

Dynamic PageRank Using Evolving Teleportation

Ryan A. Rossi and David F. Gleich

Purdue University
Department of Computer Science
305 N. University St., West Lafayette, IN 47906
{rrossi, dgleich}@purdue.edu

Abstract. The importance of nodes in a network constantly fluctuates based on changes in the network structure as well as changes in external interest. We propose an evolving teleportation adaptation of the PageRank method to capture how changes in external interest influence the importance of a node. This framework seamlessly generalizes PageRank because the importance of a node will converge to the PageRank values if the external influence stops changing. We demonstrate the effectiveness of the evolving teleportation on the Wikipedia graph and the Twitter social network. The external interest is given by the number of hourly visitors to each page and the number of monthly tweets for each user.

1 Introduction

Finding important nodes in a graph is a key task in a variety of applications: search engines [24,18], network science [17,8,14], and bioinformatics [27,22], among many others. By and large, these are global measures of node importance and one of the most well-studied measures is PageRank [24,20].

PageRank computes the importance of each node in a directed graph under a random surfer model. When at a node, the random surfer can either:

1. transition to a new node from the set of out-edges, or
2. do something else (e.g., execute a search query, use a bookmark).

The probability that the surfer performs the first action is known as the damping parameter in PageRank. We use α to denote the damping parameter. The second action is called teleporting and is modeled by the surfer picking a node at random according to a distribution called the teleportation distribution vector or personalization vector. These choices only depend on the current node and, consequently, define a Markov chain. This PageRank Markov chain always has a unique stationary distribution for any $0 \leq \alpha < 1$. The importance of a node is proportional to its stationary distribution in this Markov chain. Thus, the computation is governed by the graph, a teleportation parameter α , and a teleportation distribution vector.

The PageRank score is a simple model for the importance of a node in a graph, and there are many variations that may yield more useful scores (for instance [21] models a random walk with a back button). A common complaint

about PageRank models is that they are only defined for static graphs. Motivated by the idea of studying PageRank with dynamic graphs, we formulate a dynamic PageRank model for a static graph with a time-dependent, or evolving, teleportation vector. Intuitively, the teleportation distribution changes based on human dynamics such as recent news and seasonal preferences. For example, in our forthcoming experiments (Section 6), the time-dependent vector is the number of hourly page visits for each page from Wikipedia. We derive the model and algorithms for this dynamic version of PageRank in Section 4. The resulting algorithms scale to large graphs. Moreover, we show that the new model is a generalization of PageRank in the sense that *if the time-dependent vector stops changing then our dynamic score vector converges to the standard PageRank score*.

We make our code and data available in the spirit of reproducible research:

<http://www.cs.purdue.edu/homes/dgleich/codes/dynsyspr-waw>

2 PageRank Notation

In order to place our work in context, we first introduce some notation. Let \mathbf{A} be the adjacency matrix for a graph where $A_{i,j}$ denotes an edges from node i to node j . In order to avoid a proliferation of transposes, we define \mathbf{P} as the transposed transition matrix for a random-walk on a graph:

$$P_{j,i} = \text{probability of transitioning from node } i \text{ to node } j.$$

Hence, the matrix \mathbf{P} is *column-stochastic* instead of row-stochastic, which is the standard in probability theory. Throughout this manuscript, we utilize uniform random-walks on a graph, in which case $\mathbf{P} = \mathbf{A}^T \mathbf{D}^{-1}$ where \mathbf{D} is a diagonal matrix with the degree of each node on the diagonal. However, none of the theory is restricted to this type of random walk and any column-stochastic matrix will do. The PageRank vector \mathbf{x} is the solution of the linear system:

$$(\mathbf{I} - \alpha \mathbf{P})\mathbf{x} = (1 - \alpha)\mathbf{v}$$

for any $0 \leq \alpha < 1$ and any teleportation distribution vector \mathbf{v} such that $v_i \geq 0$ and $\sum v_i = 1$. Table 1 summarizes these notation conventions, and has a few other elements that will be discussed in the forthcoming sections.

3 Dynamic and Evolving Rankings

The PageRank literature is vast, and we now survey some of the other ideas related to incorporating graph dynamics into a PageRank vector, more general models for studying dynamic graphs, and updating PageRank vectors.

Our proposed method is related to changing the teleportation vector in the power method as its being computed. Bianchini et al. [5] noted that the power method would still converge if either the graph or the vector \mathbf{v} changed during the method, albeit to a new solution given by the new vector or graph.