



HAL
open science

On the Complexity of Optimizing PageRank

Romain Hollanders, Jean-Charles Delvenne, Raphaël M. Jungers

► **To cite this version:**

Romain Hollanders, Jean-Charles Delvenne, Raphaël M. Jungers. On the Complexity of Optimizing PageRank. 14èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (AlgoTel), 2012, La Grande Motte, France. pp.1-4. hal-00690502

HAL Id: hal-00690502

<https://hal.science/hal-00690502>

Submitted on 23 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the Complexity of Optimizing PageRank

R. Hollanders^{1†} and J.-C. Delvenne^{12‡} and R. M. Jungers^{13§}

¹Department of Mathematical Engineering, ICTEAM, UCLouvain, 4, avenue Lemaitre, B-1348 Louvain-la-Neuve, Belgium.

²Namur Complex Systems Center (NAXYS, Belgium) in Belgium and Centre for Operations Research and Econometrics (CORE, Université catholique de Louvain).

³F.R.S.-FNRS fellow.

We consider the PageRank Optimization problem in which one seeks to maximize (or minimize) the PageRank of a node in a graph through adding or deleting links from a given subset. The problem can be modeled as a Markov Decision Process and has recently received much attention. We provide provably efficient methods to solve the problem on large graphs for a number of cases of practical importance and we show using perturbation analysis that for a close variation of the problem, the same techniques have exponential worst case complexity.*[¶]

Keywords: PageRank, Policy Iteration, Markov Decision Processes, Optimization, Complexity

1 Introduction

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 [BP98]: pages with higher PageRank will then appear higher in the list of results. See [Ber05] for a survey on PageRank and its applications.

The concept of PageRank has also 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 [AL04, MV06, dKND08, CLF09, IT09, CJB11, FABG10]. 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 [AL04, dKND08]. 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) [IT09]. Similar situations also exist in economic networks in which agents choose partners in order to increase their centrality (i.e., their PageRank) [SFS⁺09] 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 [BCG09, DN07, FS09, MGS09]. It is known that small changes in the graph can substantially

[†]Corresponding author, romain.hollanders@uclouvain.be

[‡]jean-charles.delvenne@uclouvain.be

[§]raphael.jungers@uclouvain.be

* 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 responsibility rests with the authors.

[¶] Part of these results were submitted at the 2012 SIAM conference on Linear Algebra and at the 20th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2012).

change the ranking of the nodes in terms of their PageRank [LM05], hence the usefulness of PRO in these situations.

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 [CJB11]. (For more on MDP see, e.g., [Put94].) Yet in practice, MDPs are solved much more efficiently using algorithms adapted to their special structure. Among these algorithms, *Policy Iteration* (PI) [How60] 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.

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 [Fea10].

Theorem 1 (Fearnley, [Fea10]). *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 [Fea10] and discounted-cost MDPs [HDJ12]. 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 [Ye11, HMZ10] 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 [MTZ10] and for which the highest known lower bound on the number of iterations of PI is quadratic [HZ10].

In this work, we focus on the performance of PI when it is applied to PRO. As already suggested, these problems can be modeled as MDPs with unit-costs and uniform transition probability distributions. Our first contribution is to 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 provided both positive and negative costs are allowed (we refer to this modified version of PRO by ± 1 -PRO). This fact suggests that the exponential complexity example may not apply to MDPs with only positive costs, such as PROs. We provide more arguments to support this suggestion through a number of particular cases in which Policy Iteration solves PRO in polynomial time. These cases are our second contribution.

2 PI needs exponential time to solve ± 1 -PRO

First, let us define a total-cost MDP as the random process of an *agent* that evolves on a finite set of *nodes*. One of the nodes is chosen to be the initial node. At every time step, the *controller* of the process needs to choose one among the several *actions* available in the current node. The chosen action determines a *transition probability* distribution for the next node to reach as well as an immediate *cost* (or reward). The goal of the controller is to choose the right actions in each node in order to minimize the costs incurred by the agent over time until the reach of some given cost-free absorbing node that we call the target node.

A PageRank Optimization problem can be defined as an MDP with two structural constraints:

1. Transition probabilities are uniformly distributed among outgoing links and transition costs are unitary. (Note that PageRank has also been defined for non-uniform transition probabilities [NA08], but we focus on uniform PageRank.)
2. Actions consist in activating and deactivating free edges, thereby modifying the transition probabilities according to the first constraint.

The objective function of the controller is to minimize the expected return time (or cost) from the fixed target node to itself.

Our goal here is to transform the MDP example from Theorem 1 into a PRO instance in which the number of iterations of PI remains the same, and this in order to find a PRO example on which PI runs in exponential time. We show that it is possible to perform such a transformation with a polynomial number of operations, provided the first specific constraint of PRO is relaxed such that transitions can have a cost of either $+1$

or -1 . The key ideas of the construction are twofold: first we translate MDP actions into the choice of activating (or deactivating) free edges, thereby handling the second constraint of PRO. Then, we use gadgets to transform transition probabilities and costs into polynomial sized sub-structures that only make use of uniform transition probabilities and $+1$ or -1 costs. Note that the first step of the transformation requires to perturb the initial example. Nevertheless, these perturbations can be quantified and adjusted so that the number of iterations of PI does not change.

Theorem 2. *If $+1$ and -1 costs are allowed, then there exists an infinite family of PageRank Optimization problems on which the number of iterations that PI takes is lower bounded by an exponential function of the size of the problem.*

Furthermore, it seems to be impossible to get rid of the negative costs in the example from Theorem 1 without losing the exponential complexity property. This gives hope for better convergence properties of PI in the case of PRO.

3 Does PI solve PRO in polynomial time?

We provide two particular cases of practical importance in which PI provably runs 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 [Ber05]. This restarting probability is called *zapping* and can be interpreted as if the random surfer could get bored with performing its search with some probability and decide to start a new search from a new randomly chosen node.

Theorem 3. *PRO with fixed non-zero zapping probability can be solved in weakly polynomial time using PI.*

The proof of Theorem 3 relies on a result from [CJB11] that guarantees that PI runs in weakly polynomial time if the target node can be reached from any other node via a bounded length path.

2. If we assume that every free edge leaves either the target node or some other node of the process, PI can be shown to run in strongly polynomial time on PRO. This would for instance be the case of a webmaster controlling the links that leave his page or a friend's page.

Theorem 4. *If all free edges either leave some arbitrary node w or the target node v , then PI solves PRO in strongly polynomial time.*

Theorem 4 may be seen as a generalization of a result from [dKND08], which essentially provides an optimal strategy when every free edge leaves the target node. The result may possibly be extendable to larger cases through similar arguments.

References

- [AL04] K. Avrachenkov and N. Litvak. Decomposition of the Google PageRank and Optimal Linking Strategy. *Technical report, INRIA*, 2004.
- [BCG09] M. L. Bech, J. T. E. Chapman, and R. J. Garratt. Which Bank is the “Central” Bank ? *Journal of Monetary Economics*, 57:352–363, 2009.
- [Ber05] P. Berkhin. A Survey on PageRank Computing. *Internet Mathematics*, 2(1):73-120, 2005.
- [BP98] 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.
- [CJB11] B. C. Csáji, R. M. Jungers, and V. D. Blondel. PageRank Optimization by Edge Selection. *To appear in: Discrete Applied Mathematics*, 2011.

- [CLF09] D. Chaffey, C. Lake, and A. Friedlein. Search Engine Optimization - Best Practice Guide. *Econsultancy.com Ltd*, 2009.
- [dKND08] C. de Kerchove, L. Ninove, and P. Van Dooren. Maximizing PageRank via outlinks. *Linear Algebra and its Applications*, 429:1254-1276, 2008.
- [DN07] H. Degryse 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.
- [FABG10] O. Fercoq, M. Akian, M. Bouhtou, and S. Gaubert. Ergodic Control and Polyhedral approaches to PageRank Optimization. *Arxiv preprint arXiv:1011.2348*, 2010.
- [Fea10] J. Fearnley. Exponential Lower Bounds for Policy Iteration. *CoRR, abs/1003.3418*, 2010.
- [FS09] A. Fracasso and S. Schiavo. Global Imbalances, Exchange Rates Adjustment and the Crisis : Implications from Network Analysis. *Journal of Policy Modeling*, 31:601-619, 2009.
- [HDJ12] R. Hollanders, J.-C. Delvenne, and R. Jungers. The Complexity of Policy Iteration is Exponential for Discounted Markov Decision Processes. (*Submitted*), 2012.
- [HMZ10] 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.
- [How60] R.A. Howard. *Dynamic Programming and Markov Processes*. MIT Press, Cambridge, MA, 1960.
- [HZ10] T. Hansen and U. Zwick. Lower Bounds for Howard's Algorithm for Finding Minimum Mean-Cost Cycles. *ISAAC'10: Proceedings of the 21st International Symposium on Algorithms and Computation*, pages 415-426, 2010.
- [IT09] 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.
- [LM05] R. Lempel and S. Moran. Rank-Stability and Rank-Similarity of Link-Based Web Ranking Algorithms in Authority-Connected Graphs. *Information Retrieval*, 8(2):245-264, 2005.
- [MGGS09] S. Markose, S. Giansante, M. Gatkowski, and A. R. Shaghghi. 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.
- [MTZ10] 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.
- [MV06] 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.
- [NA08] D. Nemirovsky and K. Avrachenkov. Weighted Pagerank: Cluster-Related Weights. *Proceedings of the 17th Text Retrieval Conference, TREC'08*, 2008.
- [Put94] M. L. Puterman. *Markov Decision Processes*. John Wiley & Sons, 1994.
- [SFS⁺09] 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.
- [Ye11] 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.