MULTILINEAR PAGERANK[*]

DAVID F. GLEICH[†], LEK-HENG LIM[‡], AND YONGYANG YU[†]

Abstract. In this paper, we first extend the celebrated PageRank modification to a higher-order Markov chain. Although this system has attractive theoretical properties, it is computationally intractable for many interesting problems. We next study a computationally tractable approximation to the higher-order PageRank vector that involves a system of polynomial equations called multilinear PageRank. This is motivated by a novel "spacey random surfer" model, where the surfer remembers bits and pieces of history and is influenced by this information. The underlying stochastic process is an instance of a vertex-reinforced random walk. We develop convergence theory for a simple fixed-point method, a shifted fixed-point method, and a Newton iteration in a particular parameter regime. In marked contrast to the case of the PageRank vector of a Markov chain where the solution is always unique and easy to compute, there are parameter regimes of multilinear PageRank where solutions are not unique and simple algorithms do not converge. We provide a repository of these nonconvergent cases that we encountered through exhaustive enumeration and randomly sampling that we believe is useful for future study of the problem.

1. Introduction. Google devised PageRank to help determine the importance of nodes in a directed graph representing web pages (Page et al., 1999). Given a random walk on a directed graph, the PageRank modification builds a new Markov chain that always has a unique stationary distribution. This new random walk models a "random surfer" that with probability $\alpha < 1$ takes a step according to the Markov chain and with probability $1 - \alpha$ randomly jumps according to a fixed distribution. If $\boldsymbol{P}$ is a *column stochastic* matrix that represents the random walk on the original graph, then the PageRank vector $\mathbf{x}$ is unique and solves the linear system

$$\mathbf{x} = \alpha \boldsymbol{P}\mathbf{x} + (1 - \alpha)\mathbf{v},$$

where $\mathbf{v}$ is a stochastic vector and α is a probability (section 2.2 has a formal derivation). The simple Richardson iteration even converges fast for the values of α that are used in practice.

Although Google described PageRank for the web graph, the same methodology has been deployed in many applications where the importance of nodes provides insight into an underlying phenomena represented by a graph (Morrison et al., 2005; Freschi, 2007; Winter et al., 2012; Gleich, 2015). We find the widespread success of the PageRank methodology intriguing and believe that there are a few important features that contributed to PageRank's success. First and second are the uniqueness and fast

convergence. These properties enable reliable and efficient evaluation of the important nodes. Third, in most applications of PageRank, the input graph may contain modeling or sampling errors, and thus PageRank's jumps are a type of regularization. This may help capture important features in the graph despite the noise.

In this paper, we begin by developing the PageRank modification to a higher-order Markov chain (section 3). These higher-order Markov chains model stochastic processes that depend on more history than just the previous state. (We review them formally in section 2.3.) In a second-order chain, for instance, the choice of state at the next time step depends on the last two states. However, this structure corresponds to a first-order, or traditional, Markov chain on a tensor product state-space. We show that higher-order PageRank enjoys the same uniqueness and fast convergence as in the traditional PageRank problem (Theorem 3.4, Corollary 3.5), although computing these stationary distributions is prohibitively expensive in terms of memory requirements.

Recent work by Li and Ng (2013) provides an alternative: they consider a rank-1 approximation of these distributions. When we combine the PageRank modification of a higher-order Markov chain with the Li–Ng approximation, we arrive at the *multilinear PageRank problem* (section 4). For the specific case of an n-state second-order Markov chain, described by an $n \times n \times n$ transition probability table, the problem becomes finding the solution $\mathbf{x}$ of the polynomial system of equations,

$$\mathbf{x} = \alpha \boldsymbol{R}(\mathbf{x} \otimes \mathbf{x}) + (1 - \alpha)\mathbf{v},$$

where $\boldsymbol{R}$ is an $n \times n^2$ column stochastic matrix (that represents the probability table), α is a probability, $\otimes$ is the Kronecker product, and $\mathbf{v}$ is a probability distribution over the n-states encoded as an n-vector. We have written the equations in this way to emphasize the similarity to the standard PageRank equations.

One of the key contributions of our work is that the solution $\mathbf{x}$ has an interpretation as the stationary distribution of a process we describe and call the "spacey random surfer." The spacey random surfer continuously forgets its own immediate history but does remember the aggregate history and combines the current state with this aggregate history to determine the next state (section 4.1). This process provides a natural motivation for the multilinear PageRank vector in relationship to the PageRank random surfer. We build on recent advances in vertex reinforced random walks (Pemantle, 1992; Benaïm, 1997) in order to make this relationship precise.

There is no shortage of data analysis methods that involve tensors. These usually go by taking an m-way array as an order-m tensor and then performing a tensor decomposition. When $m = 2$, this is often the matrix SVD and the factors obtained give the directions of maximal variation. When $m > 2$, the solution factors lose this interpretation. Understanding the resulting decompositions may be problematic without an identifiability result such as Anandkumar et al. (2013).

Our proposal differs in that our tensor represents a probability table for a stochastic process and, instead of a decomposition, we seek an eigenvector that has a natural interpretation as a stationary distribution. In fact, a general nonnegative tensor can be regarded as a contingency table, which can be converted into a multidimensional probability table. These tables may be regarded as the probability distribution of a higher-order Markov chain, just like how a directed graph becomes a random walk. This interpretation has already been successful in a number of applied studies (Li, Ng, and Ye, 2011; Ng, Li, and Ye, 2011; Li, Ng, and Ye, 2014, 2012) with rigorous convergence analysis (Li and Ng, 2015) and even eigenvalue analysis (Chu and Wu, 2014).

For instance, Li and Ng (2015) employed theoretical tools similar to those we use to show convergence and uniqueness for a more complex model involving m stochastic tensors. Given the breadth of applications of tensors, our motivation was that the multilinear PageRank vector would be a unique, meaningful stationary distribution that we could compute quickly.

Multilinear PageRank solutions, however, are more complicated. They are not unique for any $\alpha < 1$ as was the case for PageRank, but only when $\alpha < 1/(m-1)$, where $m - 1$ is the order of the Markov chain (or m is the order of the underlying tensor) as shown in Theorem 4.3. We then consider five algorithms to solve the multilinear PageRank system: a fixed-point method, a shifted fixed-point method, a nonlinear inner-outer iteration, an inverse iteration, and a Newton iteration (section 5). These algorithms are all fast in the unique regime. Outside that range, we used exhaustive enumeration and random sampling to build a repository of problems that do not converge with our methods. Among the challenging test cases, the inner-outer algorithm and Newton's method have the most reliable convergence properties (section 6). Our codes are available for others to use and to reproduce the figures of this manuscript: https://github.com/dgleich/mlpagerank.

2. Background. The technical background for our paper includes a brief review of Li and Ng's factorization of the stationary distribution of a higher-order PageRank Markov chain, which we discuss after introducing our notation.

2.1. Notation. Matrices are bold uppercase Roman letters, as in $\boldsymbol{A}$; vectors are bold lowercase Roman letters, as in $\mathbf{x}$; and tensors are bold, underlined, uppercase Roman letters, as in $\underline{\boldsymbol{P}}$. We use $\mathbf{e}$ to be the vector of all ones. Individual elements such as A_{ij}, x_i, or $\underline{P}_{ijk}$ are always written without boldface. In some of our results, using subscripts is suboptimal, and we will use MATLAB indexing notation instead: $A(i,j)$, $x(i)$, or $\underline{P}(i,j,k)$. An order-m, n-dimensional tensor has m indices that range from 1 to n. We will use $\otimes$ to denote the Kronecker product. Throughout the paper, we call a nonnegative matrix $\boldsymbol{A}$ column-stochastic if $\sum_i A_{ij} = 1$. A stochastic tensor is a tensor that is nonnegative and where the sum over the first index i is 1. We caution our readers that what we call a "tensor" in this article really should be called a *hypermatrix*, that is, a specific coordinate representation of a tensor. See Lim (2013) for a discussion of the difference between a tensor and its coordinate representation.

We use $S_1, S_2, \ldots$ to denote a discrete time stochastic process on the state space $1, \ldots, n$. The probability of an event is denoted $\Pr(S_t = i)$ and $\Pr(S_t = i \mid S_{t-1} = j)$ is the conditional probability of the event. (For those experts in probability, we use this simplifying notation instead of the natural filtration given the history of the process.)

2.2. PageRank. In order to justify our forthcoming use of the term *higher-order PageRank*, we wish to precisely define a PageRank problem and PageRank vector. The following definition captures the discussions in Langville and Meyer (2006).

DEFINITION 2.1 (PageRank). *Let $\boldsymbol{P}$ be a column stochastic matrix, let α be a probability smaller than 1, and let $\mathbf{v}$ be a stochastic vector. A PageRank vector $\mathbf{x}$ is the* unique *solution of the linear system*

$$(2.1) \qquad \mathbf{x} = \alpha \boldsymbol{P} \mathbf{x} + (1 - \alpha)\mathbf{v}.$$

We call the set $(\alpha, \boldsymbol{P}, \mathbf{v})$ a PageRank problem.

Note that the PageRank vector $\mathbf{x}$ is equivalently a Perron vector of the matrix

$$\boldsymbol{M} = \alpha \boldsymbol{P} + (1 - \alpha)\mathbf{v}\mathbf{e}^T$$

under the normalization that $\mathbf{x} \geq 0$ and $\mathbf{e}^T\mathbf{x} = 1$. The matrix $\boldsymbol{M}$ is column stochastic and encodes the behavior of the random surfer that, with probability α, transitions according to the Markov chain with transition matrix $\boldsymbol{P}$ and, with probability $(1-\alpha)$, "teleports" according to the fixed distribution $\mathbf{v}$. When PageRank is used with a graph, then $\boldsymbol{P}$ is almost always defined as the random walk transition matrix for that graph. If a graph does not have a valid transition matrix, then there are a few adjustments available to create one (Boldi et al., 2007).

When we solve for $\mathbf{x}$ using the power method on the Markov matrix $\boldsymbol{M}$ or the Richardson iteration on the linear system (2.1), then we iterate

$$\mathbf{x}_0 = \mathbf{v}, \qquad \mathbf{x}_{k+1} = \alpha\boldsymbol{P}\mathbf{x}_k + (1-\alpha)\mathbf{v}.$$

This iteration satisfies the error bound

$$\|\mathbf{x}_k - \mathbf{x}\|_1 \leq 2\alpha^k$$

for any stochastic $\mathbf{x}_0$. For values of α between 0.5 and 0.99, which occur most often in practice, this simple iteration converges quickly.

2.3. Higher-order Markov chains. We wish to extend PageRank to higher-order Markov chains and so we briefly review their properties. An mth-order Markov chain S is a stochastic process that satisfies

$$\Pr(S_t = i_1 \mid S_{t-1} = i_2, \ldots, S_1 = i_t) = \Pr(S_t = i_1 \mid S_{t-1} = i_2, \ldots, S_{t-m} = i_{m+1}).$$

In words, this means that the future state only depends on the past m states. Although the probability structure of a higher-order Markov chain breaks the fundamental Markov assumption, any higher-order Markov chain can be reduced to a first-order, or standard, Markov chain by taking a Cartesian product of its state space. Consider, for example, a second-order n-state Markov chain S. Its transition probabilities are $\underline{P}_{ijk} = \Pr(S_{t+1} = i \mid S_t = j, S_{t-1} = k)$. We will represent these probabilities as a tensor $\underline{\boldsymbol{P}}$. The stationary distribution equation for the resulting first-order Markov chain satisfies

$$\sum_k \underline{P}_{ijk} X_{jk} = X_{ij},$$

where X_{jk} denotes the stationary probability on the product space. Here, we have induced an $n^2 \times n^2$ eigenvector problem to compute such a stationary distribution. For such first-order Markov chains, Perron–Frobenius theory (Perron, 1907; Frobenius, 1908; Varga, 1962) governs the conditions when the stationary distribution exists. However, in practice for a 100,000 × 100,000 × 100,000 tensor, we need to store 10,000,000,000 entries in $\boldsymbol{X} = [X_{ij}]$. This makes it infeasible to work with large, sparse problems.

2.4. Li and Ng's approximation. As a computationally tractable alternative to working with a first-order chain on the product state-space, Li and Ng (2013) define a new type of stationary distribution for a higher-order Markov chain. Again, we describe it for a second-order chain for simplicity. For each term X_{ij} in the stationary distribution they substitute a product $x_i x_j$, and thus for the matrix $\boldsymbol{X}$ they substitute a rank-1 approximation $\boldsymbol{X} = \mathbf{x}\mathbf{x}^T$, where $\sum_i x_i = 1$. Making this substitution and then summing over j yields an eigenvalue expression called an l^2-eigenvalue by Lim

(2005) and called a Z-eigenvalue by Qi (2005) (one particular type of tensor eigenvalue problem) for $\mathbf{x}$:

$$\sum_j \left(\sum_k \underline{P}_{ijk} x_j x_k \right) = \sum_j x_i x_j = x_i \quad \Leftrightarrow \quad \underline{P}\mathbf{x}^2 = \mathbf{x},$$

where we've used the increasingly common notational convention

$$[\underline{P}\mathbf{x}^2]_i = \sum_{jk} \underline{P}_{ijk} x_j x_k$$

from Qi (2005). All these results extend beyond second-order chains, in a relatively straightforward manner. Li and Ng present a series of theorems that govern existence and uniqueness for such stationary distributions that we revisit later.

3. Higher-order PageRank. Recall the PageRank random surfer. With probability α, the surfer transitions according to the Markov chain, and with probability $1 - \alpha$, the surfer transitions according to the fixed distribution $\mathbf{v}$. We define a higher-order PageRank by modeling a random surfer on a higher-order chain. With probability α, the surfer transitions according to the higher-order chain, and with probability $1 - \alpha$, the surfer teleports according to the distribution $\mathbf{v}$. That is, if $\underline{P}$ is the transition tensor of the higher-order Markov chain, then the higher-order PageRank chain has a transition tensor $\underline{M}$, where

$$\underline{M}(i, j, \ldots, \ell, k) = \alpha \underline{P}(i, j, \ldots, \ell, k) + (1 - \alpha)v_i.$$

Recall that any higher-order Markov chain can be reduced to a first-order chain by taking a Cartesian product of the state space. We call this the reduced form of a higher-order Markov chain, and in the following example we explore the reduced form of a second-order PageRank modification .

Example 3.1. Consider the following transition probabilities:

$$\underline{P}(\cdot, \cdot, 1) = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \\ 1 & \frac{1}{2} & 1 \end{bmatrix}; \quad \underline{P}(\cdot, \cdot, 2) = \begin{bmatrix} \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}; \quad \underline{P}(\cdot, \cdot, 3) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \end{bmatrix}.$$

Figure 1 shows the state-space transition diagram for the reduced form of the chain before and after its PageRank modification.

We define a higher-order PageRank *tensor* as the stationary distribution of the reduced Markov chain, organized so that $\underline{X}(i, j, \ldots, \ell)$ is the stationary probability associated with the sequence of states $\ell \to \cdots \to j \to i$.

DEFINITION 3.2 (higher-order PageRank). *Let $\underline{P}$ be an order-m transition tensor representing an $(m-1)$th order Markov chain, α be a probability less than 1, and $\mathbf{v}$ be a stochastic vector. Then the higher-order PageRank tensor $\underline{X}$ is the order-$(m-1)$, n-dimensional tensor that solves the linear system*

$$\underline{X}(i, j, \ldots, \ell) = \alpha \sum_k \underline{P}(i, j, \ldots, \ell, k)\underline{X}(j, \ldots, \ell, k) + (1 - \alpha)v_i \sum_k \underline{X}(j, \ldots, \ell, k).$$

For the second-order case from Example 3.1, we now write this linear system in a more traditional matrix form in order to make a few observations about its structure. Let $\boldsymbol{X}$ be the PageRank tensor (or matrix, in this case). We have

(3.1) $$\operatorname{vec}(\boldsymbol{X}) = [\alpha \boldsymbol{P} + (1 - \alpha)\boldsymbol{V}] \operatorname{vec}(\boldsymbol{X}),$$

where $\boldsymbol{P}, \boldsymbol{V} \in \mathbb{R}^{n^2 \times n^2}$, and $\boldsymbol{V} = \mathbf{e}^T \otimes \boldsymbol{I} \otimes \mathbf{v}$. In this setup, the matrix $\boldsymbol{P}$ is sparse and highly structured:

$$\boldsymbol{P} = \begin{bmatrix} 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

When $\alpha = 0.85$ and $\mathbf{v} = (1/3)\mathbf{e}$, the higher-order PageRank matrix is

$$\boldsymbol{X} = \begin{bmatrix} 0.0411 & 0.0236 & 0.0586 \\ 0.0062 & 0.0365 & 0.0397 \\ 0.0761 & 0.0223 & 0.6959 \end{bmatrix}.$$

More generally, both $\boldsymbol{P}$ and $\boldsymbol{V}$ have the following structure for the second-order case:

$$\begin{bmatrix} \underline{P}_{111}, v_1 & 0 & \cdots & 0 & \underline{P}_{112}, v_1 & 0 & \cdots & 0 & \cdots & \underline{P}_{11n}, v_1 & 0 & \cdots & 0 \\ \underline{P}_{211}, v_2 & 0 & \cdots & 0 & \underline{P}_{212}, v_2 & 0 & \cdots & 0 & \cdots & \underline{P}_{21n}, v_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \underline{P}_{n11}, v_n & 0 & \cdots & 0 & \underline{P}_{n12}, v_n & 0 & \cdots & 0 & \cdots & \underline{P}_{n1n}, v_n & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \underline{P}_{1n1}, v_1 & 0 & \cdots & 0 & \underline{P}_{1n2}, v_1 & \cdots & 0 & \cdots & 0 & \underline{P}_{1nn}, v_1 \\ 0 & \cdots & 0 & \underline{P}_{2n1}, v_2 & 0 & \cdots & 0 & \underline{P}_{2n2}, v_2 & \cdots & 0 & \cdots & 0 & \underline{P}_{2nn}, v_2 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \underline{P}_{nn1}, v_n & 0 & \cdots & 0 & \underline{P}_{nn2}, v_n & \cdots & 0 & \cdots & 0 & \underline{P}_{nnn}, v_n \end{bmatrix}.$$

In the remainder of this section, we wish to show the relationship between the reduced form of a higher-order PageRank chain and the definition of the PageRank problem (Definition 2.1). This is not as trivial as it may seem! For instance, in the second-order case, (3.1) is not of the correct form for Definition 2.1. But a slight bit of massaging produces the equivalence.

Consider the vectorized equation for the stationary distribution matrix for the second-order case (3.1) as

$$\text{vec}(\boldsymbol{X}) = \underbrace{[\alpha \boldsymbol{P} + (1-\alpha)\boldsymbol{V}]}_{\boldsymbol{M}} \text{vec}(\boldsymbol{X}).$$

Our goal is to derive a PageRank problem in the sense of Definition 2.1 to find $\text{vec}(\boldsymbol{X})$. As it turns out, $\boldsymbol{M}^2$ will give us this PageRank problem. The idea is this: in the first-order PageRank problem we lose all history after a single teleportation step by construction. In this second-order PageRank problem, we keep around one more state of history, hence, two steps of the second-order chain are required to see the effect of teleportation as in the standard PageRank problem. Formally, the matrix $\boldsymbol{M}^2$ can be written in terms of matrix $\boldsymbol{P}$ and $\boldsymbol{V}$, i.e.,

$$\boldsymbol{M}^2 = \alpha^2 \boldsymbol{P}^2 + \alpha(1-\alpha)\boldsymbol{PV} + \alpha(1-\alpha)\boldsymbol{VP} + (1-\alpha)^2 \boldsymbol{V}^2.$$

We now show that $\boldsymbol{V}^2 = (\mathbf{v} \otimes \mathbf{v})(\mathbf{e}^T \otimes \mathbf{e}^T)$ by exploiting two properties of the Kronecker product: $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD})$ and $\mathbf{a}^T \otimes \mathbf{b} = \mathbf{ba}^T$. Note that

$$\boldsymbol{V}^2 = (\mathbf{e}^T \otimes \boldsymbol{I} \otimes \mathbf{v})(\mathbf{e}^T \otimes \boldsymbol{I} \otimes \mathbf{v}) = [\mathbf{e}^T(\mathbf{e}^T \otimes \boldsymbol{I})] \otimes [(\boldsymbol{I} \otimes \mathbf{v})\mathbf{v}] = (\mathbf{e}^T \otimes \mathbf{e}^T) \otimes (\mathbf{v} \otimes \mathbf{v}).$$

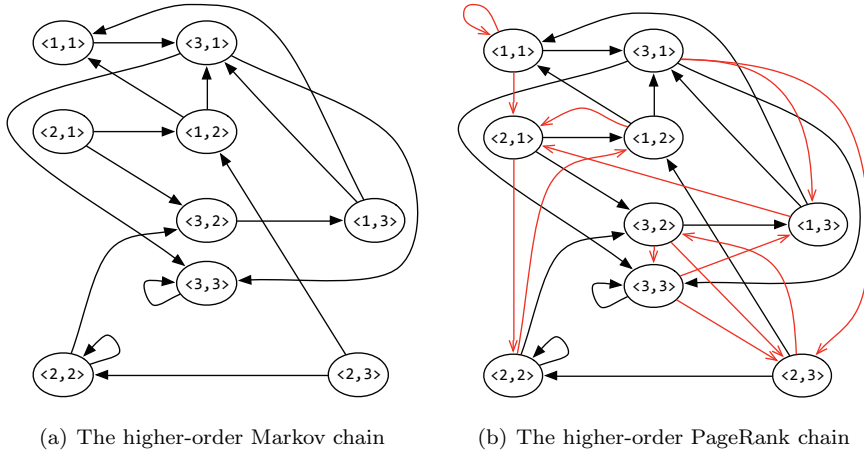(a) The higher-order Markov chain   (b) The higher-order PageRank chain

FIG. 1. *The state space transitions for a higher-order Markov chain on the product-space and the PageRank modification of that same chain with new transitions indicated in red. The transitions for both chains must satisfy $\langle j, k \rangle \to \langle i, j \rangle$. Note that, unlike the PageRank modification of a first-order Markov chain, the reduced form of the higher-order PageRank chain does not have a complete set of transitions. For instance, there is no transition between $\langle 2, 3 \rangle$ and $\langle 1, 3 \rangle$.*

This enables us to write a PageRank equation for $\text{vec}(\boldsymbol{X})$:

$$
\begin{aligned}
\text{vec}(\boldsymbol{X}) &= \boldsymbol{M}^2 \, \text{vec}(\boldsymbol{X}) \\
&= \alpha(2-\alpha) \underbrace{\left[\tfrac{\alpha}{2-\alpha} \boldsymbol{P}^2 + \tfrac{1-\alpha}{2-\alpha}(\boldsymbol{P}\boldsymbol{V} + \boldsymbol{V}\boldsymbol{P}) \right]}_{\boldsymbol{P}_{\text{pr}}} \text{vec}(\boldsymbol{X}) + (1 - 2\alpha + \alpha^2)\mathbf{v} \otimes \mathbf{v},
\end{aligned}
$$

where we used the normalization $\mathbf{e}^T \text{vec}(\boldsymbol{X}) = 1$. Thus we conclude as follows.

LEMMA 3.3. *Consider a second-order PageRank problem $\alpha, \underline{\boldsymbol{P}}, \mathbf{v}$. Let $\boldsymbol{P}$ be the matrix for the reduced form of $\underline{\boldsymbol{P}}$. Let $\boldsymbol{M} = \alpha \boldsymbol{P} + (1-\alpha)\boldsymbol{V}$ be the transition matrix for the vector representation of the stationary distribution $\boldsymbol{X}$. This stationary distribution is the PageRank vector of a PageRank problem $(2\alpha - \alpha^2, \boldsymbol{P}_{pr}, \mathbf{v} \otimes \mathbf{v})$ in the sense of Definition 2.1 with*

$$
\boldsymbol{P}_{pr} = \frac{\alpha}{2-\alpha} \boldsymbol{P}^2 + \frac{1-\alpha}{2-\alpha} \boldsymbol{P}\boldsymbol{V} + \frac{1-\alpha}{2-\alpha} \boldsymbol{V}\boldsymbol{P}.
$$

And we generalize as follows.

THEOREM 3.4. *Consider a higher-order PageRank problem $\alpha, \underline{\boldsymbol{P}}, \mathbf{v}$ where $\underline{\boldsymbol{P}}$ is an order-m tensor. Let $\boldsymbol{P}$ be the matrix for the reduced form of $\underline{\boldsymbol{P}}$. Let $\boldsymbol{M} = \alpha \boldsymbol{P} + (1-\alpha)\boldsymbol{V}$ be the transition matrix for the vector representation of the order-$(m-1)$, n-dimensional stationary distribution tensor $\underline{\boldsymbol{X}}$. This stationary distribution is equal to the PageRank vector of the PageRank problem*

$$
(1 - (1-\alpha)^{m-1}, \boldsymbol{P}_{pr}, \underbrace{\mathbf{v} \otimes \cdots \otimes \mathbf{v}}_{m-1 \; terms}), \; \text{where } \boldsymbol{P}_{pr} = \frac{\boldsymbol{M}^{m-1} - (1-\alpha)^{m-1}\boldsymbol{V}^{m-1}}{1 - (1-\alpha)^{m-1}}.
$$

Proof. We extend the previous proof as follows. The matrix $\boldsymbol{M}$ is nonnegative and has only a single recurrent class of all nodes consisting of all nodes in the *reach* of

the set of nonzero entries in v_i. Thus, the stationary distribution is unique. We need to look at the $m-1$ step transition matrix to find the PageRank problem. Consider $\underline{\boldsymbol{X}}$ as the stationary distribution eigenvector of the $m-1$ step chain:

$$\text{vec}(\underline{\boldsymbol{X}}) = \boldsymbol{M}\,\text{vec}(\underline{\boldsymbol{X}}) = \boldsymbol{M}^{m-1}\,\text{vec}(\underline{\boldsymbol{X}}).$$

The matrix $\boldsymbol{M}^{m-1}$ can be written in terms of matrix $\boldsymbol{P}$ and $\boldsymbol{V}$, i.e.,

$$\boldsymbol{M}^{m-1} = \left((\alpha\boldsymbol{P} + (1-\alpha)\boldsymbol{V})^{m-1} - (1-\alpha)^{m-1}\boldsymbol{V}^{m-1}\right) + (1-\alpha)^{m-1}\boldsymbol{V}^{m-1}.$$

The matrix $\boldsymbol{V}$ has the structure

$$\boldsymbol{V} = \mathbf{e}^T \otimes (\underbrace{\boldsymbol{I} \otimes \cdots \otimes \boldsymbol{I}}_{m-2 \text{ terms}}) \otimes \mathbf{v}.$$

We now expand $\boldsymbol{V}^{m-1}$ using the property of Kronecker products $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD})$, repeatedly:

$$\begin{aligned}
\boldsymbol{V}^{m-1} &= \left[\mathbf{e}^T \otimes (\underbrace{\boldsymbol{I} \otimes \cdots \otimes \boldsymbol{I}}_{m-2 \text{ terms}}) \otimes \mathbf{v}\right] \cdots \left[\mathbf{e}^T \otimes (\underbrace{\boldsymbol{I} \otimes \cdots \otimes \boldsymbol{I}}_{m-2 \text{ terms}}) \otimes \mathbf{v}\right] \\
&= \left[\mathbf{e}^T(\mathbf{e}^T \otimes \boldsymbol{I})(\mathbf{e}^T \otimes \boldsymbol{I} \otimes \boldsymbol{I}) \cdots (\mathbf{e}^T \otimes \underbrace{\boldsymbol{I} \otimes \cdots \otimes \boldsymbol{I}}_{m-2 \text{ terms}})\right] \\
&\qquad \otimes \left[(\underbrace{\boldsymbol{I} \otimes \cdots \otimes \boldsymbol{I}}_{m-2 \text{ terms}}\mathbf{v})(\boldsymbol{I} \otimes \boldsymbol{I} \otimes \mathbf{v})(\boldsymbol{I} \otimes \mathbf{v})\mathbf{v}\right] \\
&= (\underbrace{\mathbf{e}^T \otimes \cdots \otimes \mathbf{e}^T}_{m-1 \text{ terms}}) \otimes (\underbrace{\mathbf{v} \otimes \cdots \otimes \mathbf{v}}_{m-1 \text{ terms}}) \\
&= (\underbrace{\mathbf{v} \otimes \cdots \otimes \mathbf{v}}_{m-1 \text{ terms}})(\underbrace{\mathbf{e}^T \otimes \cdots \otimes \mathbf{e}^T}_{m-1 \text{ terms}}).
\end{aligned}$$

At this point, we are essentially done as we have shown that the *stochastic* $\boldsymbol{M}^{m-1}$ has the form $\boldsymbol{M}^{m-1} = \boldsymbol{Z} + (1-\alpha)^{m-1}(\mathbf{v} \otimes \cdots \otimes \mathbf{v})(\mathbf{e}^T \otimes \cdots \otimes \mathbf{e}^T)$. The statements in the theorem follow from splitting $\boldsymbol{M}^{m-1} = \alpha_{\text{pr}}\boldsymbol{P}_{\text{pr}} + (1-\alpha_{\text{pr}})(\mathbf{v} \otimes \cdots \otimes \mathbf{v})(\mathbf{e}^T \otimes \cdots \otimes \mathbf{e}^T)$, that is,

$$\alpha_{\text{pr}} = 1 - (1-\alpha)^{m-1}, \qquad \boldsymbol{P}_{\text{pr}} = \frac{1}{\alpha_{\text{pr}}}\left(\boldsymbol{M}^{m-1} - (1-\alpha)\boldsymbol{V}^{m-1}\right).$$

The matrix $\boldsymbol{P}_{\text{pr}}$ is stochastic because the final term in the expansion of $\boldsymbol{M}^{m-1}$ is $(1-\alpha)\boldsymbol{V}^{m-1}$; thus, the remainder is a nonnegative matrix with column sums equal to a single constant less than 1. $\quad\square$

COROLLARY 3.5. *The higher-order PageRank stationary distribution tensor* $\underline{\boldsymbol{X}}$ *always exists and is unique. Also, the standard PageRank iteration will result in a 1-norm error of* $2\left(1 - (1-\alpha)^{m-1}\right)^k$ *after* $(m-1)k$ *iterations.*

Hence, we retain all the attractive features of PageRank in the higher-order PageRank problem. The uniqueness and convergence results in this section are not overly surprising and simply clarify the relationship between the higher-order Markov chain and its PageRank modification.

4. Multilinear PageRank. The tensor product structure in the state-space and the higher-order stationary distribution make the straightforward approaches of the previous section difficult to scale to large problems, such as those encountered in modern bioinformatics and social network analysis applications. The scalability limit is the memory required. Consider an n-state, second-order PageRank chain: $(\alpha, \underline{P}, \mathbf{v})$. It requires $O(n^2)$ memory to represent the stationary distribution, which quickly grows infeasible as n scales. To overcome this scalability limitation, we consider the Li and Ng approximation to the stationary distribution with the following additional assumption.

ASSUMPTION. *There exists a method to compute $\underline{P}\mathbf{x}^2$ that works in time proportional to the memory used to represent $\underline{P}$.*

This assumption mirrors the fast matrix-vector product operator assumption in iterative methods for linear systems, although here it is critical because there must be at least n^2 nonzeros in any second-order stochastic tensor $\underline{P}$. If we could afford that storage, then the higher-order techniques from the previous section would apply and we would be under the scalability limit. We discuss how to create such fast operators from sparse datasets in section 4.5.

The Li and Ng approximation to the stationary distribution of a second-order Markov chain replaces the stationary distribution with a symmetric rank-1 factorization: $\boldsymbol{X} = \mathbf{x}\mathbf{x}^T$, where $\sum_i x_i = 1$. For a second-order PageRank chain, this transformation yields an implicit expression for $\mathbf{x}$:

$$(4.1) \qquad\qquad \mathbf{x} = \alpha\underline{P}\mathbf{x}^2 + (1-\alpha)\mathbf{v}.$$

We prefer to write this equation in terms of the Kronecker product structure of the tensor flattening along the first index. Let $\boldsymbol{R} := \underline{P}_{(1)} = \flat_1(\underline{P})$ be the n-by-n^2, stochastic flattening (see Golub and van Loan, 2013, section 12.4.5, for more on *flattenings* or *unfoldings* of a tensor, and see Draisma and Kuttler, 2014, for the $\flat$ notation):

$$
\boldsymbol{R} = \left[
\begin{array}{ccc|ccc|c|ccc}
\underline{P}_{111} & \cdots & \underline{P}_{1n1} & \underline{P}_{112} & \cdots & \underline{P}_{1n2} & \cdots & \underline{P}_{11n} & \cdots & \underline{P}_{1nn} \\
\underline{P}_{211} & \cdots & \underline{P}_{2n1} & \underline{P}_{212} & \cdots & \underline{P}_{2n2} & \cdots & \underline{P}_{21n} & \cdots & \underline{P}_{2nn} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\underline{P}_{n11} & \cdots & \underline{P}_{nn1} & \underline{P}_{n12} & \cdots & \underline{P}_{nn2} & \cdots & \underline{P}_{n1n} & \cdots & \underline{P}_{nnn}
\end{array}
\right].
$$

Then (4.1) is

$$\mathbf{x} = \alpha\boldsymbol{R}(\mathbf{x} \otimes \mathbf{x}) + (1-\alpha)\mathbf{v}.$$

Consider the tensor $\underline{P}$ from Example 3.1. The multilinear PageRank vector for this case with $\alpha = 0.85$ and $\mathbf{v} = (1/3)\mathbf{e}$ is

$$
\mathbf{x} = \begin{bmatrix} 0.1934 \\ 0.0761 \\ 0.7305 \end{bmatrix}.
$$

We generalize this second-order case to the order-m case in the following definition of the multilinear PageRank problem.

DEFINITION 4.1 (multilinear PageRank). *Let $\underline{P}$ be an order-m tensor representing an $(m-1)$th order Markov chain, α be a probability less than 1, and $\mathbf{v}$ be a stochastic vector. Then the multilinear PageRank vector is a nonnegative, stochastic solution of the following system of polynomial equations:*

$$(4.2) \quad \mathbf{x} = \alpha\underline{P}\mathbf{x}^{(m-1)} + (1-\alpha)\mathbf{v} \quad \textit{or equivalently} \quad \mathbf{x} = \alpha\boldsymbol{R}(\underbrace{\mathbf{x} \otimes \cdots \otimes \mathbf{x}}_{m-1\ terms}) + (1-\alpha)\mathbf{v},$$

where $\boldsymbol{R} := \underline{\boldsymbol{P}}_{(1)} = \flat_1(\underline{\boldsymbol{P}})$ *is an n-by-$n^{(m-1)}$ stochastic matrix of the flattened tensor along the first index.*

We chose the name *multilinear PageRank* instead of the alternative *tensor PageRank* to emphasize the multilinear structure in the system of polynomial equations rather than the tensor structure of $\underline{\boldsymbol{P}}$. Also, because the tensor structure of $\underline{\boldsymbol{P}}$ is shared with the higher-order PageRank vector, which could have then also been called a *tensor PageRank*.

A multilinear PageRank vector $\mathbf{x}$ always exists because it is a special case of the stationary distribution vector considered by Li and Ng. In order to apply their theory, we consider the equivalent problem,

$$(4.3) \qquad \mathbf{x} = (\alpha \boldsymbol{R} + (1-\alpha)\mathbf{v}\mathbf{e}^T)(\mathbf{x} \otimes \cdots \otimes \mathbf{x}) = \underline{\bar{\boldsymbol{P}}}\mathbf{x}^{m-1},$$

where $\underline{\bar{\boldsymbol{P}}}$ is the *stochastic* transition tensor whose flattening along the first index is the matrix $\alpha \boldsymbol{R} + (1-\alpha)\mathbf{v}\mathbf{e}^T$. The existence of a stochastic solution vector $\mathbf{x}$ is guaranteed by Brouwer's fixed-point theorem and, more immediately, by the stationary distributions considered by Li and Ng. The existing bulk of Perron–Frobenius theory for nonnegative tensors (Lim, 2005; Chang et al., 2008; Friedland et al., 2013), unfortunately, is not helpful with existence of uniqueness issues as it applies to problems where $\|\mathbf{x}\|_2 = 1$ instead of the 1-norm.

Although the multilinear PageRank vector always exists, it may not be unique, as shown in the following example.

Example 4.2. Let $\alpha = 0.99$, $\mathbf{v} = [0, 1, 0]^T$, and

$$\boldsymbol{R} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1/3 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1/3 & 0 & 0 \end{bmatrix}.$$

Then both $\mathbf{x} = [0, 1, 0]^T$ and $\mathbf{x} = [0.1890, 0.3663, 0.4447]^T$ solve the multilinear PageRank problem.

4.1. A stochastic process: The spacey random surfer. The PageRank vector is equivalently the stationary distribution of the random surfer stochastic process. The multilinear PageRank equation is the stationary distribution of a stochastic process with a history dependent behavior that we call the *spacey random surfer*. For simplicity, we describe this for the case of a second-order problem. Let S_t represent the stochastic process for the spacey random surfer. The process depends on the probability table for a second-order Markov chain $\underline{\boldsymbol{P}}$. Our motivation is that the spacey surfer would like to transition as the higher-order PageRank Markov chain, $\Pr(S_{t+1} = i \mid S_t = j, S_{t-1} = k) = \alpha \underline{P}_{ijk} + (1-\alpha)v_i$; however, on arriving at $S_t = j$, the surfer *spaces out* and *forgets* that $S_{t-1} = k$. Instead of using the true history state, the spacey random surfer decides to *guess* that they came from a state they've visited frequently. Let Y_t be a random state that the surfer visited in the past, chosen according to the frequency of visits to that state. (Hence, $Y_t = k$ is more likely if the surfer visited state k frequently in the past.) The spacey random surfer then transitions as

$$\Pr(S_{t+1} = i \mid S_t = j, Y_t = k) = \alpha \underline{P}_{ijk} + (1-\alpha)v_i.$$

Let us now state the resulting process slightly more formally. Let $\mathcal{F}_t$ be the natural filtration on the history of the process $S_1, \ldots, S_t$. Then

$$\Pr(Y_t = k \mid \mathcal{F}_t) = \frac{1}{t+n}\left(1 + \sum_{r=1}^{t} \mathrm{Ind}\{S_r = k\}\right),$$

where $\text{Ind}\{\cdot\}$ is the indicator event. In this definition, note that we assume that there is an initial probability of $1/n$ of Y_t taking any state. For instance, if $n = 10$ and $S_1 = 5, S_2 = 6, S_3 = 5$, and $t = 3$, then Y_t is a random variable that takes value 6 with probability $2/13$ and value 5 probability $3/13$. The stochastic process progresses as

$$\Pr(S_{t+1} = i \mid \mathcal{F}_t) = \alpha \sum_{k=1}^{n} \underline{P}(i, S_t, k) \frac{1 + \sum_{r=1}^{t} \text{Ind}\{S_r = k\}}{t + n} + (1 - \alpha)v_i.$$

This stochastic process is a new type of vertex reinforced random walk (Pemantle, 1992).

We present the following heuristic justification for the equivalence of this process with the multilinear PageRank vector. In our subsequent manuscript (Benson, Gleich, and Lim, 2015), we use results from Benaïm (1997) to make this equivalence precise and also to study the process in more depth. Suppose the process has run for a long time $t \gg 1$. Let $\mathbf{y}$ be the probability distribution of selecting any state as Y_t. The vector $\mathbf{y}$ changes slowly when t is large. For some time in the future, we can approximate the transitions as a first-order Markov chain:

$$\Pr(S_{t+1} = i \mid S_t = j) \approx \alpha \underline{P}_{i,j,k} y_k + (1 - \alpha)v_i.$$

Let $\boldsymbol{R}_k = \underline{P}(:,:,k)$ be a slice of the probability table; then the Markov transition matrix is

$$\alpha \sum_{k=1}^{n} \boldsymbol{R}_k y_k + (1 - \alpha)\mathbf{v}\mathbf{e}^T = \alpha \boldsymbol{R}(\mathbf{y} \otimes \boldsymbol{I}) + (1 - \alpha)\mathbf{v}\mathbf{e}^T = \boldsymbol{T}.$$

Given this column-stochastic matrix, the resulting stationary distribution is a stochastic vector $\mathbf{x}$ where $\boldsymbol{T}\mathbf{x} = \mathbf{x}$, that is,

$$\mathbf{x} = \alpha \boldsymbol{R}(\mathbf{y} \otimes \mathbf{x}) + (1 - \alpha)\mathbf{v}.$$

If $\mathbf{y} = \mathbf{x}$, then the distribution of $\mathbf{y}$ will not change in the future, whereas if $\mathbf{y} \neq \mathbf{x}$, then the distribution of $\mathbf{y}$ *must change* in the future. Hence, we must have $\mathbf{x} = \mathbf{y}$ at stationarity and any stationary distribution of the spacey random surfer must be a solution of the multilinear PageRank problem.

4.2. Sufficient conditions for uniqueness. In this section, we provide a sufficient condition for the multilinear PageRank vector to be unique. Our original conjecture was that this vector would be unique when $\alpha < 1$, which mirrors the case of the standard and higher-order PageRank vectors; however, we have already seen an example where this was false. Throughout this section, we shall derive and prove the following result.

THEOREM 4.3. *Let $\underline{P}$ be an order-m stochastic tensor and $\mathbf{v}$ be a nonnegative vector. Then the multilinear PageRank equation*

$$\mathbf{x} = \alpha \underline{P}\mathbf{x}^{(m-1)} + (1 - \alpha)\mathbf{v}$$

has a unique solution when $\alpha < \frac{1}{m-1}$.

To prove this statement, we first prove a useful lemma about the norm of the difference of the Kronecker products between two stochastic vectors with respect to the difference of each part. We suspect this result is known but were unable to find an existing reference.

LEMMA 4.4. *Let* $\mathbf{a}, \mathbf{b}, \mathbf{c},$ *and* $\mathbf{d}$ *be stochastic vectors where* $\mathbf{a}$ *and* $\mathbf{c}$ *have the same size. The* 1*-norm of the difference of their Kronecker products satisfies the following inequality:*

$$\|\mathbf{a} \otimes \mathbf{b} - \mathbf{c} \otimes \mathbf{d}\|_1 \leq \|\mathbf{a} - \mathbf{c}\|_1 + \|\mathbf{b} - \mathbf{d}\|_1.$$

Proof. This proof is purely algebraic and begins by observing

$$\mathbf{a} \otimes \mathbf{b} - \mathbf{c} \otimes \mathbf{d} = \frac{1}{2}(\mathbf{a} - \mathbf{c}) \otimes (\mathbf{b} + \mathbf{d}) + \frac{1}{2}(\mathbf{a} + \mathbf{c}) \otimes (\mathbf{b} - \mathbf{d}).$$

If we separate the bound into pieces we must bound terms such as $\|(\mathbf{a} - \mathbf{c}) \otimes (\mathbf{b} + \mathbf{d})\|_1$. But by using the stochastic property of the vectors, this term equals $\sum_{ij}(b_j + d_j)|a_i - c_i| = 2\|\mathbf{a} - \mathbf{c}\|_1$. ∎

This result is essentially tight. Let us consider two stochastic vectors of two dimensions, $\mathbf{x} = [x_1, 1 - x_1]^T$ and $\mathbf{y} = [y_1, 1 - y_1]^T$, where $x_1 \neq y_1$. Then,

$$\frac{\|\mathbf{x} \otimes \mathbf{x} - \mathbf{y} \otimes \mathbf{y}\|_1}{\|\mathbf{x} - \mathbf{y}\|_1} = \frac{1}{2}|x_1 + y_1| + |1 - (x_1 + y_1)| + \frac{1}{2}|2 - (x_1 + y_1)|.$$

The ratio of $\|\mathbf{x} \otimes \mathbf{x} - \mathbf{y} \otimes \mathbf{y}\|_1 / \|\mathbf{x} - \mathbf{y}\|_1$ approaches 2 as $x_1 + y_1 \to 0$ or $x_1 + y_1 \to 2$. However, this bound cannot be achieved.

The conclusion of Lemma 4.4 can be easily extended to the case where there are multiple Kronecker products between vectors.

LEMMA 4.5. *For stochastic vectors* $\mathbf{x}_1, \ldots, \mathbf{x}_m$ *and* $\mathbf{y}_1, \ldots, \mathbf{y}_m$ *where the size of* $\mathbf{x}_i$ *is the same as the size of* $\mathbf{y}_i$, *then*

$$\|\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_m - \mathbf{y}_1 \otimes \cdots \otimes \mathbf{y}_m\|_1 \leq \sum_i \|\mathbf{x}_i - \mathbf{y}_i\|_1.$$

Proof. Let us consider the case of $m = 3$. Let $\mathbf{a} = \mathbf{x}_1 \otimes \mathbf{x}_2$, $\mathbf{c} = \mathbf{y}_1 \otimes \mathbf{y}_2$, $\mathbf{b} = \mathbf{x}_3$, and $\mathbf{d} = \mathbf{y}_3$. Then

$$\|\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3 - \mathbf{y}_1 \otimes \mathbf{y}_2 \otimes \mathbf{y}_3\|_1 = \|\mathbf{a} \otimes \mathbf{b} - \mathbf{c} \otimes \mathbf{d}\|_1 \leq \|\mathbf{a} - \mathbf{c}\|_1 + \|\mathbf{x}_3 - \mathbf{y}_3\|_1$$

by using Lemma 4.4. But by recurring on $\mathbf{a} - \mathbf{c}$, we complete the proof for $m = 3$. It is straightforward to apply this argument inductively for $m > 3$. ∎

This result makes it easy to show uniqueness of the multilinear PageRank vectors.

LEMMA 4.6. *The multilinear PageRank equation has a unique solution when* $\alpha < 1/2$ *for third-order tensors.*

Proof. Assume there are two distinct solutions to the multilinear PageRank equation,

$$\mathbf{x} = \alpha \mathbf{R}(\mathbf{x} \otimes \mathbf{x}) + (1 - \alpha)\mathbf{v},$$
$$\mathbf{y} = \alpha \mathbf{R}(\mathbf{y} \otimes \mathbf{y}) + (1 - \alpha)\mathbf{v},$$
$$\mathbf{x} - \mathbf{y} = \alpha \mathbf{R}(\mathbf{x} \otimes \mathbf{x} - \mathbf{y} \otimes \mathbf{y}).$$

We simply apply Lemma 4.4,

$$\|\mathbf{x} - \mathbf{y}\|_1 = \|\alpha \mathbf{R}(\mathbf{x} \otimes \mathbf{x} - \mathbf{y} \otimes \mathbf{y})\|_1 \leq 2\alpha\|\mathbf{R}\|_1\|\mathbf{x} - \mathbf{y}\|_1 < \|\mathbf{x} - \mathbf{y}\|_1,$$

which is a contradiction (recall that $\mathbf{R}$ is stochastic). Thus, the multilinear PageRank equation has the unique solution when $\alpha < 1/2$. ∎

The proof of the general result in Theorem 4.3 is identical, except that it uses the general bound Lemma 4.5 on the order-m problem.

4.3. Uniqueness via Li and Ng's results. Li and Ng's recent paper (2013) tackled the same uniqueness question for the general problem:

$$\underline{P}\mathbf{x}^{m-1} = \mathbf{x}.$$

We can also write our problem in this form as in (4.3) and apply their theory. In the case of a third-order problem, or $m = 3$, they define a quantity to determine uniqueness:

$$\beta = \min_{S \subset \langle n \rangle} \left\{ \min_{k \in \langle n \rangle} \left(\min_{j \in S} \sum_{i \in \bar{S}} \underline{P}_{ijk} + \min_{j \in \bar{S}} \sum_{i \in S} \underline{P}_{ijk} \right) + \min_{j \in \langle n \rangle} \left(\min_{k \in S} \sum_{i \in \bar{S}} \underline{P}_{ijk} + \min_{k \in \bar{S}} \sum_{i \in S} \underline{P}_{ijk} \right) \right\}.$$

For any tensor where $\beta > 1$, the vector $\mathbf{x}$ that solves

$$\underline{P}\mathbf{x}^2 = \mathbf{x}$$

is unique. In Appendix A, we show that $\beta > 1$ is a stronger condition than $\alpha < 1/2$. We defer this derivation to the appendix as it is slightly tedious and does not result in any new insight into the problem.

4.4. PageRank and higher-order PageRank. We conclude this section by establishing some relationships between multilinear PageRank, higher-order PageRank, and PageRank for a special tensor. In the case when there is no higher-order structure present, then the multilinear PageRank, higher-order PageRank, and PageRank are all equivalent. The precise condition is where $\boldsymbol{R} = \mathbf{e}^T \otimes \boldsymbol{Q}$ for a stochastic matrix $\boldsymbol{Q}$, which models a higher-order random surfer with behavior that is *independent* of the last state. Thus, we'd expect that none of our higher-order modifications would change the properties of the stationary distribution.

PROPOSITION 4.7. *Consider a second-order multilinear PageRank problem with a third-order stochastic tensor where the flattened matrix $\boldsymbol{R} = \mathbf{e}^T \otimes \boldsymbol{Q}$ has dimension $n \times n^2$ and where $\boldsymbol{Q}$ is an $n \times n$ column stochastic matrix. Then for all $0 < \alpha < 1$ and stochastic vectors $\mathbf{v}$, the multilinear PageRank vector is the same as the PageRank vector of $(\alpha, \boldsymbol{Q}, \mathbf{v})$. Also, the marginal distribution of the higher-order PageRank solution matrix, $\boldsymbol{X}\mathbf{e}$, is the same as well.*

Proof. If $\boldsymbol{R} = \mathbf{e}^T \otimes \boldsymbol{Q}$, then any solution of (4.2) is also the unique solution of the standard PageRank equation:

$$\mathbf{x} = \alpha(\mathbf{e}^T \otimes \boldsymbol{Q})(\mathbf{x} \otimes \mathbf{x}) + (1 - \alpha)\mathbf{v} = \alpha\boldsymbol{Q}\mathbf{x} + (1 - \alpha)\mathbf{v}.$$

Thus, the two solutions must be the same and the multilinear PageRank problem has a unique solution as well. Now consider the solution of the second-order PageRank problem from (3.1):

$$\text{vec}(\boldsymbol{X}) = [\alpha\boldsymbol{P} + (1 - \alpha)\boldsymbol{V}]\,\text{vec}(\boldsymbol{X}).$$

Note that $\boldsymbol{R} = (\mathbf{e}^T \otimes \boldsymbol{I})\boldsymbol{P}$. Consider the marginal distribution: $\mathbf{y} = \boldsymbol{X}\mathbf{e} = (\mathbf{e}^T \otimes \boldsymbol{I})\,\text{vec}(\boldsymbol{X})$. The vector $\mathbf{y}$ must satisfy

$$\mathbf{y} = (\mathbf{e}^T \otimes \boldsymbol{I})\,\text{vec}(\boldsymbol{X}) = (\mathbf{e}^T \otimes \boldsymbol{I})[\alpha\boldsymbol{P} + (1 - \alpha)\boldsymbol{V}]\,\text{vec}(\boldsymbol{X}) = \alpha\boldsymbol{R}\,\text{vec}(\boldsymbol{X}) + (1 - \alpha)\mathbf{v}.$$

But $\boldsymbol{R}\,\text{vec}(\boldsymbol{X}) = (\mathbf{e}^T \otimes \boldsymbol{Q})\,\text{vec}(\boldsymbol{X}) = \boldsymbol{Q}\mathbf{y}.$ $\quad\Box$

4.5. Fast operators from sparse data. The last detail we wish to mention is how to build a fast operator $\underline{P}\mathbf{x}^{m-1}$ when the input tensor is highly sparse. Let $\underline{Q}$ be the tensor that models the original *sparse* data, where $\underline{Q}$ has far fewer than n^{m-1} nonzeros and cannot be stochastic. Nevertheless, suppose that $\underline{Q}$ has the following property:

$$\underline{Q}(i, j, \ldots, k) \geq 0 \quad \text{and} \quad \sum_i \underline{Q}(i, j, \ldots, k) \leq 1 \text{ for all } j, \ldots, k.$$

This could easily be imposed on a set of nonnegative data in time and memory proportional to the nonzeros of $\underline{Q}$ if that were not originally true. To create a fast operator for a fully stochastic problem, we generalize the idea behind the *dangling indicator* correction of PageRank. (The following derivation is entirely self-contained, but the genesis of the idea is identical to the dangling correction in PageRank problems (Boldi et al., 2007). Let $\boldsymbol{S}$ be the flattening of $\underline{Q}$ along the first index. Let $\mathbf{d}^T = \mathbf{e}^T - \mathbf{e}^T \boldsymbol{S} \geq 0$, and let $\mathbf{u}$ be a stochastic vector that determines what the model should do on a *dangling case*. Then

$$\boldsymbol{R} = \boldsymbol{S} + \mathbf{u}\mathbf{d}^T$$

is a column stochastic matrix, which we interpret as the flattening of $\underline{P}$ along the first index. If $\mathbf{x}$ is a stochastic vector, then we can evaluate

$$\boldsymbol{R}\mathbf{x} = \underbrace{\boldsymbol{S}\mathbf{x}}_{\mathbf{z}} + \mathbf{u}(\mathbf{e}^T\mathbf{x} - \mathbf{e}^T\boldsymbol{S}\mathbf{x}) = \mathbf{z} + (1 - \mathbf{e}^T\mathbf{z})\mathbf{u},$$

which only involves work proportional to the nonzeros of $\boldsymbol{S}$ or the nonzeros of $\underline{Q}$. Thus, given any sparse tensor data, we can create a fully stochastic model.

5. Algorithms for multilinear PageRank. At this point, we begin our discussion of algorithms to compute the multilinear PageRank vector. In the following section, we investigate five different methods to compute it. The methods are all inspired by the fixed-point nature of the multilinear PageRank solution. They are
1. a fixed-point iteration, as in the power method and the Richardson method;
2. a shifted fixed-point iteration, as in SS-HOPM (Kolda and Mayo, 2011);
3. a nonlinear inner-outer iteration, akin to Gleich et al. (2010);
4. an inverse iteration, as in the inverse power method; and
5. a Newton iteration.
We will show that the first four of them converge in the case that $\alpha < 1/(m-1)$ for an order-m tensor. For Newton, we show it converges quadratically fast for a third-order tensor when $\alpha < 1/2$. We also illustrate a few empirical advantages of each method.

The test problems. Throughout the following section, the following two problems help illustrate the methods:

$$\boldsymbol{R}_1 = \left[\begin{array}{ccc|ccc|ccc} 1/3 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 1/3 & 0 & 1/2 & 1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 & 1/3 & 1 & 1/2 & 0 & 1 & 0 \end{array} \right],$$

$$\boldsymbol{R}_2 = \left[\begin{array}{cccc|cccc|cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1/2 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1/2 & 1 & 1/2 \end{array} \right]$$

with $\mathbf{v} = \mathbf{e}/n$. The parameter α will vary through our experiments, but we are most interested in the regime where $\alpha > 1/2$ to understand how the algorithms behave outside of the region where we can prove they converge. We derived these problems by using exhaustive and randomized searches over the space of $2 \times 2 \times 2$, $3 \times 3 \times 3$, and $4 \times 4 \times 4$ binary-valued tensors, which we then normalized to be stochastic. Problems $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ were made from the database of problems we consider from the next section (section 6).

The residual of a problem and a potential solution $\mathbf{x}$ is the 1-norm:

$$(5.1) \qquad \text{residual} = \|\alpha\underline{\boldsymbol{P}}\mathbf{x}^{m-1} + (1-\alpha)\mathbf{v} - \mathbf{x}\|_1.$$

We seek methods that cause the residual to drop below 10^{-8}. For all the examples in this section, we ran the method out to 10,000 iterations to ensure there was no delayed convergence, although we show only 1000 iterations.

5.1. The fixed-point iteration. The multilinear PageRank problem seeks a fixed-point of the following nonlinear map:

$$f(\mathbf{x}) = \alpha\underline{\boldsymbol{P}}\mathbf{x}^{m-1} + (1-\alpha)\mathbf{v}.$$

We first show convergence of the iteration implied by this map in the case that $\alpha < 1/(m-1)$.

THEOREM 5.1. *Let $\underline{\boldsymbol{P}}$ be an order-m stochastic tensor, let $\mathbf{v}$ and $\mathbf{x}_0$ be stochastic vectors, and let $\alpha < 1/(m-1)$. The fixed-point iteration*

$$\mathbf{x}_{k+1} = \alpha\underline{\boldsymbol{P}}\mathbf{x}_k^{m-1} + (1-\alpha)\mathbf{v}$$

will converge to the unique solution $\mathbf{x}$ of the multilinear PageRank problem (4.2) *and also*

$$\|\mathbf{x}_k - \mathbf{x}\|_1 \leq [\alpha(m-1)]^k\|\mathbf{x}_0 - \mathbf{x}\|_1 \leq 2[\alpha(m-1)]^k.$$

Proof. Note first that this problem has a unique solution $\mathbf{x}$ by Theorem 4.3 and also that $\mathbf{x}_k$ remains stochastic for all iterations. This result is then, essentially, an implication of Lemma 4.5. Let $\boldsymbol{R}$ be the flattening of $\underline{\boldsymbol{P}}$ along the first index. Then using that lemma,

$$\|\mathbf{y} - \mathbf{x}_{k+1}\|_1 = \|\alpha\boldsymbol{R}(\mathbf{y} \otimes \cdots \otimes \mathbf{y} - \mathbf{x}_k \otimes \cdots \otimes \mathbf{x}_k)\|_1 \leq \alpha(m-1)\|\mathbf{y} - \mathbf{x}_k\|_1.$$

Thus, we have established a contraction. $\square$

Li and Ng treat the same iteration in their paper and they show a more general convergence result that implies our theorem, thus providing a more refined understanding of the convergence of this iteration. However, their result needs a difficult-to-check criteria. In earlier work by Rabinovich, Sinclair, and Wigderson (1992), they show that the fixed-point iteration will always converge when a certain symmetry property holds; however, they do not have a rate of convergence. Nevertheless, it is still easy to find PageRank problems that will not converge with this iteration. Figure 2 shows the result of using this method on $\boldsymbol{R}_1$ with $\alpha = 0.95$ and $\alpha = 0.96$. The former converges nicely and the later does not.
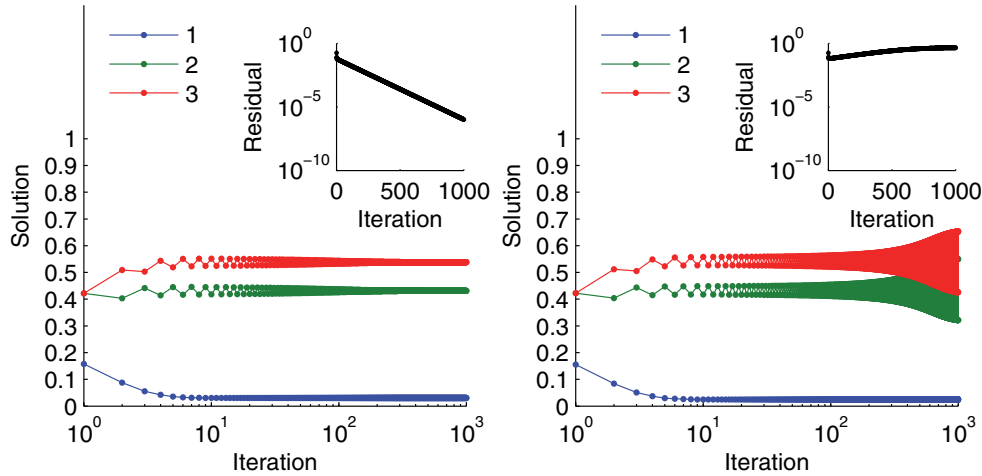
FIG. 2. *At left, the components of the iterates of the fixed-point iteration for $\boldsymbol{R}_1$ with $\alpha = 0.95$ show that it converges to a solution. (See the inset residual in the upper right.) At right, the result of the fixed-point iteration with $\alpha = 0.96$ illustrates a case that does not converge.*

5.2. The shifted fixed-point iteration. Kolda and Mayo (2011) noticed a similar phenomenon for the convergence of the symmetric higher-order power method and proposed the *shifted* symmetric higher-order power method (SS-HOPM) to address these types of oscillations. They were able to show that their iteration always converges monotonically for an appropriate shift value. For the multilinear PageRank problem, we study the iteration given by the equivalent fixed-point:

$$(1 + \gamma)\mathbf{x} = \left[\alpha \underline{\boldsymbol{P}} \mathbf{x}^{m-1} + (1 - \alpha)\mathbf{v} \right] + \gamma \mathbf{x}.$$

The resulting iteration is what we term the *shifted fixed-point iteration*

$$\mathbf{x}_{k+1} = \frac{\alpha}{1 + \gamma} \underline{\boldsymbol{P}} \mathbf{x}_k^{m-1} + \frac{1 - \alpha}{1 + \gamma} \mathbf{v} + \frac{\gamma}{1 + \gamma} \mathbf{x}_k.$$

It shares the property that an initial stochastic approximation $\mathbf{x}_0$ will remain stochastic throughout.

THEOREM 5.2. *Let $\underline{\boldsymbol{P}}$ be an order-m stochastic tensor, let $\mathbf{v}$ and $\mathbf{x}_0$ be stochastic vectors, and let $\alpha < 1/(m - 1)$. The shifted fixed-point iteration*

$$(5.2) \qquad \mathbf{x}_{k+1} = \frac{\alpha}{1 + \gamma} \underline{\boldsymbol{P}} \mathbf{x}_k^{m-1} + \frac{1 - \alpha}{1 + \gamma} \mathbf{v} + \frac{\gamma}{1 + \gamma} \mathbf{x}_k$$

will converge to the unique solution $\mathbf{x}$ of the multilinear PageRank problem (4.2) *and also*

$$\|\mathbf{x}_k - \mathbf{x}\|_1 \leq \left(\frac{\alpha(m - 1) + \gamma}{1 + \gamma} \right)^k \|\mathbf{x}_0 - \mathbf{x}\|_1 \leq 2 \left(\frac{\alpha(m - 1) + \gamma}{1 + \gamma} \right)^k.$$

The proof of this convergence is, in essence, identical to the previous case and we omit it for brevity.

This result also suggests that choosing $\gamma = 0$ is optimal and we should not shift the iteration at all. That is, we should run the fixed-point iteration. This analysis,
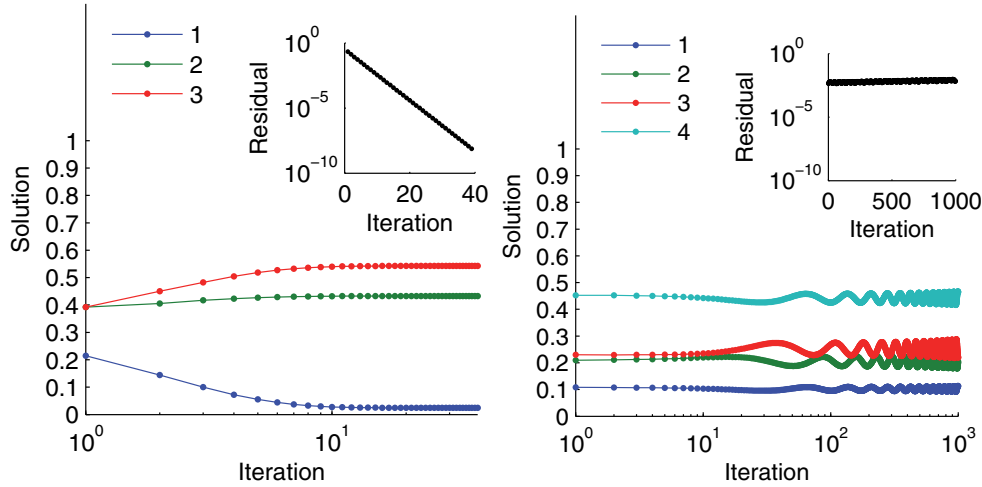
FIG. 3. *When we use a shift $\gamma = 1/2$, then at left, the iterates of the shifted iteration for $\boldsymbol{R}_1$ with $\alpha = 0.96$ shows that it quickly converges to a solution, whereas this same problem did not converge with the fixed-point method. At right, the result of the shifted iteration on $\boldsymbol{R}_2$ with $\alpha = 0.97$ again shows a case that does not converge.*

however, is misleading, as illustrated in Figure 3. There, we show the iterates from solving $\boldsymbol{R}_1$ with $\alpha = 0.96$, which did not converge with the fixed-point iteration but converges nicely with $\gamma = 1/2$. However, $\gamma < (m-2)/2$ will not guarantee convergence and the same figure shows that $\boldsymbol{R}_2$ with $\alpha = 0.97$ will not converge. We now present a necessary analysis that shows this method may not converge if $\gamma < (m-2)/2$ when $\alpha > 1/(m-1)$.

On the necessity of shifting. To derive this result, we shall restate the multilinear PageRank problem as the limit point of an ODE. There are other ways to derive this result as well, but this one is familiar and relatively straightforward. Consider the ODE

$$(5.3) \qquad \frac{d\mathbf{x}}{dt} = \alpha \underline{\boldsymbol{P}} \mathbf{x}^{m-1} + (1 - \alpha)\mathbf{v} - \mathbf{x}.$$

A forward Euler discretization yields the iteration

$$\mathbf{x}_{k+1} = \alpha h \underline{\boldsymbol{P}} \mathbf{x}_k^{m-1} + (1 - \alpha)h\mathbf{v} + (1 - h)\mathbf{x}_k,$$

which is identical to the shifted iteration (5.2) with $h = \frac{1}{1+\gamma}$. To determine if forward Euler converges, we need to study the Jacobian of the ODE. Let $\boldsymbol{R}$ be the flattening of $\underline{\boldsymbol{P}}$ along the first index; then the Jacobian of the ODE (5.3) is

$$\boldsymbol{J}(\mathbf{x}) = \alpha \boldsymbol{R}(\boldsymbol{I} \otimes \mathbf{x} \otimes \cdots \otimes \mathbf{x} + \mathbf{x} \otimes \boldsymbol{I} \otimes \mathbf{x} \otimes \cdots \otimes \mathbf{x} + \mathbf{x} \otimes \cdots \otimes \mathbf{x} \otimes \boldsymbol{I}) - \boldsymbol{I}.$$

A necessary condition for the forward Euler method to converge is that it is absolutely stable. In this case, we need $|1 - h\rho(\boldsymbol{J})| \leq 1$, where ρ is the spectral radius of the Jacobian. For all stochastic vectors $\mathbf{x}$ generated by iterations of the algorithm, $\rho(\boldsymbol{J}(\mathbf{x})) \leq (m-1)\alpha + 1 \leq m$. Thus, $h \leq 2/m$ is necessary for a general convergence result when $\alpha > 1/(m-1)$. This, in turn, implies that $\gamma \geq (m-2)/2$. In the case

that $\alpha < 1/(m-1)$, then the Jacobian already has eigenvalues within the required bounds and no shift is necessary.

Remark 5.3. Based on this analysis, we always recommend the shifted iteration with $\gamma \geq (m-2)/2$ for any problem with $\alpha > 1/(m-1)$.

5.3. An inner-outer iteration. We now develop a nonlinear iteration scheme that uses multilinear PageRank, in the convergent regime, as a subroutine. To derive this method, we use the relationship between multilinear PageRank and the multilinear Markov chain formulation discussed in section 2.4. Let $\bar{\boldsymbol{R}} = \alpha\boldsymbol{R} + (1-\alpha)\mathbf{v}\mathbf{e}^T$, then note that the Markov chain form of the problem is

$$\bar{\boldsymbol{R}}(\underbrace{\mathbf{x}\otimes\cdots\otimes\mathbf{x}}_{m-1 \text{ terms}}) = \mathbf{x} \quad\Leftrightarrow\quad [\alpha\boldsymbol{R} + (1-\alpha)\mathbf{v}\mathbf{e}^T](\underbrace{\mathbf{x}\otimes\cdots\otimes\mathbf{x}}_{m-1 \text{ terms}}) = \mathbf{x}.$$

Equivalently, we have

$$\frac{\alpha}{m-1}\bar{\boldsymbol{R}}(\underbrace{\mathbf{x}\otimes\cdots\otimes\mathbf{x}}_{m-1 \text{ terms}}) + \left(1 - \frac{\alpha}{m-1}\right)\mathbf{x} = \mathbf{x}.$$

From here, the nonlinear iteration emerges:

$$(5.4)\qquad \mathbf{x}_{k+1} = \frac{\alpha}{m-1}\bar{\boldsymbol{R}}(\underbrace{\mathbf{x}_{k+1}\otimes\cdots\otimes\mathbf{x}_{k+1}}_{m-1 \text{ terms}}) + \left(1 - \frac{\alpha}{m-1}\right)\mathbf{x}_k.$$

Each iteration involves solving a multilinear PageRank problem with $\bar{\boldsymbol{R}}, \alpha/(m-1)$, and $\mathbf{x}_k$. Because $\alpha < 1$, then $\alpha/(m-1) < 1/(m-1)$ and the solution of these subproblems is unique, and thus the method is well-defined. Not surprisingly, this method also converges when $\alpha < 1/(m-1)$.

THEOREM 5.4. *Let $\underline{\boldsymbol{P}}$ be an order-m stochastic tensor, let $\mathbf{v}$ and $\mathbf{x}_0$ be stochastic vectors, and let $\alpha < 1/(m-1)$. Let $\boldsymbol{R}$ be the flattening of $\underline{\boldsymbol{P}}$ along the first index and let $\bar{\boldsymbol{R}} = \alpha\boldsymbol{R} + (1-\alpha)\mathbf{v}\mathbf{e}^T$. The inner-outer multilinear PageRank iteration*

$$\mathbf{x}_{k+1} = \frac{\alpha}{m-1}\bar{\boldsymbol{R}}(\underbrace{\mathbf{x}_{k+1}\otimes\cdots\otimes\mathbf{x}_{k+1}}_{m-1 \ terms}) + \left(1 - \frac{\alpha}{m-1}\right)\mathbf{x}_k$$

converges to the unique solution $\mathbf{x}$ of the multilinear PageRank problem and also

$$\|\mathbf{x}_k - \mathbf{x}\|_1 \leq \left(\frac{1 - \alpha/(m-1)}{1 - \alpha^2}\right)^k\|\mathbf{x}_0 - \mathbf{x}\|_1 \leq 2\left(\frac{1 - \alpha/(m-1)}{1 - \alpha^2}\right)^k.$$

Proof. Recall that this is the regime of α when the solution is unique. Note that

$$\mathbf{x}_{k+1} - \mathbf{x} = \frac{\alpha}{m-1}\bar{\boldsymbol{R}}(\underbrace{\mathbf{x}_{k+1}\otimes\cdots\otimes\mathbf{x}_{k+1}}_{m-1 \text{ terms}} - \underbrace{\mathbf{x}\otimes\cdots\otimes\mathbf{x}}_{m-1 \text{ terms}}) + \left(1 - \frac{\alpha}{m-1}\right)(\mathbf{x}_k - \mathbf{x})$$

$$= \frac{\alpha^2}{m-1}\boldsymbol{R}(\underbrace{\mathbf{x}_{k+1}\otimes\cdots\otimes\mathbf{x}_{k+1}}_{m-1 \text{ terms}} - \underbrace{\mathbf{x}\otimes\cdots\otimes\mathbf{x}}_{m-1 \text{ terms}}) + \left(1 - \frac{\alpha}{m-1}\right)(\mathbf{x}_k - \mathbf{x}).$$

By using Lemma 4.5, we can bound the norm of the difference of the $m-1$ term Kronecker products by $(m-1)\|\mathbf{x}_{k+1} - \mathbf{x}\|_1$. Thus,

$$\|\mathbf{x}_{k+1} - \mathbf{x}\|_1 \leq \alpha^2\|\mathbf{x}_{k+1} - \mathbf{x}\|_1 + \left(1 - \frac{\alpha}{m-1}\right)\|\mathbf{x}_k - \mathbf{x}\|_1,$$
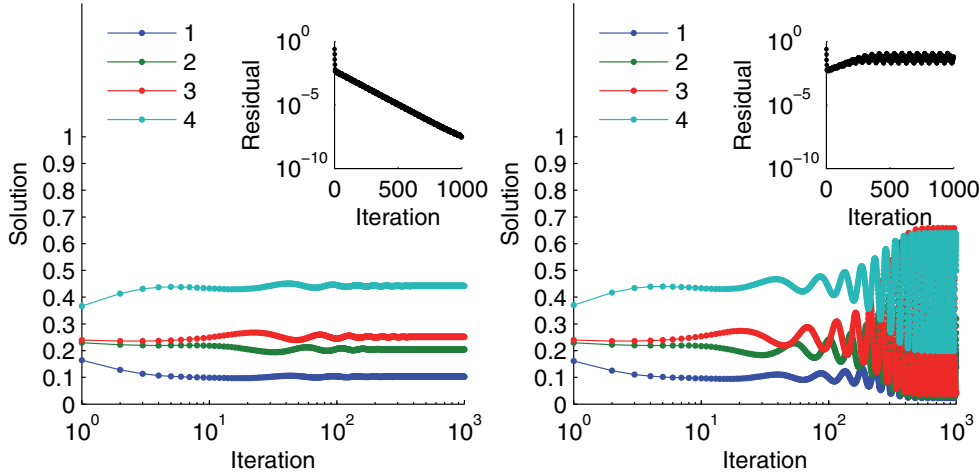
FIG. 4. *At left, the iterates of the inner-outer iteration for $\boldsymbol{R}_2$ with $\alpha = 0.97$ show that it converges to a solution, whereas this same problem did not converge with the shifted method. At right, the result of the shifted iteration on $\boldsymbol{R}_2$ with $\alpha = 0.90$ again shows an example that doesn't converge.*

and the scheme converges linearly with rate $\frac{1-\alpha/(m-1)}{1-\alpha^2} < 1$ when $\alpha < 1/(m-1)$. $\qquad\square$

In comparison with the shifted method, each iteration of the inner-outer method is far more expensive and involves solving a multilinear PageRank method. However, if $\underline{\boldsymbol{P}}$ is available only through a fast operator, this may be the only method possible. In Figure 4, we show that the inner-outer method converges in the case that the shifted method failed to converge. Increasing α to 0.99, however, now generates a problem where the inner-outer method will not converge.

5.4. An inverse iteration. Another algorithm we consider is given by our interpretation of the multilinear PageRank solution as a stochastic process. Observe, for the second-order case,

$$\alpha\boldsymbol{R}(\mathbf{x}\otimes\mathbf{x}) = \tfrac{\alpha}{2}\boldsymbol{R}(\mathbf{x}\otimes\boldsymbol{I} + \boldsymbol{I}\otimes\mathbf{x}) = \alpha\left[\tfrac{1}{2}\boldsymbol{R}(\mathbf{x}\otimes\boldsymbol{I}) + \tfrac{1}{2}\boldsymbol{R}(\boldsymbol{I}\otimes\mathbf{x})\right].$$

Both matrices

$$\boldsymbol{R}(\mathbf{x}\otimes\boldsymbol{I}) \quad \text{and} \quad \boldsymbol{R}(\boldsymbol{I}\otimes\mathbf{x})$$

are stochastic. Let $\boldsymbol{S}(\mathbf{x}) = \tfrac{1}{2}\boldsymbol{R}(\mathbf{x}\otimes\boldsymbol{I}) + \tfrac{1}{2}\boldsymbol{R}(\boldsymbol{I}\otimes\mathbf{x})$ be the stochastic sum of these two matrices. Then the multilinear PageRank vector satisfies

$$\mathbf{x} = \alpha\boldsymbol{S}(\mathbf{x})\mathbf{x} + (1-\alpha)\mathbf{v}.$$

This equation has a subtle interpretation. The multilinear PageRank vector *is* the PageRank vector of a solution-dependent Markov process. The stochastic process presented in section 4.1 shows this in a slightly different manner. The iteration that arises is a simple fixed-point idea using this interpretation:

$$\mathbf{x}_{k+1} = \alpha\boldsymbol{S}(\mathbf{x}_k)\mathbf{x}_{k+1} + (1-\alpha)\mathbf{v}.$$

Thus, at each step, we solve a PageRank problem given the current iterate to produce the subsequent vector. For this iteration, we could then leverage a fast PageRank

solver if there is a way of computing $\boldsymbol{S}(\mathbf{x}_k)$ effectively or $\boldsymbol{S}(\mathbf{x}_k)\mathbf{x}$ effectively. The method for a general problem is the same, except for the definition of $\boldsymbol{S}$. In general, let

$$(5.5)$$
$$\boldsymbol{S}(\mathbf{x}) = \frac{1}{m-1}\boldsymbol{R}(\boldsymbol{I} \otimes \underbrace{\mathbf{x}_k \otimes \cdots \otimes \mathbf{x}_k}_{m-2 \text{ terms}} + \mathbf{x}_k \otimes \boldsymbol{I} \otimes \underbrace{\mathbf{x}_k \otimes \cdots \otimes \mathbf{x}_k}_{m-3 \text{ terms}} + \cdots + \underbrace{\mathbf{x}_k \otimes \cdots \otimes \mathbf{x}_k}_{m-2 \text{ terms}} \otimes \boldsymbol{I}).$$

This iteration is guaranteed to converge in the unique solution regime.

THEOREM 5.5. *Let $\underline{\boldsymbol{P}}$ be an order-m stochastic tensor, let $\mathbf{v}$ and $\mathbf{x}_0$ be stochastic vectors, and let $\alpha < 1/(m-1)$. Let $\boldsymbol{S}(\mathbf{x}_k)$ be an $n \times n$ stochastic matrix defined via (5.5). The inverse multilinear PageRank iteration*

$$\mathbf{x}_{k+1} = \alpha \boldsymbol{S}(\mathbf{x}_k)\mathbf{x}_{k+1} + (1-\alpha)\mathbf{v}$$

converges to the unique solution $\mathbf{x}$ of the multilinear PageRank problem and also

$$\|\mathbf{x}_k - \mathbf{x}\|_1 \le \left(\frac{(m-2)\alpha}{1-\alpha}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|_1 \le 2\left(\frac{(m-2)\alpha}{1-\alpha}\right)^k.$$

Proof. We complete the proof using the terms involved in the fourth-order case ($m = 4$) because it simplifies the indexing tremendously, although the terms in our proof will be entirely general. Consider the error at the $(k+1)$th iteration:

$$\begin{aligned}
\mathbf{x}_{k+1} - \mathbf{x} &= \frac{\alpha}{m-1}\boldsymbol{R}[(\boldsymbol{I} \otimes \mathbf{x}_k \otimes \mathbf{x}_k + \mathbf{x}_k \otimes \boldsymbol{I} \otimes \mathbf{x}_k + \mathbf{x}_k \otimes \mathbf{x}_k \otimes \boldsymbol{I})\mathbf{x}_{k+1} \\
&\quad - (\boldsymbol{I} \otimes \mathbf{x} \otimes \mathbf{x} + \mathbf{x} \otimes \boldsymbol{I} \otimes \mathbf{x} + \mathbf{x} \otimes \mathbf{x} \otimes \boldsymbol{I})\mathbf{x}] \\
&= \frac{\alpha}{m-1}\boldsymbol{R}[(\mathbf{x}_{k+1} \otimes \mathbf{x}_k \otimes \mathbf{x}_k + \mathbf{x}_k \otimes \mathbf{x}_{k+1} \otimes \mathbf{x}_k + \mathbf{x}_k \otimes \mathbf{x}_k \otimes \mathbf{x}_{k+1}) \\
&\quad - (\mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x} + \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x} + \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x})].
\end{aligned}$$

At this point, it suffices to prove that terms of the form $\|\mathbf{x}_{k+1} \otimes \mathbf{x}_k \otimes \mathbf{x}_k - \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x}\|_1$ are bounded by $(m-1)\|\mathbf{x}_{k+1} - \mathbf{x}\|_1 + (m-1)(m-2)\|\mathbf{x}_k - \mathbf{x}\|_1$. Showing this for one term also suffices because all these terms are equivalent up to a permutation.

We continue by Lemma 4.5, which yields

$$\|\mathbf{x}_{k+1} \otimes \mathbf{x}_k \otimes \mathbf{x}_k - \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x}\|_1 \le \|\mathbf{x}_{k+1} - \mathbf{x}\|_1 + 2\|\mathbf{x}_k - \mathbf{x}\|_1$$

in the third-order case and $\|\mathbf{x}_{k+1} - \mathbf{x}\|_1 + (m-2)\|\mathbf{x}_k - \mathbf{x}\|_1$ in general. Since there are $m - 1$ of these terms, we are done. $\square$

In comparison to the inner-outer iteration, this method requires detailed knowledge of the operator $\underline{\boldsymbol{P}}$ in order to form $\boldsymbol{S}(\mathbf{x}_k)$ or even matrix-vector products $\boldsymbol{S}(\mathbf{x}_k)\mathbf{z}$. In some applications this may be easy. In Figure 5, we illustrate the convergence of the inverse iteration on the problems that the inner-outer method's illustration used. The convergence pattern is the same.

5.5. Newton's method. Finally, consider Newton's method for solving the non-linear equation

$$\mathbf{f}(\mathbf{x}) = \alpha \boldsymbol{R}(\mathbf{x} \otimes \mathbf{x}) + (1-\alpha)\mathbf{v} - \mathbf{x} = 0.$$

The Jacobian of this operator is

$$\boldsymbol{J}(\mathbf{x}) = \alpha \boldsymbol{R}(\mathbf{x} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \mathbf{x}) - \boldsymbol{I}.$$
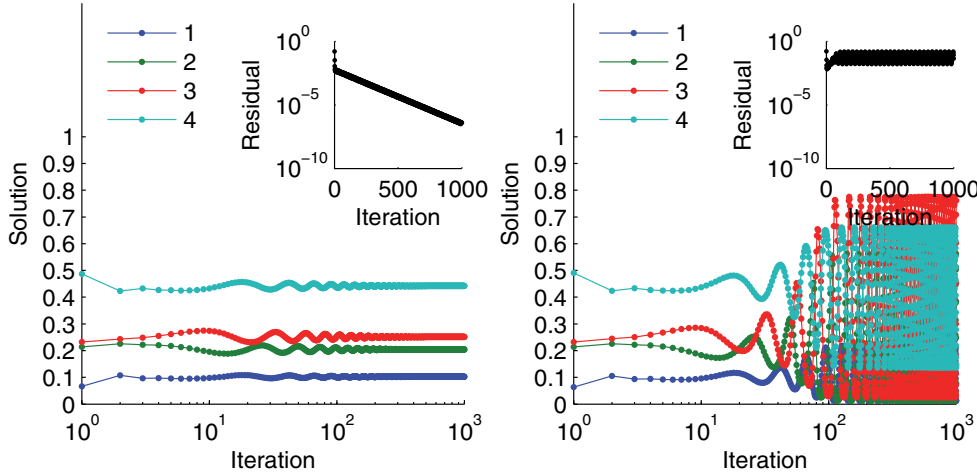
FIG. 5. *At left, the iterates from the inverse iteration to solve problem $\boldsymbol{R}_2$ with $\alpha = 0.97$, and at right, the iterates to solve problem $\boldsymbol{R}_2$ with $\alpha = 0.99$. Both show similar convergence behavior as to the inner-outer method.*

We now prove the following theorem about the convergence of Newton's method.

THEOREM 5.6. *Let $\underline{\boldsymbol{P}}$ be a third-order stochastic tensor, let $\mathbf{v}$ be a stochastic vector, and let $\alpha < 1/2$. Let $\boldsymbol{R}$ be the flattening of $\underline{\boldsymbol{P}}$ along the first index. Let*

$$\mathbf{f}(\mathbf{x}) = \alpha \boldsymbol{R}(\mathbf{x} \otimes \mathbf{x}) + (1 - \alpha)\mathbf{v} - \mathbf{x} = 0.$$

Newton's method to solve $\mathbf{f}(\mathbf{x}) = 0$, and hence compute the unique multilinear Page-Rank vector, is the iteration

$$(5.6) \qquad [\boldsymbol{I} - \alpha \boldsymbol{R}(\mathbf{x}_k \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \mathbf{x}_k)]\, \mathbf{p}_k = \mathbf{f}(\mathbf{x}_k), \qquad \mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{p}_k, \qquad \mathbf{x}_0 = 0.$$

It produces a unique sequence of iterates where

$$\mathbf{f}(\mathbf{x}_k) \geq 0, \quad \mathbf{e}^T \mathbf{f}(\mathbf{x}_k) = \frac{\alpha (\mathbf{e}^T \mathbf{f}(\mathbf{x}_{k-1}))^2}{(1 - 2\alpha)^2 + 4\alpha \mathbf{e}^T \mathbf{f}(\mathbf{x}_{k-1})} \leq \alpha (1 - \alpha)^2 \frac{1}{4^{k-1}}, \quad k \geq 1,$$

that also converges quadratically in the $k \to \infty$ limit.

This result shows that Newton's method always converges quadratically fast when solving multilinear PageRank vectors inside the unique regime.

Proof. We outline the following sequence of facts and lemmas we provide to compute the result. The key idea is to use the result that second-order multilinear PageRank is a second degree polynomial, and hence we can use Taylor's theorem to derive an *exact* prediction of the function value at successive iterations. We first prove this *key fact*. Subsequent steps of the proof establish that the sequence of iterates is unique and well-defined (that is, that the Jacobian is always nonsingular). This involves showing, additionally, that $\mathbf{x}_k \geq 0$, $\mathbf{e}^T \mathbf{x}_k \leq 1$, and $\mathbf{f}(\mathbf{x}_k) \geq 0$. Let $f_k = \mathbf{e}^T \mathbf{f}$, since $\mathbf{f}(\mathbf{x}_k) \geq 0$, showing that $f_k \to 0$ suffices to show convergence. Finally, we derive a recurrence:

$$(5.7) \qquad f_{k+1} = \frac{\alpha f_k^2}{(1 - 2\alpha)^2 + 4\alpha f_k}.$$

Key fact. Let $\mathbf{p}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$. If the Jacobian $\boldsymbol{J}(\mathbf{x}_k)$ is nonsingular in the kth iteration, then $\mathbf{f}(\mathbf{x}_{k+1}) = \alpha \boldsymbol{R}(\mathbf{p}_k \otimes \mathbf{p}_k)$. To prove this, we use an exact version of Taylor's theorem around the point $\mathbf{x}_k$:

$$\mathbf{f}(\mathbf{x}_k + \mathbf{p}) = \mathbf{f}(\mathbf{x}_k) + \boldsymbol{J}(\mathbf{x}_k)\mathbf{p} + \tfrac{1}{2}\underline{\boldsymbol{T}}\mathbf{p}^2,$$

where $\underline{\boldsymbol{T}}\mathbf{p}^2 = \alpha \boldsymbol{R}(\boldsymbol{I} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{I})(\mathbf{p} \otimes \mathbf{p})$ is *independent* of the current point. Note also that Newton's method chooses $\mathbf{p}$ such that $\mathbf{f}(\mathbf{x}_k) + \boldsymbol{J}(\mathbf{x}_k)\mathbf{p} = 0$. Then

$$\mathbf{f}(\mathbf{x}_{k+1}) = \mathbf{f}(\mathbf{x}_k + \mathbf{p}_k) = \alpha \boldsymbol{R}(\mathbf{p}_k \otimes \mathbf{p}_k).$$

Well-defined sequence. We now show that $\boldsymbol{J}(\mathbf{x}_k)$ is nonsingular for all $\mathbf{x}_k$, and hence, the Newton iteration is well-defined. It is easy to do so if we establish that

$$(5.8) \qquad \mathbf{f}(\mathbf{x}_k) \geq 0, \mathbf{x}_k \geq 0 \text{ and } z_k = \mathbf{e}^T\mathbf{x}_k \leq 1$$

also hold at each iteration. Clearly, these properties hold for the initial iteration where $\mathbf{f}(\mathbf{x}_0) = (1-\alpha)\mathbf{v}$. Thus, we proceed inductively. Note that if $\mathbf{x}_k \geq 0$ and $z_k \leq 1$, then the Jacobian is nonsingular because $-\boldsymbol{J}(\mathbf{x}_k) = [\boldsymbol{I} - \alpha \boldsymbol{R}(\mathbf{x}_k \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \mathbf{x}_k)]$ is a strictly diagonally dominant matrix, M-matrix when $\alpha < 1/2$. (In fact, both $\boldsymbol{R}(\mathbf{x}_k \otimes \boldsymbol{I})$ and $\boldsymbol{R}(\boldsymbol{I} \otimes \mathbf{x}_k)$ are nonnegative matrices with column norms equal to $z_k = \mathbf{e}^T\mathbf{x}_k$.) Thus, $\mathbf{x}_{k+1}$ is well-defined and it remains to show that $\mathbf{f}(\mathbf{x}_{k+1}) \geq 0$, $\mathbf{x}_{k+1} \geq 0$, and $z_{k+1} \leq 1$. Now, by the definition of Newton's method,

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \boldsymbol{J}(\mathbf{x}_k)^{-1}\mathbf{f}(\mathbf{x}_k),$$

but $-\boldsymbol{J}$ is an M-matrix, and so $\mathbf{x}_{k+1} \geq 0$. This also shows that $\mathbf{p}_k = \mathbf{x}_{k+1} - \mathbf{x}_k \geq 0$, from which we can use our key fact to derive that $\mathbf{f}(\mathbf{x}_{k+1}) \geq 0$. What remains to show is that $z_{k+1} \leq 1$. By taking summations on both sides of (5.6), we have

$$(1 - 2\alpha z_k)(z_{k+1} - z_k) = \alpha z_k^2 + (1 - \alpha) - z_k.$$

A quick, but omitted, calculation confirms that $z_{k+1} > 1$ implies $z_k > 1$. Thus, we completed our inductive conditions for (5.8).

Recurrence. We now show that (5.7) holds. First, observe that

$$f_k = \alpha(\mathbf{e}^T\mathbf{p}_k)^2, \qquad \mathbf{e}^T\mathbf{p}_k = \frac{f_k}{1 - 2\alpha z_k}, \qquad \alpha z_k^2 + (1 - \alpha) - z_k - f_k = 0.$$

We now solve for z_k in terms of f_k. This involves picking a root for the quadratic equation. Since $z_k \leq 1$, this makes the choice the negative root in

$$z_k = \frac{1 - \sqrt{(1 - 2\alpha)^2 + 4\alpha f_k}}{2\alpha} \leq 1.$$

Assembling these pieces yields (5.7).

Convergence. We have an easy result that $f_{k+1} \leq \tfrac{1}{4}f_k$ by ignoring the term $(1 - 2\alpha)^2$ in the denominator. Also, by direct evaluation, $f_1 = \alpha(1 - \alpha)^2$. Thus,

$$f_k \leq \frac{1}{4^{k-1}}f_1 = \frac{1}{4^{k-1}}\alpha(1 - \alpha)^2,$$

which is one side of the convergence rate. The sequence for f_k also converges quadratically in the limit because $\lim_{k \to \infty} \frac{f_{k+1}}{f_k^2} = \frac{\alpha}{(1-2\alpha)^2}$. $\square$
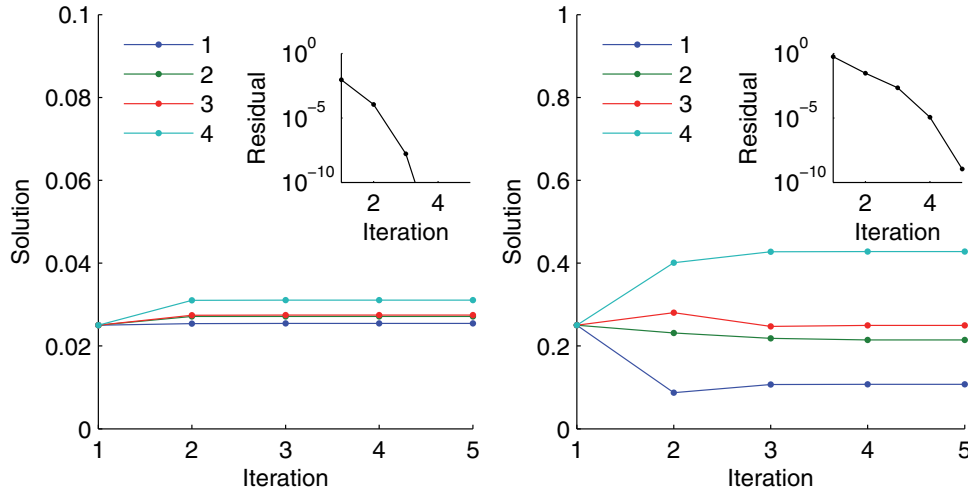
FIG. 6. *For problem $\mathbf{R}_2$ with $\alpha = 0.99$, the Newton iteration from Theorem 5.6 (left figure) converges to a nonstochastic solution—note the scale of the solution axis. The always-stochastic iteration (5.9) (right figure) converges for this problem. Thus, we recommend the iteration (5.9) when $\alpha > \frac{1}{m-1}$.*

A practical, always-stochastic Newton iteration. The Newton iteration from Theorem 5.6 begins as $\mathbf{x}_0 = 0$ and, when $\alpha < 1/2$, gradually grows the solution until it becomes stochastic and attains optimality. For problems when $\alpha > 1/2$, however, this iteration often converges to a fixed-point where $\mathbf{x}$ is not stochastic. (In fact, it always did this in our brief investigations.) To make our codes practical for problems where $\alpha > 1/2$, then, we enforce an explicit stochastic normalization after each Newton step:

$$(5.9) \quad [\boldsymbol{I} - \alpha\boldsymbol{R}(\mathbf{x}_k \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \mathbf{x}_k)]\mathbf{p}_k = \mathbf{f}(\mathbf{x}_k), \quad \mathbf{x}_{k+1} = \mathrm{proj}(\mathbf{x}_k + \mathbf{p}_k), \quad \mathbf{x}_0 = (1-\alpha)\mathbf{v},$$

where

$$\mathrm{proj}(\mathbf{x}) = \max(\mathbf{x}, 0)/\mathbf{e}^T \max(\mathbf{x}, 0)$$

is a projection operator onto the probability simplex that sets negative elements to zero and then normalizes those left to sum to one. We found this iteration superior to a few other choices, including using a proximal point projection operator to produce always stochastic iterates (Parikh and Boyd, 2014, section 6.2.5). We illustrate an example of the difference in Figure 6, where the always-stochastic iteration solves the problem and the iteration without this projection converges to a nonstochastic fixed-point. Note that, like a general instance of Newton's method, the system $[\boldsymbol{I} - \alpha\boldsymbol{R}(\mathbf{x}_k \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \mathbf{x}_k)]$ may be singular. We never ran into such a case in our experiments and in our study. In the regime when $\alpha < 1/2$, then the Jacobian is always nonsingular. Also, in this regime, we observed that stochastic iterates stay stochastic. Analyzing the convergence of a stochastic sequence of iterates when $\alpha < 1/2$ is thus an interesting future direction.

We further illustrate the behavior of Newton's method on $\boldsymbol{R}_2$ with $\alpha = 0.97$ and $\boldsymbol{R}_2$ with $\alpha = 0.99$ in Figure 7. The second of these problems did not converge for either the inner-outer or the inverse iteration. Newton's method solves it in just a
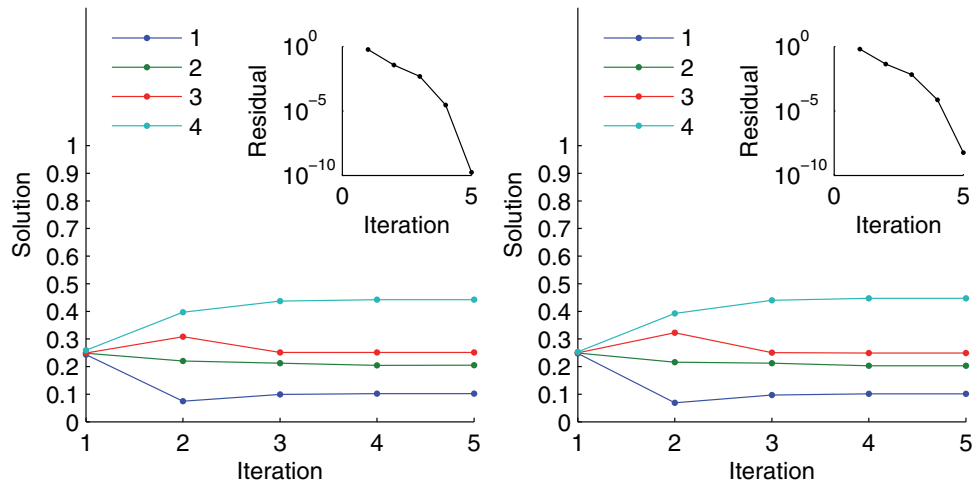
FIG. 7. *At left, the iterates of Newton's method to solve problem $\boldsymbol{R}_2$ with $\alpha = 0.97$, and at right, the iterates to solve problem $\boldsymbol{R}_2$ with $\alpha = 0.99$. Both sequences converge unlike the inner-outer and inverse iterations.*

few iterations. In comparison to both the inner-outer and inverse iterations, however, Newton's method requires even more direct access to $\underline{\boldsymbol{P}}$ in order to solve for the steps with the Jacobian.

6. Experimental results. To evaluate these algorithms, we create a database of problematic tensors. We then use this database to address two questions:

1. What value of the shift is most reliable?
2. Which method has the most reliable convergence?

In terms of reliability, we wish for the method to generate a residual smaller than 10^{-8} before reaching the maximum iteration count where the residual for a potential solution is given by (5.1). In many of the experiments, we run the methods for between 10,000 to 100,000 iterations. If we do not see convergence in this period, we deem a particular trial a failure. The value of $\mathbf{v}$ is always $\mathbf{e}/n$ but α will vary between our trials. Before describing these results, we begin by discussing how we created the test problems.

6.1. Problems. We used exhaustive enumeration to identify $2 \times 2 \times 2$ and $3 \times 3 \times 3$ binary tensors, which we then normalized to stochastic tensors, which exhibited convergence problems with the fixed-point or shifted methods. We also randomly sampled many $4 \times 4 \times 4$ binary problems and saved those that showed slow or erratic convergence for these same algorithms. We used $\alpha = 0.99$ and $\mathbf{v} = \mathbf{e}/n$ for these studies. The $6 \times 6 \times 6$ problems were constructed randomly in an attempt to be adversarial. Tensors with strong "directionality" seemed to arise as interesting cases in much of our theoretical study (this is not presented here). By this, we mean, for instance, tensors where a single state has many incoming links. We created a random procedure that generates problems where this is true (the exact method is in the online supplement) and used this to generate $6 \times 6 \times 6$ problems. In total, we have

the following problems:

$$3 \times 3 \times 3, \qquad 5 \text{ problems,}$$
$$4 \times 4 \times 4, \qquad 19 \text{ problems,}$$
$$6 \times 6 \times 6, \qquad 5 \text{ problems.}$$

The full list of problems is in given in Appendix B.

We used the MATLAB symbolic toolbox to compute a set of exact solutions to these problems. These $6 \times 6 \times 6$ problems often had multiple solutions, whereas the smaller problems only had a single solution (for the values of α we considered). While it is possible there are solutions missed by this tool, prior research found symbolic computation a reliable means of solving these polynomial systems of equations (Kolda and Mayo, 2011).

6.2. Shifted iteration. We begin our study by looking at a problem where the necessary shift suggested by the ODE theory ($\gamma = 1/2$ for third-order data) does not result in convergence. We are interested in whether varying the shift will alter the convergence behavior. This is indeed the case. For the problem $\boldsymbol{R}_{4,11}$ from the appendix with $\alpha = 0.99$, we show the convergence of the residual as the shift γ varies in Figure 8. When $\gamma = 0.5$, the iteration does not converge. There is a point somewhere between $\gamma = 0.554$ and $\gamma = 0.5545$ where the iteration begins to converge. When we set $\gamma = 1$, the iteration converged rapidly.

In the next experiment, we wished to understand how the reliability of the method depended on the shift γ. In Table 1, we vary α and the shift γ and look at how many of the 29 test problems the shifted method can solve within 10,000 iterations. Recall that a method solves a problem if it pushes the residual below 10^{-8} within the iteration bound. The results from that table show that $\gamma = 1$ or $\gamma = 2$ results in the most reliable method. When $\gamma = 10$, then the method was less reliable. This is likely due to the shift delaying convergence for too long. Note that we chose many of the problems based on the failure of the shifted method with $\gamma = 0$ or $\gamma = 1/2$ and so the poor performance of these choices may not reflect their true reliability. Nevertheless, based on the results of this table, we recommend a shift of $\gamma = 1$ for a third-order problem, or a shift of $\gamma = m - 2$ for a problem with an order-m tensor.

6.3. Solver reliability. In our final study, we utilize each method with the following default parameters:

F	fixed-point	10,000 maximum iterations, $\mathbf{x}_0 = \mathbf{v}$
S	shifted	10,000 maximum iterations, $\gamma = 1$, $\mathbf{x}_0 = \mathbf{v}$
IO	inner-outer	1,000 outer iterations, internal tolerance ε, $\mathbf{x}_0 = \mathbf{v}$
Inv	inverse	1,000 iterations, $\mathbf{x}_0 = \mathbf{v}$
N	Newton	1,000 iterations, projection step, $\mathbf{x}_0 = (1 - \alpha)\mathbf{v}$

We also evaluate each method with 10 times the default number of iterations.

The results of the evaluation are shown in Table 2 as α varies from 0.7 to 0.99. The fixed-point method has the worst performance when α is large. Curiously, when $\alpha = 0.99$ the shifted method outperforms the inverse iteration, but when $\alpha = 0.95$ the inverse iteration outperforms the shifted iteration. This implies that the behavior and reliability of the methods are not monotonic in α. While this fact is not overly surprising, it is pleasing to see a concrete example that might suggest some tweaks to the methods to improve their reliability. Overall, the inner-outer and Newton's methods have the most reliable convergence on these difficult problems.
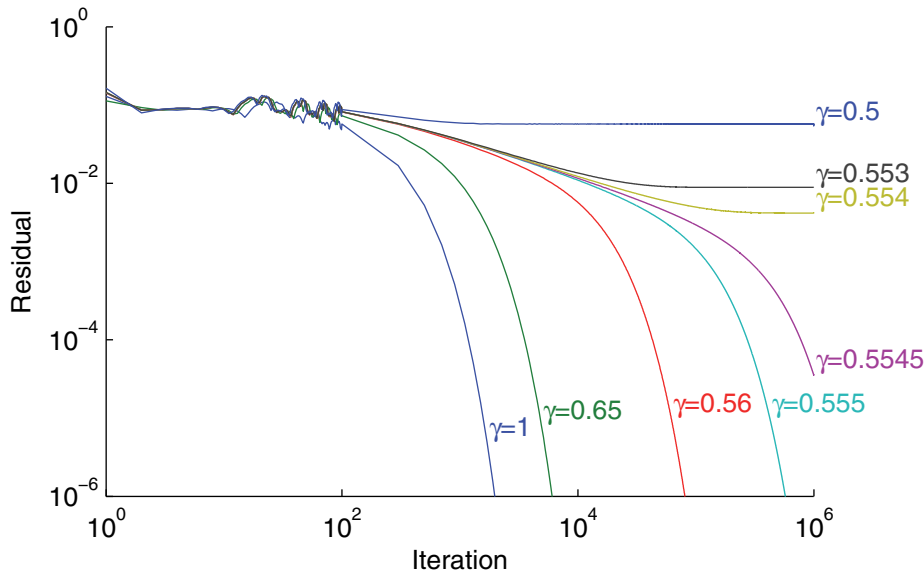
FIG. 8. For the problem $\boldsymbol{R}_{4,11}$ with $\alpha = 0.99$ and $\mathbf{v} = \mathbf{e}/n$, the shifted method will not converge unless γ is slightly larger than 0.554. As γ becomes larger, the convergence rate increases.

TABLE 1
Each row of this table reports the number of problems successfully solved by the shifted iteration as the shift varies from 0 to 10. The values are reported for each value of α considered, as well as broken down into the different problem sizes considered.

α	n	Shifts γ						
		0	1/4	1/2	3/4	1	2	10
0.70	3	5	5	5	5	5	5	5
	4	19	19	19	19	19	19	19
	6	5	5	5	5	5	5	5
		29	29	29	29	29	29	29
0.85	3	5	5	5	5	5	5	5
	4	19	19	19	19	19	19	19
	6	5	5	5	5	5	5	5
		29	29	29	29	29	29	29
0.90	3	5	5	5	5	5	5	5
	4	18	19	19	19	19	19	19
	6	5	5	5	5	5	5	5
		28	29	29	29	29	29	29
0.95	3	5	5	5	5	5	5	5
	4	7	11	13	13	16	19	18
	6	5	5	5	5	5	5	5
		17	21	23	23	26	29	28
0.99	3	4	5	5	5	5	5	5
	4	0	1	1	2	2	2	2
	6	1	1	1	2	2	2	1
		5	7	7	9	9	9	8

TABLE 2
Each row of this table reports the number of problems successfully solved by the various iterative methods in two cases: with their default parameters and with 10 times the standard number of iterations. The values are reported for each value of α considered, as well as broken down into the different problem sizes considered. The columns are F for the fixed-point, S for the shifted method, IO for the inner-outer, Inv for the inverse iteration, and N for Newton's method.

α	n	Method (defaults)					Method (Extra iteration)				
		F	S	IO	Inv	N	F	S	IO	Inv	N
0.70	3	5	5	5	5	5	5	5	5	5	5
	4	19	19	19	19	19	19	19	19	19	19
	6	5	5	5	5	5	5	5	5	5	5
		29	29	29	29	29	29	29	29	29	29
0.85	3	5	5	5	5	5	5	5	5	5	5
	4	19	19	19	19	19	19	19	19	19	19
	6	5	5	5	5	5	5	5	5	5	5
		29	29	29	29	29	29	29	29	29	29
0.90	3	5	5	5	5	5	5	5	5	5	5
	4	18	19	19	19	19	18	19	19	19	19
	6	5	5	5	5	5	5	5	5	5	5
		28	29	29	29	29	28	29	29	29	29
0.95	3	5	5	5	5	5	5	5	5	5	5
	4	7	16	18	19	19	8	16	19	19	19
	6	5	5	5	5	5	5	5	5	5	5
		17	26	28	29	29	18	26	29	29	29
0.99	3	4	5	5	5	5	4	5	5	5	5
	4	0	2	15	1	19	0	2	17	1	19
	6	1	2	3	1	4	2	3	4	3	4
		5	9	23	7	28	6	10	26	9	28

Newton's method, in fact, solves all but one instance: $\boldsymbol{R}_{6,3}$ with $\alpha = 0.99$. We explore this problem in slightly more depth in Figure 9. This problem only has a single unique solution (based on our symbolic computation). However, none of the iterations will find it using the default settings—all the methods are attracted to a point with a small residual and an indefinite Jacobian. We were able to find the true solution by using Newton's method with random starting points. It seems that the iterates need to approach the solution from a rather precise trajectory in order to overcome an indefinite region. This problem should be a useful case for future algorithmic studies on the problem.

6.4. A four-mode example. We exhaustively enumerated stochastic normalizations of binary $2 \times 2 \times 2 \times 2$ and $2 \times 2 \times 2 \times 2 \times 2$ problems and found that the fixed-point method converged in all instances. (There are only about 4 billion such objects, so this takes only a few days on a modern computer.) We then began exhaustively enumerating $3 \times 3 \times 3 \times 3$ problems that are stochastically normalized and have the dangling adjustment using the uniform vector. Among the first few billion of these, the fixed-point method always converges, although it does so slowly in some cases. We list an example in Appendix B.4 that has 15 nonzeros and show the convergence with $\alpha = 0.99$ in Figure 10. The convergence is smooth but extremely slow with a rate of approximately 0.9998. By way of comparsion, this is approximately $\alpha^{1/42} = 0.99^{1/42} \approx 0.9998$. If we ran the power method on the associated higher-order Markov chain, the convergence rate would be $\alpha^{1/3}$.

The point	The eigenvalues	The true solution
0.199907259533067	0.980000000000000	0.043820721946272
0.006619352098700	0.000064771773360	0.002224192630620
0.116429656827957	-1.786544142144891	0.009256490884022
0.223220491129316	-0.575965838505486	0.819168263512464
0.079958855790239	-0.575965838505486	0.031217440669761
0.373864384620721	-1.438690261635567	0.094312890356862

The Jacobian

-0.9712	0.2246	0.3496	0.1944	0.3395	0.7435
0	-0.7299	0.0131	0	0.0824	0
0.4781	0.1851	-0.9505	0	0.4621	0.2408
0.0288	0.1851	0.0495	0.1822	0	0.4453
0	0.4192	0.3701	0.0857	-0.5939	0.1581
1.4443	0.6960	1.1482	0.5176	0.6899	-0.6077

FIG. 9. *Newton's method on the nonconvergent case of $\boldsymbol{R}_{6,3}$ with $\alpha = 0.99$. The method repeatedly drops to a small residual before moving away from the solution. This happens because when the residual becomes small, then the Jacobian acquires a nontrivial positive eigenvalue as exemplified by the point with the red-circled residual. We show the Jacobian and eigenvalues at this point. It appears to be a pseudosolution that attracts all the algorithms. Using random starting points, Newton's method will sometimes generate the true solution, which is far from the the attracting point.*

6.5. A large example. We conclude this section with an example using a large tensor. The `wiki-talk` graph (Leskovec, Huttenlocher, and Kleinberg, 2010) provided in the SNAP repository (https://snap.stanford.edu) is a collection of 2,394,385 editors on Wikipedia and their relationships. We have a directed edge between user i and user j if user i ever edited the "talk" page associated with user j. There are 5,021,410 of these directed relationships. Based on the ideas from Klymko, Gleich, and Kolda (2014) that we later extended in Benson, Gleich, and Leskovec (2015), we consider finding the multilinear PageRank vector of the tensor of directed 3-cycles. Let $\underline{\boldsymbol{C}}$ be an $n \times n \times n$ symmetric tensor where

$$\underline{C}(i,j,k) = \begin{cases} 1, & \text{there is a directed three-cycle between nodes } i,j,k, \\ 0 & \text{otherwise.} \end{cases}$$

This tensor has 15,416,112 directed three-cycles. Let $\boldsymbol{S}$ be the 1-mode unfolding of $\underline{\boldsymbol{C}}$ that has been normalized to be a substochastic matrix by scaling the columns. Let $\boldsymbol{M} = \boldsymbol{A}^T \boldsymbol{D}^+$ be the standard random-walk normalization of the directed adjacency matrix. Finally, let dangling($\cdot$) be the dangling correction vector as discussed in section 4.5. We consider using the tensor with unfolded matrix

$$\boldsymbol{R} = \beta[\boldsymbol{S} + \mathbf{v} \cdot \text{dangling}(\boldsymbol{S})] + (1 - \beta)[\boldsymbol{M} + \mathbf{v} \cdot \text{dangling}(\boldsymbol{M})] \otimes \mathbf{e}^T.$$

We can compute products $\boldsymbol{R}(\mathbf{x} \otimes \mathbf{x})$ in time proportional to the number of directed three-cycles by using the previously described algorithm that corrects for the missing sum of associated dangling vectors. Enumerating these cycles and computing this product are both fast operations and take at most a few seconds for our unoptimized code.

The number of iterations required for the fixed-point without any shift is given in Table 3. We fixed $\mathbf{v} = \mathbf{e}/n$ as the uniform stochastic vector. We found that the simple fixed-point method always converges quickly for this problem as we vary both α and β. As β tended toward 1, the resulting tensor becomes so sparse that the method
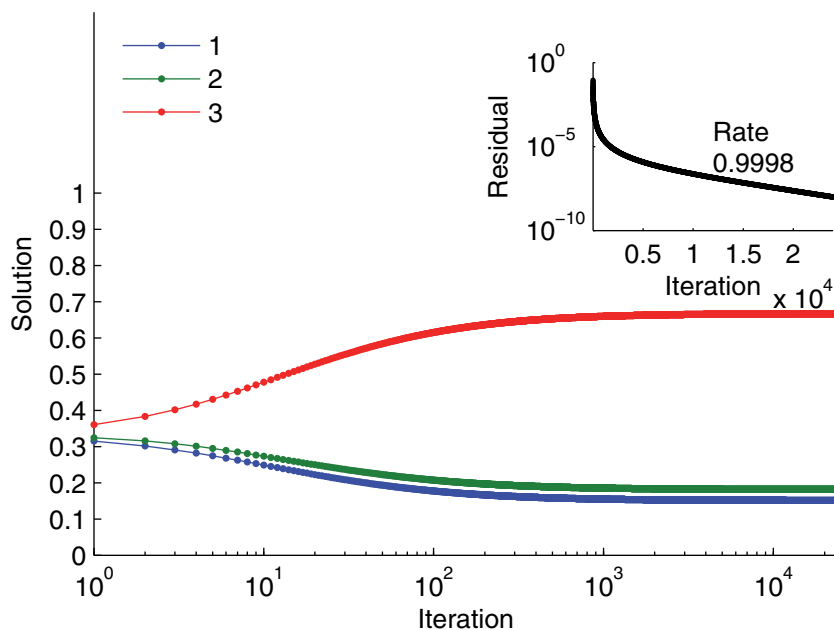
FIG. 10. *The convergence of the fixed-point iteration on a four-mode, three-dimensional problem is regular but extremely slow. The convergence rate over the last* 10,000 *iterations was* 0.9998 *where* $\alpha = 0.99$.

TABLE 3
The number of iterations required for a problem with $n = 2,394,385$ *using the* `wiki-Talk` *dataset. The stochastic tensor is a combination of directed three-cycles and edges of the original graph that varies with the parameter* β *such that* $\beta = 1$ *uses only three cycles and* $\beta = 0$ *reduces to the traditional PageRank case. Because of the sparsity of the three-cycle information, the iteration converges more quickly as* $\beta \to 1$.

α	β		
	0.50	0.85	0.99
0.50	22	12	6
0.85	35	15	7
0.99	42	16	7

always converges quickly. This happens because matrix $\boldsymbol{R} \approx \mathbf{v}(\mathbf{e}^T \otimes \mathbf{e}^T)$. That said, there is additional information in these solutions that can be useful for identifying clusters of states in the Markov chain, as we discuss in Benson, Gleich, and Leskovec (2015).

7. Discussion. In this manuscript, we studied the higher-order PageRank problem as well as the multilinear PageRank problem. The higher-order PageRank problem behaves much like the standard PageRank problem: we always have guaranteed uniqueness and fast convergence. The multilinear PageRank problem, in contrast, only has uniqueness and fast convergence in a more narrow regime. Outside of that regime, existence of a solution is guaranteed, although uniqueness is not. But note that the multilinear PageRank problem is interesting only for massive problems. If $O(n^2)$ memory is available, then the higher-order PageRank vector should be used unless there is a modeling reason to choose the multilinear formulation.

For the multilinear PageRank problem, convergence of simple iterative methods outside of the uniqueness regime is highly dependent on the data. We created a test set based on problems where both the fixed-point and shifted fixed-point methods fail. On these tough problems, both the inner-outer and Newton iterations had the best performance. This result suggests a two-phase approach to solving the problems: first try the simple shifted method. If that does not seem to converge, then use either a Newton or an inner-outer iteration. Our current findings are focused to the third-order case. In our online codes, we have conducted experiments with all stochastic normalizations of binary $2 \times 2 \times 2 \times 2$ (4-mode) and $2 \times 2 \times 2 \times 2 \times 2$ (5-mode) problems. This exhausive test of over 4 billion problems did not reveal any nonconvergent data.

We also briefly explored optimization-based methods to find fixed-points by minining the norm of the residual squared as well as methods to solve nonlinear equations. Experiments with the MATLAB optimization toolbox showed that these approches converged for all the difficult problems we found. We leave deriving the equivalent convergence guarantees and exploring the experimental trade-offs to future work as this paper is focused on simple algorithms and understanding what is and is not possible with the multilinear PageRank problem.

Based on our theoretical results, we note that there seems to be a key transition for *all* the algorithms and theory that arise at the uniqueness threshold: $\alpha < 1/(m-1)$. Curiously, the same type of bound exists for the more complicated model studied by Li and Ng (2015) involving m-stochastic tensors and we conjecture that their model is a special case, or closely related case, to our multilinear PageRank model. We are currently trying to find algorithms with guaranteed convergence when $\alpha > 1/(m-1)$ but have not been successful yet. We plan to explore using sum-of-squares programming for this task in the future. Such an approach has given one of the first algorithms with good guarantees for the tensor eigenvalue problem (Nie and Wang, 2014). In closing, however, let us note that this seems to be a largely theoretical issue. We explored a few large-scale applications of multilinear PageRank in Benson, Gleich, and Leskovec (2015) and never observed any issues with convergence using the fixed-point method. It is possible that a smoothed analysis approach could be used to make this observation formal.

Appendix A. Applying Li and Ng's results to multilinear PageRank. For the third-order tensor problem

$$\underline{P}\mathbf{x}^2 = \mathbf{x}$$

Li and Ng (2013) define a quantity called β to determine if the solution is unique. (Their quantity was γ, but we use β here to avoid confusion with the shifting parameter γ.) When $\beta > 1$, then the solution is unique, and in this section, we show that $\beta > 1$ is a stronger condition than $\alpha < 1/2$. The scalar value β ($0 \le \beta \le 4$) is defined

$$\beta = \min_{S \subset \langle n \rangle} \left\{ \underbrace{\min_{k \in \langle n \rangle} \left(\min_{j \in S} \sum_{i \in \bar{S}} \underline{P}_{ijk} + \min_{j \in \bar{S}} \sum_{i \in S} \underline{P}_{ijk} \right)}_{\beta_1} + \underbrace{\min_{j \in \langle n \rangle} \left(\min_{k \in S} \sum_{i \in \bar{S}} \underline{P}_{ijk} + \min_{k \in \bar{S}} \sum_{i \in S} \underline{P}_{ijk} \right)}_{\beta_2} \right\},$$

where $\langle n \rangle \equiv \{1, 2, \ldots, n\}, S \subset \langle n \rangle$, and $\bar{S} = \langle n \rangle \setminus S$. Note that we divide this into two components, β_1 and β_2, that both depend on the set S.

When we apply their theory to multilinear PageRank, we study the problem

$$\bar{\underline{P}}\mathbf{x}^2 = \mathbf{x}, \text{ where } \bar{\underline{P}}_{ijk} = \alpha\underline{P}_{ijk} + (1-\alpha)v_i.$$

The value of β is a function of $\underline{P}$ and clearly $\beta(\omega\underline{P}) = \omega\beta(\underline{P})$, where ω is a scalar. Generally, $\beta(\underline{P} + \underline{Q}) \neq \beta(\underline{P}) + \beta(\underline{Q})$ for arbitrary tensors $\underline{P}$ and $\underline{Q}$. However, the equation $\beta(\underline{P} + \underline{Q}) = \beta(\underline{P}) + \beta(\underline{Q})$ holds for the construction of $\bar{\underline{P}}$ as we now show.

Let $\underline{Q}$ be the tensor where $\underline{Q}_{ijk} = (1 - \alpha)v_i$ and define $q_i = (1 - \alpha)v_i$ to simplify the notation. Then $\bar{\underline{P}} = \alpha\underline{P} + \underline{Q}$. Let us first consider β_1:

$$\beta_1(\alpha\underline{P} + \underline{Q}) = \min_{k \in \langle n \rangle} \left[\min_{j \in S} \sum_{i \in \bar{S}} \left(\alpha\underline{P}_{ijk} + \underline{Q}_{ijk} \right) + \min_{j \in \bar{S}} \sum_{i \in S} \left(\alpha\underline{P}_{ijk} + \underline{Q}_{ijk} \right) \right]$$

$$= \min_{k \in \langle n \rangle} \left[\min_{j \in S} \sum_{i \in \bar{S}} \left(\alpha\underline{P}_{ijk} + q_i \right) + \min_{j \in \bar{S}} \sum_{i \in S} \left(\alpha\underline{P}_{ijk} + q_i \right) \right]$$

$$= \min_{k \in \langle n \rangle} \left[\left(\min_{j \in S} \sum_{i \in \bar{S}} \alpha\underline{P}_{ijk} + \min_{j \in \bar{S}} \sum_{i \in S} \alpha\underline{P}_{ijk} \right) + \sum_i q_i \right]$$

$$= \beta_1(\alpha\underline{P}) + \beta_1(\underline{Q}).$$

By the same derivation,

$$\beta_2(\alpha\underline{P} + \underline{Q}) = \beta_2(\alpha\underline{P}) + \beta_2(\underline{Q}).$$

Now note that, because $\beta_1(\underline{Q}) = \beta_2(\underline{Q}) = 1 - \alpha$ independently of the set S, we have

$$\beta(\alpha\underline{P} + \underline{Q}) = \alpha\beta(\underline{P}) + 2(1 - \alpha).$$

We are interested in the case that $\beta(\alpha\underline{P}+\underline{Q}) > 1$ to apply the uniqueness theorem. Note that $\beta > 1$ can be true even if $\alpha > 1/2$. However, $\alpha < 1/2$ implies that $\beta > 1$. Thus, the condition $\beta > 1$ is stronger.

Appendix B. The tensor set. The following problems gave us the tensors $\underline{P}$ for our experiments, after they were normalized to be column stochastic matrices.

B.1. $3 \times 3 \times 3$.

$$\boldsymbol{R}_{3,1} = \left[\begin{array}{ccc|ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \end{array} \right]$$

$$\boldsymbol{R}_{3,2} = \left[\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{array} \right]$$

$$\boldsymbol{R}_{3,3} = \left[\begin{array}{ccc|ccc|ccc} 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \end{array} \right]$$

$$\boldsymbol{R}_{3,4} = \left[\begin{array}{ccc|ccc|ccc} 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{array} \right]$$

$$\boldsymbol{R}_{3,5} = \left[\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

B.2. $4 \times 4 \times 4$.

$$\boldsymbol{R}_{4,1} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{array} \right]$$

$$R_{4,2} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&1&1&1&0&0&1&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&1&0&1&1&0&0&1&0 \\ 1&1&1&1&0&0&0&1&0&0&0&0&1&0&0&1 \end{array}\right]$$

$$R_{4,3} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&1&0&0&1&0&0&1 \\ 0&0&0&0&1&1&1&0&0&1&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&1&0&0&1&0 \\ 1&1&1&1&0&0&0&1&1&0&0&0&0&1&0&0 \end{array}\right]$$

$$R_{4,4} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1 \\ 0&0&0&0&0&1&0&1&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0&0&1&1&0&0&0&1&0 \\ 1&1&1&1&1&0&0&0&1&0&0&1&1&0&0&0 \end{array}\right]$$

$$R_{4,5} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&1&0&0&0&1&0&0&1 \\ 0&0&0&0&0&1&0&1&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0&0&1&1&0&0&0&0&0 \\ 1&1&1&1&1&0&0&0&0&0&0&1&0&0&1&1 \end{array}\right]$$

$$R_{4,6} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0 \\ 0&0&0&0&0&1&0&1&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0&0&1&1&0&0&0&1&0 \\ 1&1&1&1&1&0&0&0&1&0&0&0&0&0&1&1 \end{array}\right]$$

$$R_{4,7} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&1 \\ 0&0&0&0&0&1&0&1&0&1&0&0&1&1&0&0 \\ 0&0&0&0&0&0&1&0&0&1&1&0&0&0&0&0 \\ 1&1&1&1&1&0&0&0&1&0&0&1&0&0&1&0 \end{array}\right]$$

$$R_{4,8} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&0&0&1&0&0&1&0 \\ 0&0&0&0&0&1&0&1&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0&0&1&1&0&0&0&0&0 \\ 1&1&1&1&1&0&0&0&1&0&0&0&1&0&0&1 \end{array}\right]$$

$$R_{4,9} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&0 \\ 0&0&0&0&0&1&0&1&0&1&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0&0&1&1&0&0&0&0&0 \\ 1&1&1&1&1&0&0&0&1&0&0&0&0&0&0&1 \end{array}\right]$$

$$R_{4,10} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&1&0&0&0&0&1&0&0&1 \\ 0&0&0&0&0&1&1&0&0&1&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&1&0&1&1&0 \\ 1&1&1&1&1&0&0&0&1&0&0&0&0&0&0&0 \end{array}\right]$$

$$R_{4,11} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1 \\ 0&0&0&0&0&1&0&1&0&1&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0&0&1&1&1&0&0&0&0 \\ 1&1&1&1&1&0&0&0&1&0&0&0&1&1&1&1 \end{array}\right]$$

$$R_{4,12} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&1&0&0&0&0&1&0&0&1 \\ 0&0&0&0&0&1&1&0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&1&0&0&1&0 \\ 1&1&1&1&1&0&0&0&1&0&0&1&1&1&0&1 \end{array}\right]$$

$$R_{4,13} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1 \\ 0&0&0&0&0&1&0&1&0&1&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0&1&1&1&0&0&0&0&0 \\ 1&1&1&1&1&0&0&0&0&0&0&1&1&0&1&0 \end{array}\right]$$

$$R_{4,14} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&1&0&0&0&0&1&0&0 \\ 0&0&0&0&1&1&1&0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&1&0&0&1&0 \\ 1&1&1&1&0&0&0&1&0&0&0&0&1&0&1&1 \end{array}\right]$$

$$R_{4,15} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&1&0&1&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0&1&1&1&0&0&0&1&0 \\ 1&1&1&1&1&0&0&0&0&0&0&1&1&0&0&1 \end{array}\right]$$

$$R_{4,16} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&1 \\ 0&0&0&0&0&1&1&0&0&1&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&1&1&1&1&0&1&0 \\ 1&1&1&1&1&0&0&1&1&0&0&0&0&1&0&0 \end{array}\right]$$

$$R_{4,17} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&1&0&0&0&0&0&0&0&0&1&0&1 \\ 0&0&0&0&0&1&1&0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&1&1&0&1&1&0 \\ 1&1&1&1&0&0&0&1&1&0&0&0&1&1&1&0 \end{array}\right]$$

$$R_{4,18} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&1&0&0&0&0&0&0&0&1&1&0&1 \\ 0&0&0&0&0&1&1&0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&1&0&0&1&0 \\ 1&1&1&1&0&0&0&1&1&1&0&0&0&0&0&0 \end{array}\right]$$

$$R_{4,19} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&1 \\ 0&0&0&0&0&1&0&1&1&0&0&0&1&1&0&0 \\ 0&0&0&0&0&0&1&0&0&1&1&0&1&0&0&0 \\ 1&1&1&1&1&0&0&0&1&0&0&1&1&0&1&0 \end{array}\right]$$

B.3. $6 \times 6 \times 6$.

$$R_{6,1} = \left[\begin{array}{cccccc|cccccc|cccccc|cccccc|cccccc|cccccc} 0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&1&0&0&0&1&0&1&0&0&0&0&1&1&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&1&0&0&1&0&0&0&0&1&0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&1&1&0&1&0&1&0&0&0&0&0&0&1&0&1&0&1&0&1&0&0&0&0&0&0&0&1&1&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&1&0&1&0&0&1&0&0&0&0 \\ 1&1&1&1&1&1&1&0&0&1&0&1&0&0&0&1&0&1&0&1&0&0&0&0&0&1&0&1&0&0&0&1&0&0&0&1 \end{array}\right]$$

$$R_{6,2} = \left[\begin{array}{cccccc|cccccc|cccccc|cccccc|cccccc|cccccc} 0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&1&0&1&0&0&0&1 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&1&0&0&1&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&1&0&0&0&1&0 \\ 0&0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&1&0&1&0&0&0&1&0&0&0&1&0&1&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&0&1&1&0&0&0&0&0&0&0 \\ 1&1&1&1&1&1&0&1&1&1&1&0&0&1&1&0&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0 \end{array}\right]$$

$$R_{6,3} = \left[\begin{array}{cccccc|cccccc|cccccc|cccccc|cccccc|cccccc} 0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0&1&0&1&0&0&1&1&0&0&0&0&0&0&0&0&1 \\ 0&0&0&0&0&0&0&0&1&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&1&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&1&0&1&1&0&0&1&0 \\ 0&0&0&0&0&0&0&0&1&0&0&0&1&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0&0&0&1&0&1&0&0 \\ 0&0&0&0&0&0&0&1&0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&0&1&0&1&0 \\ 1&1&1&1&1&1&1&0&0&1&1&0&1&0&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0 \end{array}\right]$$

$$R_{6,4} = \left[\begin{array}{cccccc|cccccc|cccccc|cccccc|cccccc|cccccc} 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&1&0&1&0&0&0&0&0&1&1&1&0&0&0&0&1&1&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&1&0&0&0&0&0&0&0&0&1&1&0&1&0&1&0 \\ 0&0&0&0&0&0&0&0&1&0&1&0&1&0&0&0&1&0&1&0&1&1&1&1&0&0&0&1&1&1&0&0&0&1&0&1 \\ 0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&1&0 \\ 1&1&1&1&1&1&1&1&0&1&0&1&0&1&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0&1&0&0&0&1&0&1 \end{array}\right]$$

$$R_{6,5} = \left[\begin{array}{cccccc|cccccc|cccccc|cccccc|cccccc|cccccc} 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&1&0&0&0&0&0&0&0&1&1&0&0&0&0&1&1&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&1&0&0&0&0&0&0&1&1&0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0&1&0&0&1&1&0&1&0&1&0&0&1&0&0&0&1&0&1&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&1&0&0&1&1&1&0&0&0&0&0 \\ 1&1&1&1&1&1&1&0&0&0&1&1&0&1&0&1&0&0&1&1&1&1&0&1&0&0&1&0&1&0&0&0&0&0&0&1 \end{array}\right]$$

B.4. $3 \times 3 \times 3 \times 3$.

The final example needs to have the column normalization and dangling correction applied. It consists of only 15 nonzeros in the original tensor. Because there is only one, we use the canonical $\underline{\boldsymbol{A}}$.

$$\underline{A}(3,3,3,2) = 1 \qquad \underline{A}(3,3,1,3) = 1 \qquad \underline{A}(2,3,2,3) = 1$$
$$\underline{A}(2,1,1,3) = 1 \qquad \underline{A}(1,1,2,3) = 1 \qquad \underline{A}(3,3,2,3) = 1$$
$$\underline{A}(1,2,1,3) = 1 \qquad \underline{A}(1,2,2,3) = 1 \qquad \underline{A}(3,1,3,3) = 1$$
$$\underline{A}(2,2,1,3) = 1 \qquad \underline{A}(2,2,2,3) = 1 \qquad \underline{A}(3,2,3,3) = 1$$
$$\underline{A}(2,3,1,3) = 1 \qquad \underline{A}(1,3,2,3) = 1 \qquad \underline{A}(3,3,3,3) = 1$$

REFERENCES

A. Anandkumar, D. Hsu, M. Janzamin, and S. Kakade (2013), *When are Overcomplete Topic Models Identifiable? Uniqueness of Tensor Tucker Decompositions with Structured Sparsity*, arXiv:1308.2853, 2013.

M. Benaïm (1997), *Vertex-reinforced random walks and a conjecture of Pemantle*, Ann. Probab., 25, pp. 361–392; also available online from http://projecteuclid.org/euclid.aop/1024404292.

A. Benson, D. F. Gleich, and J. Leskovec (2015), *Tensor spectral clustering for partitioning higher-order network structures*, in Proceedings of the 2015 SIAM International Conference on Data Mining, Vancouver, BC, pp. 118–126; doi:10.1137/1.9781611974010.14.

A. Benson, D. F. Gleich, and L.-H. Lim (2015), *The Spacey Random Walk: A stochastic Process for Higher-Order Data*, In preparation.

P. Boldi, R. Posenato, M. Santini, and S. Vigna (2007), *Traps and pitfalls of topic-biased PageRank*, in Proceedings of the Fourth International Workshop on Algorithms and Models for the Web-Graph WAW2006, pp. 107–116; doi:10.1007/978-3-540-78808-9_10.

K. C. Chang, K. Pearson, and T. Zhang (2008), *Perron-Frobenius theorem for nonnegative tensors*, Commun. Math. Sci. 6, pp. 507–520; doi:10.4310/CMS.2008.v6.n2.a12.

M. T. Chu and S.-J. Wu (2014), *On the Second Dominant Eigenvalue Affecting the Power Method for Transition Probability Tensors*, http://www4.ncsu.edu/~mtchu/Research/Papers/power_tpt_02.pdf.

J. Draisma and J. Kuttler (2014), *Bounded-rank tensors are defined in bounded degree*, Duke Math. J., 163, pp. 35–63; doi:10.1215/00127094-2405170.

V. Freschi (2007), *Protein function prediction from interaction networks using a random walk ranking algorithm*, in Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering (BIBE 2007), pp. 42–48; doi:10.1109/BIBE.2007.4375543.

S. Friedland, S. Gaubert, and L. Han (2013), *Perron-Frobenius theorem for nonnegative multilinear forms and extensions*, Linear Algebra Appl., 438, pp. 738–749; doi:10.1016/j.laa.2011.02.042.

G. Frobenius (1908), *Über matrizen aus positiven elementen*, 1, Sitzungsber. Königl. Preuss. Akad. Wiss., pp. 471–476.

D. F. Gleich (2015), *PageRank beyond the web*, SIAM Rev., 57, pp. 321–363; doi:10.1137/140976649.

D. F. Gleich, A. P. Gray, C. Greif, and T. Lau (2010), *An inner-outer iteration for PageRank*, SIAM J. Sci. Comput., 32, pp. 349–371; doi:10.1137/080727397.

G. H. Golub and C. van Loan (2013), *Matrix Computations*, Johns Hopkins University Press, Baltimore, 2013.

C. Klymko, D. F. Gleich, and T. G. Kolda (2014), *Using triangles to improve community detection in directed networks*, in Proceedings of the ASE BigData Conference; also available online from http://www.ase360.org/handle/123456789/104.

T. G. Kolda and J. R. Mayo (2011), *Shifted power method for computing tensor eigenpairs*, SIAM J. Matrix Anal. Appl., 32, pp. 1095–1124; doi:10.1137/100801482.

A. N. Langville and C. D. Meyer (2006), *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, NJ.

J. Leskovec, D. Huttenlocher, and J. Kleinberg (2010), *Governance in social media: A case study of the Wikipedia promotion process*, in Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, pp. 98–105; http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1485/1841.

W. Li and M. K. Ng (2013), *On the limiting probability distribution of a transition probability tensor*, Linear Multilinear Algebra, pp. 1–24; doi:10.1080/03081087.2013.777436.

X. Li and M. Ng (2015), *Solving sparse non-negative tensor equations: Algorithms and applications*, Front. Math. China, 10, pp. 649–680; doi:10.1007/s11464-014-0377-3.

X. Li, M. Ng, and Y. Ye (2011), *Finding Stationary Probability Vector of a Transition Probability Tensor Arising from a Higher-Order Markov Chain*, Technical report, Hong Kong Baptist University, http://www.math.hkbu.edu.hk/ mng/tensor-research/report1.pdf.

X. Li, M. Ng, and Y. Ye (2014), *Multicomm: Finding community structure in multi-dimensional networks*, IEEE Trans. Knowledge Data Engineering, 26, pp. 929–941; doi:10.1109/TKDE.2013.48.

X. Li, M. K. Ng, and Y. Ye (2012), *HAR: Hub, authority and relevance scores in multi-relational data for query search*, in Proceedings of the 2012 SIAM International Conference on Data Mining, pp. 141–152; doi:10.1137/1.9781611972825.13.

L.-H. Lim (2005), *Singular values and eigenvalues of tensors: A variational approach*, in Proceedings of the 1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, pp. 129–132; doi:10.1109/CAMAP.2005.1574201.

L.-H. Lim (2013), *Tensors and hypermatrices*, in Handbook of Linear Algebra, 2nd ed., Chapman and Hall/CRC, Boca Raton, FL, pp. 231–260; doi:10.1201/b16113-19.

J. L. Morrison, R. Breitling, D. J. Higham, and D. R. Gilbert (2005), *GeneRank: Using search engine technology for the analysis of microarray experiments*, BMC Bioinformatics, 6, p. 233; doi:10.1186/1471-2105-6-233.

M. K.-P. Ng, X. Li, and Y. Ye (2011), *Multirank: Co-ranking for objects and relations in multi-relational data*, in Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1217–1225; doi:10.1145/2020408.2020594.

J. Nie and L. Wang (2014), *Semidefinite Relaxations for Best Rank-1 Tensor Approximations*, SIAM J. Matrix Anal. Appl., 35, pp. 1155–1179; doi:10.1137/130935112.

L. Page, S. Brin, R. Motwani, and T. Winograd (1999), *The PageRank Citation Ranking: Bringing Order to the Web*, Technical report 1999-66, Stanford University, Stanford, CA, http://dbpubs.stanford.edu:8090/pub/1999-66.

N. Parikh and S. Boyd (2014), *Proximal algorithms*, Found. Trends Optim., 1, pp. 127–239; doi:10.1561/2400000003.

R. Pemantle (1992), *Vertex-reinforced random walk*, Probab. Theory Related Fields, 92, pp. 117–136; doi:10.1007/BF01205239.

O. Perron (1907), *Zur theorie der matrices*, Math. Ann., 64, pp. 248–263; doi:10.1007/BF01449896.

L. Qi (2005), *Eigenvalues of a real supersymmetric tensor*, J. Symbolic Comput., 40, pp. 1302–1324; doi:10.1016/j.jsc.2005.05.007.

Y. Rabinovich, A. Sinclair, and A. Wigderson (1992), *Quadratic dynamical systems*, in Proceedings of the 33rd Annual Symposium on Foundations of Computer Science, pp. 304–313; doi:10.1109/SFCS.1992.267761.

R. S. Varga (1962), *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ.

C. Winter, G. Kristiansen, S. Kersting, J. Roy, D. Aust, T. Knösel, P. Rümmele, B. Jahnke, V. Hentrich, F. Rückert, M. Niedergethmann, W. Weichert, M. Bahra, H. J. Schlitt, U. Settmacher, H. Friess, M. Büchler, H.-D. Saeger, M. Schroeder, C. Pilarsky, and R. Grützmann (2012), *Google goes cancer: Improving outcome prediction for cancer patients by network-based ranking of marker genes*, PLoS Comput. Bio., 8, e1002511; doi:10.1371/journal.pcbi.1002511.