

Local Approximation of PageRank and Reverse PageRank*

Ziv Bar-Yossef

Department of Electrical Engineering
Technion, Haifa, Israel

and

Google Haifa Engineering Center
Haifa, Israel

zivby@ee.technion.ac.il

Li-Tal Mashiach

Department of Computer Science
Technion, Haifa, Israel

litalma@cs.technion.ac.il

February 5, 2008

Abstract

We consider the problem of approximating the PageRank of a target node using only local information provided by a link server. This problem was originally studied by Chen, Gan, and Suel (CIKM 2004), who presented an algorithm for tackling it. We prove that local approximation of PageRank, even to within modest approximation factors, is infeasible in the worst-case, as it requires probing the link server for $\Omega(n)$ nodes, where n is the size of the graph. The difficulty emanates from nodes of high in-degree and/or from slow convergence of the PageRank random walk.

We show that when the graph has bounded in-degree and admits fast PageRank convergence, then local PageRank approximation can be done using a small number of queries. Unfortunately, natural graphs, such as the web graph, are abundant with high in-degree nodes, making this algorithm (or any other local approximation algorithm) too costly. On the other hand, *reverse* natural graphs tend to have low in-degree while maintaining fast PageRank convergence. It follows that calculating *Reverse PageRank* locally is frequently more feasible than computing PageRank locally.

We demonstrate that Reverse PageRank is useful for several applications, including computation of hub scores for web pages, finding influencers in social networks, obtaining good seeds for crawling, and measurement of semantic relatedness between concepts in a taxonomy.

1 Introduction

1.1 Local PageRank approximation

The vast majority of algorithms for computing PageRank, whether they are centralized, parallel, or decentralized, have focused on *global* computation of the PageRank vector. That is, PageRank scores for *all* the graph's nodes are computed. While in many applications of PageRank a global computation is needed, there are situations in which one is interested in computing PageRank scores for just a small subset of the nodes.

Consider, for instance, a business that owns a website and would like to promote the website in search engine rankings in order to attract traffic of potential clients. As PageRank is well-known to be an important signal used by the scoring functions of major search engines, tracking

*This work was funded by the European Commission Marie Curie International Re-integration Grant.