

Ordering graphs in a normalized singular value measure

Charles R. Johnson, Brian Lins, Victor Luo and Sean Meehan

Ordering graphs in a normalized singular value measure

Charles R. Johnson, Brian Lins, Victor Luo and Sean Meehan

(Communicated by Joshua Cooper)

A proposed measure of network cohesion for graphs arising from interrelated economic activity is studied. The measure is the largest singular value of a rowstochastic matrix derived from the adjacency matrix. It is shown here that among graphs on n vertices, the star universally gives the (strictly) largest measure. Other universal comparisons among graphs with larger measures are difficult to make, but one is conjectured, and a selection of empirical evidence is given.

1. Introduction

In [\[Cavalcanti et al. 2012;](#page-12-0) [2013\]](#page-12-1) the authors studied the role of network "cohesion" in the equilibration of economic or other activity among agents whose interaction is governed by a particular graph. An example is the one in which adjacency is the bordering relationship among countries. Giannitsarou and Johnson (personal communication, 2011) proposed a particular numerical measure of network cohesion and raised the question of which graph on n vertices resulted in the highest measure. That measure may be described as follows. Let A be the adjacency matrix of a graph G, define $B = A + I$, and let D be the positive diagonal matrix whose diagonal entries are the row sums of B. If $R = D^{-1}B$, then R is row-stochastic, and $\sigma(G)$, the measure of cohesion, is the largest singular value of R. Recall that the singular values of R are the square roots of the eigenvalues of RR^T . Another application where the matrix R has appeared is in [\[Echenique and Fryer 2007\]](#page-12-2), where it is referred to as the matrix of social interactions.

Here, we show that, for any n, $\sigma(G)$ is maximized by the star S_n . The measure $\sigma(G)$ is 1 if and only if G is regular, and 1 is the smallest possible value [\(Section 2,](#page-3-0) [Proposition 1\)](#page-3-1). Using our methods, it is difficult to determine, in advance, the relative position in this order of other graphs. Indeed, for graphs naturally defined on any number of vertices, the position often changes with n . However, we do

Keywords: graph singular values, graph measure, network cohesion.

This work was partially supported by NSF grant DMS-0751964.

MSC2010: 05C40, 15A18.

conjecture that the star plus an edge that connects two of the pendant vertices is next after the star, based, in part, on empirical evidence. After that, however, there may be no universal third place independent of n .

In the next section we mention known results that we use, and develop some new ideas that are important for our observations. In particular, the entries of RR^T have a nice and useful interpretation. Then, we show the star yields the highest measure by showing that a lower bound for the square of its largest singular value beats an upper bound for that of any other graph. Finally, in an [Appendix,](#page-9-0) we give a selection of empirical information of interest [\(Table 1](#page-9-1) and Figures [2,](#page-10-0) [3,](#page-10-1) [4,](#page-11-0) [5\)](#page-11-1).

2. Background and tools

Given a graph G on *n* vertices, let A be the adjacency matrix of G. Unless otherwise noted, our notation follows [\[West 1996\]](#page-12-3). Let $R = D^{-1}(A + I)$, where D is the unique positive diagonal matrix such that R is row-stochastic. Let $\lambda(G)$ denote the maximum eigenvalue of RR^T , and note that $\sigma(G) = \sqrt{\lambda(G)}$.

Proposition 1. *For any connected graph* G *on n vertices*, $\sigma(G) \geq 1$ *, and* $\sigma(G) = 1$ *if and only if* G *is regular.*

Proof. Note that G is regular if and only if R is doubly stochastic. If R is doubly stochastic, then it is a convex combination of permutation matrices by Birkhoff's theorem [\[Horn and Johnson 1990,](#page-12-4) Theorem 8.1.7], and therefore the operator norm of R, which equals the maximum singular value, is 1. Let $e \in \mathbb{R}^n$ denote the vector with 1 in every entry. By the Cauchy–Schwarz inequality, $||e^T R||_2 \ge \frac{1}{\pi}$ the vector with 1 in every entry. By the Cauchy–Schwarz inequality, $||e^T R||_2 \ge$
 $\langle e^T R, e/\sqrt{n} \rangle = \sqrt{n} = ||e^T||_2$, with equality if and only if $e^T R$ is a multiple of e^T . Therefore, when R is row-stochastic but not doubly stochastic, the operator norm of R is strictly greater than one. It follows that $\sigma(G) > 1$ when G is not regular. \Box

Note that $D = diag({d_i + 1}_{i \in 1,...,n})$, where d_i is the degree of vertex i in G. Let $C = (A + I)(A + I)^{T}$. The (i, j) entry of C, which we denote by c_{ij} , is the number of vertices that are adjacent to both vertex i and vertex j , with the convention that two adjacent vertices are common neighbors of each other, that is, $c_{ij} = |N[i] \cap N[j]|$. In particular $c_{ii} = d_i + 1$. Thus the entries of RR^T are

$$
r_{ij} = \frac{c_{ij}}{(d_i + 1)(d_j + 1)}.
$$
 (1)

Lemma 1. Let RR^T be defined as above and assume that $n > 2$. When $i \neq j$, *the largest possible values of* r_{ij} *are* $\frac{1}{3}$ *and* $\frac{1}{4}$ *. If* $r_{ij} = \frac{1}{3}$ *for some* $i \neq j$ *, then* $d_i = d_j = 2$ *with* $c_{ij} = 3$ *or* $\{d_i, d_j\} = \{1, 2\}$ *with* $c_{ij} = 2$ (*see* [Figure 1\)](#page-4-0)*.*

Figure 1. Possible adjacency graphs when $r_{ij} = \frac{1}{3}$.

Proof. We may assume that $d_j \geq d_i$. Note that $c_{ij} \leq d_i + 1$; thus $r_{ij} \leq 1/(d_j + 1)$. If $r_{ij} > \frac{1}{4}$ $\frac{1}{4}$, then $d_j = 1$ or $d_j = 2$. In the former case, $d_i = d_j = 1$, which can only happen if $n = 2$, since G is assumed to be connected. In the latter case, $d_i = 1$ or $d_i = 2$ while $d_j = 2$. If $d_i = 1$ and $d_j = 2$, then $r_{ij} = c_{ij} / 6 \in \{0, \frac{1}{6}\}$ $\frac{1}{6}, \frac{1}{3}$ $\frac{1}{3}$, depending on the value of c_{ij} . If $d_i = d_j = 2$, then $r_{ij} = c_{ij} / 9 \in \{0, \frac{1}{9}\}$ $\frac{1}{9}$, $\frac{2}{9}$ $\frac{2}{9}, \frac{1}{3}$ $\frac{1}{3}$ \Box

Suppose that G is a connected graph with n vertices such that every vertex has degree 1 (is pendant) except for a single central vertex with degree $n - 1$. We refer to any such graph as a *star* on *n* vertices, denoted by S_n . We may assume without loss of generality that vertex 1 is the central vertex of the star. Using [\(1\),](#page-3-2) we see that, for the star,

$$
RRT = \begin{bmatrix} \frac{g}{n} & \frac{1}{n} & \cdots & \cdots & \frac{1}{n} \\ \frac{g}{n} & \frac{1}{2} & \frac{1}{4} & \cdots & \frac{1}{4} \\ \vdots & \frac{1}{4} & \frac{1}{2} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \frac{1}{4} \\ \frac{g}{n} & \frac{1}{4} & \cdots & \frac{1}{4} & \frac{1}{2} \end{bmatrix}
$$

:

Note that $RR^T - \frac{1}{4}$ $\frac{1}{4}I$ is of rank 2, and therefore it is possible to explicitly calculate the characteristic polynomial of this matrix. Recall [\[Horn and Johnson 1990,](#page-12-4) Theorem 1.2.12] that the characteristic polynomial of a matrix is given by

$$
p(t) = tn - E1tn-1 + E2tn-2 + \dots + (-1)n En,
$$

where each E_k is the sum of the k-by-k principal minors of the matrix. For $RR^T - \frac{1}{4}$ $\frac{1}{4}I$, only the 1-by-1 and 2-by-2 principal minors can be nonzero. Thus the characteristic equation for $RR^T - \frac{1}{4}$ $\frac{1}{4}I$ is

$$
p(t) = tn - \left(\frac{1}{n} + \frac{1}{4}(n-1)\right)t^{n-1} + \left(\frac{n-4}{4n^2}\right)(n-1)t^{n-2}.
$$

The nonzero roots of this polynomial are

$$
\frac{\frac{1}{4}(n-1) + \frac{1}{n} \pm \sqrt{\left(\frac{1}{4}(n-1) + \frac{1}{n}\right)^2 - \frac{n-4}{n^2}}}{2},
$$

and therefore the maximum eigenvalue of RR^T for the star on *n* vertices is

$$
\lambda(S_n) = \frac{1}{4} + \frac{\frac{1}{4}(n-1) + \frac{1}{n} + \sqrt{\left(\frac{1}{4}(n-1) + \frac{1}{n}\right)^2 - \frac{n-4}{n^2}}}{2}.
$$

3. The star is a maximum

We seek to estimate the maximum eigenvalue $\lambda(G)$ of RR^T . The row sums of RR^{T} place constraints on $\lambda(G)$. By [\[Horn and Johnson 1990,](#page-12-4) Theorem 8.1.22],

$$
\min_{i} \left\{ \sum_{j} r_{ij} \right\} \leq \lambda(G) \leq \max_{i} \left\{ \sum_{j} r_{ij} \right\}.
$$
 (2)

For the star on *n* vertices, $RR^T - \frac{1}{4}$ $\frac{1}{4}I$ contains an $(n-1)$ -by- $(n-1)$ submatrix with all entries equal to $\frac{1}{4}$. It follows from the inclusion principle [\[Horn and Johnson](#page-12-4) [1990,](#page-12-4) Theorem 4.3.15] that $\lambda(S_n) \geq \frac{1}{4}$ $\frac{1}{4}n$. Combining this with the maximum row sum, we see that $\frac{1}{4}n \leq \lambda(S_n) \leq \frac{1}{4}$ $\frac{1}{4}n + \frac{1}{n}$.

The following observation is an immediate consequence of [Lemma 1:](#page-3-3)

Lemma 2. *Suppose that* $n > 2$ *, and consider the rows of RR^T. If row i has* diagonal entry $r_{ii} = \frac{1}{k}$ with $k \ge 4$ and no off-diagonal entry equals $\frac{1}{3}$, then the sum *of the entries in row i is at most* $\frac{1}{k} + \frac{1}{4}(n-1)$.

Let us make a basic observation which we will use in the proofs of several subsequent propositions.

Lemma 3. Let $c > 0$. The function $x \mapsto 1/(x+1) + cx$ is concave up for all $x > 0$, *and therefore its maximum on any interval* $[a, b] \subset (0, \infty)$ *is attained at one of the endpoints.*

The following observations about the row sums of RR^T cover the cases when [Lemma 2](#page-5-0) does not apply:

Lemma 4. Suppose that $n > 3$. If row i has diagonal entry $r_{ii} = \frac{1}{2}$ and G is not *the star, then the sum of the entries in row i is at most* $-\frac{1}{6} + \frac{1}{n} + \frac{1}{4}n$.

Proof. Since $r_{ii} = \frac{1}{2}$, $d_i = 1$. Let j denote the vertex adjacent to i. The sum of the entries in row i is then

$$
r_{ii} + r_{ij} + \sum_{m \neq i,j} r_{im} = \frac{1}{2} + \frac{1}{d_j + 1} + \sum_{m \neq i,j} \frac{c_{im}}{d_m + 1}.
$$

Note that $c_{im}=1$ if there is an edge connecting vertex j to vertex m and $c_{im}=0$ otherwise. Therefore we have the following upper bound for the sum of entries in row i :

$$
r_{ii} + r_{ij} + \sum_{m \neq i,j} r_{im} \leq \frac{1}{2} + \frac{1}{d_j + 1} + \sum_{m \in N(j)} \frac{1}{2(d_m + 1)}.
$$

If $d_i = n - 1$, and the graph is not the star, then there must be at least two vertices m_1 and m_2 such that $d_{m_1} > 1$ and $d_{m_2} > 1$. In this case an upper bound for the sum of the entries in row i is

$$
\frac{1}{2} + \frac{1}{n} + \frac{1}{4}(d_j - 3) + 2\frac{1}{6} = -\frac{1}{6} + \frac{1}{n} + \frac{1}{4}n.
$$

If $d_i < n-1$, then

$$
r_{ii} + r_{ij} + \sum_{m \neq i,j} r_{im} \leq \frac{1}{2} + \frac{1}{d_j + 1} + \sum_{m \in N(j)} \frac{1}{2(d_m + 1)}
$$

$$
\leq \frac{1}{2} + \frac{1}{d_j + 1} + \frac{1}{4}(d_j - 1).
$$

Since $2 \le d_i < n-1$, we use [Lemma 3](#page-5-1) to see that an upper bound for this expression is

$$
\max\Big\{\frac{13}{12}, -\frac{1}{4} + \frac{1}{n-1} + \frac{1}{4}n\Big\}.
$$

For $n > 3$,

$$
\max\left\{\frac{13}{12}, -\frac{1}{4} + \frac{1}{n-1} + \frac{1}{4}n\right\} \le -\frac{1}{6} + \frac{1}{n} + \frac{1}{4}n.
$$

Lemma 5. Suppose $n > 3$. If row i has diagonal entry $r_{ii} = \frac{1}{3}$, then the sum of the *entries in row i is less than* $-\frac{1}{6} + \frac{1}{n} + \frac{1}{4}n$.

Proof. Since $r_{ii} = \frac{1}{3}$, $d_i = 2$. Let j and k denote the two vertices adjacent to i. *Case I*. If there is an edge connecting j and k, then $c_{ij} = c_{ik} = 3$. If $m \neq i$ is a vertex adjacent to both j and k , then

$$
r_{im} = \frac{2}{3(d_m + 1)} \le \frac{2}{9}.
$$

If *m* is only adjacent to one of j or k , then

$$
r_{im} = \frac{1}{3(d_m+1)} \leq \frac{1}{6}.
$$

Let $d = \max\{d_i, d_k\}$ and $D = \max\{d_i, d_k\}$. There are at most $d - 2$ vertices other than *i* that are common neighbors of both *j* and *k*, and there are at most $D - d$ remaining vertices other than i that could be adjacent to exactly one of j or k . Therefore the sum of the entries in row i is at most

$$
r_{ii} + r_{ij} + r_{ik} + \frac{2}{9}(d-2) + \frac{1}{6}(D-d) \le -\frac{1}{9} + \frac{1}{(d+1)} + \frac{1}{(D+1)} + \frac{1}{18}d + \frac{1}{6}D.
$$

In this case, $2 \le d \le D \le n-1$. By [Lemma 3,](#page-5-1) it follows that the possible maximum values in the expression above occur when either $d = D = 2$, or $d = 2$, $D = n - 1$, or $d = D = n - 1$. The corresponding upper bounds on the row sum are

1,
$$
\frac{1}{6} + \frac{1}{n} + \frac{1}{6}n
$$
, $-\frac{1}{3} + \frac{2}{n} + \frac{2}{9}n$.

Each of these bounds is less than $-\frac{1}{6} + \frac{1}{n} + \frac{1}{4}n$ for all $n > 3$.

Case II. If there is no edge connecting j with k, then $c_{ij} = c_{ik} = 2$. If $m \neq i$ is a vertex adjacent to both i and k , then

$$
r_{im}=\frac{2}{3(d_m+1)}\leq \frac{2}{9}.
$$

If *m* is only adjacent to one of *j* or k , then

$$
r_{im} = \frac{1}{3(d_m+1)} \leq \frac{1}{6}.
$$

Let $d = \max\{d_i, d_k\}$ and $D = \max\{d_i, d_k\}$. There are at most $d - 1$ vertices other than *i* that are common neighbors of both *j* and *k*, and there are at most $D - d$ remaining vertices other than i that could be adjacent to exactly one of j or k . Therefore the sum of the entries in row i is at most

$$
r_{ii} + r_{ij} + r_{ik} + \frac{2}{9}(d-1) + \frac{1}{6}(D-d) \le \frac{1}{9} + \frac{2}{3(d+1)} + \frac{2}{3(D+1)} + \frac{1}{18}d + \frac{1}{6}D.
$$

We know that $1 \le d \le D \le n - 2$. By [Lemma 3,](#page-5-1) it follows that the possible maximum values in the expression above occur when either $d = D = 1$, or $d = 1$, $D = n - 2$, or $d = D = n - 2$. The corresponding upper bounds on the row sum are

1,
$$
\frac{1}{6} + \frac{2}{3(n-1)} + \frac{1}{6}n
$$
, $\frac{-1}{3} + \frac{4}{3(n-1)} + \frac{2}{9}n$.

Once again, each of these bounds is less than $-\frac{1}{6} + \frac{1}{n} + \frac{1}{4}n$ for all $n > 3$.

Lemma 6. Suppose $n > 3$. If row i contains an off-diagonal entry $r_{ij} = \frac{1}{3}$, then the *sum of the entries in row i is at most* $-\frac{1}{6} + \frac{1}{n} + \frac{1}{4}n$.

Proof. There are three possible cases, depending on the possible degrees of i and j given by [Lemma 1.](#page-3-3)

Case I. If $d_i = 1$ and $d_j = 2$, then there is only one other vertex, aside from i and j, that can share a common neighbor with i. Call that vertex k . The sum of entries in row i is

$$
r_{ii} + r_{ij} + r_{ik} = \frac{1}{2} + \frac{1}{3} + \frac{1}{2(d_k + 1)} \le \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \frac{13}{12},
$$

which is less than or equal to $-\frac{1}{6} + \frac{1}{n} + \frac{1}{4}n$ for all $n > 3$ (equality occurs only when $n = 4$).

Case II. If $d_i = 2$ and $d_j = 1$, then [Lemma 5](#page-6-0) implies that the sum of the entries in row *i* is less than $-\frac{1}{6} + \frac{1}{n} + \frac{1}{4}n$.

Case III. If $d_i = d_j = 2$, then by [Lemma 1,](#page-3-3) $c_{ij} = 3$. Let k denote the third common neighbor of i and j . The sum of the entries in row i is then

$$
r_{ii} + r_{ij} + r_{ik} + \sum_{m \neq i, j, k} r_{im} = \frac{1}{3} + \frac{1}{3} + \frac{1}{d_k + 1} + \sum_{m \neq i, j, k} \frac{1}{3(d_m + 1)}
$$

$$
\leq \frac{2}{3} + \frac{1}{d_k + 1} + \frac{1}{6}(d_k - 2)
$$

$$
\leq \frac{2}{3} + \frac{1}{n - 1} + \frac{1}{6}(n - 4)
$$

$$
= \frac{1}{6}n + \frac{1}{n - 1}.
$$

This upper bound is less than $-\frac{1}{6} + \frac{1}{n} + \frac{1}{4}n$ for all $n > 3$.

Theorem 1. *Of all connected graphs on* n *vertices*, *the star attains the maximum value of* σ .

Proof. Suppose that G is not S_n . The contents of Lemmas [2,](#page-5-0) [4,](#page-5-2) [5,](#page-6-0) and [6](#page-8-0) show that the maximum row sum of RR^T is less than or equal to $-\frac{1}{6} + \frac{1}{n} + \frac{1}{4}n$. If $n > 6$, then this upper bound is less than $\frac{1}{4}n$, and, by the comment after [\(2\),](#page-5-3) we conclude that $\lambda(G) < \lambda(S_n)$ and therefore $\sigma(G) < \sigma(S_n)$. When $3 < n \le 6$, we can verify by explicit computation that $-\frac{1}{6} + \frac{1}{n} + \frac{1}{4}n < \lambda(S_n)$. When $n = 3$, the theorem can be verified directly since there are only two connected graphs on 3 vertices. \Box

Appendix

Here we present the values of $\sigma(G)$ for every connected graph up to 6 vertices. The graphs are given in graph6 string format [McKay 1981; 2005], and the values of $\sigma(G)$ are given to 8 decimal places. The values for the stars are given in boldface.

Table 1. The value of $\sigma(G)$ (to 8 decimal places) for every connected graph with at most 6 vertices, with the values of stars given in boldface.

Figure 2. The graphs with the four highest singular values for $n = 5$.

Figure 3. The graphs with the four highest singular values for $n = 6$.

Figure 4. The graphs with the four highest singular values for $n = 7$.

Figure 5. The graphs with the four highest singular values for $n = 8$.

References

- [Cavalcanti and Giannitsarou 2013] T. V. V. Cavalcanti and C. Giannitsarou, ["Growth and human](http://www.econ.cam.ac.uk/faculty/giannitsarou/net_growth.pdf) [capital: a network approach",](http://www.econ.cam.ac.uk/faculty/giannitsarou/net_growth.pdf) April 23, 2013, available at http://www.econ.cam.ac.uk/faculty/ giannitsarou/net_growth.pdf. Submitted. Formerly titled "Network structure and human capital dynamics".
- [Cavalcanti et al. 2012] T. V. V. Cavalcanti, C. Giannitsarou, and C. R. Johnson, ["Network cohesion",](http://www.econ.cam.ac.uk/faculty/giannitsarou/cohesion.pdf) 2012, available at http://www.econ.cam.ac.uk/faculty/giannitsarou/cohesion.pdf.

[Echenique and Fryer 2007] F. Echenique and R. G. Fryer, ["A measure of segregation based on social](http://dx.doi.org/10.1162/qjec.122.2.441) [interactions",](http://dx.doi.org/10.1162/qjec.122.2.441) *Q. J. Econ.* 122:2 (2007), 441–485.

[Horn and Johnson 1990] R. A. Horn and C. R. Johnson, *[Matrix analysis](http://dx.doi.org/10.1017/CBO9780511810817)*, Cambridge University Press, 1990. Corrected reprint of the 1985 original. [MR 91i:15001](http://msp.org/idx/mr/91i:15001) [Zbl 0704.15002](http://msp.org/idx/zbl/0704.15002)

[McKay 1981] B. D. McKay, ["Practical graph isomorphism",](http://cs.anu.edu.au/~bdm/papers/pgi.pdf) pp. 45–87 in *Proceedings of the Tenth Manitoba Conference on Numerical Mathematics and Computing* (Winnipeg, MB, 1980), edited by D. S. Meek and H. C. Williams, Congr. Numer. 30, Utilitas Mathematical Publishing, Winnipeg, MB, 1981. [MR 83e:05061](http://msp.org/idx/mr/83e:05061) [Zbl 0521.05061](http://msp.org/idx/zbl/0521.05061)

[McKay 2005] B. D. McKay, ["Description of graph6 and sparse6 encodings",](http://cs.anu.edu.au/~bdm/data/formats.txt) 2005, available at http://cs.anu.edu.au/~ bdm/data/formats.txt.

[West 1996] D. B. West, *Introduction to graph theory*, Prentice Hall, Upper Saddle River, NJ, 1996. [MR 96i:05001](http://msp.org/idx/mr/96i:05001) [Zbl 0845.05001](http://msp.org/idx/zbl/0845.05001)

involve

msp.org/involve

EDITORS

MANAGING EDITOR

Kenneth S. Berenhaut, Wake Forest University, USA, berenhks@wfu.edu

BOARD OF EDITORS

PRODUCTION

Silvio Levy, Scientific Editor

See inside back cover or msp.org/involve for submission instructions. The subscription price for 2015 is US \$140/year for the electronic version, and \$190/year (+\$35, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

<http://msp.org/>

© 2015 Mathematical Sciences Publishers

involve vol. 8

