Dirichlet PageRank and Trust-based Ranking Algorithms

Fan Chung[⋆], Alexander Tsiatas, and Wensong Xu

Department of Computer Science and Engineering
University of California, San Diego
{fan,atsiatas,w4xu}@cs.ucsd.edu

Abstract. Motivated by numerous models of representing trust and distrust within a graph ranking system, we examine a quantitative vertex ranking with consideration of the influence of a subset of nodes. An efficient algorithm is given for computing *Dirichlet PageRank* vectors subject to Dirichlet boundary conditions on a subset of nodes. We then give several algorithms for various trust-based ranking problems using Dirichlet PageRank with boundary conditions, showing several applications of our algorithms.

1 Introduction

PageRank has proven to be a useful tool for ranking nodes in a graph in many contexts, but it is clear that many refinements can be made for specific purposes. For example, PageRank can be manipulated by link spammers to boost certain nodes' ranking, and it interprets all links between nodes as positive votes for importance even if some links are meant to show distrust.

As an illustrative example, consider the following problem. Suppose an agent v in a social network wants to compute a quantitative ranking among all nodes, but v has a subset of neighbors X who v trusts. v wants to compute a personal ranking of the nodes in the network, but v wants to make sure that there is no large disparity with its trusted neighbors.

Another trust that ranking systems do not take into account arises from a distinction between types of social networks. Some networks, such as Facebook, model closer relationships between people: a social friendship where edges form presumably only between people who know each other personally. This is in direct contrast with systems such as Twitter where the act of "following" someone indicates no such connection. The close-knit network indicates a higher degree of trust.

When agents participate in both networks, it is useful to be able to rank the nodes of the larger network based on the smaller. A node v can compute a ranking among its own personal trust network and then use this to calculate a ranking of nodes in the larger network, many of which do not appear in the

[⋆] Research supported in part by ONR MURI N000140810747, and AFOR's complex networks program.

more personal network. Current ranking mechanisms such as PageRank compute a global ranking and therefore are not suitable for this situation.

In this paper, we consider a tool that can model these and many other scenarios: Dirichlet PageRank vectors with boundary conditions. Given a graph G and a subset of nodes S, we first compute a PageRank vector pr subject to Dirichlet boundary conditions $\mathrm{pr}(v) = 0$ for vertices v on the boundary of S. For example, Dirichlet boundary conditions can be used to model and propagate the distrust of specified vertices in the graph. We then generalize Dirichlet PageRank to use arbitrary boundary conditions $\mathrm{pr}(v) = \sigma(v)$ for boundary nodes v, providing a general framework for a variety of ranking models. Our algorithms are efficient, giving approximate solutions.

1.1 Related Work

The idea of ranking nodes in a graph has a rich history starting from the introduction of PageRank by Brin and Page [5]. The original PageRank was meant for Web Search, but many researchers have developed more tailored ranking systems such as personalized PageRank [13, 14], giving a ranking relative to some specified starting distribution $\boldsymbol{s}$.

One pitfall with PageRank as a ranking system is the fact that all edges contribute positively. In practice, an edge such as a link from one Web page to another can also represent a negative interaction or distrust between the nodes. Several related mathematical models of propagating trust and distrust in a network ranking system are given [11]. There are numerous empirical results. Another algorithm [12] relies on a small hand-picked set of trusted nodes, but one must be careful not to allow malicious nodes to be included.

There are many other algorithms derived from PageRank that use specific tweaks to model trust or distrust in ranking schemes. [1] considers axioms that a ranking system should satisfy and develops several ranking systems accordingly. [4] and [16] systematically model distrust by modifying the PageRank equations to consider negatively-weighted edges, and [15] gives an algorithm with a similar flavor using random walks. Many of these algorithms are closely related, but rigorous analysis is desired for capturing specific phenomena. We will show that these related models can be represented by Dirichlet PageRank with appropriate boundary conditions.

Another area of research concerns spam nodes when they are identified. It has been shown that if agents can collude [2] or easily create pseudonyms [6], they can artificially boost their ranking in PageRank and other ranking systems. There is some work done in how to penalize these vertices [3], and our Dirichlet PageRank can be efficiently used to achieve the same goal.

1.2 Results in this Paper

Motivated by the continual development of new PageRank-based algorithms and analysis of Dirichlet eigenvectors in [9], we develop and analyze Dirichlet PageRank algorithms. For a connected graph G, we examine a Dirichlet PageRank equation and compute the unique solution with Dirichlet boundary conditions $\mathrm{pr}(v) = 0$ for vertices v on the boundary of a specified vertex subset S.

After giving the algorithm for computing Dirichlet PageRank vectors, we generalize the boundary conditions to arbitrary values $\mathrm{pr}(v) = \sigma(v)$ for boundary vertices v. We give an efficient algorithm ApproxDirichPR to compute approximate Dirichlet PageRank vectors with any boundary condition σ. We also give a full analysis leading to the following theorem, where $|\cdot|$ is the L_1-norm, and $\mathrm{vol}(S)$ is twice the number of edges in the specified vertex subset. (Detailed definitions are given in Section 2.)

Theorem 1. *For any $\epsilon \in (0,1)$ and any jumping constant $\alpha \in (0,1)$, the algorithm ApproxDirichPR outputs an ϵ-approximate Dirichlet PageRank vector $\widetilde{\mathrm{pr}}_S$ in time $O(\frac{\mathrm{vol}(S)\log\frac{1}{\epsilon}}{\alpha})$, which, compared to the exact Dirichlet PageRank pr_S, satisfies:*

$$\widetilde{\mathrm{pr}}_S(v) \le \mathrm{pr}_S(v), \forall v \in S$$

$$|\mathrm{pr}_S - \widetilde{\mathrm{pr}}_S| < \frac{\epsilon\mathrm{vol}(S)}{\alpha}.$$

We illustrate several applications of Dirichlet PageRank with boundary conditions below. Many of the specific PageRank variations are covered by this general framework, and we will show how its use can allow the efficient consideration of several models in [1, 3, 4].

1.3 Several applications of Dirichlet PageRank

• **Diminishing known spammers' influence.** A first application concerns the adjustment of PageRank in the presence of known spammers. For example, on the Web, many web pages can be identified as spammers based on content or user reports. We would like a network ranking scheme to take this into account, penalizing these nodes and others that heavily link to them. We will show that Dirichlet PageRank is useful for this problem.

• **Considering trusted friends' opinions.** While adjusting PageRank to take into account known spammers is relatively simplistic, there are plently of more subtle or complex problems that Dirichlet PageRank can handle well. For example, a single node in a graph may want to compute a personalized ranking of nodes, but it has a set of trusted friends or neighbors whose own rankings are to be valued accordingly. While previously-studied personalized PageRank vectors do not take this effect into account, we can use Dirichlet PageRank with boundary conditions to compute a more informed ranking. We will present an algorithm PRTrustedFriends for this problem.

• **Validating rank for newly-created nodes.** The previous application of adjusting PageRank to account for trusted friends can be logically extended to validating a new vertex's ranking within a larger network. Suppose that a new person enters a social network but is unsure about who to trust and distrust within the network. Personalized PageRank is a useful tool for quantitative information, but it raises the question of whether or not this ranking is susceptible to unknown spammers. Without a specific set of trusted friends, it may seem hopeless for the newcomer, but the new vertex can use Dirichlet PageRank with boundary conditions to validate and adjust its own ranking with a randomly-selected pool of established nodes within the network. We will give the details in an algorithm PRValidation.

• **Reconciling rank in personal and global social networks.** Additional interesting questions arise when analyzing different types of social networks. While social networks are often treated on their own, there are many different types of networks with wide participation. Some, like Facebook, offer a more personal viewpoint on network structure, where edges are formed only by mutual consent between two people who usually know each other. This is in contrast with a network such as Twitter, where connections are much more impersonal. Here, people "follow" each other based not only on friendship, but also interest in subject matter, celebrity appeal, advertising, and many conceivable other reasons.

With the vast array of information available on a network such as Twitter, it is important for a user to know who is trustworthy or worth following. This is a difficult problem, but a user does have some information at hand; its own, more close-knit, smaller social network or even a trusted subgraph of the larger network. Using Dirichlet PageRank, a user can compute a ranking on the smaller network, and then use boundary conditions appropriately to infer a ranking on the remaining nodes in the larger network, taking its personal associations into account. We will develop an algorithm PRTrustNetwork for this problem.

Finally, we can use similar ideas to tackle the problem in reverse: suppose a global ranking of the nodes in a larger, loose social network such as Twitter is known, and a user wants to develop a personalized ranking for a small subgraph or its own trusted network taking the global ranking into account. We again can use Dirichlet PageRank with appropriate boundary conditions, as outlined in an algorithm PRInferRanking.

The rest of the paper proceeds as follows. In Section 2, we outline necessary background on PageRank and Dirichlet boundary conditions. Section 3 develops the theory of PageRank with Dirichlet boundary conditions, and Section 4 extends this theory for arbitrary boundary conditions σ. We develop and analyze the ApproxDirichPR algorithm in Section 5 and give algorithms for the previously-discussed applications in Section 6.

2 Preliminaries

For a connected undirected graph $G = (V, E)$ with n vertices and m edges, let A be the *adjacency matrix* and D be the *diagonal degree matrix* where D_{ii} is the degree of the i-th vertex. The typical random walk on G is defined by the *transition probability matrix* $D^{-1}A$.

The *normalized Laplacian* $\mathcal{L}$ is defined as:

$$\mathcal{L} = D^{-1/2}(D - A)D^{-1/2} = I - D^{-1/2}AD^{-1/2}.$$

Let S denote a subset of V consisting of vertices in G. The *vertex boundary* δS is defined as: $\delta S = \{v | v \notin S, (u, v) \in E, \text{where } u \in S\}$, and the *edge boundary* ∂S is defined as: $\partial S = \{(u, v) | u \in S, v \notin S\}$. The *volume* $\mathrm{vol}(S)$ denotes the sum of the degrees of vertices in S.

The *restricted Laplacian* $\mathcal{L}_S$ is the submatrix of $\mathcal{L}$ restricted to $S \times S$. And the *restricted Green's function* $\mathcal{G}_{S,\beta}$ is defined as: $\mathcal{G}_{S,\beta}(\beta I_S + \mathcal{L}_S) = I_S$, where $\beta \geq 0$. Note that $\mathcal{L}_S$ is positive definite [9], so $\mathcal{G}_{S,\beta}$ is well-defined. The *PageRank vector* pr is defined as:

$$\mathrm{pr} = \alpha s + (1 - \alpha)\mathrm{pr}W,$$

where α is called the *jumping constant*, s is the *seed vector*, and $W = \frac{1}{2}(I + D^{-1}A)$ is called the *lazy random walk transition matrix*. PageRank was first introduced in [5] to measure the importance of Web pages, and recently it has been applied to many fields, such as measuring trust in social networks [1].

3 PageRank with Dirichlet Boundary Conditions

Let S be a subset of G; for a function (or vector) $f : V \to \mathbb{R}$, we say f satisfies the Dirichlet boundary condition if $f(v) = 0$ for all $v \in \delta S$.

The PageRank vector satisfying the Dirichlet boundary conditions is defined by the equations:

$$\mathrm{pr}(v) = \begin{cases} \alpha s(v) + (1 - \alpha) \sum_{u \in V} \mathrm{pr}(u) W_{uv} & \text{if } v \in S \\ 0 & \text{otherwise .} \end{cases} \tag{1}$$

Let pr_S, s_S denote the vectors pr and s restricted to S, and W_S, D_S, A_S denote the respective matrices restricted to $S \times S$. Dirichlet boundary conditions are also applied to PageRank vectors in [7].

Theorem 2. *For a connected graph, the above PageRank equation (1) has one and only one solution. With $\beta = \frac{2\alpha}{1-\alpha}$, it is given by*

$$\mathrm{pr}_S = \beta s_S D_S^{-1/2} \mathcal{G}_{S,\beta} D_S^{1/2}.$$

Proof. Since $\mathrm{pr}(v) = 0$ when $v \notin S$, the PageRank equation (1) is equivalent to

$$\mathrm{pr}_S = \alpha s_S + (1 - \alpha)\mathrm{pr}_S W_S.$$

And since

$$W = \frac{1}{2}(I + D^{-1}A) = I - \frac{1}{2}(D^{-1/2}\mathcal{L}D^{1/2}),$$

and D is diagonal matrix, we have

$$\mathrm{pr}_S = \alpha s_S + (1 - \alpha)\mathrm{pr}_S(I - \frac{1}{2}(D_S^{-1/2}\mathcal{L}_S D_S^{1/2})).$$

Solving for pr_S gives the theorem. □

Let pr$'$ be the original PageRank vector with no boundary conditions. Suppose G_S is the subgraph of G consisting of vertices of S and only those edges between vertices in S. Let pr$''$ be the PageRank of G_S.

It is easy to see that for every $v \in S$, $\mathrm{pr}(v) \leq \mathrm{pr}'(v)$; however, since we only care about the relative ranking of the vertices and not the value itself, and the L_1-norm of these vectors is arbitrary (depending on the boundary condition), it is more interesting to compare the relative magnitudes of the PageRank values of different sets of nodes among these 3 definitions of PageRank on the subset S.

Define

$$S_o = \{v \in S | \exists u \notin S : (u, v) \in E\}, S_i = S \setminus S_o,$$

and assume that neither of these two sets are empty.

Let W_{ii} be W restricted to $S_i \times S_i$, W_{0i} be W restricted to $S_o \times S_i$, and

$$w_0 = (1-\alpha)\mathbf{1}W_{0i}^T, w_i = \mathbf{1} - (1-\alpha)\mathbf{1}W_{ii}^T.$$

Lemma 1.

$$\frac{\mathrm{pr}'_{S_o} w_0^T}{\mathrm{pr}'_{S_i} w_i^T} \geq \frac{\mathrm{pr}_{S_o} w_0^T}{\mathrm{pr}_{S_i} w_i^T} \geq \frac{\mathrm{pr}''_{S_o} w_0^T}{\mathrm{pr}''_{S_i} w_i^T}$$

Proof. Let W'' be the lazy random walk transition probability matrix of G_S; then the following equations hold:

$$\mathrm{pr}'_{S_i} = \alpha s_{S_i} + (1-\alpha)\left(\mathrm{pr}'_{S_i} W_{ii} + \mathrm{pr}'_{S_o} W_{0i}\right), \tag{2}$$

$$\mathrm{pr}_{S_i} = \alpha s_{S_i} + (1-\alpha)\left(\mathrm{pr}_{S_i} W_{ii} + \mathrm{pr}_{S_o} W_{0i}\right), \tag{3}$$

$$\mathrm{pr}''_{S_i} = \alpha s_{S_i} + (1-\alpha)\left(\mathrm{pr}''_{S_i} W''_{ii} + \mathrm{pr}''_{S_o} W''_{0i}\right). \tag{4}$$

Let $c_1 = \alpha s_{S_i}\mathbf{1}^T$. Then we have

$$\frac{\mathrm{pr}_{S_o} w_0^T}{\mathrm{pr}_{S_i} w_i^T} = \frac{\mathrm{pr}_{S_i} w_i^T - c_1}{\mathrm{pr}_{S_i} w_i^T}.$$

(2) - (3) gives

$$\left(\mathrm{pr}'_{S_i} - \mathrm{pr}_{S_i}\right)\left(I - (1-\alpha)W_{ii}\right) = \left(\mathrm{pr}'_{S_o} - \mathrm{pr}_{S_o}\right)\left((1-\alpha)W_{0i}\right)$$

Let $c_2 = \left(\mathrm{pr}'_{S_i} - \mathrm{pr}_{S_i}\right)\left(I - (1-\alpha)W_{ii}\right)\mathbf{1}^T$. Since $\forall v \in S, \mathrm{pr}(v) \leq \mathrm{pr}'(v)$ implies that $\mathrm{pr}'_{S_i} - \mathrm{pr}_{S_i}$ is nonnegative, c_2 is also nonnegative. Then we have

$$\frac{\mathrm{pr}'_{S_o} w_0^T}{\mathrm{pr}'_{S_i} w_i^T} = \frac{\mathrm{pr}'_{S_i} w_i^T - c_1}{\mathrm{pr}'_{S_i} w_i^T} = \frac{\mathrm{pr}_{S_i} w_i^T + c_2 - c_1}{\mathrm{pr}_{S_i} w_i^T + c_2}.$$

Since for every $v \in S_o$, the degree decreases by at least 1 from G to G_S, $W''_{v,u} \geq W_{v,u}(1 + \frac{1}{d})$, where d is the maximum degree of vertices in S_o of G_S. And for $v \in S_i$, there is no change in degree from G to G_S, so $W_{ii} = W''_{ii}$. Hence we have

$$\begin{aligned}\mathrm{pr}''_{S_i}\left(I - (1-\alpha)W_{ii}\right) &= \alpha s_{S_i} + \mathrm{pr}''_{S_o}\left((1-\alpha)W''_{0i}\right) \\ &\geq \alpha s_{S_i} + \mathrm{pr}''_{S_o}\left((1-\alpha)W_{0i}\right)(1 + \frac{1}{d}).\end{aligned}$$

So

$$\frac{\mathrm{pr}''_{S_o} w_0^T}{\mathrm{pr}''_{S_i} w_i^T} \leq \frac{\mathrm{pr}_{S_i} w_i^T - c_1}{(1+\frac{1}{d})\mathrm{pr}_{S_i} w_i^T},$$

and the lemma follows from

$$\frac{\mathrm{pr}_{S_i} w_i^T + c_2 - c_1}{\mathrm{pr}_{S_i} w_i^T + c_2} \geq \frac{\mathrm{pr}_{S_i} w_i^T - c_1}{\mathrm{pr}_{S_i} w_i^T} \geq \frac{\mathrm{pr}_{S_i} w_i^T - c_1}{(1+\frac{1}{d})\mathrm{pr}_{S_i} w_i^T}$$

□

Intuitively, since pr'' ignores all the boundary edges, pr''_{S_o} is inaccurate and underestimated. On the other hand, we want to decrease the influence of boundary nodes, so we want the result PageRank on S_o not to be overestimated compared with pr'_{S_o}. Lemma 1 shows that pr_{S_o} is bounded between pr''_{S_o} and pr'_{S_o}; therefore it is preferred in such cases.

4 Dirichlet PageRank with Given Boundary Conditions

In some cases, we want to further decrease or increase the influence of the boundary nodes, or we already have some estimation of the PageRank on the boundary of some vertex set S and want to approximate the PageRank of S very quickly. Then instead of setting $p(v) = 0$ for all $v \notin S$, we can set them according to arbitrary boundary conditions σ.

The Dirichlet PageRank with given boundary conditions σ is defined by the equations:

$$\mathrm{pr}(v) = \begin{cases} \alpha s(v) + (1-\alpha)\sum_{u \in V} \mathrm{pr}(u) W_{uv} & \text{if } v \in S \\ \sigma(v) & \text{otherwise} \end{cases}, \qquad (5)$$

where $\sigma(v) \geq 0, \forall v$ and $|\sigma| \leq 1$.

Let $W_{\delta S}$ denote W restricted to $\delta S \times S$.

Theorem 3. *For a connected graph, the above PageRank equation (5) has one and only one solution. With* $\beta = \frac{2\alpha}{1-\alpha}$, *it is given by*

$$\mathrm{pr}_S = (\beta s_S + 2\sigma_{\delta S} W_{\delta S}) D_S^{-1/2} \mathcal{G}_{S,\beta} D_S^{1/2}.$$

Proof. Notice that

$$\begin{aligned} \mathrm{pr}_S &= \alpha s_S + (1-\alpha)\left(\mathrm{pr}_S W_S + \sigma_{\delta S} W_{\delta S}\right) \qquad (6) \\ &= \frac{1-\alpha}{2}(\beta s_S + 2\sigma_{\delta S} W_{\delta S}) + (1-\alpha)\mathrm{pr}_S W_S. \end{aligned}$$

Then it is straightforward to see the theorem holds by following the steps as in the proof of Theorem 2. □

5 Algorithms and Analysis

To solve the PageRank equations (1, 5) with boundary conditions, as implied by Theorem 2 and Theorem 3, all it requires are vector-matrix multiplication and solving a linear system:

$$x(\beta I_S + \mathcal{L}_S) = y.$$

Since D is diagonal, the complexity of solving the PageRank is just the complexity of solving the linear system. And since the matrix $\beta I_S + \mathcal{L}_S$ is diagonally dominant, it can be solved approximately in nearly linear time with a Spielman-Teng Solver [17]. Here, we show a simpler algorithm ApproxDirichPR to solve the PageRank equations with boundary conditions approximately, which is faster and has a better approximation ratio as long as α is not too small.

The basic outline of our algorithm ApproxDirichPR is as follows: we initialize pr_S as $\mathbf{0}$ and maintain a residue r, which is the difference between the right side and left side of equation 6. Then we gradually move mass from r to pr_S while maintaining the following invariant:

$$\mathrm{pr}_S + r = \alpha s_S + (1-\alpha)\left(\mathrm{pr}_S W_S + \sigma_{\delta S} W_{\delta S}\right)$$

until for every $v \in S$, $r(v) \le \epsilon' d_v$. In the beginning, we set $\epsilon' = 1$; after every iteration, we decrease ϵ' by half until $\epsilon' \le \epsilon$ which is the given desired approximation ratio.

Algorithm 1 ApproxDirichPR

```
Input: $G$, $S$, $\alpha$, $s$, $\sigma$, $\epsilon$
Output: $\mathrm{pr}_S$
  $\mathrm{pr}_S \Leftarrow \mathbf{0}$, $\epsilon' \Leftarrow 1$, $r \Leftarrow \alpha s_S + (1-\alpha)\sigma_{\delta S} W_{\delta S}$
  while $\epsilon' > \epsilon$ do
    while $r(v) \ge \epsilon' d_v$ for some $v$ do
      $\mathrm{pr}_S(v) \Leftarrow \mathrm{pr}_S(v) + r(v)$
      For each neighbor $u$ of $v$, $r(u) \Leftarrow r(u) + (1-\alpha) r(v)/2d_v$
      $r(v) \Leftarrow (1-\alpha) r(v)/2$
    end while
    $\epsilon' \Leftarrow \epsilon'/2$
  end while
```

Now that we have presented our algorithm, we will prove Theorem 1.

Proof. (of Theorem 1) Since a FIFO queue can be used to store every vertex v such that $r(v) \ge \epsilon' d_v$, each iteration of the inner loop can be done in $O(d_v)$ time. For each iteration of the outer loop, since $|r| \le 2\epsilon' \mathrm{vol}(S)$ at the beginning, and $r(v)$ will decrease at least $\alpha\epsilon' d_v$, let T be the number of iterations and $v_i, 1 \le i \le T$ be the vertex selected at the i-th step, we have

$$\sum_{i=1}^{T} \alpha \epsilon' d_{v_i} \le 2\epsilon' \mathrm{vol}(S),$$

so

$$\sum_{i=1}^{T} d_{v_i} \leq \frac{2\mathrm{vol}(S)}{\alpha}.$$

There are $\log \frac{1}{\epsilon}$ iterations of the outer loop, so the running time is $O(\frac{\mathrm{vol}(S)\log\frac{1}{\epsilon}}{\alpha})$.

The output $\widetilde{\mathrm{pr}}_S$ satisfies:

$$\widetilde{\mathrm{pr}}_S + r = \alpha s_S + (1-\alpha)\left(\widetilde{\mathrm{pr}}_S W_S + \sigma_{\delta S} W_{\delta S}\right),$$

where $0 \leq r(v) < \epsilon d_v \forall v \in S$, and the exact solution pr_S satisfies:

$$\mathrm{pr}_S = \alpha s_S + (1-\alpha)\left(\mathrm{pr}_S W_S + \sigma_{\delta S} W_{\delta S}\right).$$

Subtracting these two equations gives:

$$\mathrm{pr}_S - \widetilde{\mathrm{pr}}_S = \alpha\frac{r}{\alpha} + (1-\alpha)\left((\mathrm{pr}_S - \widetilde{\mathrm{pr}}_S)W_S\right),$$

which is also a PageRank equation. Since r is nonnegative,

$$\widetilde{\mathrm{pr}}_S(v) \leq \mathrm{pr}_S(v), \forall v \in S.$$

And by the properties of PageRank that $|\mathrm{pr}_{\alpha,s}| \leq |s|$ and $|\mathrm{pr}_S| \leq |\mathrm{pr}|$, we have

$$|\mathrm{pr}_S - \widetilde{\mathrm{pr}}_S| \leq |\frac{r}{\alpha}| < \frac{\epsilon\mathrm{vol}(S)}{\alpha}.$$

□

6 Applications of Dirichlet PageRank

6.1 Adjusting Spammers' Influence

One downfall of pure link-based ranking systems such as PageRank is that they interpret all nodes as honest agents and all links as votes or validation between nodes. However, real-world networks such as the World Wide Web often contain malicious nodes or spammers. It then becomes an important question to find ranking systems that better represent the true, honest ranking of nodes in the graph.

There are many schemes developed to try to combat this problem [1,3,4,6, 11,12,15,16], but it turns out that many of them can be modeled using Dirichlet PageRank with different boundary conditions. This will allow for the efficient consideration of many different models by simply considering different boundary conditions. For example, [3] outlines an algorithm SpamRank which penalizes spam nodes. Using Dirichlet boundary conditions, we can penalize known spammers v by enforcing the condition $\mathrm{pr}(v) = 0$. One can adjust their ranking even further by enforcing $\mathrm{pr}(v) = -1$.

Another paper [4] concerns propagating trust and distrust within a network, using a weighted random walk W with a trusted seed vertex s. Here, the authors start by assigning rank $\mathrm{pr}(s) = 1$, a condition covered by Dirichlet PageRank with the boundary condition $\sigma(s) = 1$. There is a subtle difference in the way distrust is handled (the original algorithm does not allow for the propagation of trust scores less than 0), but it should be clear that Dirichlet PageRank allows us to efficiently consider these and many other models.

6.2 Adjusting Rank Based on Trust

While it is interesting to be able to devise ranking systems that take known spammers into account, it is also important to calculate a ranking based on various concepts of trust in a network. There are numerous scenarios to consider, and Dirichlet PageRank with boundary conditions will be a useful algorithmic tool.

Consider the following problem: in a network G, node v wants to compute a personalized ranking of the nodes, but v trusts its own friends and wants its ranking on the top ρ fraction of nodes to be similar to its friends'. Presumably one's friends' actions carry a lot of weight in one's own decisions. Vertex v can efficiently compute a personalized PageRank vector as its ranking function using algorithms from [10], but PageRank alone will not take into account the implied trust between v and its friends. But using Dirichlet PageRank with boundary conditions, we can take v's trusted friends into account. We illustrate this in the algorithm PRTrustedFriends.

Algorithm 2 PRTrustedFriends

```
Input: $G = (V, E)$, $v$, $\alpha$, $\rho$, $\epsilon$
Output: $\boldsymbol{p}$
  $\boldsymbol{p} \Leftarrow \mathrm{SharpApproximatePR}(v, \alpha, \epsilon)$ [10]
  $\boldsymbol{p'} \Leftarrow \frac{1}{\sum_{u\sim v} p(u)} \sum_{u\sim v} p(u)\mathrm{SharpApproximatePR}(u, \alpha, \epsilon)$
  $S \Leftarrow \arg\max_{S\subseteq V, |S|\leq \rho|V|} \sum_{s\in S} p(s)$
  $\boldsymbol{p} \Leftarrow \mathrm{ApproxDirichPR}(G, V \setminus S, \alpha, v, \boldsymbol{p'}, \epsilon)$
```

A natural extension of PRTrustedFriends is a similar problem where v is a newcomer to a network and is therefore unsure about what other nodes are trustworthy. In such a scenario, the only available information to v is the network itself. For ranking purposes, v can select a small sample of nodes to compare with its own ranking; if these nodes are well distributed, they provide a good control to ensure that v's own ranking function is too distorted by the presence of nearby spam or malicious nodes. We give the algorithm PRValidation.

A third, more complex situation arises in the context of different types of social networks. Although the problem setup here appears rather complicated, it is a natural model for a common social phenomenon: a distinction between different types of social graphs.

Suppose that a vertex v is part of two networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ with $V_1 \subseteq V_2$. We interpret G_1 as a closely-knit social network where

Algorithm 3 PRValidation

```
Input: $G = (V, E)$, $v$, $k$, $\alpha$, $\rho$, $\epsilon$
Output: $\boldsymbol{p}$
  $\boldsymbol{p} \Leftarrow$ SharpApproximatePR$(v, \alpha, \epsilon)$ [10]
  $v_1, \ldots, v_k \Leftarrow$ i.i.d. samples from $V$ according to $\boldsymbol{p}$
  $\boldsymbol{p'} \Leftarrow \frac{1}{\sum_{i=1}^{k} p(v_k)} \sum_{i=1}^{k} p(v_k)$SharpApproximatePR$(u, \alpha, \epsilon)$
  $S \Leftarrow \arg\max_{S \subseteq V, |S| \leq \rho |V|} \sum_{s \in S} p(s)$
  $\boldsymbol{p} \Leftarrow$ ApproxDirichPR$(G, V \setminus S, \alpha, v, \boldsymbol{p'}, \epsilon)$
```

edges represent a deeper connection with the implication that the endpoints share mutual trust for one another. G_2 is a larger network where nodes form edges for looser reasons; for example, acquaintanceship or curiosity. We assume that v does not know much about the many sources of information present in G_2. An important question for v is: which nodes in G_2 are trustworthy? Is there some way to rank the vertices of G_2?

One effective way of finding such a ranking of vertices in G_2 for a node v is by computing the ranking on G_1 and then computing Dirichlet PageRank on G_2 using G_1's ranking as the boundary condition. This is outlined in the algorithm PRTrustNetwork.

Algorithm 4 PRTrustNetwork

```
Input: $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$, $v \in V_1 \cap V_2$, $\alpha$, $\epsilon$
Output: $\boldsymbol{q}$
  $\boldsymbol{p} \Leftarrow$ SharpApproximatePR$(v, \alpha, \epsilon)$ [10] for $G_1$
  $\boldsymbol{q} \Leftarrow$ ApproxDirichPR$(G_2, V_2 \setminus V_1, \alpha, v, \boldsymbol{p}, \epsilon)$
```

Dirichlet PageRank can also be used to solve a related problem if a global ranking for G_2 is already known or pre-computed. Suppose that such a ranking exists, and $v \in G_2$ wants to be able to rank its own more personal network or neighborhood G_1 taking this into account. One way to do this is to compute a Dirichlet PageRank vector for G_1, but using the nodes adjacent to G_1 as a boundary with rank given by the global ranking on G_2. This procedure is given in the following algorithm PRInferRanking.

Algorithm 5 PRInferRanking

```
Input: $G_2 = (V_2, E_2)$, $G_1 = (V_1, E_1) \subseteq G_2$, $v \in V_1$, $\boldsymbol{p}$, $\alpha$, $\epsilon$
Output: $\boldsymbol{q}$
  $\partial E_1 \Leftarrow \{(w, x) \in E_2 | w \in V_1, x \notin V_1\}$
  $\partial V_1 \Leftarrow \{w \in V_2 \setminus V_1 | w$ is an endpoint of an $e \in \partial E_1\}$
  $\boldsymbol{q} \Leftarrow$ ApproxDirichPR$((V_1 \cup \partial V_1, E_1 \cup \partial E_1), V_1, \alpha, v, \boldsymbol{p}, \epsilon)$
```

From the examples above, we see that Dirichlet PageRank with boundary conditions is a useful tool, especially in modeling trust and distrust in a network ranking system. It is of interest to further take advantage of the efficient computation and approximation of Dirichlet PageRank vectors. More applications and directions remain to be explored.

References

1. R. Andersen, C. Borgs, J. Chayes, U. Feige, A. Flaxman, A. Kalai, V. Mirrokni, and M. Tennenholtz. Trust-based recommendation systems: an axiomatic approach. In *WWW* 2008.
2. R. Baeza-Yates, C. Castillo and V. López. PageRank increase under different collusion topologies. In Proceedings of the 1st International Workshop on Adversarial Information Retrieval on the Web, 2005.
3. A. Benczur, K. Csalogany, T. Sarlos and M. Uher. SpamRank — fully automatic link spam detection. In Proceedings of the 1st International Workshop on Adversarial Information Retrieval on the Web, 2005.
4. C. Borgs, J. Chayes, A.T. Kalai, A. Malekian and M. Tennenholtz. A novel approach to propagating distrust. WINE 2010.
5. S. Brin and L. Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems*, **30 (1-7)**, (1998), 107-117.
6. A. Cheng and E. Friedman, Sybilproof reputation mechanisms. In Proceedings of Third Workshop on Economics of Peer-to-Peer Systems, 2005.
7. F. Chung, PageRank as a discrete Green's function, *Geometry and Analysis* I, ALM 17 (2010), 285–302.
8. F. Chung, *Spectral Graph Theory*, AMS Publications, 1997.
9. F. Chung and S.-T. Yau, Discrete Green's functions, *J. Combinatorial Theory (A)* **91** (2000), 191–214.
10. F. Chung and W. Zhao. A sharp PageRank algorithm with applications to edge ranking and graph sparsification. *Proceedings of Workshop on Algorithms and Models for the Web Graph* (WAW 2010), *Lecture Notes in Computer Science* **6516**, 2–14.
11. R. Guha, R. Kumar, P. Raghavan and A. Tomkins, Propagation of trust and distrust. In *WWW* 2004.
12. Z. Gyöngyi, H. Garcia-Molina and J. Pedersen, Combating Web spam with TrustRank. In *VLDB* 2004.
13. T. Haveliwala, Topic-sensitive PageRank: A context-sensitive ranking algorithm for Web search, *IEEE Transactions on Knowledge and Data Engineering* **15** (2004), 784–796.
14. G. Jeh and J. Widom, Scaling personalized Web search. In *WWW* 2003.
15. S. Kamvar, M. Schlosser and H. Garcia-Molina. The EigenTrust algorithm for reputation management in P2P networks. WWW 2003.
16. C. de Kerchove and P. Dooren. The PageTrust algorithm: how to rank Web pages when negative links are allowed? In Proceedings of the SIAM International Conference on Data Mining (2008).
17. D. A. Spielman and S. -H. Teng, Nearly-Linear Time Algorithms for Preconditioning and Solving Symmetric, Diagonally Dominant Linear Systems, 2008. Available at `http://arxiv.org/abs/cs.NA/0607105`.