Stochastic Stability

John R. Wicks

May 10, 2005

1 Introduction

Google is a well–known search engine that employs the PageRank algorithm [9] to rank the relative importance of webpages based on the link structure of the World–Wide Web. Intutively, rankings are based on a "random–surfer" model. Despite its tremendous success, Langville and Meyer [8] have pointed out some of its drawbacks. As the web continues to grow the computation of page ranks becomes more and more expensive, leading to less frequent recomputation (currently, on the order of weeks), which in turn leads to "stale" values being used in Web queries. In order to insure convergence, the algorithm introduces random perturbations, which corrupts the information inherent in the Web's link–structure. Anecdotally, it is also well–known that individuals may collude to increase their ranking by creating additional links to one another [1]. In addition, it tends to create multiple references to what is ultimately the same document because it gives excessive weight to "navigational" pages, such as the results of queries to other web search–engines, relative to the actual pages referenced. Kamvar, et. al. [4] have suggested exploiting the hierarchical structure of the web to make the computation of page ranks more efficient. However, since they are ultimately interested in computing the same PageRank values, their approach does not solve any of the other problems.

In this paper, we will analyze the PageRank algorithm to explain exactly how and why PageRank values may be manipulated. We will show how the "random–surfer" interpretation is related to a voting model. In particular, we will propose an alternative model of "weighted–voters" and suggest why this may be superior. We will also provide a theoretical justification for Kamvar, et. al.'s work as a "hierarchical voting system". By choosing appropriate "personalization" vectors and avoiding their final PageRank iteration, we can address the remaining deficiencies in the Brin and Page ranking system. This analysis will depend on a number of new techical results regarding (perturbed) Markov processes [2, 11].

Specifically, we will describe a quotient construction which may be used both to eliminate transient states or to collape any closed class down to a proper subset of itself. We will show that this construction is natural, in the sense of category theory, which will make it easy to compute in a number of important specific cases. We will then show how this construction relates both to the long–term, stable distribution of a Markov process and the stochastically stable distribution of a perturbed Markov process. We will conclude by demonstrating the connection between perturbed Markov processes, the PageRank alorithm, Kamav's BlockRank algorithm, and (hierarchical) voting systems.

2 Markov Processes and Quotients

We first begin with some basic definitions and facts regarding Markov processes. Let J denote a row vector of 1's. We say that an $n \times n$ matrix M is *Markov* iff $M \geq 0$ and $JM = J$. A distribution is a vector $v \geq 0$ s.t. $Jv = 1$. A discrete–time, finite–state Markov *process* may be uniquely characterized by a Markov matrix and an initial distribution vector, v_0. A *stable distribution* of M is one which is also an eigenvector with eigenvalue 1 (i.e., $Mv = v$). If $\dim M = n$ and σ is the standard n–simplex, the set of stable distributions of M, $\mathrm{Stab}\,(M) = \ker\,(M - I) \cap \sigma$. It is well known that if $M > 0$, $|\mathrm{Stab}\,(M)| = 1$ and that for any initial distribution, the associated Markov process converges to the unique stable distribution. In Section 4, we will show that this result generalizes, in that we may uniquely associate a stable distribution with any Markov process, and if that if M is *irreducible*, this distribution is independent of inital distribution.

Observe that a Markov matrix M may be viewed as a weighted, directed graph whose weights are in $[0, 1]$ such that the sum of the outgoing edges from any node is 1. We can then partition the vertices into *communicating classes*, corresponding to its strongly connected components (SCC). Alternatively, communicating classes are maximal sets of states for which there is a positive probability of eventually reaching any state from any other one. We will say that a subset of states is *invariant* under M iff the probability of ever leaving the set is 0. A *closed* class corresponds to a SCC which has no outgoing edges to any other SCC, that is, an invariant, communicating class. A non-closed class and its states are said to be *transient*. We will say that a Markov matrix is *regular* iff it has a single closed class.

We will define two Markov matrices, M_1 and M_2, as *equivalent* iff:

$$\ker\,(M_1 - I) = \ker\,(M_2 - I)\,.$$

It will turn out that two matrices are equivalent iff they have the same stable distributions. Notice that, given a Markov matrix, M, $M_\epsilon = (1 - \epsilon)\, I + \epsilon M$ gives a parameterized family of equivalent Markov matrices for small enough ϵ. In fact, since $M_\epsilon = I + \epsilon\,(M - I)$ it suffices to take $0 < \epsilon \leq \min_i (1 - M_{i,i})^{-1}$. Intuitively, M_ϵ represents the associated process where time has been "slowed down"; each step of the new process consists of deciding whether we will transition (with probability ϵ), and if so, choosing the transition according to the original process, M. Mathematically, this is just a convex combination of M and I; since the set of Markov matrices is convex, this is automatically Markov.

While $\ker(M_\epsilon - I) = \ker(M - I)$ for even larger values of ϵ, M_ϵ would not be positive (i.e., Markov), and for large enough values, will no longer have 1 as a dominant eigenvalue. In general, the eigenvectors of M and M_ϵ are identical, but an eigenvalue, λ, of M corresponds to an eigenvalue, $(1-\epsilon)\,1 + \epsilon\lambda$, of M_ϵ. Similarly, given any matrix, N, with positive off–diagonal entries such that $JN = 0$, $M_\epsilon = I + N\epsilon$ is Markov for sufficiently small ϵ.

We will need the following technical result from linear algebra:

Lemma 2.1 *Consider the submatrix* $\overline{M}$ *of a Markov matrix* M *on a set of states* $\mathcal{P}$*. If* $\mathcal{P}$ *does not contain a closed class of* M*, then* $I - \overline{M}$ *is invertible,* $\overline{M}^k \to 0$*, and* $\left(I - \overline{M}\right)^{-1} = \lim_{i\to\infty} \sum_{j=0}^{i-1} \overline{M}^j$*.*

Proof By a permutation similarity, we may assume that $\mathcal{P} = \{1, \ldots, m\}$. Since the column sums of $\overline{M}$ are less than or equal to 1, $\left|\overline{M}\right|_1 \le 1$. We can, in fact, show that $c = \left|\overline{M}^n\right|_1 < 1$. This is sufficient to complete the proof, as follows: $\left|\overline{M}^{k+1}\right|_1 \le \left|\overline{M}^k\right|_1 \left|\overline{M}\right|_1 \le \left|\overline{M}^k\right|_1$, so $\left|\overline{M}^k\right|_1$ is a decreasing sequence, containing the subsequence $\left|\overline{M}^{nk}\right|_1 \le \left(\left|\overline{M}^n\right|_1\right)^k = c^k$, which goes to 0 exponentially. Thus, $\overline{M}^k \to 0$ exponentially, $\lim_{i\to\infty} \sum_{j=0}^{i-1} \overline{M}^j$ exists, and, by the usual argument, equals $\left(I - \overline{M}\right)^{-1}$.

To show that $\left|\overline{M}^n\right|_1 < 1$, view M as a directed, weighted graph. If two states i and j are in the same communicating class, there is a path from i to j and vice versa. That is, there is a cycle in the directed graph containing i and j. If the class contains k vertices, there is such a cycle of length no more than k. In particular, there is a path of length no more than $k-1$ between any two vertices. Since each transient class contains a state with a non-zero probability of exiting the class, there is a non-zero probability of exiting the class in less than k steps. Notice that there cannot be any cycles *between* the classes (or all the classes visited in such a cycle would form a single communicating class). Thus, if the cardinalities of the transient classes are $k_1, \ldots, k_m$, then there is a non-zero probability of reaching a closed class in less than $\sum_{i=1}^m k_i + 1 \le n$ steps, and the probability of staying within the set of transient classes in a walk of length n is strictly less than 1. That is, the column sums of $\overline{M}^n$ are strictly less than 1. ■

We will show that given any set of states, $\mathcal{P}$, which does not contain an entire closed class of a Markov matrix, M, we may eliminate the transitions into $\mathcal{P}$, while changing the set of stable distributions in a predictable manner. If $M = \begin{pmatrix} \widetilde{M} & \overline{N} \\ \widetilde{N} & \underline{M} \end{pmatrix}$ is partitioned via the states in $\mathcal{P}$, by column elimination on $M - I$, we may obtain a new process $M_{\widehat{\mathcal{P}}} = \begin{pmatrix} \widehat{M} & \overline{N} \\ 0 & \underline{M} \end{pmatrix}$ such that $Stab\,(M)$ is in 1-1 correspondence with $Stab\left(M_{\widehat{\mathcal{P}}}\right)$. Now column elimination is equivalent to a factorization of the form $(M-I)\begin{pmatrix} I & 0 \\ A & I \end{pmatrix} = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$, or $\begin{pmatrix} \widetilde{M}-I & \overline{N} \\ \widetilde{N} & \overline{M}-I \end{pmatrix}\begin{pmatrix} I & 0 \\ A & I \end{pmatrix} = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$. By Lemma 2.1, $I-\overline{M}$ is invertible, so that $C = \overline{N}$, $D = \overline{M} - I$, $A = \left(I-\overline{M}\right)^{-1}\widetilde{N}$, and $B = \widetilde{M} + \overline{N}\left(I-\overline{M}\right)^{-1}\widetilde{N} - I = \widehat{M} - I$. Notice that this implies that:

$$\begin{aligned} \ker(M-I) &= \begin{pmatrix} I & 0 \\ \left(I-\overline{M}\right)^{-1}\widetilde{N} & I \end{pmatrix} \ker \begin{pmatrix} \widehat{M}-I & \overline{N} \\ 0 & \overline{M}-I \end{pmatrix} \\ &= \begin{pmatrix} I \\ \left(I-\overline{M}\right)^{-1}\widetilde{N} \end{pmatrix} \ker\left(\widehat{M}-I\right). \end{aligned}$$

Since $\overline{N}\left(I-\overline{M}\right)^{-1}\widetilde{N} = \overline{N}\left(\sum_{j=0}^{\infty}\overline{M}^j\right)\widetilde{N}$, represents the probablility of going from one state to another, via an arbitrary number of steps through states in $\mathcal{P}$, $\left(\widehat{M}\right)_{i,j}$ represents the probability of going from state i to state j via a path whose interior states are only in $\mathcal{P}$ (including the trivial case of single edge). Thus, we may view this construction as forcing transitions between the states in $\mathcal{P}$ to occur “instantaneously”. In particular, this construction preserves all path connections between the remaining states, so that:

Theorem 2.2 *If M is regular, so is $\widehat{M}$.*

If we let:

$$p = \begin{pmatrix} I & \overline{N}\left(I-\overline{M}\right)^{-1} \end{pmatrix} \text{ and } i = \begin{pmatrix} I \\ \left(I-\overline{M}\right)^{-1}\widetilde{N} \end{pmatrix},$$

then i is a positive matrix which maps $\ker\left(\widehat{M}-I\right)$ to $\ker(M-I)$; in particular, the stable distributions of $\widehat{M}$ are mapped (up to normalization) to those of M. In addition, p is a mapping between simplices, i.e., is positive and has columns which sum to 1. It suffices to show that $J\overline{N}\left(I-\overline{M}\right)^{-1} = J$ (where, as usual, we ambiguously write J for a row of 1's of the appropriate dimension) or $J\overline{N} = J\left(I-\overline{M}\right) = JI - J\overline{M} = J - J\overline{M}$, or $J\overline{M} + J\overline{N} = J$. Since the columns sums of M are 1, this equality is clear. Since:

$$(M-I)\,i = \begin{pmatrix} \widetilde{M}-I & \overline{N} \\ \widetilde{N} & \overline{M}-I \end{pmatrix}\begin{pmatrix} I \\ \left(I-\overline{M}\right)^{-1}\widetilde{N} \end{pmatrix} = \begin{pmatrix} \widehat{M}-I \\ 0 \end{pmatrix},$$

we have:

$$p\,(M-I)\,i + I = \begin{pmatrix} I & \overline{N}\left(I-\overline{M}\right)^{-1} \end{pmatrix}\begin{pmatrix} \widehat{M}-I \\ 0 \end{pmatrix} = \left(\widehat{M}-I\right) + I = \widehat{M}.$$

That is, up to translations by the identity, the elimination construction may be accomplished by composition with i and p.

Theorem 2.3 *An irreducible Markov matrix, M, has a unique invariant distribution, v, which spans $ker(M-I)$. Moreover, $v>0$.*

Proof If $\dim M = 1$, then the result is immediate. Otherwise, eliminate all states in $\mathcal{P} = \{2, \ldots, \dim M - 1\}$. Since $\dim \widehat{M} = 1$, the unique column of $i = (w)$ spans $\ker(M-I)$. Since $i \geq 0$, normalizing w yields a stable distribution v which spans $\ker(M-I)$. We show that $v>0$ by contradiction. Assume without loss of generality that $v = \begin{pmatrix} 0 \\ \overline{v} \end{pmatrix}$, with $\overline{v} \neq 0$, and $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, so that $\begin{pmatrix} 0 \\ \overline{v} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 0 \\ \overline{v} \end{pmatrix}$. Since $0 \leq B, \overline{v}$, while $B\overline{v} = 0$ and $\overline{v} \neq 0$, we must have $B = 0$, contradicting irreducibility of M. ∎

Given a Markov matrix, M, if we let $\mathcal{P}$ be all states but one member of each closed class and compute the associated matrices p and i, we will refer to the operator $A \to \widehat{A} = p(A-I)i + I$ as the *quotient with respect to M*; this is a convex operator on Markov processes that preserves the identity matrix. Since there are no paths between closed classes, $\widehat{M} = I$, that is:

Theorem 2.4 *The quotient of M with respect to itself is I and i gives a 1-1 correspondence between $R^{\dim \ker(M-I)}$ and $\ker(M-I)$.*

Corollary 2.5 *If a Markov matrix is partitioned $M = \begin{pmatrix} \widetilde{M} & \overline{N} \\ \widetilde{N} & \overline{M} \end{pmatrix}$ according to a set of states $\mathcal{P}$ containing all but one member of each closed class, then $\overline{N}\left(I-\overline{M}\right)^{-1}\widetilde{N} = I - \widetilde{M}$ is a diagonal matrix. In particular, $p(M-I) = (M-I)i = 0$*

Proof Since $I = \widehat{M} = \widetilde{M} + \overline{N}\left(I-\overline{M}\right)^{-1}\widetilde{N}$, $\overline{N}\left(I-\overline{M}\right)^{-1}\widetilde{N} = I - \widetilde{M}$. Since $\overline{\mathcal{P}}$ contains exactly one member of each closed class, $\widetilde{M}$ is diagonal. ∎

More generally, given a set of states, $\mathcal{P}$, which does not contain a closed class, we define the *quotient of M with respect to $\mathcal{P}$* as the triple $\left(\widehat{M}, p, i\right)$. This construction is natural, in the sense that:

Theorem 2.6 *If $\mathcal{P} = \mathcal{P}_\infty \sqcup \mathcal{P}_\in$, (M_1, p_1, i_1) is the quotient of M with respect to $\mathcal{P}_1$, (M_2, p_2, i_2) is the quotient of M_1 with respect to $\mathcal{P}_2$, and $\left(\widehat{M}, p, i\right)$ is the quotient of M with respect to $\mathcal{P}$, then $p = p_2 p_1$ and $i = i_1 i_2$.*

Proof Let $T(\pi, \mathcal{P}) = \min\{i | \pi_i \in \overline{\mathcal{P}}\}$ be the first time the trajectory, π, leaves $\mathcal{P}$ (since $\mathcal{P}$ does not contain a closed class, this is well-defined) and let $S(\pi, \mathcal{P}) = \pi_{T(\pi,\mathcal{P})}$. Then $p_{i,j} = Pr\{S(\pi, \mathcal{P}) = i | \pi_0 = j\}$ and represents the probability that the first state in $\overline{\mathcal{P}}$ along a trajectory, π, from j is i. Since $\overline{\mathcal{P}} \subset \overline{\mathcal{P}}_1$, any trajectory from j to $\overline{\mathcal{P}}$ must enter $\overline{\mathcal{P}}_1$ at some point k. In particular, there is exactly one first such point and $p_{i,j} = \Sigma_k Pr\{S(\pi, \mathcal{P}) = i, S(\pi, \mathcal{P}_1) = k | \pi_0 = j\} = \Sigma_k (p_2)_{i,k} (p_1)_{k,j}$. Thus, $p = p_2 p_1$.

The inclusion mapping of the quotient operator does not have such a natural interpretation, so we must resort to direct, algebraic calculation. Partitioning the states by $\mathcal{P}$, via an appropriate permutation, we may assume that $M = \begin{pmatrix} \widetilde{M} & \overline{N} \\ \widetilde{N} & \overline{M} \end{pmatrix}$. Further partitioning by $\mathcal{P}_2$ and $\mathcal{P}_1$ gives $\widetilde{N} = \begin{pmatrix} \widetilde{N}_2 \\ \widetilde{N}_1 \end{pmatrix}$, $\overline{N} = \begin{pmatrix} \overline{N}_2 & \overline{N}_1 \end{pmatrix}$, and $\overline{M} = \begin{pmatrix} \widetilde{M}_1 & \overline{N}_{1,2} \\ \widetilde{N}_{1,2} & \overline{M}_1 \end{pmatrix}$. Then $i = \begin{pmatrix} I \\ \left(I - \overline{M}\right)^{-1} \widetilde{N} \end{pmatrix}$, while $i_1 = \begin{pmatrix} I & 0 \\ 0 & I \\ \left(I - \overline{M}_1\right)^{-1} \widetilde{N}_1 & \left(I - \overline{M}_1\right)^{-1} \widetilde{N}_{1,2} \end{pmatrix}$ and:

$$\begin{aligned} M_1 &= \begin{pmatrix} \widetilde{M} + \overline{N}_1 \left(I - \overline{M}_1\right)^{-1} \widetilde{N}_1 & \overline{N}_2 + \overline{N}_1 \left(I - \overline{M}_1\right)^{-1} \widetilde{N}_{1,2} \\ \widetilde{N}_2 + \overline{N}_{1,2} \left(I - \overline{M}_1\right)^{-1} \widetilde{N}_1 & \widetilde{M}_1 + \overline{N}_{1,2} \left(I - \overline{M}_1\right)^{-1} \widetilde{N}_{1,2} \end{pmatrix} \\ &\equiv \begin{pmatrix} M_{1,1} & M_{1,2} \\ M_{2,1} & M_{2,2} \end{pmatrix}. \end{aligned}$$

Therefore:

$$i_2 = \begin{pmatrix} I \\ \left(I - M_{2,2}\right)^{-1} M_{2,1} \end{pmatrix},$$

and:

$$i_1 i_2 = \begin{pmatrix} I \\ \left(I - M_{2,2}\right)^{-1} M_{2,1} \\ \left(I - \overline{M}_1\right)^{-1} \widetilde{N}_1 + \left(I - \overline{M}_1\right)^{-1} \widetilde{N}_{1,2} \left(I - M_{2,2}\right)^{-1} M_{2,1} \end{pmatrix}.$$

It then suffices to show that:

$$\left(I - \overline{M}\right)^{-1} \widetilde{N} = \begin{pmatrix} \left(I - M_{2,2}\right)^{-1} M_{2,1} \\ \left(I - \overline{M}_1\right)^{-1} \widetilde{N}_1 + \left(I - \overline{M}_1\right)^{-1} \widetilde{N}_{1,2} \left(I - M_{2,2}\right)^{-1} M_{2,1} \end{pmatrix}$$

or:

$$\begin{pmatrix} \widetilde{N}_2 \\ \widetilde{N}_1 \end{pmatrix} = \left(I - \overline{M}\right) \begin{pmatrix} \left(I - M_{2,2}\right)^{-1} M_{2,1} \\ \left(I - \overline{M}_1\right)^{-1} \widetilde{N}_1 + \left(I - \overline{M}_1\right)^{-1} \widetilde{N}_{1,2} \left(I - M_{2,2}\right)^{-1} M_{2,1} \end{pmatrix}$$

Since $\left(I - \overline{M}\right) = \begin{pmatrix} I - \widetilde{M}_1 & -\overline{N}_{1,2} \\ -\widetilde{N}_{1,2} & I - \overline{M}_1 \end{pmatrix}$, we may multiply out to see it suffices to show that:

$$\begin{aligned} \widetilde{N}_2 &= \left(I - \widetilde{M}_1\right) \left(I - M_{2,2}\right)^{-1} M_{2,1} \\ &\quad - \overline{N}_{1,2} \left(I - \overline{M}_1\right)^{-1} \left(\widetilde{N}_1 + \widetilde{N}_{1,2} \left(I - M_{2,2}\right)^{-1} M_{2,1}\right) \end{aligned} \tag{1}$$

and:

$$\widetilde{N}_1 = -\widetilde{N}_{1,2} \left(I - M_{2,2}\right)^{-1} M_{2,1} + \widetilde{N}_1 + \widetilde{N}_{1,2} \left(I - M_{2,2}\right)^{-1} M_{2,1} \tag{2}$$

Since the first and last terms on the right-hand side of equation (2) cancel, it is obviously true. Returning to equation (1), if we multiply out, it becomes:

$$\begin{aligned}\widetilde{N}_2 &= \left(I-\widetilde{M}_1\right)(I-M_{2,2})^{-1}M_{2,1} \\ &\quad -\overline{N}_{1,2}\left(I-\overline{M}_1\right)^{-1}\widetilde{N}_1-\overline{N}_{1,2}\left(I-\overline{M}_1\right)^{-1}\widetilde{N}_{1,2}(I-M_{2,2})^{-1}M_{2,1}\end{aligned}$$

or:

$$\begin{aligned}M_{2,1} &= \widetilde{N}_2+\overline{N}_{1,2}\left(I-\overline{M}_1\right)^{-1}\widetilde{N}_1 \\ &= \left[I-\widetilde{M}_1-\overline{N}_{1,2}\left(I-\overline{M}_1\right)^{-1}\widetilde{N}_{1,2}\right](I-M_{2,2})^{-1}M_{2,1}\end{aligned}$$

which is true, since:

$$I-\widetilde{M}_1-\overline{N}_{1,2}\left(I-\overline{M}_1\right)^{-1}\widetilde{N}_{1,2}=I-M_{2,2}$$

by definition of $M_{2,2}$ ■

Notice that a Markov matrix M restricts to an irreducible Markov matrix on each of its closed classes. Thus, by Theorem 2.3, we may refer to the stable distribution associate with each closed class. By padding with 0's, each such stable distribution of the restriction yields a stable distribution of M, s, such that $\text{supp}_s \equiv \{i|s_i>0\}$ is the corresponding closed class.

Fix a closed class and associated stable distribution, s, and let $\mathcal{P}$ be all but one member of the class. Assuming that there are no transient states, we may give explicit formulas for the quotient. We will assume, for convenience, that $\mathcal{P}=\{n-k+1,\ldots,n\}$, $\text{supp}_s=\{n-k,\ldots,n\}$, and $s=\begin{pmatrix}0\\1-t\\t\overline{s}\end{pmatrix}$. Thus, we may decompose $M=\begin{pmatrix}\widetilde{M} & \overline{N}\\ \widetilde{N} & \overline{M}\end{pmatrix}$ further as $M=\begin{pmatrix}\widetilde{M}_1 & 0 & 0\\ 0 & \tilde{m}_2 & \overline{n}\\ 0 & \tilde{n} & \overline{M}\end{pmatrix}$. Therefore, $\begin{pmatrix}0\\1-t\\t\overline{s}\end{pmatrix}=\begin{pmatrix}\widetilde{M}_1 & 0 & 0\\ 0 & \tilde{m}_2 & \overline{n}\\ 0 & \tilde{n} & \overline{M}\end{pmatrix}\begin{pmatrix}0\\1-t\\t\overline{s}\end{pmatrix}=\begin{pmatrix}0\\(1-t)\,\tilde{m}_2+t\overline{n}\,\overline{s}\\(1-t)\,\tilde{n}+t\overline{M}\,\overline{s}\end{pmatrix}$, $\left(I-\overline{M}\right)^{-1}\tilde{n}=\frac{t}{1-t}\overline{s}$, and $i=\begin{pmatrix}I & 0\\0 & 1\\0 & \frac{t}{1-t}\overline{s}\end{pmatrix}$.

Obviously, $\overline{N}\left(I-\overline{M}\right)^{-1}$ is 0 in all but the $(n-k)$th row, and since it has unit column sums, this must be a row of all 1's, i.e., $p=\begin{pmatrix}I & 0 & 0\\0 & 1 & J\end{pmatrix}$. Finally:

$$\widehat{M} = p\,(M-I)\,i+I$$

$$= \begin{pmatrix} I & 0 & 0 \\ 0 & 1 & J \end{pmatrix} \begin{pmatrix} \widetilde{M_1} - I & 0 & 0 \\ 0 & \tilde{m}_2 - 1 & \overline{n} \\ 0 & \tilde{n} & \overline{M} - I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & 1 \\ 0 & \frac{t}{1-t}\overline{s} \end{pmatrix} + I$$

$$= \begin{pmatrix} \widetilde{M_1} - I & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & 1 \\ 0 & \frac{t}{1-t}\overline{s} \end{pmatrix} + I$$

$$= \begin{pmatrix} \widetilde{M_1} - I & 0 \\ 0 & 0 \end{pmatrix} + I = \begin{pmatrix} \widetilde{M_1} & 0 \\ 0 & 1 \end{pmatrix}$$

That is, the quotient effectively collapes the closed class to a single state, represented by s.

Notice also that when $\mathcal{P}$ is all the transient states, and $M = \begin{pmatrix} \widetilde{M} & \overline{N} \\ \widetilde{N} & \overline{M} \end{pmatrix}$, $\widetilde{N} = 0$, $i = \begin{pmatrix} I \\ 0 \end{pmatrix}$, and $\widehat{M} = \widetilde{M}$. Taking these two observations together, we may characterize the quotient of M with respect to itself, as the composite of two smaller quotients. Namely, first eliminate all the transient states, $\mathcal{P}_1$, to obtain $\left(\widetilde{M}, \begin{pmatrix} I & \overline{N}\left(I - \overline{M}\right)^{-1} \end{pmatrix}, \begin{pmatrix} I \\ 0 \end{pmatrix}\right)$, then eliminate all but one state from each of the closed classes. Since $\widetilde{M} = \begin{pmatrix} M_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & M_k \end{pmatrix}$ is block-diagonal, with the blocks corresponding to each closed class, by our previous observation and naturality of the construction, this is given by:

$$\left(I, \begin{pmatrix} J & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & J \end{pmatrix}, \begin{pmatrix} \frac{s_1}{(s_1)_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{s_k}{(s_k)_1} \end{pmatrix} \right).$$

Taken together, we see that the quotient of M with respect to M is:

$$\left(I, \begin{pmatrix} J & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & J \end{pmatrix} \begin{pmatrix} I & \overline{N}\left(I - \overline{M}\right)^{-1} \end{pmatrix}, \begin{pmatrix} \frac{s_1}{(s_1)_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{s_k}{(s_k)_1} \\ 0 & 0 & 0 \end{pmatrix} \right).$$

We may now generalize Theorem 2.3:

Theorem 2.7 *The set of stable distributions of M is the convex closure of the stable distributions of its closed classes and the eigenspace $ker(M - I)$ is spanned by this set of stable distributions.*

Proof By Theorem 2.4, $\ker(M - I)$ is spanned by the columns of i, which we have just shown are (up to multiples) the stable distributions of its closed classes. ■

Note: While this result is well–known (cf. the proof of Theorem 2.1 on p. 4 of Karlin and Taylor [6]), the quotient construction provides a conceptually satisfying, constuctive proof.

Corollary 2.8 *M_1 and M_2 are equivalent iff they have the same set of stable distributions.*

Proof If M_1 and M_2 are equivalent, $\ker(M_1 - I) = \ker(M_2 - I)$. Therefore, $\mathrm{Stab}(M_1) = \ker(M_1 - I) \cap \sigma = \ker(M_2 - I) \cap \sigma = \mathrm{Stab}(M_2)$. Conversely, if $\mathrm{Stab}(M_1) = \mathrm{Stab}(M_2)$, $\ker(M_1 - I) = \mathrm{span}\ \mathrm{Stab}(M_1) = \mathrm{span}\ \mathrm{Stab}(M_2) = \ker(M_2 - I)$, so that M_1 and M_2 are equivalent. ■

3 Markov Processes and Determinants

Let M be Markov and consider the basis $\{\overline{e}_{p,q} \equiv e_{p,q} - e_{q,q} | p \neq q\}$, where $\{e_{p,q}\}$ is the standard basis for $n \times n$ matrices. Notice that $N = \sum_{p \neq q} M_{p,q} \overline{E}_{p,q} = M - I$, i.e., we have a canonical basis expansion for any Markov matrix. If M is regular, by Theorem 2.7, $\dim\ker(N) = 1$. Since $0 = \det(N) I = N\mathrm{ad}(N)$, we have that, after normalization, any non-zero column of $\mathrm{ad}(N)$ gives the unique stable distribution of M, v_∞. **Note**: Since $\mathrm{rank}(N) = n - 1$, there must be at least one such non-zero column. Likewise, since $0 = \det(N) I = \mathrm{ad}(N) N$, every row must be a multiple of J, i.e., all entries must be equal. In general, let w be the vector consisting of the diagonal entries of $\mathrm{ad}(N)$. If M is regular, this equals any of the columns of $\mathrm{ad}(N)$. In particular, $w \propto v_\infty$. Each entry w_p is then equal to the pth principal minor of N. If $R_p(N)$ is the result of replacing the pth column of N by the standard basis vector e_p, we have $w_p = |R_p(N)|$.

If M is regular with closed class, $\mathcal{P}$, then $w_p \neq 0 \Leftrightarrow p \in \mathcal{P}$. In particular, if M is irreducible, each entry of w is non-zero, which implies that *every* $(n-1) \times (n-1)$ submatrix of N is non-singular. To see this, as before, we write $M - I = \begin{pmatrix} \overline{M} - I & \widetilde{N} \\ 0 & \widetilde{M} - I \end{pmatrix}$. If $p \in \mathcal{P}$, $R_p(N) = \begin{pmatrix} R_p(\overline{M} - I) & \widetilde{N} \\ 0 & \widetilde{M} - I \end{pmatrix}$, and $w_p = |R_p(N)| = \left|R_p(\overline{M} - I)\right| \left|\widetilde{M} - I\right|$; otherwise, $w_p = \left|R_p\left(\widetilde{M} - I\right)\right| = \left|\overline{M} - I\right| \left|R_p\left(\widetilde{M} - I\right)\right| = 0$. In fact, we see that, up to a multiple, w is just the inclusion of the stable distribution of $\overline{M}$ into n dimensions. By the same argument, we see that if M has more than one closed class, $w = 0$. To summarize:

Theorem 3.1 *If $w_i = ad(M - I)_{i,i}$ for a Markov matrix,M, then $w = 0$ iff M has more than one closed class. Moreover, if M is regular, $supp_w$ is that class, and $span\{w\} = ker(M - I)$.*

Let $\{\overline{e}_{p,q} \equiv e_p - e_q | p \neq q\}$, where $\{e_p\}$ is the standard basis for n-dimensional column vectors. Using the wedge product, we may write:

$$w_k = \sum_{p \neq 1} M_{p,1} \overline{e}_{p,1} \wedge \sum_{p \neq 2} M_{p,2} \overline{e}_{p,2} \wedge \cdots \wedge e_k \wedge \cdots \wedge \sum_{p \neq n} M_{p,n} \overline{e}_{p,n}.$$

Using multi-linearity of the wedge product, we may express this as a sum over terms of the form $\left(\Pi_{p\neq k} M_{\sigma(p),p}\right) \overline{e}_{\sigma(1),1} \wedge \overline{e}_{\sigma(2),2} \wedge \cdots \wedge e_k \wedge \cdots \wedge \overline{e}_{\sigma(n),n}$, where $\sigma : \{1, \ldots, n\} - \{k\} \to \{1, \ldots, n\}$ is a mapping such that $\sigma(i) \neq i$. Such a mapping corresponds to a spanning subgraph with $n-1$ edges of the associated graph, such that k has out-degree 0. Strikingly, each term is either 0 or $(-1)^{n-1}$, depending on whether or not σ represents a spanning tree, rooted at k, i.e., the subgraph is connected.

That such terms are 0 or non-zero follows from our previous observation. Each term corresponds to a Markov process with only 0 and 1 entries and the subgraph associated with σ is the entire graph minus one edge from k to any other vertex. If the graph is not connected, there is a component that does not contain k, which contains at least one strongly-connected component $\mathcal{P}$. We make take the "missing" edge to go from k to any vertex in $\mathcal{P}$, so that k is transient. By our previous observation, this implies that $w_k = 0$. On the other hand, if the graph is connected, it is a tree rooted at k. Thus, if we take the "missing" edge to be a loop at k, $\{k\}$ is the unique closed class and all other states are transient. In particular, $w_k \neq 0$. In this case, however, we may be even more precise. By following the paths from k to every other vertex, we may use column elimination to eliminate the 1's in all columns other than the kth, leaving a diagonal matrix with -1 in all diagonal entries except the kth. Thus, the determinant is $(-1)^{n-1}$.

Thus, we have the following formula for the stable distribution of a regular Markov matrix.

Theorem 3.2 *If w is as in Theorem 3.1, then $w_k = (-1)^{n-1} \sum_\sigma \Pi_{p\neq k} M_{\sigma(p),p}$, where the sum is taken over all mappings $\sigma : \{1, \ldots, n\} - \{k\} \to \{1, \ldots, n\}$ which specify the parent relation of a spanning subtree of the graph of M with root k.*

Note: This sharpens a result proven by Freidlin and Wentzell [2] for irreducible Markov matrices.

Notice that this means that $(v_\infty)_p$ may be viewed as the conditional distribution that a randomly selected directed subtree is rooted at p, where the relative probability of each tree is given by product of the weights of its edges. This suggests that, if we could efficiently simulate this distribution on directed subgraphs, we could obtain a Monte Carlo algorithm for computing v_∞.

4 Markov Processes and Long-Term Distributions

Now, if $I_N = U(0, N-1)$ is a discrete uniform random variable on $\{0, \ldots, N-1\}$, we will show that:

Theorem 4.1 *As $N \to \infty$, $E(X_{I_N})$ converges to a stable distribution, v_∞, which depends only on the initial distribution of X_0. This can be thought of* *as the distribution of X_∞, and we will call it the* long-term *distribution of the process.*

First, notice that $E\left(X_{I_N}\right) = \frac{1}{N}\sum_{i=0}^{N-1} M^i v_0$. Since v_0 can be any vector, we must have convergence of the the sum $\frac{1}{N}\sum_{i=0}^{N-1} M^i$. Theorem 4.1 will follow from the following more detailed result:

Theorem 4.2 *If M is Markov, $M^\infty = lim_{N\to\infty}\frac{1}{N}\sum_{j=0}^{N-1} M^j$ is a well-defined Markov process which maps any initial distribution to its corresponding long-term distribution, i.e., $v^\infty = M^\infty v_0$. Moreover, $M^\infty = i(pi)^{-1}p$, where i and p are the quotient operators wrt M.*

As we will see in the proof, $(pi)^{-1}$ is a diagonal matrix whose entries are the sums of the columns of i, so that $\tilde{\imath} = i\,(pi)^{-1}$ is just a normalized version of i whose columns are the stable distributions of the closed classes of M, and $M^\infty = \tilde{\imath}p$. We observed earlier that $p = \begin{pmatrix} J & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & J \end{pmatrix} \begin{pmatrix} I & \overline{N}\left(I - \overline{M}\right)^{-1} \end{pmatrix}$. Thus, the passage from initial condition to equilibrium may be broken into two stages: use the quotient which eliminates transient states to determine how the initial distribution eventually maps into the closed classes, summing the results in each closed class to discover the components of each corresponding stable distribution; the inclusion $\tilde{\imath}$ then uniquely specifies the long-term distribution from these components.

Note: Since $M^N v_0$ also represents the probability that a path of length N terminates in a given state, we may also interpret v^∞ as the probability that a randomly chosen path ends in a given state.

Now observe how, once we show that M^∞ exists, it follows immediately that M^∞ is Markov and $v^\infty = M^\infty v_0$. Moreover:

Corollary 4.3 *If M is Markov, the image of the standard simplex under M^∞ yields the set of stable distributions of M. In particular, if M is regular, v^∞ is independent of the initial condition; that is, $M^\infty = v^\infty J$, since its image must be v^∞, the unique, stable distribution of M.*

Proof Notice that:

$$\begin{aligned} MM^\infty &= \lim_{k\to\infty} M\left(\frac{1}{k}\sum_{i=0}^{k-1} M^i\right) \\ &= \lim_{k\to\infty}\frac{k+1}{k}\left(\frac{1}{k+1}\sum_{i=0}^{k} M^i - \frac{I}{k+1}\right) \\ &= 1\left(\lim_{k\to\infty}\frac{1}{k}\sum_{i=0}^{k-1} M^i - 0\right) = M^\infty. \end{aligned}$$

This proves the columns are all stable distributions of M. Conversely, any stable distribution is an eigenvector of M^∞, and hence in its image. ■

For the proof of the Theorem 4.2, we will need the following technical Lemma.

Lemma 4.4 *If s_i is a sequence such that $\lim_{i\to\infty} s_i = 0$, then:*

$$\lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} s_i = 0.$$

Proof Given $\epsilon > 0$, take N so that $|s_i| < \frac{\epsilon}{2}$ for $i \geq N$. Let $s = \max_i |s_i|$, and we may assume, without loss of generality, that $s \geq \frac{\epsilon}{2}$ (otherwise, simply take a smaller N). Then $\left|\frac{1}{k}\sum_{i=0}^{k-1} s_i\right| \leq \frac{1}{k}\sum_{i=0}^{k-1}|s_i| \leq \frac{N}{k}s + \frac{(k-N)\epsilon}{2k} = \frac{\epsilon}{2} + \frac{N(s-\epsilon/2)}{k} < \epsilon$ for $k > \left\lceil N\left(\frac{2s}{\epsilon} - 1\right)\right\rceil$. Thus, $\lim_{k\to\infty}\frac{1}{k}\sum_{i=0}^{k-1} s_i = 0$. ■

Proof (of Theorem 4.2) We will prove that M^∞ is well-defined by case analysis. Consider first the case when M is irreducible. **Note**: This portion of the proof is adapted from Horn and Johnson, [3]. Let v^∞ be the unique, stable distribution of M and let $L = v^\infty J = \tilde{\imath}p$. We will show that $M^\infty = L$. First, notice that, $LL = v^\infty J v^\infty J = v^\infty J = L$ and $LM = v^\infty JM = v^\infty J = L$, so that $L(M-L) = 0$. In addition, if $(M-L)w = w$, then $L(M-L)w = Lw$; since $L(M-L) = 0$, $Lw = 0$, and $Mw = w$. By Theorem 2.3, we must then have $w = \alpha v^\infty$, so that $w = (M-L)w = (M-L)\alpha v^\infty = \alpha(Mv^\infty - Lv^\infty) = \alpha(v^\infty - v^\infty) = 0$. In other words, $\ker(I-(M-L)) = 0$ and $I-(M-L)$ is invertible. Now since LL, ML, and LM all equal L, $(M-L)^k = M^k - L$. Therefore, $\frac{1}{n}\sum_{j=0}^{n-1} M^j = L + \frac{1}{n}\sum_{j=0}^{n-1}\left(M^j - L\right) = L + \frac{1}{n}\sum_{j=0}^{n-1}(M-L)^j$, which may we simplify as $L + \frac{1}{n}\left(I-(M-L)^n\right)\left(I-(M-L)\right)^{-1}$. Since $|I-(M-L)^n| = |I - M^n + L| \leq |I| + |M|^n + |L| \leq 3$, the limit is L.

Now assume that M is reducible with no transient states, i.e., up to permutation similarity $M = \begin{pmatrix} M_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & M_k \end{pmatrix}$. Then, $M^\infty = \begin{pmatrix} M_1^\infty & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & M_k^\infty \end{pmatrix} = \begin{pmatrix} v_1^\infty J & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & v_k^\infty J \end{pmatrix} = \begin{pmatrix} v_1^\infty & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & v_k^\infty \end{pmatrix}\begin{pmatrix} J & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & J \end{pmatrix} = \tilde{\imath}p$, by the previous case. More generally, if M includes transient states, partition M in terms of its transient and non-transient states, $M = \begin{pmatrix} \widetilde{M} & \overline{N} \\ 0 & \overline{M} \end{pmatrix}$. We can then write $M^i = \begin{pmatrix} \widetilde{M}^i & \sum_{j=0}^{i-1}\widetilde{M}^{i-j-1}\overline{N}\overline{M}^j \\ 0 & \overline{M}^i \end{pmatrix}$ and:

$$M^\infty = \begin{pmatrix} \widetilde{M}^\infty & \lim_{k\to\infty}\frac{1}{k}\sum_{i=0}^{k-1}\sum_{j=0}^{i-1}\widetilde{M}^{i-j-1}\overline{N}\overline{M}^j \\ 0 & 0 \end{pmatrix}.$$

By the previous case, we know that $\widetilde{M}^{\infty}$ is well-defined, and since we know that $I - \overline{M}$ is invertible, we can simplify this further.

Lemma 4.5 $\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \sum_{j=0}^{i-1} \widetilde{M}^{i-j-1} \overline{N}\overline{M}^{j} = \widetilde{M}^{\infty} \overline{N} \left(I - \overline{M}\right)^{-1}$

Proof By change of variables in the summation $\frac{1}{k} \sum_{i=0}^{k-1} \sum_{j=0}^{i-1} \widetilde{M}^{i-j-1} \overline{N}\overline{M}^{j} = \frac{1}{k} \sum_{s=1}^{k-1} \widetilde{M}^{s-1} \overline{N} \sum_{j=0}^{s} \overline{M}^{j}$. Since $\left(I - \overline{M}\right)^{-1} = \lim_{i\to\infty} \sum_{j=0}^{i-1} \overline{M}^{j}$ and $\left|\widetilde{M}\right| \leq 1$, $\lim_{s\to\infty} \widetilde{M}^{s-1} \overline{N} \left(\sum_{j=0}^{s} \overline{M}^{j} - \left(I - \overline{M}\right)^{-1}\right) = 0$, so by Lemma 4.4 the limit of the previous expression equals the limit of $\frac{1}{k} \sum_{s=1}^{k-1} \widetilde{M}^{s-1} \overline{N} \left(I - \overline{M}\right)^{-1}$, which is just $\widetilde{M}^{\infty} \overline{N} \left(I - \overline{M}\right)^{-1}$. ■

Thus:

$$\begin{aligned} M^{\infty} &= \begin{pmatrix} \widetilde{M}^{\infty} & \widetilde{M}^{\infty} \overline{N} \left(I - \overline{M}\right)^{-1} \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \widetilde{M}^{\infty} \\ 0 \end{pmatrix} \begin{pmatrix} I & \overline{N} \left(I - \overline{M}\right)^{-1} \end{pmatrix} \\ &= \begin{pmatrix} v_1^{\infty} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & v_k^{\infty} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} J & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & J \end{pmatrix} \begin{pmatrix} I & \overline{N} \left(I - \overline{M}\right)^{-1} \end{pmatrix}, \end{aligned}$$

which we recognize as $\tilde{i}p$. ■

5 Stochastic Stability

A *perturbed Markov process* [11], M_ϵ, is a family of Markov processes for sufficiently small $\epsilon \geq 0$ and whose entries converge exponentially. That is, either $(M_\epsilon)_{i,j} = 0$ or we may write $(M_\epsilon)_{i,j} = \epsilon^{r_{i,j}} c_{i,j}$, for some integers $r_{i,j}$ and continuous, strictly positive functions of $\epsilon \geq 0$, $c_{i,j}$; by convention, if $(M_\epsilon)_{i,j} = 0$, we will set $r_{i,j} = \infty$ and $\epsilon^{\infty} = 0$. We will call $r_{i,j}$ the *degree* (or *resistance*) of the (i, j)-th edge.

If M_ϵ are regular for $\epsilon > 0$, and so have a unique stationary distribution v_ϵ, we say that this process is *regular.* **Note**: This generalizes the usual definition, which requires that M_ϵ be irreducible.

Lemma 5.1 *If M_ϵ is a regular perturbed Markov process, then v_ϵ is a continuous function of ϵ. In particular, we may define $v_0 \equiv \lim_{\epsilon\to 0} v_\epsilon$, the* stochastically stable distribution of *M_ϵ.*

Proof Since M_ϵ is regular, we may define w_ϵ as in Theorem 3.2, and we know that its kth entry is a sum of products of the form $\Pi_{p\neq k} M_{\sigma(p),p} = \Pi_{p\neq k} \epsilon^{r_{\sigma(p),p}} c_{\sigma(p),p}$, where σ is the parent relation of a spanning subtree rooted at

k. In particular, since $v_\epsilon = \frac{w_\epsilon}{\sum_p (w_\epsilon)_p}$, $(w_\epsilon)_p$ all have the same sign, and $w_\epsilon \neq 0$, v_ϵ is continuous for $\epsilon > 0$. Moreover, if we let $m = \min_k \min_\sigma \sum_{p \neq k} r_{\sigma(p),p}$, m is the smallest integer such that $\lim_{\epsilon\to 0} \epsilon^{-m} w_\epsilon$ exists and is non-zero. If we let $u_\epsilon \equiv (-1)^{n-1} \epsilon^{-m} w_\epsilon$, $v_\epsilon = \frac{u_\epsilon}{\sum_p (u_\epsilon)_p}$, and v_ϵ is continuous at 0 as well. ∎

We are interested in characterizing this stochastically stable distribution. We will say that two perturbed Markov processes are *equivalent* if they have the same stochastically stable distribution. Notice also that without any increase in generality, we may allow the $r_{i,j}$ to be rational, since by reparameterization, we may obtain an equivalent perturbed Markov process with integer exponents.

Lemma 5.2 *The stochastically stable distribution of a regular perturbed Markov process, M_ϵ, depends only on the leading coefficients, $c_{i,j}(0)$. In particular, it is equivalent to one in which all the off-diagonal $c_{i,j}$ are constant.*

Proof By the proof of Lemma 5.1, we know that the stochastically stable distribution $v_0 = \frac{u_0}{\sum_p u_{0p}}$. Moreover, for each of the terms in $(u_\epsilon)_p$ corresponding to a subtree rooted at p given by σ, if $\sum_{p\neq k} r_{\sigma(p),p} > m$, then the corresponding product goes to 0 in the limit and does not affect v_0. That is, if we restrict attention to those trees in $\mathcal{S}_k = \{\sigma | m = \sum_{p \neq k} r_{\sigma(p),p}\}$ (i.e., of minimal weight), the kth entry of u_0 is $\sum_{\sigma \in \mathcal{S}_\|} \Pi_{p\neq k} c_{\sigma(p),p}(0)$.

By the discussion preceeding Lemma 2.1, if necessary, we may replace M_ϵ by $\frac{I+M_\epsilon}{2}$ to guarantee that the diagonal entries of M_ϵ contain a non-zero constant term. We may then replace $c_{i,j}$ by $c_{i,j}(0)$ in all the off-diagonal entries, adjusting the diagonal accordingly, to obtain an equivalent process. ∎

Corollary 5.3 *A regular perturbed Markov process is equivalent to one of the form $M_\epsilon = M_0 + N_\epsilon \epsilon$, where N_ϵ is a polynomial in ϵ, such that $M_0 \neq I$ and $I + N_\epsilon$ are Markov with the off-diagonal, non-zero entries of M_0 and N_ϵ in distinct positions.*

Proof By Lemma 5.2, any perturbed Markov process is equivalent to M_ϵ whose off-diagonal entries are of the form $\epsilon^{r_{i,j}} c_{i,j}$ for some integers $r_{i,j}$ and positive constants $c_{i,j}$, such that $\sum_{\{i | r_{i,j}=0, i\neq j\}} c_{i,j} < 1$, for all j. The diagonal entries are then $1 - \sum_{i\neq j} \epsilon^{r_{i,j}} c_{i,j}$. Let $k = \min_{i\neq j} r_{i,j}$. Grouping entries by degree, $M_\epsilon = I + N\epsilon^k + N_\epsilon \epsilon^{k+1}$ with $N \neq 0$. Since $JM_\epsilon = J$, $JN = \lim_{\epsilon\to 0} J(M_\epsilon - I)\epsilon^{-k} = 0$. In addition, $JN_\epsilon = J\left(M_\epsilon - I - N\epsilon^k\right)\epsilon^{-k-1} = 0$. $I + N\epsilon, I + N_\epsilon \epsilon \geq 0$ for small enough ϵ. Since the off-diagonal entries of N and N_ϵ are all positive, by considering $\left(1 - \delta\epsilon^{-k}\right) I + \delta\epsilon^{-k} M_\epsilon$ for sufficiently small δ, we may assume without loss of generality that $k = 0$ and that $M_0 = I + N$ and $I + N_\epsilon$ are Markov. ∎

We now prove one last technical Lemma:

Lemma 5.4 *Let $M_\epsilon = M_0 + N_\epsilon \epsilon$ be a regular perturbed Markov process, as in Corollary 5.3. Let $\mathcal{P}_/$ be all states except one representative of each closed class* *of M_0. Then $\widehat{M}_\epsilon$, obtained by eliminating the states in $\mathcal{P}_\prime$ from M_ϵ, and $\widehat{\widehat{M}}_\epsilon$, the quotient of M_ϵ with respect to M_0 are equivalent. $\widehat{\widehat{M}}_\epsilon$ is a regular perturbed Markov process and the inclusion of the quotient maps the stochastically stable distribution of $\widehat{\widehat{M}}_\epsilon$ to (a multiple of) that of M_ϵ.*

Proof Partitioning M_ϵ according to the set $\mathcal{P}_\prime$ gives $M_\epsilon = \begin{pmatrix} \widetilde{M}_\epsilon & \overline{N}_\epsilon \\ \widetilde{N}_\epsilon & \overline{M}_\epsilon \end{pmatrix}$. Graph theoretically, M_0 contains all the 0-degree edges. Eliminating the states in $\mathcal{P}_\prime$ from M_ϵ yields $\widehat{M}_\epsilon = p_\epsilon \left(M_\epsilon - I\right) i_\epsilon + I = \widetilde{M}_\epsilon + \overline{N}_\epsilon \left(I - \overline{M}_\epsilon\right)^{-1} \widetilde{N}_\epsilon$ representing paths whose interior states are only in $\mathcal{P}_\prime$, where $p_\epsilon = \begin{pmatrix} I & \overline{N}_\epsilon \left(I - \overline{M}_\epsilon\right)^{-1} \end{pmatrix}$ and $i_\epsilon = \begin{pmatrix} I \\ \left(I - \overline{M}_\epsilon\right)^{-1} \widetilde{N}_\epsilon \end{pmatrix}$. Since the states in $\overline{\mathcal{P}_\prime}$ have no 0-degree paths between them, all the off-diagonal entries of $\widehat{M}_\epsilon$ are of higher degree.

On the other hand:

$$
\begin{aligned}
\widehat{\widehat{M}}_\epsilon &= p_0 \left(M_\epsilon - I\right) i_0 + I \\
&= \begin{pmatrix} I & \overline{N}_0 \left(I - \overline{M}_0\right)^{-1} \end{pmatrix} \begin{pmatrix} \widetilde{M}_\epsilon - I & \overline{N}_\epsilon \\ \widetilde{N}_\epsilon & \overline{M}_\epsilon - I \end{pmatrix} \begin{pmatrix} I \\ \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0 \end{pmatrix} + I \\
&= \begin{pmatrix} I & \overline{N}_0 \left(I - \overline{M}_0\right)^{-1} \end{pmatrix} \begin{pmatrix} \widetilde{M}_\epsilon - I + \overline{N}_\epsilon \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0 \\ \widetilde{N}_\epsilon + \left(\overline{M}_\epsilon - I\right) \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0 \end{pmatrix} + I \\
&= \widetilde{M}_\epsilon + \overline{N}_\epsilon \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0 + \overline{N}_0 \left(I - \overline{M}_0\right)^{-1} [\widetilde{N}_\epsilon + \left(\overline{M}_\epsilon - I\right) \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0] \\
&= \widetilde{M}_\epsilon + \overline{N}_\epsilon \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0 + \overline{N}_0 \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_\epsilon \\
&\quad + \overline{N}_0 \left(I - \overline{M}_0\right)^{-1} \left(\overline{M}_\epsilon - I\right) \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0
\end{aligned}
$$

Notice that the 0-degree component of $\widehat{\widehat{M}}_\epsilon$ is given by:

$$
\begin{aligned}
& \widetilde{M}_0 + \overline{N}_0 \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0 + \overline{N}_0 \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0 \\
& + \overline{N}_0 \left(I - \overline{M}_0\right)^{-1} \left(\overline{M}_0 - I\right) \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0 \\
= \; & \widetilde{M}_0 + \overline{N}_0 \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0
\end{aligned}
$$

By Theorem 2.4, this is I, i.e., all the off-diagonal entries of $\widehat{\widehat{M}}_\epsilon$ are of higher degree, just like $\widehat{M}_\epsilon$. Since $\overline{N}_0 \left(I - \overline{M}_0\right)^{-1} (-I) \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0$ only contributes to the 0-degree terms of $\widehat{\widehat{M}}_\epsilon$, the off-diagonal entries of $\widehat{\widehat{M}}_\epsilon$ are the same as the higher-degree terms of the off-diagonal entries of $\widetilde{M}_\epsilon + \overline{N}_\epsilon \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0 + \overline{N}_0 \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_\epsilon + \overline{N}_0 \left(I - \overline{M}_0\right)^{-1} \overline{M}_\epsilon \left(I - \overline{M}_0\right)^{-1} \widetilde{N}_0$. Each higher-degree, off-diagonal entry of $\widehat{\widehat{M}}_\epsilon$ corresponds to the edges in $\overline{\mathcal{P}_\prime}$ plus the paths with exactly one higher-degree edge. By comparison, each off-diagonal entry of $\widehat{M}_\epsilon$

corresponds to the edges in $\overline{\mathcal{P}_{\prime}}$ plus the paths with at least one higher-degree edge. In particular, the off-diagonal entries differ by higher-order terms. Thus, by Lemma 5.2, $\widehat{\widehat{M}}_\epsilon$ and $\widehat{M}_\epsilon$ are equivalent.

Since the entries of $\widehat{\widehat{M}}_\epsilon$ are linear combinations of the entries of M_ϵ, they converge exponentially. Since M_ϵ is regular, so is $\widehat{M}_\epsilon$. Since $\widehat{\widehat{M}}_\epsilon$ has the same paths as $\widehat{M}_\epsilon$, it is also regular. If v_ϵ is the stable distribution of M_ϵ and $\hat{v}_\epsilon$ is the stable distribution of $\widehat{M}_\epsilon$, we know that $v_\epsilon \sim i_\epsilon \hat{v}_\epsilon$, since i_ϵ preserves the set of stable distributions. In particular, $v_0 \sim i_0 \hat{v}_0 = i_0 \hat{\hat{v}}_0$. ■

It follows imediately that:

Corollary 5.5 *If M_ϵ is a regular perturbed Markov process, by inductively taking the quotient of M_ϵ with respect to its lowest-order terms, and i is the composition of the inclusions, then $v_0 \sim i$.*

Corollary 5.5 corresponds to the following algorithm for computing the stochastically stable distribution of a regular perturbed Markov process:

Given: A regular perturbed Markov matrix, M_ϵ.

Goal: Calculate the stochastically stable distribution v_0 of M_ϵ.

Step 1: Set $r \leftarrow I$, an identity matrix with the same dimensions as M_ϵ.

Step 2: Apply Corollary 5.3 to determine an equivalent process of the form $\overline{M}_\epsilon \equiv M_0 + \sum_{j=1}^{k} N_j \epsilon^j$ and set $M_\epsilon \leftarrow \overline{M}_\epsilon$.

Step 3: Determine a set of representatives from each closed class of M_0, $\overline{\mathcal{P}_{\prime}}$.

Step 4: Calculate the quotient (I, p, i) of M_0 with respect to $\mathcal{P}_{\prime}$.

Step 5: $M_\epsilon \leftarrow pM_\epsilon i + I - pi$ and $r \leftarrow ri$.

Step 6: If $\dim M_\epsilon = 1$, go to Step 7, otherwise go to Step 2.

Step 7: $v_0 \leftarrow \frac{1}{|r|_1} r$.

Using Corollary 5.5, we may also conveniently identify supp_{v_0}, the *stochastically stable states of M_ϵ*. If we define the set of *eventually transient* states as those states which become transient in one of the iterated quotients, then by construction:

Corollary 5.6 *If M_ϵ is a regular perturbed Markov process, the stochastically stable states of M_ϵ are precisely those states which are* not *eventually transient.*

The quotient construction provides insight into Young's proof [10] of a similar result. In the proof, he partitions the state space according to the closed class of the 0th–order terms of M_ϵ and constructs an associated directed graph on the subsets of the partition, effectively constructing the graph of $\widehat{\widehat{M}}_\epsilon$ of Lemma 5.4.

However, in so doing, he loses the underlying structure of a Markov process, because he only weights the edges of the resulting graph by the *resistance* (i.e., degree) of the corresponding entry of $\widehat{\widehat{M}}_\epsilon$. He is still able to characterize the stochastically stable states of the original system in terms of this directed graph, via the result of Freidlin and Wentzell [2], in terms of the set of all directed, spanning subtrees of this graph. However, because his construction does not produce the leading coefficients of this process, he cannot compute the exact stable distribution.

6 Application to Hierachical Voting Systems

All of our work to this point allows us to easily prove the following:

Corollary 6.1 *If M is Markov and $\overline{v} > 0$, then $M_\epsilon = (1-\epsilon)M + \epsilon\overline{v}J$ is a regular perturbed Markov process with stochastically stable distribution $M^\infty\overline{v}$.*

Proof Since $\overline{v} > 0$, $M_\epsilon = M + \epsilon(\overline{v}J - M)$ is irreducible, and hence a regular perturbed Markov process. By Theorem 4.2, $M^\infty\overline{v} = i(pi)^{-1}p\overline{v}$, where i and p are the quotient operators wrt M. By the proof of Lemma 5.4, we may consider the quotient of M_ϵ wrt M, $\widehat{\widehat{M}}_\epsilon = I + \epsilon\,(p\overline{v}Ji - pi)$, which is equivalent to $\widehat{M} = I + c^{-1}\,(p\overline{v}Ji - pi)$ for $c = \max_j (pi)_{j,j}$ (i.e., large enough to guarantee that $\widehat{M}$ is positive). This has $\tilde{v} = (pi)^{-1}p\overline{v}$ as an eigenvector with eigenvalue 1, since $\widehat{M}\tilde{v} = \tilde{v} + c^{-1}\,(p\overline{v}Ji\tilde{v} - pi\tilde{v}) = \tilde{v} + c^{-1}\,(p\overline{v}JM^\infty\overline{v} - pM^\infty\overline{v}) = \tilde{v} + c^{-1}(p\overline{v} - p\overline{v}) = \tilde{v}$. By Lemma 5.4, $i\tilde{v} = M^\infty\overline{v}$ is proportional to the stochastically stable distribution of M_ϵ. Since it is a distribution, it must equal the the stochastically stable distribution. ■

As in the Eigentrust algorithm [5], $(M)_{i,j}$ may be viewed as representing the degree to which page j's recommends page i. Unlike other algorithms which require irreducibility (i.e., independence of initial conditions), we view $\overline{v}$ as an initial allocation of proxy votes, representing each page's apriori reputation or "clout". $M^\infty\overline{v}$ then represents the ultimate distribution of votes, given that a recommendation corresponds to a reallocation of proxy votes, i.e., one page imparting some of its "clout" to another. Given this interpretation, Corollary 6.1 says that we may simulate such a voting system by a perturbed Markov process, where proxy votes may be transferred infinitely often.

Corollary 6.1 is closely related to the PageRank algorithm [9]. Let W be the total set of web pages, P_i be the set of pages to which page i links, $f_{i,j}$ be the number of links to page i from page j, and $s(j) = \sum_{i\in P_j} f_{i,j}$. The PageRank algorithm considers the stable distribution of the (irreducible) Markov process given by $M_{i,j} = 0.85\,\tilde{f}_{i,j} + 0.15\,\delta$, where $\tilde{f}_{i,j} = \begin{cases} \frac{f_{i,j}}{s(j)} & \text{if } s(j) \neq 0 \\ \delta_{i,j} & \text{otherwise} \end{cases}$, and $\delta = \frac{1}{\#W}$. Notice that this is of the form of Corollary 6.1 with $\epsilon = 0.15$, so we may view its ranking as an approximation to the results of a proxy voting.

This interpretation suggests how, in practice, the PageRank algorithm can be manipulated by web–designers. Specifically, by adding a sufficiently many pages to one's site that only link to one designated original page, it is possible to artificially increase the ranking of the designated page by as much as one wants. Intutively, this increases the number of proxy votes concentrated on that page after one time step, which results in artificially boosting its ranking.

For the type of perturbed process in Corollary 6.1, we can actually give a formula for v_ϵ when $\epsilon > 0$.

Theorem 6.2 *If $\epsilon > 0$, $v_\epsilon = E\left(M^I \overline{v}\right) = \epsilon\left[1-(1-\epsilon)M\right]^{-1}\overline{v}$ is the stable distribution of $M_{\epsilon,\overline{v}}$, where I is geometric with $E(I) = \frac{1}{\epsilon}$.*

Proof Since $M_{\epsilon,\overline{v}}$ is irreducible, $v_\epsilon = {M_\epsilon}^{\infty}\overline{v}$. For large N:

$$M_{\epsilon,\overline{v}}^{\infty}\overline{v} \approx \frac{1}{N}\sum_{u=0}^{N-1} {M_{\epsilon,\overline{v}}}^{u}\overline{v} = E\left(M_{\epsilon,\overline{v}}^{U}\overline{v}\right),$$

where U is uniformly distributed on the first N non–negative integers. However, we may view $M_{\epsilon,\overline{v}}$ as a process which usually applies M, but which may apply $\overline{v}J$ with probability ϵ. Thus, we may instead consider the process which generates a sequence of Bernoulli trials of length N with probability of "success", ϵ, and applies the product of the corresponding sequence of M and $\overline{v}J$ to $\overline{v}$. Notice that such a product equals $M^I\overline{v}$, where I is the number of M factors before the first $\overline{v}J$. Thus, $E\left(M_{\epsilon,\overline{v}}^{U}\overline{v}\right) = E\left(M^I\overline{v}\right)$, where I is distributed as a "truncated" geometric (i.e., all the distribution in the tail $I \geq N$ of the usual geometric distribution is concentrated at $I = N$). In particular, as $N \to \infty$, $M_{\epsilon,\overline{v}}^{\infty}\overline{v} = E\left(M^I\overline{v}\right)$, where I is a true geometric with $E(I) = \frac{1}{\epsilon}$. ∎

Theorem 6.2 suggests that we may interpret PageRank as the result of a proxy vote, where votes are discounted by 85% as they are transferred from one page to another. One could argue that this is more realistic that the model of Corollary 6.1, where votes are infinitely transferrable. To highlight the problems of the latter model, consider the effect of a page without out–links, a so–called "rank sink", in that model. A recommendation of such a page (i.e., a link) is ultimately a willingness to surrender all one's "clout" to that page. Moreover, a recommendation of a page which links to such a page is an *e*qually serious commitment. In fact, *a*ny chain of recommendations to such a page will result in 0 ranking for all pages in the chain, except the last. When one links to a page, it is because one believes that the information content of that page and its immediate links are valuable. But rarely would one give equal credence to every subsequent linked page.

Using Theorem 6.2, we may explicity compute the effect of adding k spurious pages, in the scheme described above. Specifically, if M represents the original web graph, adding k pages pointing to page 1 corresponds to replacing M by $\widetilde{M} = \begin{pmatrix} M & N \\ 0 & 0 \end{pmatrix}$, where $\tilde{N} = \begin{pmatrix} J \\ 0 \end{pmatrix}$ is a $n \times k$ matrix.

Theorem 6.3 *Let $\overline{v} > 0$ be the perturbation vector for M and $\hat{v} = \begin{pmatrix} \frac{n}{n+k}\overline{v} \\ \frac{k}{n+k}w \end{pmatrix}$ be the perturbation for $\widetilde{M}$ for some k–dimensional distribution, $w > 0$. If v_ϵ is the stable distribution of $M_{\epsilon,\overline{v}}$, e_ϵ is the stable distribution of M_{ϵ,e_1} (where e_1 is the first standard basis vector), and $\tilde{v}_\epsilon$ is the normalized component of the stable distribution of $\widetilde{M}_{\epsilon,\hat{v}}$ over the first n states, $\tilde{v}_\epsilon = \frac{n}{n+k\epsilon}v_\epsilon + \frac{k\epsilon}{n+k\epsilon}e_\epsilon$.*

Proof Notice that when $i > 0$:

$$\begin{aligned}
\widetilde{M}^i\hat{v} &= \begin{pmatrix} M^i & M^{i-1}N \\ 0 & 0 \end{pmatrix}\hat{v} = \begin{pmatrix} \frac{n}{n+k}M^i\overline{v} + \frac{k}{n+k}M^{i-1}Nw \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \frac{n}{n+k}M^i\overline{v} + \frac{k}{n+k}M^{i-1}e_1 \\ 0 \end{pmatrix}.
\end{aligned}$$

Notice that:

$$\begin{aligned}
E\left(\widetilde{M}^I\hat{v}\right) &= (1-\epsilon)\left[\hat{v} + \begin{pmatrix} \frac{n}{n+k}\sum_{i=1}^\infty M^i\overline{v}\epsilon^i + \frac{k}{n+k}\sum_{i=1}^\infty M^{i-1}e_1\epsilon^i \\ 0 \end{pmatrix}\right] \\
&= \frac{n}{n+k}\begin{pmatrix} E\left(M^I\hat{v}\right) \\ 0 \end{pmatrix} + \frac{k}{n+k}\begin{pmatrix} \sum_{i=1}^\infty M^{i-1}e_1\epsilon^i(1-\epsilon) \\ (1-\epsilon)w \end{pmatrix} \\
&= \frac{n}{n+k}\begin{pmatrix} v_\epsilon \\ 0 \end{pmatrix} + \frac{k}{n+k}\begin{pmatrix} \epsilon E\left(M^I e_1\right) \\ (1-\epsilon)w \end{pmatrix} \\
&= \frac{n}{n+k}\begin{pmatrix} v_\epsilon \\ 0 \end{pmatrix} + \frac{k}{n+k}\begin{pmatrix} \epsilon e_\epsilon \\ (1-\epsilon)w \end{pmatrix}.
\end{aligned}$$

Restricting attention to the original n states and normalizing, by Theorem 6.2, we obtain $\tilde{v}_\epsilon = \frac{n}{n+k\epsilon}v_\epsilon + \frac{k\epsilon}{n+k\epsilon}e_\epsilon$. ■

It now suffices to observe that page 1 has the highest ranking in e_ϵ, so that for large enough k, it will have the highest ranking in the modified web.

Lemma 6.4 *Using the notation of Theorem 6.3, $(e_\epsilon)_1 > (e_\epsilon)_k$ for $1 < k \leq n$, assuming that $\epsilon < 0.5$.*

Proof By Theorem 6.2, $e_\epsilon = E\left(M^I e_1\right) = (1-\epsilon)e_1 + \sum_{i=1}^\infty M^i e_1\epsilon^i(1-\epsilon)$. Since each entry of $M^i e_1$ is less than 1, $(e_\epsilon)_k \leq \sum_{i=1}^\infty \epsilon^i(1-\epsilon) = \epsilon$ for $k > 1$. In particular, $(e_\epsilon)_1 \geq 1-\epsilon > \epsilon \geq (e_\epsilon)_k$ for $1 < k \leq n$. ■

As mentioned in Section 1, using $\epsilon > 0$ is objectionable on esthetic grounds in that it corrupts the information inherent in the Web. From Theorem 6.3, we see that the PageRank algorithm this is also precisely the reason why it may be manipulated by web–designers. There are some serious objections, however, to letting $\epsilon \to 0$.

First, determining the stochastically stable distribution may be computationally too expensive. If a closed class is large, exact methods for determining its stable distribution may take too long, while iterative methods may converge too slowly. In addition, Theorem 5.6 implies that transient states will automatically have weight 0 in the stochastically stable distribution, i.e., be pruned from the ranking. Estimates of the Web, such as those made by Kumar, et. al. [7], suggest that at least 55% of the Web consists of transient states. Specifically, non–transient states can only reside in the regions of the Web which they refer to as OUT and TENDRILS. Thus, this may prune out too much of the Web. In addition, the property of being closed is not numerically stable with respect to the Web graph. Thus, a small perturbation of the Web graph (i.e., the introduction of one link out of a closed class) may cause a large change in the stochastically stable distribution (i.e., the ranking of all pages in the class will drop to 0).

One way to address both of these problems is to use a hierarchical voting system. Kamvar, et. al. [4] have attempted something along these lines to speed up the original PageRank algorithm. While they were able to demonstrate a marked improvement in computational efficiency, the resulting rankings are still vulnerable to manipulation.

The key idea is to recognize that the Web is partitioned into domains. Thus, we may distinguish between intra– and inter–domain links. Let $\mathcal{D}$ be the set of domains, $D(i)$ be the domain of page i, $f_{i,j}$ be the number of links to page i from page j, $s(j) = \sum_{i \text{ s.t. } D(i)=D(j)} f_{i,j}$, $\overline{s}(j) = \sum_{i \text{ s.t. } D(i)\neq D(j)} f_{i,j}$, $\tilde{\delta}_j = \frac{1}{\#D(j)}$, $\hat{\delta}_i = \frac{1}{\#D(i)\#\mathcal{D}}$, $\tilde{f}_{i,j} = \begin{cases} \frac{f_{i,j}}{s(j)} & \text{if } s(j) \neq 0 \\ \delta_{i,j} & \text{otherwise} \end{cases}$, $\hat{f}_{i,j} = \begin{cases} \frac{f_{i,j}}{\overline{s}(j)} & \text{if } \overline{s}(j) \neq 0 \\ 0 & \text{otherwise} \end{cases}$, and $\hat{\epsilon}_j = \begin{cases} \epsilon & \text{if } \overline{s}(j) \neq 0 \\ 0 & \text{otherwise} \end{cases}$. We may then refine the PageRank process as follows:

$$M_{i,j} = \begin{cases} (0.85 - 0.85\hat{\epsilon}_j - 0.15\epsilon)\,\tilde{f}_{i,j} + 0.15\tilde{\delta}_j + 0.15\epsilon\hat{\delta}_i & \text{if } D(i) = D(j) \\ 0.85\hat{\epsilon}_j\hat{f}_{i,j} + 0.15\epsilon\hat{\delta}_i & \text{if } D(i) \neq D(j) \end{cases}$$

In this process, a random–surfer at page j chooses to either follow an intra–domain link (with probability $0.85 - 0.85\hat{\epsilon}_j - 0.15\epsilon$), jump randomly within the domain (with probability 0.15), follow an inter–domain link (with probability $0.85\hat{\epsilon}$), or jump to a random page (with probability 0.15ϵ).

To understand this process, first ignore the ϵ terms. Notice that the resulting process reduces to the PageRank algorithm on each domain. This defines a local ranking of pages within each domain. **Note**: At this level, one may not care too much about the manipulability of PageRank, since a domain administrator should be able to have total control over such a local ranking. If we apply the quotient construction from Corollary 5.5 and normalize the result via Corollary 5.3, we are left with a process whose states now correspond to domains and whose edges correspond to the inter–domain links aggregated via the local rankings, plus a uniform perturbation. In other words, we have another vote among the domains to determine the "clout" of each domain. Corollary 5.5 then implies that the stochastically stable distribution of this process gives a weight to each page which is the product of its local weight times the weight of its domain.

In [4], Kamvar, et. al. use such a product as an initial approximation to the PageRank distribution. Our analysis provides a theoretical justification for using this product ranking *i*nstead of the usual PageRank. While link–spamming may be used to affect local rankings, it cannot significantly affect the ranking of domains. Thus, it can only have a limited effect on the global page rankings. While we have specified uniform vote distributions both locally and globally, we are free to specify any distribution we wish. In particular, by elliminating initial votes for navigational domains, such as "www.google.com", we can decrease their ranking.

References

[1] Nadav Eiron, Kevin S. McCurley, and John A. Tomlin. Ranking the web frontier. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 309–318. ACM Press, 2004.

[2] Mark Friedlin and Alexander Wentzell. *Random Perturbations of Dynamical Systems*. Springer–Verlag, Berlin, 1984.

[3] Roger Horn and Charles Johnson. *Matrix Analysis*. Cambridge University Press, New York, 1985.

[4] S. Kamvar, T. Haveliwala, C. Manning, and G. Golub. Exploiting the block structure of the web for computing pagerank. Technical report, Stanford Digital Library Technologies Project, 2003.

[5] S. Kamvar, M. Schlosser, and H. Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th International World Wide Web Conference (WWW)*, pages 640–651, 2003.

[6] Samuel Karlin and Howard M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, New York, 1981.

[7] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, D. Sivakumar, Andrew Tompkins, and Eli Upfal. The web as a graph. In *PODS '00: Proceedings of the nineteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 1–10. ACM Press, 2000.

[8] A. Langville and C. Meyer. A survey of eigenvector methods of web information retrieval. *SIAM Review (Forthcoming)*, 2004.

[9] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.

[10] H. Peyton Young. The evolution of conventions. *Econometrica*, 61:57–84, 1993.

[11] H. Peyton Young. *Individual strategy and social structure : an evolutionary theory of institutions*. Princeton University Press, Princeton, c1998.