

A Note on Rankability:

Current Progress regarding Measure and the SVD

Thomas R. Cameron

June 2018

Abstract

This note is an organization of the author's thoughts on the measure of rankability and the SVD. At this point, our major result is a cheap measure of the rankability based on the eigenvalues of the covariance matrix associated with the adjacency matrix of a set of data.

1 Introduction

In essence, the ranking of a data set is a one-dimensional compression of the data. Therefore, it is reasonable to suspect that the SVD will provide useful information in determining how rankable a set of data is. To gain some intuition for what information the SVD will provide we begin by analyzing the graphs given in Figure 5.1 of [1]. Below is a table with the graphs considered and the singular values of the adjacency matrix associated with each graph.

Dominance Graph	$\{3.5, 1.2, 0.76, 0.59, 0.52, 0.\}$
Dominance with small perturbation	$\{3.4, 1.3, 1., 0.63, 0.52, 0.\}$
Tree	$\{1.4, 1.4, 1.4, 0., 0., 0.\}$
Completely Connected	$\{5., 1., 1., 1., 1., 1.\}$
Cyclic	$\{1., 1., 1., 1., 1., 1.\}$
Completely Decomposable	$\{1., 1., 1., 1., 1., 1.\}$

Figure 1: Singular Values of Extreme Cases

In Section 2 we formalize several observations realized from the table in Figure 1. In Section 3 we note that the singular values of the centered matrix or the eigenvalues of the covariance matrix may be more meaningful. Furthermore, we propose a cheap rankability measure based on these values.

2 Initial Observations

The first observations we make is that all rankable graphs correspond to a singular adjacency matrix. We formalize this observation in the following theorem.

Theorem 2.1. *If the adjacency matrix of a graph is non-singular, then the corresponding graph must have a contradiction, i.e., there exists nodes P_i and P_j such that $P_i \rightarrow P_j$ and $P_j \rightarrow P_i$.*

Proof. The adjacency matrices of interest are $(0, 1)$ matrices with 0's along the main diagonal. Therefore, there are $(n - 1)$ choices to place a 1 in the first row. Note that 1 must be placed on the first row for the adjacency matrix to be non-singular. Given the choice of a 1 in the first row, there are $(n - 2)$ choices to place a 1 in the second row and not have a singular matrix. Resuming this construction we have $(n - n)$ choices to place a 1 in the last row and have a non-singular matrix and the result follows. \square

Remark 2.2. *It follows from Theorem 2.1 that if the smallest singular value of an adjacency matrix is positive, then the corresponding graph can be classified as unrankable. If we are willing to allow ties, then a non-singular adjacency matrix may indicate a cluster or group or equivalence class of objects that should be rated the same. However, this does not readily imply a metric to determine how rankable or unrankable a graph is.*

To gain intuition for what the SVD can tell us about the rankability of a data set, we consider the questions proposed in [1, Section 5.1] Specifically, we consider the minimum number of changes (edges added or removed) that must be made to a graph in order to transform it into a dominance graph. Then, given this minimum number of changes, how many rankings can be produced. To relate this question to the SVD we make use of the following well-known result, a proof of which can be found in [5, Theorem 4.2.15].

Theorem 2.3. *Let $A \in \mathbb{R}^{n \times m}$ with $\text{rank}(A) = r > 0$. Let $A = U \sum V^T$ be the SVD of A , with singular values $\sigma_1 > \sigma_2 > \dots > \sigma_r > 0$. For $k = 1, \dots, r-1$, define $A_k = U \sum_k V^T$, where $\sum_k \in \mathbb{R}^{n \times m}$ is the diagonal matrix $\text{diag}\{\sigma_1, \dots, \sigma_k, 0, \dots, 0\}$. Then $\text{rank}(A_k) = k$, and*

$$\sigma_{k+1} = \|A - A_k\|_2 = \min\{\|A - B\|_2 : \text{rank}(B) \leq k\}. \quad (1)$$

That is, of all matrices of rank k or less, A_k is closest to A .

When changing the graph (adding or removing an edge) we are essentially considering rank 1 perturbations of the adjacency matrix. Therefore, we have the following corollary.

Corollary 2.4. *Let A be the adjacency matrix of a graph with n nodes and let D be the adjacency matrix corresponding to a dominance graph. Denote by $\sigma_1, \dots, \sigma_r > 0$ the singular values of $(A - D)$. Then, the graph of A requires at least r changes in order to translate it into the dominance graph of D .*

Proof. Applying Theorem 2.3 to the matrix $(A - D)$ we have

$$\min\{\|(A - D) - B\|_2 : \text{rank}(B) \leq (r - 1)\} > 0.$$

Thus, $(A - D)$ cannot be approximated by a rank $(r - 1)$ matrix. Since the sum of k rank 1 matrices is at most rank k , it follows that at least r rank 1 perturbations must be added to A in order to get D . \square

Remark 2.5. As a result of Corollary 2.4, we can determine the minimum number of changes, denoted \hat{k} , and the corresponding \hat{p} . Indeed, simply apply (1) to $(A - D)$ for all dominance matrices. Then apply Corollary 2.4 to obtain \hat{k} . Then tabulate all dominance matrices that correspond to \hat{k} to obtain \hat{p} . It follows that we can approximate the rankability measure given in [1] by computing

$$\frac{\hat{k}\hat{p}}{k_{max}p_{max}}. \quad (2)$$

However, it is clear that this method has a factorial complexity and therefore no advantages over the algorithms discussed in [1].

3 Connections to PCA

To obtain a cheaper measure of the rankability of a dataset, we consider connections to PCA (Principal component analysis).

The objective of ranking is to sort objects in a dataset according to some criteria. As a graph problem, we are finding the order or rank of vertices in a (weighted) directed graph. We are interested in defining a measure for the degree of rankability of a dataset.

Let A denote the adjacency matrix associated with some unweighted directed graph. The j th column of A , denoted a_j , shows the objects that are superior to object j with respect to some criterion. Therefore, the column mean

$$\mu = \frac{1}{n} \sum_{j=1}^n a_j \quad (3)$$

is a column vector whose i th component is the percentage of objects that object i is superior to with respect to some criterion.

Remark 3.1. The mean is a simple measurement that delivers a good amount of information. The mean for each graph given in Figure 5.1 of [1] is below.

Dominance Graph	{0.83, 0.67, 0.5, 0.33, 0.17, 0.}
Dominance with small perturbation	{0.83, 0.5, 0.67, 0.33, 0.17, 0.}
Tree	{0.29, 0.29, 0.29, 0., 0., 0.}
Completely Connected	{0.83, 0.83, 0.83, 0.83, 0.83, 0.83}
Cyclic	{0.17, 0.17, 0.17, 0.17, 0.17, 0.17}
Completely Decomposable	{0.17, 0.17, 0.17, 0.17, 0.17, 0.17}

Comparing the magnitude of each entry in the mean implies a potential ranking for the data. In fact, for these simple examples, the implied ranking is exactly what we would expect: $[1, 2, 3, 4, 5, 6]$ for the Dominance

Graph and [1, 3, 2, 4, 5, 6] for the Dominance graph with small perturbation. The mean even tells us that the Completely Connected, Cyclic, and Completely Decomposable graphs are made up of objects with equal rank. Perhaps the mean can be used as a starting point to get more sophisticated measures of rank and rankability?

We now consider the centered and covariance matrix associated with our adjacency matrix A , which are defined as

$$X = A - \mu e^T \quad (4)$$

and

$$C_X = \frac{1}{n} X X^T, \quad (5)$$

respectively.

In order to make comparisons to PCA, we consider X to be a data matrix with zero mean. In particular, we view each row of X as a zero mean measurement of a particular type. In this case, the i th row is a zero mean measurement of which objects are inferior to object i .

Furthermore, the (i, j) th element of C_X is the dot product between the vector of the i th measurement type and the vector of the j th measurement type. We summarize several properties of C_X from [4] below:

- C_X is a $n \times n$ symmetric matrix.
- The diagonal entries of C_X are the *variance* of particular measurement types.
- The off-diagonal entries of C_X are the *covariance* between measurement types.

Remark 3.2. *It is worth noting that the covariance measurements in C_X can pick up inconsistencies in the data. For instance, the covariance matrix of the Dominance with small perturbation is given below.*

$$\begin{pmatrix} 0.138889 & 0.0833333 & -0.0555556 & 0.0555556 & 0.0277778 & 0. \\ 0.0833333 & 0.25 & 0.166667 & 0.166667 & 0.0833333 & 0. \\ -0.0555556 & 0.166667 & 0.222222 & 0.111111 & 0.0555556 & 0. \\ 0.0555556 & 0.166667 & 0.111111 & 0.222222 & 0.111111 & 0. \\ 0.0277778 & 0.0833333 & 0.0555556 & 0.111111 & 0.138889 & 0. \\ 0. & 0. & 0. & 0. & 0. & 0. \end{pmatrix}$$

Note the negative entry in the (1, 3) and (3, 1) position of the matrix.

The goal of PCA is to minimize redundancy, measured by the magnitude of the covariance, and maximize the signal, measured by the variance. Since C_X is symmetric, the Spectral Theorem implies that C_X is orthogonally diagonalizable, see [3, Section 7.1] or [2, Section 6.6]. In addition, C_X is positive semidefinite, and it follows that we can diagonalize C_X in such a way that its diagonal entries occur in decreasing order. Thus, there

exists a change of basis $PX = Y$ such that

$$\begin{aligned} C_Y &= C_{PX} \\ &= \frac{1}{n}(PX)(PX)^T \\ &= \frac{1}{n}PXX^TP^T \\ &= PC_XP^T \end{aligned}$$

is a diagonal matrix with non-negative diagonal entries that occur in decreasing order.

The columns of P correspond to eigenvectors of C_X and the diagonal entries of C_Y correspond to eigenvalues of C_X . The eigenvectors correspond to directions of maximal variance. However, I am not sure of the connection to ranking at this time. The eigenvalues, on the other hand, have a clear connection to the rankability of the original dataset.

Remark 3.3. *To have a clearly defined (unique) ranking in a dataset, the variance should be mostly in one direction. Therefore, when looking at the diagonal entries of C_Y one would expect to see the first eigenvalue as (much) larger than the second. This is how we define the rankability of a data set:*

$$r = \frac{\lambda_1}{\lambda_2},$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are the eigenvalues of C_X .

Actually, it is not good numerical practice to form the “normalized” matrix XX^T and then compute its eigenvalues [5, Section 5.8] Therefore, the related rankability measure may be preferred:

$$r = \frac{\sigma_1}{\sigma_2},$$

where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ are the singular values of X .

Pseudo code for the rankability measure noted in the previous remark is given below.

Algorithm 1 Rankability Measure

Require: adjacency matrix A

form the column mean: $\mu = \frac{1}{n} \sum_{j=1}^n a_j$

form the centered matrix: $X = A - \mu e^T$

compute the singular values of X : $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$

return $\frac{\sigma_1}{\sigma_2}$

The rankability measure for each graph given in Figure 5.1 of [1] is below.

$$\left(\begin{array}{ll} \text{Dominance Graph} & 1.93185 \\ \text{Dominance with small perturbation} & 1.64147 \\ \text{Tree} & 1. \\ \text{Completely Connected} & 1. \\ \text{Cyclic} & 1. \\ \text{Completely Decomposable} & 1. \end{array} \right)$$

Theorem 3.4. *Let A denote any adjacency matrix. Then the rankability measure of A as defined in Remark 3.3 is invariant under permutation similarity.*

Proof. Let P be any permutation matrix and denote by X the centered matrix of A and by X_P the centered matrix of $P^T A P$. Then it follows that

$$\begin{aligned} X_P &= P^T A P \left(I - ee^T/n \right) \\ &= P^T A \left(P - P e e^T / n \right) \\ &= P^T A \left(I - ee^T/n \right) P \\ &= P X P^T. \end{aligned}$$

Therefore, the singular values of X_P and X are the same. \square

The importance of Theorem 3.4 is clear as it guarantees that the rankability of a dataset cannot change by reordering its objects. Perhaps under certain constraints this ranking is maximized for dominance graphs?

References

- [1] P. E. ANDERSON, T. P. CHARTIER, AND A. N. LANGVILLE, *The rankability of data*, Working Paper, (2018).
- [2] S. H. FRIEDBERG, A. J. INSEL, AND L. E. SPENCE, *Linear Algebra*, Pearson Education, Upper Saddle River, NJ, 4th ed., 2003.
- [3] D. C. LAY, S. LAY, AND J. McDONALD, *Linear Algebra and its applications*, Pearson Education, New York, NY, 5th ed., 2016.
- [4] J. SHLENS, *A tutorial on principal component analysis*. arXiv:1404.1100 [cs.LG], 2014.
- [5] D. S. WATKINS, *Fundamentals of Matrix Computations*, John Wiley and Sons, Hoboken, NJ, 3rd ed., 2010.