ON THE COMPLEXITY OF OPTIMIZING PAGERANK*

R. HOLLANDERS†, J.-C. DELVENNE‡, AND R.M. JUNGERS§

Key words. PageRank, Policy Iteration, Markov Decision Processes, Optimization, Complexity

AMS subject classifications. 90C40, 90C27, 68Q25

EXTENDED ABSTRACT. In search engines, it is critical to be able to compare webpages according to their relative importance, with as few as possible computational resources. This is done by computing the *PageRank* of every webpage from the web [4] (i.e., the average portion of time spent in the webpage during an infinite and uniform random walk on the web). Pages with higher PageRank will then appear higher in the list of results. See [3] for a survey on PageRank and its applications.

The concept of PageRank has generated a large number of questions and challenges. Among these, the problem of optimizing the PageRank of webpages raises increasing interest, as evidenced by the growing literature on the subject [1, 20, 7, 5, 17, 6, 10]. In the PageRank Optimization problem (PRO) that we study, one aims to maximize (or minimize) the PageRank of some target node when control is granted on some subset of *free* edges that may be chosen to be activated or deactivated.

A typical application of PRO is the so-called *webmaster problem* in which a webmaster tries to maximize the PageRank of one of his webpage by determining which links under his control (i.e. on his website, or on an allied website for instance) he should activate and which links he should not [1, 7]. The same tools may be used to find how much the PageRank of some nodes can vary when the presence or absence of some links is uncertain (e.g. because a link is broken, the server is down or because of traffic problems) [17]. Similar situations also exist in economic networks in which agents choose partners in order to increase their centrality (i.e., their PageRank) [22] or decrease the centrality of other agents. For instance, a government might want to prevent a bank from acquiring excessive influence in order to limit the sensitivity of the bank network to a possible bankruptcy, and it should therefore allow or reject some transactions [2, 8, 11, 19].

Csáji et al. proposed a way of modeling PRO as a *Markov Decision Process* (MDP), thereby showing that an exact solution of the problem could be found in weakly polynomial time using linear programming [6]. (For more on MDP see, e.g., [21].) Yet in practice, MDPs are solved much more efficiently using algorithms adapted to their special structure. Among these algorithms, *Policy Iteration* (PI) [16] performs very well in practice and is guaranteed to converge to the optimal solution in a finite number of iterations. However, even though PI usually converges in a few iterations, its actual complexity remains unclear.

*This work was supported by the ARC grant 'Large Graphs and Networks' from the French Community of Belgium and by the IAP network 'Dysco' funded by the office of the Prime Minister of Belgium. The scientific responsablity rests with the authors. Additional details and proofs can be found in [15].

†Corresponding author, romain.hollanders@uclouvain.be. The authors are with Department of Mathematical Engineering, ICTEAM, UCLouvain, 4, avenue Lemaitre, B-1348 Louvain-la-Neuve, Belgium. J.-C. D. is also a member of Namur Complex Systems Center (NAXYS, Belgium) in Belgium and the Centre for Operations Research and Econometrics (CORE, Universit catholique de Louvain). R. M. J. is also an F.R.S.-FNRS fellow.

‡jean-charles.delvenne@uclouvain.be

§raphael.jungers@uclouvain.be

There is a significant research effort for understanding the complexity of PI. Recently, an example has been proposed on which PI runs in exponential time [9].

THEOREM 1 (Fearnley, [9]). *There exists an infinite family of MDPs on which the number of iterations that PI takes is lower bounded by an exponential function of the size of the MDP.* This result holds for most of the main classes of MDPs, namely total-cost, average-cost [9] and discounted-cost MDPs [14]. However, special cases exist for which the example from Theorem 1 does not work, as for instance the case of discounted-cost MDPs with a fixed discount factor [23, 13] for which PI runs in strongly polynomial time, thereby showing that the gap between exponential and strongly polynomial time guarantees is sometimes narrow. Deterministic MDPs are another interesting case for which strongly polynomial time algorithms exist [18] and for which the highest known lower bound on the number of iterations of PI is quadratic [12].

In this work, we focus on the performance of PI when it is applied to PRO. We show that the exponential complexity example from Theorem 1 in which PI needs an exponential number of iterations to converge can be extended to PRO if 0 and -1 costs are allowed, thereby suggesting that both positive and negative costs are required for the exponential complexity example to work in general. We provide more arguments to support this suggestion through two particular cases of practical importance in which PI solves the unit-cost version of PRO in polynomial time:

1. In many applications, such as web search engines, it is assumed that the random walk used to compute PageRank can be interrupted at any time with some fixed probability and start again from an arbitrary node of the graph [3]. This restarting probability is called *zapping* and can be interpreted as if one could get bored of surfing with some probability and decide to start a new search from a new randomly chosen page. We show that in the case of a fixed zapping probability, PI solves PRO in weakly polynomial time.
2. We show that if all free edges either leave some arbitrary node and/or the target node, then PRO can be solved in strongly polynomial time using PI.

REFERENCES

[1] K. Avrachenkov and N. Litvak. Decomposition of the Google PageRank and Optimal Linking Strategy. *Technical report, INRIA*, 2004.
[2] M. L. Bech, J. T. E. Chapman, and R. J. Garratt. Which Bank is the "Central" Bank ? *Journal of Monetary Economics*, 57:352363, 2009.
[3] P. Berkhin. A Survey on PageRank Computing. *Internet Mathematics*, 2(1):73-120, 2005.
[4] S. Brin and L. Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, 1998.
[5] D. Chaffey, C. Lake, and A. Friedlein. Search Engine Optimization - Best Practice Guide. *Econsultancy.com Ltd*, 2009.
[6] B. C. Csáji, R. M. Jungers, and V. D. Blondel. PageRank Optimization by Edge Selection. *To appear in: Discrete Applied Mathematics*, 2011.
[7] C. de Kerchove, L. Ninove, and P. Van Dooren. Maximizing PageRank via outlinks. *Linear Algebra and its Applications*, 429:1254-1276, 2008.
[8] H. Degrysea and G. Nguyen. Interbank Exposures: An Empirical Examination of Contagion Risk in the Belgian Banking System. *International Journal of Central Banking*, 3:123-169, 2007.
[9] J. Fearnley. Exponential Lower Bounds for Policy Iteration. *CoRR, abs/1003.3418*, 2010.
[10] O. Fercoq, M. Akian, M. Bouhtou, and S. Gaubert. Ergodic Control and Polyhedral approaches to PageRank Optimization. *Arxiv preprint arXiv:1011.2348*, 2010.
[11] A. Fracasso and S. Schiavo. Global Imbalances, Exchange Rates Adjustment and the Crisis : Implications from Network Analysis. *Journal of Policy Modeling*, 31:601619, 2009.
[12] T. Hansen and U. Zwick. Lower Bounds for Howards Algorithm for Finding Minimum Mean-

Cost Cycles. *ISAAC'10: Proceedings of the 21st International Symposium on Algorithms and Computation*, pages 415–426, 2010.

[13] T. D. Hansen, P. B. Miltersen, and U. Zwick. Strategy Iteration is Strongly Polynomial for 2-Player Turn-Based Stochastic Games with a Constant Discount Factor. *Arxiv preprint arXiv:1008.0530*, 2010.

[14] R. Hollanders, J.-C. Delvenne, and R. Jungers. The Complexity of Policy Iteration is Exponential for Discounted Markov Decision Processes. *(Submitted)*, 2012.

[15] R. Hollanders, J.-C. Delvenne, and R.M. Jungers. Policy Iteration is Well Suited to Optimize PageRank. *Arxiv preprint arXiv:1108.3779*, 2011.

[16] R.A. Howard. *Dynamic Programming and Markov Processes.* MIT Press, Cambridge, MA, 1960.

[17] H. Ishii and R. Tempo. Computing the PageRank Variation for Fragile Web Data. *SICE Journal of Control, Measurement, and System Integration*, 2(1):1-9, 2009.

[18] O. Madani, M. Thorup, and U. Zwick. Discounted Deterministic Markov Decision Processes and Discounted All-Pairs Shortest Paths. *ACM Transactions on Algorithms, TALG*, 6(2):33, 2010.

[19] S. Markose, S. Giansante, M. Gatkowski, and A. R. Shaghaghi. Too Interconnected to Fail : Financial Contagion and Systemic Risk in Network Model of CDS and Other Credit Enhancement Obligations of US Banks. *Economics Discussion Papers*, 883, 2009.

[20] F. Mathieu and L. Viennot. Local Aspects of the Global Ranking of Web Pages. *Proceedings of the 6th International Workshop on Innovative Internet Community Systems*, 44, 2006.

[21] M. L. Puterman. *Markov Decision Processes.* John Wiley & Sons, 1994.

[22] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White. Economic Networks : The New Challenges. *Science*, 325:422-425, 2009.

[23] Y. Ye. The Simplex and Policy-Iteration Methods are Strongly Polynomial for the Markov Decision Problem with a Fixed Discount Rate. *Mathematics of Operations Research*, 2011.