ARTICLE IN PRESS

Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Pairwise comparisons in spectral ranking

Ying Tang [1,*], Yinrun Li

Department of Information Science and Technology, Chengdu University of Technology, Chengdu 610059, People's Republic of China

ARTICLE INFO

Article history:
Received 14 May 2015
Received in revised form
11 May 2016
Accepted 4 August 2016
Communicated by H. Yu

Keywords:
Spectral ranking
HITS
Eigenvector
Ordinary differential equation

ABSTRACT

This paper addresses the problem of pairwise comparisons in spectral ranking from an ordinary differential equation view. Given a nonnegative symmetric matrix $\mathbf{A}$ of order n, we provide an $O(1)$ algorithm for single pairwise comparison without computing the exact value of the principal eigenvector of $\mathbf{A}$ if assuming $\mathbf{A}$ and $\mathbf{A}^2$ have been constructed offline, which further leads to an $O(\nu^2)$ algorithm for ranking any subset of size ν, or an $O(kn)$ algorithm for the top k selection. We prove that in ER graphs the correct rate of pairwise comparisons converges to one as n approaches infinity. We also experimentally demonstrate the high correct rate on various artificial and real-world graphs.

1. Introduction

Eigenvector inspired algorithms have found countless applications in machine learning [1–3]. In spectral clustering, eigenvectors of similarity matrices mirror the cluster information of dataset [4–6]. In dimension reduction, they represent the intrinsic coordinates of high-dimensional data that obeys a low-dimensional manifold structure [7]. In particular, the principle eigenvector (**PE**, the eigenvector corresponding to the largest eigenvalue) is especially attractive in spectral ranking [8], e.g., PageRank [9,10], HITS [11,12], or IsoRank [13], where the **PE**s of the matrices describing the link relationship between nodes specify importance score to each node.

There exist plenty of iterative methods for extracting **PE**s. In practice, the most preferred one possibly belongs to the family of power iterations [14,15]. There also exist some other choices, especially for fast approximating the PageRank vector, e.g., the Monte Carlo sampling based on short random walks [16], subgraph based method [17], ϵ-approximation method [5], and linear system solver related method [18]. However, it is important to note that in many applications, we are often interested with the **PE** induced order, instead of its exact value. The traditional methods thus do much more than necessary.

This paper addresses the general problem of pairwise comparisons among the **PE** of nonnegative symmetric matrices without computing the **PE**'s exact value. To the best of our knowledge, this is the first method explicitly avoiding the computation of eigenvectors in spectral ranking, which leads to an $O(1)$ pairwise comparison algorithm developed under an ordinary differential equation (ODE) based framework. We strictly prove the converge of the proposed algorithm in sufficiently large Erdős–Rényi (ER) graphs (a type of random graphs with each edge independently generated), and experimentally show its effectiveness and efficiency on various artificial and real-world networks. In addition, some direct extensions are also discussed, e.g., ranking an subset of **PE** in $O(\nu^2)$ with ν being the size of the subset, or the top k selection in $O(nk)$ with n being the order of matrices, both of which have common interests in practice. The algorithm of this paper makes it possible for spectral ranking algorithms such as HITS to deal with large-scale datasets in real time. Throughout the paper, we indicate matrices and vectors using bold faced letters.

2. Motivation from the 2-D case

This section provides one useful lemma, and uses the 2-D case to show the basic idea behind the algorithm. The starting point of this paper refers to the Oja's rule [19,20], whose typical application lies in Principal Component Analysis, thus naturally bridged to **PE** related problems. As a neural network based eigenvector solver, the Oja's rule tends to be quantified using ODEs. In this paper, we consider the following differential system:

$$\dot{\mathbf{x}}(t) = \mathbf{x}(t)^T\mathbf{x}(t)\mathbf{A}\mathbf{x}(t) - \mathbf{x}(t)^T\mathbf{A}\mathbf{x}(t)\mathbf{x}(t), \tag{1}$$

where $\mathbf{A}$ is a real symmetric matrix and $\dot{\mathbf{x}}(t)$ denotes the derivative of $\mathbf{x}(t)$ w.r.t. t. The above system falls in the class of generalized Oja's rules in [21] for extracting **PE**s. Also, note that reversing the

* Corresponding author.
E-mail addresses: mathtygo@gmail.com (Y. Tang), liyinrun26@163.com (Y. Li).
[1] Postal address: Room 5602, NanYi Building, Chengdu University of Technology, ErXianQiao East 3rd Road No. 1, Chengdu, Sichuan 610059, China.

http://dx.doi.org/10.1016/j.neucom.2016.08.029
0925-2312/

sign of the right side of (1) leads to the Minor Component Analysis algorithm in [22].

Throughout the paper, we will assume that $\mathbf{A} = (a_{ij})$ is nonnegative and has an unrepeated largest eigenvalue. Moreover, we will use **PE** to mean the nonnegative **PE** since spectral ranking algorithms always stand on the nonnegative **PE** (also note from the Perron-Frobenius theorem that any nonnegative **A** is sure to have a nonnegative **PE**). Let PE_i, $x_i(t)$, $\dot{x}_i(t)$, and $\ddot{x}_i(t)$, be the ith element of **PE**, $\mathbf{x}(t)$, $\dot{\mathbf{x}}(t)$, and $\ddot{\mathbf{x}}(t)$ (the second-order derivative of $\mathbf{x}(t)$), respectively.

Although the convergence analysis of (1) or general cases has been carried out in [21] or [22], here we provide a different, but more straightforward convergence analysis for (1), from which we also obtain necessary equations for the next section that are unavailable in [21,22].

Lemma 1. *Given any positive starting point* $\mathbf{x}(0)$ *(thus the projection of* $\mathbf{x}(0)$ *on* ***PE*** *is positive), then except a positive scale, the solution of (1),* $\mathbf{x}(t)$*, converges to* ***PE*** *(see proof in Appendix).*

Next, the 2-D case is used to show what bridges the dynamical behavior of (1) and the purpose of sorting **PE**. Firstly, note that $\mathbf{x}(t)$ has invariant norm since $\frac{\mathrm{d}[\mathbf{x}(t)^T\mathbf{x}(t)]}{\mathrm{d}t} = 2\mathbf{x}(t)^T\dot{\mathbf{x}}(t) = 2\mathbf{x}(t)^T\mathbf{x}(t)\mathbf{x}(t)^T\mathbf{A}\mathbf{x}(t) - 2\mathbf{x}(t)^T\mathbf{A}\mathbf{x}(t)\mathbf{x}(t)^T\mathbf{x}(t) = 0$. Thus, we have $\|\mathbf{x}(t)\| = \|\mathbf{x}(0)\|$ for $\forall\, t$, where $\|\cdot\|$ denotes the Frobenius norm. Thus, $\mathbf{x}(t)$, $t = 0, 1, 2, \ldots$, lives on a circle centered at the original point. If letting $\mathbf{x}(0) = [a, a]^T$, $\forall\, a > 0$, $\mathbf{x}(t)$ will move along the circle clockwise or anticlockwise until meeting **PE**. The reason why $\mathbf{x}(t)$ cannot cross **PE** lies in: (1) $\mathbf{x}(t)$ has zero speed when meeting **PE**; (2) $\mathbf{x}(t)$ cannot change its moving direction from clockwise to anticlockwise (or vice verse) since $\mathbf{x}(t)$, as a solution of an ODE, can not generate intersections. In a word, there are only two possible trajectories for $\mathbf{x}(t)$ starting from $[a, a]^T$, as plotted in Fig. 1(a) and (b). Thus, the task to compare PE_1 with PE_2 can be achieved in a one-step fashion: let $\mathbf{x}(0) = \frac{1}{\sqrt{2}}[1, 1]^T$ and update $\mathbf{x}(0)$ to $\mathbf{x}(1)$ using (1), then without any effort for computing **PE** we can claim: $\mathrm{PE}_1 \gtreqless \mathrm{PE}_2$ if $x_1(1) \gtreqless x_2(1)$. Note that comparing $x_1(1)$ with $x_2(1)$ can be reduced to comparing $\dot{x}_1(0)$ with $\dot{x}_2(0)$ since $x_1(0) = x_2(0)$, and from (1) this can be further simplified to checking which element in $\mathbf{A}\mathbf{x}(0)$ is bigger since both $\mathbf{x}(0)^T\mathbf{x}(0)$ and $\mathbf{x}(0)^T\mathbf{A}\mathbf{x}(0)$ are real numbers, and $x_1(0) = x_2(0)$.

For the general case of $n > 2$, let $\mathbf{P}_{ij}(\mathbf{x}(t))$ be the projection of the curve $\mathbf{x}(t) \in \mathbf{R}^n$ on the plane $\overline{x_i x_j}$, where $\overline{x_i x_j}$ denotes the plane spanned by the ith and jth axes in $\mathbf{R}^n$. The key point supporting the analysis in the 2-D case is that with any $\mathbf{x}(0)$ chosen on the line $x_1 = x_2 > 0$, then $\mathbf{x}(t)$, $t > 0$, is guaranteed not to cross the line $x_1 = x_2$ again. In the case of $n > 2$, one may have expected that $\mathbf{P}_{ij}(\mathbf{x}(t))$ would have the same property on the plane $\overline{x_i x_j}$ as above, such that the analysis in the 2-D case can be directly generalized.

Unfortunately, $\mathbf{P}_{ij}(\mathbf{x}(t))$ may cross the line $x_i = x_j$ more than once even if $\mathbf{x}(0)$ is chosen to be positive and satisfies $x_i(0) = x_j(0)$. However, we will show in the next section that with probability one the sign of $\mathrm{PE}_i - \mathrm{PE}_j$, $i \neq j$, $i, j = 1, \ldots, n$, can be determined in $O(1)$ based on the similar motivation as above when the underlying graph follows the ER model with n sufficiently large. In the same section, we will provide numerical evidence to explain the high correct rate of the proposed algorithm on various types of model based graphs (e.g., scale-free or small-world networks). Implementation details are discussed in Section 4, where two direct extensions based on the pairwise comparisons are also developed for the subset ranking and top k selection. Section 5 provides experimental results and Section 6 concludes the paper.

3. Determining the sign of $(\mathrm{PE}_i - \mathrm{PE}_j)$ in general cases

Fig. 1(c)–(h) plots six typical $\mathbf{P}_{ij}(\mathbf{x}(t))$'s on the $\overline{x_i x_j}$ plane. Intuitively, $\mathbf{P}_{ij}(\mathbf{x}(t))$'s shown in Fig. 2(g) and (h) are more revealing due to the following facts: with higher probability those two curves will not cross the line $x_i = x_j$ again for $t > 0$ since both are tangent to the line $x_i = x_j$ at $t = 0$, and will (locally) move away from the line $x_i = x_j$ soon since both have unequal acceleration along axes. On the contrary, $\mathbf{P}_{ij}(\mathbf{x}(t))$'s in Fig. 1(e) and (f) will not immediately diverge from the line $x_i = x_j$ since $\dot{x}_i(0) = \dot{x}_j(0)$ and $\ddot{x}_i(0) = \ddot{x}_j(0)$, indicating that it is impossible to compare PE_i with PE_j only based on $\mathbf{x}(1)$. As for two curves in Fig. 1(c) and (d), the plotted trajectories are unpredictable in the sense that intuitively we have no confidence to predict whether they will cross the line $x_i = x_j$ at some $t > 0$ or not, thus are not our desired cases either.

Here, we do not depict more trajectories, e.g., the analogues to Fig. 1(g) and (h) but corresponding to $\dot{x}_i(0) = \dot{x}_j(0) < 0$. Nevertheless, it is not hard to summarize the following conditions for positive $\mathbf{x}(0)$ such that $\mathbf{P}_{ij}(\mathbf{x}(t))$ agrees with those shown in Fig. 2 (g), (h), as well as their analogues mentioned as above:

$$x_i(0) = x_j(0); \quad \dot{x}_i(0) = \dot{x}_j(0); \quad \ddot{x}_i(0) \neq \ddot{x}_j(0). \tag{2}$$

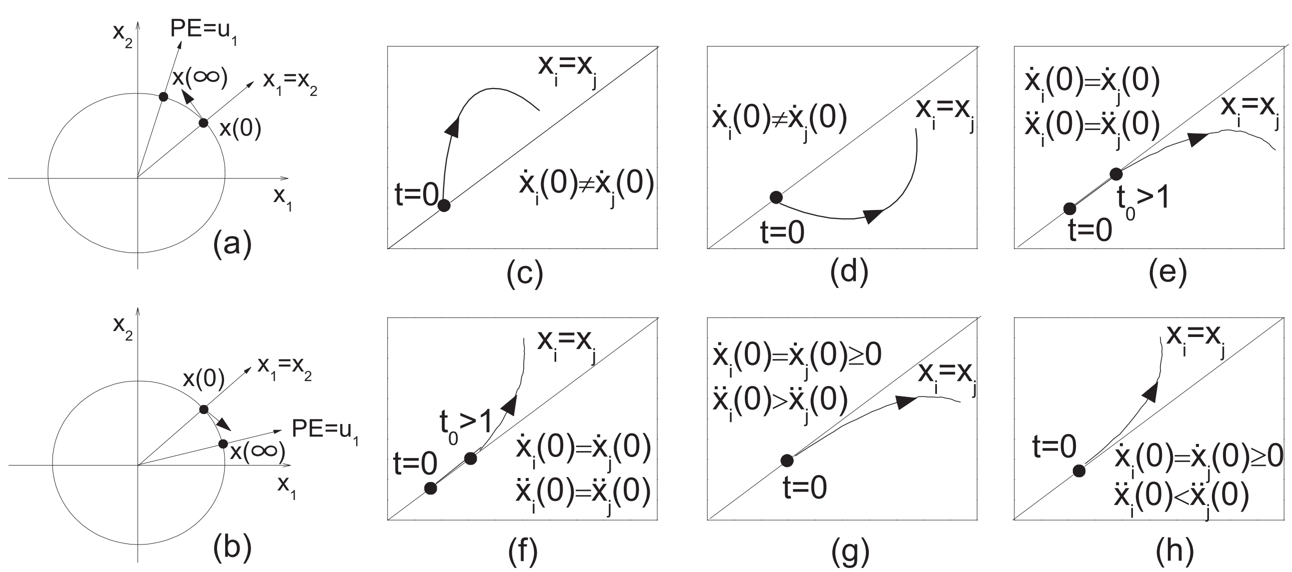

Fig. 1. (a) and (b) Two trajectories of $\mathbf{x}(t)$ in the 2-D case with $x_1(0) = x_2(0) > 0$ move to **PE** along circles. (c)–(h) Typical projections of $\mathbf{x}(t)$ on the $\overline{x_i x_j}$ plane with $x_i(0) = x_j(0)$. The last two are tangent to $x_i = x_j$ with nonequal acceleration at $t = 0$, thus are our desired cases.