CPSC 532L Project
Development and Axiomatization of a Ranking System

Catherine Gamroth
cgamroth@cs.ubc.ca

Hammad Ali
hammada@cs.ubc.ca

April 22, 2009

Abstract

Ranking systems are central to many internet applications including, notably, Google's PageRank algorithm for ranking web pages. Ranking systems are a special case of a social choice problem in which the set of agents and the set of outcomes coincide. In this paper we consider PageRank as a particular ranking system and we present two axiomatizations that allow PageRank to be studied from a social choice perspective. The first axiomatization is normative and leads to the theory that no ranking system can simultaneously satisfy two desirable properties. Despite this discouraging result, PageRank is nonetheless highly successful and the second, descriptive, axiomatization characterizes PageRank uniquely among ranking systems. We conclude by suggesting areas for further research, including axioms describing vulnerability to manipulation and preferences submitted directly by human users.

1 Introduction

Ranking systems play an important role in the domain of e-commerce and internet technologies by creating global rankings of websites or on-line trade partners. Two well-known examples are eBay's reputation system and Google's PageRank algorithm. Although these systems have been very successful, very little work was initially done to formally characterize their performance and understand their theoretical bases. In response, Altman and Tennenholtz studied page ranking systems in a social choice framework ([1], [2]). In particular, they addressed the questions of what properties a ranking system should have, and which properties of a particular page ranking algorithm differentiate it from other page ranking algorithms.

In a scenario with multiple self-interested agents and several possible outcomes to be chosen from, it is often desirable to collect the agent's preferences for these outcomes and aggregate these to make a final choice that reflects the preferences of the agent population. The most common example of such a problem would be voters voting for their most-preferred candidate, where ideally we want to eventually select a candidate who best reflects the collective preferences of the voters. This general class or problems is referred to as social choice problems. Social choice theory is the study of how to solve this problem in a way so as to satisfy certain desirable properties. Social choice theory is nonstrategic - agents have preferences, and no agent tries to hide their true preferences in order to manipulate the outcome. The agents report their preferences as an ordering over the outcomes, which expresses their order of preference over these.

By Arrow's impossibility theorem, there is no social choice function that satisfies some minimal requirements that we would want such functions to have. However, an appropriate social choice function can be achieved by relaxing any one of these requirements. As we will explore in this paper, a similar set of properties and results can be derived for the special case of page ranking.

In this paper we will look at an axiomatic approach that allows us to formally define the PageRank ranking system in the context of social choice theory. In the next section we introduce ranking systems as a special case of social choice and in section 3 we define the PageRank ranking system. Section 4 introduces normative axioms that one may require for any ranking system, and descriptive axioms that relate to PageRank in particular. Further discussion of these axiomatizations is presented in section 5.

2 Ranking Systems

One interesting special case of the social choice problem occurs when the set of agents and the set of outcomes coincide. In such a setting, the agents express their evaluations of each other by submitting a two-level preference (vote or no vote) for each other agent. This case is referred to as the problem of ranking systems, and has several interesting features. For one, agents voting for any subset of the other agents introduces transitive effects - being voted for by an agent who has been voted for by others is deemed more important than being voted for by an agent who has not been voted for.

Ranking systems are not governed by Arrow's impossibility theorem [3]. In a ranking system, agents partition outcomes into only two sets while Arrow's theorem is defined in a setting with arbitrary preferences and three or more possible outcomes. As we shall see in the following sections there are properties similar to Arrow's axioms that we would like to hold in a ranking systems setting.

Ranking systems share some commonalities with approval voting. In approval voting, each agent can cast a single vote for as many of the outcomes as he wishes, and the outcome with the most votes is selected. While the concept of each agent voting for a number of outcomes is similar to ranking systems, it is important to note that in ranking systems the set of agents and outcomes coincide, which is not the case with approval voting.

3 PageRank

The PageRank algorithm models the link structure of the internet as a directed graph $G = (V, E)$ where $v \in V$ is a web page and $(v_1, v_2) \in E$ is a link from page 1 to page 2 and represent a vote for page 2 by page 1. The basic idea is that the rank of a page is calculated based on the number of votes a page received and the ranking of the voters.

An intuitive model that can help understand PageRank is the Random Surfer Model [4], which corresponds to the standing probability distribution of a random walk along the graph G. This models the behaviour of a random surfer who clicks on successive links at random. Each link from a given page is assigned an equal probability. This captures the idea that pages that have been linked to more often will be reached by the random surfer with a higher probability.

One point worthy of mention is that if a real web surfer is caught in a small loop of web pages, it is unlikely that the surfer will continue in the loop forever. Instead, he or she will just jump to some other page. This behaviour is also reflected the Random Surfer Model; with some probability the surfer stops following successive links and moves to some random page.

We now define the PageRank matrix which captures this idea formally. We begin with the following definition:

Definition 3.1 *Let $G = (V, E)$ be some graph and $v \in V$ be some vertex. Let $P_G(v) = \{u|(u, v) \in E\}$ and $S_G(v) = \{u|(u, v) \in E\}$ denote the predecessor and successor sets of v in G respectively. When G is understood from context, we will use $P(v)$ and $S(v)$.*

The *PageRank Matrix* A_G (of dimension $n \times n$) is defined as:

$$[A_G]_{i,j} = \begin{cases} 1/|S_G(v_j)| & (v_i, v_j) \in E \\ 1 & \text{otherwise} \end{cases}$$

Using the random surfer model, we may define PageRank as follows:

Definition 3.2 *Let $G = (V, E)$ be some strongly connected graph, and assume $V = \{v_1, v_2, ..., v_n\}$. Let $\boldsymbol{r}$ be the unique solution of the system $A_G \cdot \boldsymbol{r} = \boldsymbol{r}$ where $r_1 = 1$. The **PageRank** $PR_G(v_i)$ of a vertex $v_1 \in V$ is defined as $PR_G(v_i) = r_i$. The **PageRank** ranking system is a ranking system that for the vertext set V maps G to $\succeq_G^{PR}$, where $\succeq_G^{PR}$ is defined as: for all $v_i, v_j \in V$: $v_i \succeq_G^{PR} v_j$ if and only if $PR_G(v_i) \leq PR_G(v_j)$. (Reproduced from [1])*

To achieve this, a rank is then calculated for each page, based on the sum of the links to that page. As illustrated in Figure 1, this calculation takes into consideration the fact that links from less important pages have less weight than links from more important pages. It also divies the weight of a page equally among its links to capture the idea that the fewer successors a page has, the more strongly it is recommending each of them.

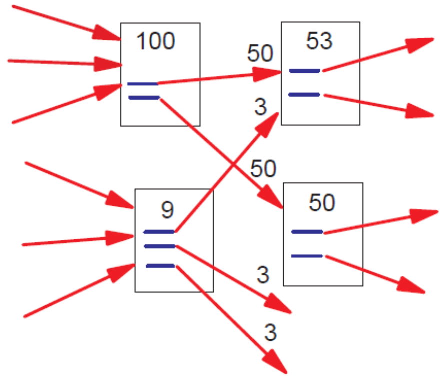


Figure 1: Simplified PageRank Calculation (reproduced from [4])

$$PR(v_1) = \sum_{v \in P(v_1)} \frac{PR(v)}{|S(v)|} \tag{1}$$

We should mention here that PageRank as described in [4] and [5] is vulnerable to manipulation. One way in which it can be manipulated is that an entity is free to introduce as many agents into the graph as he chooses, and is able to control their preferences [6]. This and many other considerations are accounted for in the PageRank algorithm currently used in the Google search engine, although details of that particular implementation have not been published.

4 Axiomatizing PageRank

We will examine two approaches to axiomatizing PageRank: normative and descriptive. A normative approach starts by defining properties that are desirable in a social choice function, and attempts to identify functions that satisfy these properties. Sometimes this normative approach leads to an impossibility theorem that proves that no such function exists, such as Arrow's impossibility theorem. By contrast, a descriptive approach starts by analyzing a particular function and attempts to fully characterize it using a minimal set of axioms. These axioms should be selected in a way such that any other function that satisfies them must coincide with the function under consideration. A set of axioms that has these two properties is termed a representation theorem, and can be an effective tool in the understanding and comparison of different social choice functions. In this section we present an overview of the axioms and resulting theorems. We do not intend to reproduce the proofs of the theorems; the interested reader is referred to the original papers for details.

4.1 Normative Axioms

We have seen that Arrow's impossibility theorem does not apply to page ranking, but it turns out that another, similar impossibility theorem exists. This theorem is general to all ranking systems and considers two properties: transitivity and ranked independence of irrelevant alternatives (RIIA).

The first property, transitivity, is an important idea in the PageRank algorithm; by this property, a vote from a highly ranked page is given higher importance that a vote from a lower-ranked page. Thus, a vote for page a is indirectly also a vote for all pages linked to by a.

Axiom 4.1.1 (Strong Transitivity) *Let F be a ranking system. We say that F satisfies strong transitivity if for all graphs $G = (V,E)$ and for all vertices $v_1, v_2 \in V$: Assume that there is a 1-1 mapping (but not necessarily onto) $f : P(v_1) \mapsto P(v_2)$ such that for all $v \in P(v_1) : v \succeq f(v)$. Then $v_1 \succeq v_2$. Further assume that either f is not onto or for some $v \in P(v_1) : v \succ f(v)$. Then $v_1 \succ v_2$. (Reproduced from [4])*

Strong transitivity states that page a should be ranked higher than page b if page a received at least as many votes as page b and each vote for b can be paired in a 1-1 manner with a vote for a from a higher or equally-ranked voter. We also wish to define a weaker requirement that transitivity hold when comparing agents whose predecessors have an equal number of successors.

Axiom 4.1.2 (Weak Transitivity) *Let F be a ranking system. We say that F satisfies weak transitivity if for all graphs $G = (V,E)$ and for all vertices $v_1, v_2 \in V$: Assume that there is a 1-1 mapping $f : P(v_1) \mapsto P(v_2)$ such that for all $v \in P(v_1) : v \succeq f(v)$ and $|S(v)| = |S(f(v))|$. Then $v_1 \succeq v_2$. Further assume that either f is not onto or for some $v \in P(v_1) : v \succ f(v)$. Then $v_1 \succ v_2$. (Reproduced from [4])*

We see that PageRank does not satisfy strong transitivity since the relative importance of a link from page a is inversely proportional to $|S(a))|$ by equation 1. However, according to [4], strong transitivity is a less desirable property than weak transitivity in this application. By the following equation, we see that weak transitivity holds:

$$PR(v_1) = \sum_{v \in P(v_1)} \frac{PR(v)}{|S(v)|} \leq \sum_{v \in P(v_1)} \frac{PR(v)}{|S(f(v))|} \leq \sum_{v \in P(v_2)} \frac{PR(v)}{|S(v)|} = PR(v_2) \quad (2)$$

Note that PageRank relies on mathematical calculations, but this is a particular implementation, not necessarily a property of the family of algorithms that satisfy weak transitivity.

The second normative property, RIIA, is similar to Arrow's IIA, except that the ranking system considers the rank, but not the identity, of the voters.

Axiom 4.1.3 (Ranked Independence of Irrelevant Alternatives (RIIA)) *A ranking rule satisfies RIIA if the relative rank between pairs of outcomes is always determined according to the same rule, this rule depends only on: the number of votes each outcome received; and the relative ranks of these voters. (Reproduced from [3])*

It can be shown by counter-example that PageRank does not satisfy RIIA [2]. Approval voting, however, does satisfy RIIA since by approval voting the candidate with the greatest number of votes receives the highest ranking. Of course approval voting does not satisfy strong or weak transitivity since the rankings of the voters are not considered. Therefore, these axioms can be satisfied independently and the question then becomes whether there exists a class of ranking system that simultaneously satisfy both transitivity and RIIA. The details are omitted here, but it is shown in [2] that no such ranking system exists.

Theorem 4.1 *No ranking system can simultaneously satisfy weak transitivity and RIIA [2].*

4.2 Descriptive Axioms

Since PageRank is such a widely-used and highly influential ranking system, it is beneficial to have a descriptive axiomatization as a means of characterizing and evaluating the ranking system. What follows are five desirable properties that describe PageRank (reproduced from [1]).

The first axiom states that the ranking procedure should be independent of the names of the pages.

Axiom 4.1 (Isomorphism) *A ranking system F satisfies isomorphism if for every isomorphism function $\varphi : V_1 \mapsto V_2$, and two isomorphic graphs $G \in \mathbb{G}_{V_1}$, $\varphi(G) \in \mathbb{G}_{V_2} : \succeq^F_{\varphi(G)} = \succeq^F_G$.*

The second axiom states that if a does not link to itself and a has a rank at least as high as b, then adding a link from a to itself should result in a being ranked higher than b. The relative ranking of all pages other than a should remain unchanged by the addition of the link from a to itself.

Axiom 4.2 (Self Edge) *Let F be a ranking system. F satisfies the self edge axiom if for every vertex set V and for every vertex $v \in V$ and for every graph $G = (V, E) \in \mathbb{G}_V$ s.t. $(v, v) \notin E$, and for every $v_1, v_2 \in V \ \{v\}$: Let G' =[insert text!] If $v_1 \succeq^F_G v$ then $v \not\preceq^F_{G'} v_1$; and $v_1 \succeq^F_G v_2$ iff $v_1 \succeq^F_{G'} v_2$.*

The third axiom states that the relative ranking of all pages should be unchanged if a set of pages is inserted between page a and a's successors, and the following changes to the links are made: all original links from a are removed; one link is inserted from a to each of the new pages; and one link is added from each of the new pages to all pages previously linked to by a. This is illustrated in Figure 2.

Axiom 4.3 (Vote by Committee) *Let F be a ranking system. F satisfies vote by committee if for every vertex set V, for every vertex $v \in V$, for every graph $G = (V, E) \in \mathbb{G}_V$, for every $v_1, v_2 \in V$, and for every $m \in \mathbb{N}$: Let $G' = (V \cup \{u_1, u_2, ..., u_m\}, E \backslash \{(v, x)|x \in S_G(v)\} \cup \{(v, u_i)|i = 1, ...m\} \cup \{(u_i, x)|x \in S_G(v), i = 1, ..., m\})$, where $\{u_1, u_2, ..., u_m\} \cap V = \emptyset$. Then, $v_1 \succeq^F_G v_2$ iff $v_1 \succeq^F_{G'} v_2$.*

The fourth axiom states that if pages a and b have no predecessors in common and a and b link to the same set of pages, then the relative ranking of all pages except a and b are unchanged if a and b are removed and replaced by a single node, say c, which is linked to by all of a and b's predecessors, and which links to all of a and b's sucessors. This is illustrated in Figure 2.

Axiom 4.4 (Collapsing) *Let F be a ranking system. F satisfies collapsing if for every vertex set V, for every $v, v' \in V$, for every $v_1, v_2 \in V \backslash \{v, v'\}$, and for every graph $G = (V, E) \in \mathbb{G}_V$ for which $S_G(v) = S_G(v')$, $P_G(v) \cap P_G(v') = \emptyset$, and$[P_G(v) \cup P_G(v')] \cap \{v, v'\} = \emptyset$: Let $G' = (V \backslash \{v'\}, E \backslash \{(v', x)|x \in S_G(v')\} \backslash \{(x, v')|x \in P_G(v')\} \cup \{(x, v)|x \in P_G(v')\})$. Then $v_1 \succeq^F_G v_2$ iff $v_1 \succeq^F_{G'} v_2$.*

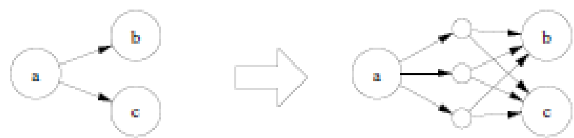
The final axiom states that if a has k predecessors, all with equal ranks, and a also has k sucessors, then the relative ranking of all pages excluding a should be unchanged if a is removed and each of a's predecessors is linked to one of a's successors in a 1-1 manner. This is illustrated in Figure 2.

Axiom 4.5 (Proxy) *Let F be a ranking system. F satisfies proxy if for every vertex set V, for every $v \in V$, for every $v_1, v_2 \in V \backslash \{v\}$, and for every graph $G = (V, E) \in \mathbb{G}_V$ for which $|P_G(v)| = |S_G(v)|$, for all $p \in P_G(v) : S_G(p) = \{v\}$, and for all $p, p' \in P_G(v)$: $p \simeq^F_G p'$: Assume $P_G(v) = \{p_1, p_2, ..., p_m\}$ and $S_G(v) = \{s_1, s_2, ..., s_m\}$. Let $G' = (V \backslash \{v\}, E \backslash \{(x, v), (v, x)|x \in V\} \cup \{(p_i, s_i)|i \in \{1, ..., m\})$. Then $v_1 \succeq^F_G v_2 iff v_1 \succeq^F_{G'} v_2$.*

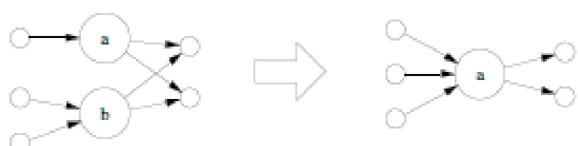
Theorem 4.2 *The PageRank algorithm presented in [4] satisfies isomorphism, self edge, vote by committee, collapsing, and proxy. In addition, every ranking system that satisfies these properties coincides with PageRank.*

5 Discussion

Although the "true" ranking of any webpage is inherently subjective, the goal of automatically generating rankings requires objective measures of success. By modeling page ranking systems in a graph-theoretic way, we are able to reason



(a) Vote by Committee



(b) Collapsing

(c) Proxy

Figure 2: Visualization of Axioms (reproduced from [1])

from a social choice perspective. Axioms provide us with objective requirements for our system and allow for characterization and comparison.

Although it is discouraging to learn that no ranking system can satisfy the two basic axioms of transitivity and RIIA, we have clear evidence from the success of Google that page ranking systems can nonetheless be valuable and can provide satisfying results to human users. By studying transitivity and RIIA, we can understand the trade-offs that are made in designing a ranking system, and also better understand the differences between approaches such as approval voting and PageRank. Among other things, this helps us improve existing ranking systems and understand appropriate settings for their use.

The normative approach allowed us to identify that PageRank satisfies weak transitivity, but did not characterize the algorithm uniquely among other models that might also satisfy weak transitivity. The descriptive approach provides more detail on the properties that hold when PageRank is represented as a graph that satisfies weak transitivity. This provides us with insight into the working of PageRank and gives us a framework in which to analyse any modifications that are made to the PageRank algorithm due to practical considerations.

It should be noted that the descriptive axioms discussed here are based on the PageRank algorithm as described in [4]. It is known that Google's PageRank accounts for other factors such as relevance of search words on a page, susceptability to manipulation, and actual visits to the page. Since these details have not been published and are likely being regularly updated, it is reasonable to restrict the social choice analysis to the basic graph-theoretical model. However, in practice, these axioms are likely not a complete characterization.

In particular, the axiomatic treatment of ranking system in [1] and [2] does not address vulnerability to manipulation. In practice, this is such an important consideration that it has driven many changes to the PageRank algorithm. As such, it may be desirable to create a set of normative axioms to formally describe different ways that ranking systems could be manipulated and desirable properties of a ranking system to minimise this problem.

Many of the ways that the true PageRank algorithm differs from that presented in [4] are the result of one fundamental assumption: that links are a direct indicator of the relevance of a page. Of course, this is not the case. Although we rank pages based on the votes of a set of agents (pages), we are hoping to provide a ranking that satisfies a different set of agents (human users). This is interesting since at no point in the ranking process have we directly received preferences from the users we are trying to satisfy. This consideration has not yet been formally addressed in othe social choice literature. Although this can't be incorporated into a descriptive model without more information on the actual implementation, it may be beneficial to consider a normative approach whereby the user's site visits and result click-throughs are treated as a form of approval voting that occurs after the initial ranking. This would allow for a formal model of preferences submitted by human users to be compared to the results presented to those users.

6 Conclusions

Page ranking is a special case of social choice where the set of agents coincides with the set of outcomes and each agent ranks the others on a two-level preference system. As such, page ranking is not governed by Arrow's impossibility theorem, but a similar theorem exists stating that no ranking system can simultaneously satisfy weak transitivity and RIIA. PageRank is a well-known and widely-used page ranking system that satisfies weak transitivity but not RIIA. PageRank is uniquely characterized by the axioms of isomorphism, self edge, vote by committee, collapsing, and proxy. Any ranking system that satisfies these five axioms coincides with page rank. In practice, Google implements a ranking system more complex than PageRank as presented in [4] which is likely not fully characterized by the above axioms. The current axiomatizations of page ranking systems do not address the following two ideas: that resistance to manipulation is an important property of a ranking system; and that approval voting by human users in the form of page visits or link clicks is a valuable source of preference declarations.

References

[1] A. Altman and M. Tennenholtz. Ranking systems: The pagerank axioms. *EC: Proceedings of the ACM Conference on Electronic Commerce*, pages 1–8, 2005.

[2] A. Altman and M. Tennenholtz. Axiomatic foundations for ranking systems. *JAIR: Journal of Artificial Intelligence Research*, 31:473–495, 2008.

[3] Y. Shoham and K. Leyton-Brown. *Mulitagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations.* 2009.

[4] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. *Technical Report. Stanford InfoLab.*, 1998.

[5] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30:107–117, 1998.

[6] Vincent Conitzer, Jérôme Lang, and Tuomas Sandholm. How many candidates are needed to make elections hard to manipulate? In *TARK '03: Proceedings of the 9th conference on Theoretical aspects of rationality and knowledge*, pages 201–214, New York, NY, USA, 2003. ACM.