\
&\qquad m_k = \min(m, k)\\
&\qquad F_k = G(u_k) - u_k\\
&\qquad \text{Minimize } \| \textstyle\sum_{j=0}^{m_k} \alpha_j^k F_{k-m_k+j} \| \text{ subject to}\\
&\qquad \textstyle\sum_{j=0}^{m_k} \alpha_j^k = 1.\\
&\qquad u_{k+1} = (1-\beta_k) \textstyle\sum_{j=0}^{m_k} \alpha_j^k u_{k-m_k+j} + \beta_k \sum_{j=0}^{m_k} \alpha_j^k G(u_{k-m_k+j})\\
&\quad \textbf{end for}
\end{aligned}
$$

5. Data

We use three datasets provided by Ovchinnikov G.

- 1st — Twitter users social graph (approximately 300k nodes)
- 2nd — vk.com users social graph (approximately 1 M nodes)
- 3rd — vk.com users social graph (approximately 300 M nodes)

6. Related work

[1] Ovchinnikov, G. V., D. A. Kolesnikov, and Ivan V. Oseledets. "Algebraic reputation model RepRank and its application to spambot detection.

7. Scope

The outcome is an implemented RepRank algorithm with anderson acceleration. The phases are:

- Related literature overview (papers, books etc.)
- Data preprocessing
- Algorithm Implementation
- Algorithm testing on available datasets
- Project presentation preparation
- Summarizing the results at the written report

8. Evaluation

Our team is going to evaluate accuracy and execution time of the algorithm.

9. References

[2] Zolt´an Gy¨ongyi, Hector Garcia-Molina, and Jan Pedersen. Combating web spam with trustrank. In Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30, VLDB ’04, pages 576–587. VLDB Endowment, 2004.

[3] Vijay Krishnan. Web spam detection with anti-trust rank. In In AIRWEB, pages 37–40, 2006.

[4] Xinyue Liu, You Wang, Shaoping Zhu, and Hongfei Lin. Combating web spam through trust-distrust propagation with confidence. Pattern Recogn. Lett., 34(13):1462–1469, October 2013.

[5] Xianchao Zhang, You Wang, Nan Mou, and Wenxin Liang. Propagating both trust and distrust with target differentiation for combating link-based web spam. ACM Trans. Web, 8(3):15:1–15:33, July 2014.