Ensuring Strong Dominance
of the Leading Eigenvalues for Cluster Ensembles

H. T. Kung

School of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02478

Bruce W. Suter

Air Force Research Laboratory/RITB
525 Brooks Road
Rome, NY

Abstract—**Spectral analysis is a popular mathematical tool in analyzing a variety of network and distributed systems. For a special class of networks, called cluster ensembles, which are made of interconnected clusters, we can characterize those which exhibit strong dominance of the leading eigenvalues in terms of the cluster structure. For such systems, only these leading eigenvalues and their corresponding eigenvectors will need to be examined in studying important properties of the underlying system. This paper establishes several bounds on eigenvalue separation ratios in terms of the number of clusters, their sizes and cluster interconnection topologies.**

Keywords-eigenvalue;eigenvector; spectral analysis; network; cluster

I.   INTRODUCTION

Spectral analysis is a widely used mathematical tool in analyzing network and distributed systems (see, e.g., [2]). Given such a system, e.g., a sensor-target measurement system (see, e.g., [5]), it's often sufficient to examine only those dominating eigenvectors for which the corresponding eigenvalues are much larger than the rest of eigenvalues. That is, these dominating eigenvectors alone can reveal important properties of the system, e.g., its clusters (see, e.g., [5]). From the resulting clustering structure one can detect, for example, malfunctioning or malicious sensor nodes in sensor network applications [5].

More precisely, for a given system, let A denote the adjacency matrix of the graph reflecting the underlying network topology. We are interested in the eigen structure of $A^T A$, and in particular, the dominance of the leading eigenvalues of $A^T A$.

In this paper we consider a class of network-based systems, called *cluster ensembles*, which are collections of clusters connected via some *cluster interconnection network* as depicted in Figure 1. We show that for some cluster ensembles with regular structures, such as alpha-beta cluster ensembles defined in Section III, we can characterize those ensembles which exhibit strong dominance of the leading eigenvalues in terms of their underlying network topologies. These characterizations can provide guidelines in designing cluster ensembles where such strong dominance is desired.
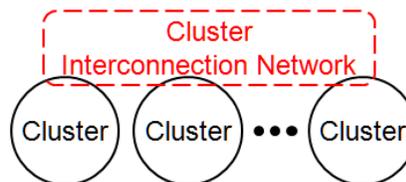

Figure 1. Cluster ensemble

II.   AN EXAMPLE OF CLUSTER ENSEMBLE

To provide a concrete example, we consider a cluster ensemble composed of three clusters interconnected via a clique network, as shown in Figure 2. We call those cluster nodes which are part of the cluster interconnection network *hub nodes*. Thus, in this example, nodes 10, 11 and 12 are hub nodes.
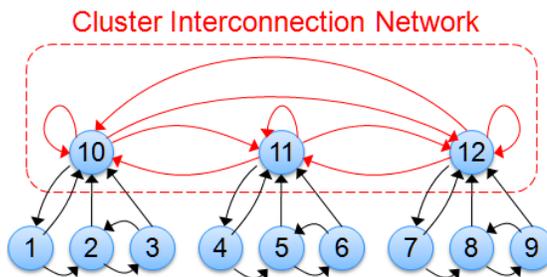

Figure 2. A three-cluster ensemble, where the cluster interconnection network is a clique

A more readable drawing of the same cluster ensemble is depicted Figure 3, where the cluster interconnection network is placed in the center.

III.   ALPHA-BETA CLUSTER ENSEMBLE

We focus on a special class of cluster ensembles, called *alpha-beta cluster ensembles*, which satisfy the following properties:

P1.   The cluster ensemble is composed of α identical clusters.

P2. Each cluster contains one hub node and β non-hub nodes. This means that the size of each cluster is β + 1.

P3. In the adjacency matrix C of each cluster, columns of non-hub nodes are orthogonal to each other. It is possible to relax this orthogonality condition so $C^T C$ goes from a diagonal matrix to a diagonally dominant one as described in [4].

The alpha-beta clusters represent a natural model of connecting multiple clusters. Figure 3 depicts an instance of an alpha-beta cluster ensemble with α = 3 and β = 3.

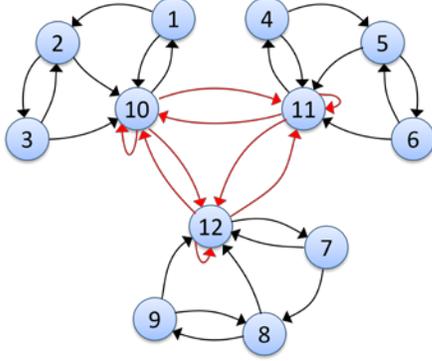Figure 3. Another drawing of the cluster ensemble in Figure 2. It is an alpha-beta cluster ensemble with α = 3 and β = 3.

In what follows, the eigenvalues λ_i of a symmetric matrix $P \in R^{n \times n}$ are always ordered such that $\lambda_1 \le \lambda_2 \le ... \le \lambda_n$. When it is useful to indicate that they are the eigenvalues on a specific matrix P, we write $\lambda_1(P) \le \lambda_2(P) \le ... \le \lambda_n(P)$ instead.

IV. PREVIEW OF AN EXEMPLAR RESULT

To have a quick reading on the work of this paper, we preview an exemplar result here.

Note that for an alpha-beta cluster ensemble, the total number of nodes is α (β +1). We use m to denote α (β + 1). Moreover, we use μ to denote the largest in-degree of a non-hub node. For the network topology in Figure 3, we can check that nodes 2, 5 and 8 are those non-hub nodes which have the largest in-degree value of 2. Thus in this case μ = 2.

Let A be the $m \times m$ adjacency matrix of the entire ensemble. Then we can show the following (Corollary 4 of Section VI):

Exemplar Theorem.

$$\frac{\lambda_{m-\alpha+1}(A^T A)}{\lambda_{m-\alpha}(A^T A)} \ge \frac{\beta}{\mu}$$

We describe some system implication of the Theorem. Suppose that we increase β, while keeping α and μ fixed. Then, by the Theorem, the separation ratio of the α leading eigenvalues and the rest of non-zero eigenvalues will increase by a factor of at least β/u. This means that when β is large, there are α dominating leading eigenvalues and the number of these leading eigenvalues is exactly equal to the number of clusters in the alpha-beta cluster ensemble. Thus we have succeeded in relating dominance of the leading eigenvalues to cluster size.

To establish the Theorem and other similar results of this paper, we will carry out the following steps:

S1. Consider a general m-node, k-cluster ensemble.

S2. Partition the adjacency matrix A of the cluster ensemble and introduce the cluster interconnection submatrix B.

S3. Derive bounds for the eigenvalues of $A^T A$ in terms of those of B.

S4. Consider the special case of alpha-beta cluster ensembles, and derive bounds for the eigenvalues of B in terms of α and β.

S5. Derive bounds on the separation ratio for the α leading eigenvalues of an alpha-beta cluster ensemble, including the result of the Exemplar Theorem described above.

In the rest of the paper, we present our work in the order of these steps.

V. PARTITIONING $A^T A$ AND CLUSTER INTERCONNECTION SUBMATRIX

For a given m-node, k-cluster ensemble, we consider its adjacency matrix A. Since the eigenvalues of $A^T A$ are invariant under similarity transformations, we can permute the columns of A so that its last k columns correspond to the k hub nodes. We express $A^T A$ in a partitioned form:

$$A^T A = \begin{pmatrix} T & X \\ X^T & B \end{pmatrix}$$

where T, X and B are $(m-k) \times (m-k)$, $(m-k) \times k$ and $k \times k$, respectively. The B submatrix is of particular interest, as it reflects the interconnections among the hub nodes and those between the hub nodes and other nodes in their respective clusters. We call B the *cluster interconnection submatrix*. In the rest of this section, we will show that the eigenvalues of B can contribute to upper and lower bounds for those of $A^T A$ (see Corollary 1).

Theorem 1. (Cauchy's Interlacing Theorem). Let U be an $m \times m$ symmetric matrix and V a $(m-k) \times (m-k)$ principal submatrix of U, for $1 \le k < m$. Then, for $j = 1, ... , m-k$,

$$\lambda_j(U) \le \lambda_j(V) \le \lambda_{k+j}(U)$$

Proof: See, for example, page 189 of [3].

2

Theorem 2. Let G be an $m \times m$ positive semi-definite symmetric matrix partitioned as

$$G = \begin{pmatrix} H & X \\ X^T & F \end{pmatrix}$$

where H is $(m-k) \times (m-k)$ for some $0 < k < m$. Then, for $j = 0, \ldots, k-1$,

$$\lambda_{k-j}(F) \leq \lambda_{m-j}(G) \leq \lambda_{k-j}(F) + \lambda_{m-k}(H)$$

Proof: The stated lower bound on $\lambda_{m-j}(G)$ follows from Theorem 10.1.1 of [6]. The stated upper bound on $\lambda_{m-j}(G)$ can be derived from Aronszajn's inequality on page 64 of [1].

Corollary 1. Suppose that $A^T A$ is an $m \times m$ symmetric positive semi-definite matrix and that T and B are $(m-k) \times (m-k)$ and $k \times k$ sub-matrices, respectively, defined by

$$A^T A = \begin{pmatrix} T & X \\ X^T & B \end{pmatrix}$$

Then for $j = 1, \ldots, m-k$,

$$\lambda_j(A^T A) \leq \lambda_j(T) \leq \lambda_{k+j}(A^T A)$$

and for $j = 0, \ldots, k-1$,

$$\lambda_{k-j}(B) \leq \lambda_{m-j}(A^T A) \leq \lambda_{k-j}(B) + \lambda_{m-k}(T)$$

Proof: The first inequality follows from Theorem 1. The second inequality follows from Theorem 2.

Corollary 2. Suppose that

$$A^T A = \begin{pmatrix} T & X \\ X^T & B \end{pmatrix}$$

where $A^T A$, T and B are $m \times m$, $(m-k) \times (m-k)$ and $k \times k$, respectively. Then

$$\frac{\lambda_{m-k+1}(A^T A)}{\lambda_{m-k}(A^T A)} \geq \frac{\lambda_1(B)}{\lambda_{m-k}(T)}$$

Proof: By letting $j = m - k$ in the first inequalities expression of Corollary 1, we have

$$\lambda_{m-k}(A^T A) \leq \lambda_{m-k}(T)$$

By letting $j = k - 1$ in the second inequalities expression of Corollary 1, we have

$$\lambda_1(B) \leq \lambda_{m-k+1}(A^T A)$$

Corollary 2 follows from the above two inequalities.

Corollary 3. Suppose that

$$A^T A = \begin{pmatrix} D & X \\ X^T & B \end{pmatrix}$$

where $A^T A$, D and B are $m \times m$, $(m-k) \times (m-k)$ and $k \times k$ matrices, respectively, and D is a diagonal matrix $D = diag(d^{(1)}, \ldots, d^{(m-k)})$ with $d^{(1)} \leq \ldots \leq d^{(m-k)}$. Then

$$\frac{\lambda_{m-k+1}(A^T A)}{\lambda_{m-k}(A^T A)} \geq \frac{\lambda_1(B)}{d^{(m-k)}}$$

Proof: The result follows from Corollary 2, by noting that $\lambda_{m-k}(T) = d^{(m-k)}$ in this case.

VI. ALPHA-BETA CLUSTER ENSEMBLES AND THEIR CLUSTER INTERCONNECTION NETWORKS

To establish further bounds on eigenvalues of $A^T A$, we consider alpha-beta cluster ensembles as defined in Section III. Note that an alpha-beta cluster ensemble is a m-node, k-cluster ensemble with $m = \alpha(\beta+1)$ and $k = \alpha$. Let A be the adjacency matrix. Then $A^T A$ is $\alpha(\beta+1) \times \alpha(\beta+1)$, and can be written as:

$$A^T A = \begin{pmatrix} D & X \\ X^T & B \end{pmatrix}$$

where D, X and B are $\alpha\beta \times \alpha\beta$, $\alpha\beta \times \alpha$, and $\alpha \times \alpha$ matrices, respectively, and D is a diagonal matrix $D = diag(d^{(1)}, \ldots, d^{(\alpha\beta)})$ with $d^{(1)} \leq \ldots \leq d^{(\alpha\beta)}$. Since each diagonal element of D is the in-degree of a non-hub node, we have

$$d^{(\alpha\beta)} = \mu$$

where μ is the largest in-degree of a non-hub node.

The cluster interconnection submatrix B reflects the topology of the interconnection network connecting hub nodes of the clusters. We consider here two specific structures, which represent two design extremes. In the first case, hub nodes are densely connected as a clique, i.e., every hub node has a directed edge to every other hub node, including itself. In the second case, hub nodes are sparsely connected as a ring, i.e., if the clusters are labeled as cluster 1 through α then, for each $i = 1, \ldots \alpha$, the hub node in cluster i has a directed edge to the hub node in cluster $(i+1) \bmod \alpha$, and also a directed edge to itself. These two cases are depicted in Figure 4.
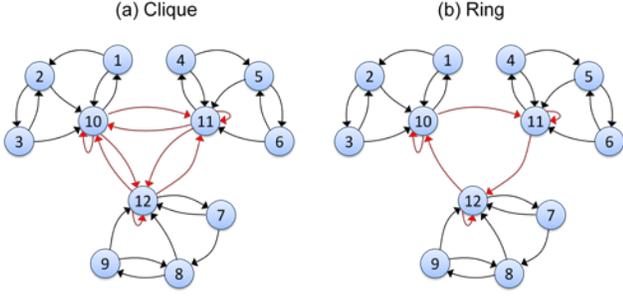
Figure 4. Two cluster interconnection networks: (a) clique and (b) ring. Note that the alpha-beta cluster ensemble of Figure 3 uses the clique interconnection network

Theorem 3. Suppose that the hub nodes are connected as a clique. Then, for $j = 1,\dots,\alpha-1$,

$$\lambda_j(\boldsymbol{B}) = \beta$$

and

$$\lambda_\alpha(\boldsymbol{B}) = \alpha^2 + \beta$$

Proof: When the hub nodes are connected as a clique, we have:

$$\boldsymbol{B} = \alpha\boldsymbol{J}_\alpha + \beta\boldsymbol{I}_\alpha,$$

where each entry of $\boldsymbol{J}_\alpha \in \boldsymbol{R}^{\alpha x \alpha}$ is equal to one and $\boldsymbol{I}_\alpha \in \boldsymbol{R}^{\alpha x \alpha}$ is the identity matrix.

Note the vector $\boldsymbol{u} = (1, 1, \dots, 1)^T$ with all of its components equal to 1 is an eigenvector of the cluster interconnection submatrix $\boldsymbol{B}$ with the corresponding eigenvalue being $\alpha^2 + \beta$. Let $\boldsymbol{v}$ be any other eigenvector of $\boldsymbol{B}$. Since $\boldsymbol{v}$ must be orthogonal to $\boldsymbol{u}$, we have $\boldsymbol{Cv} = 0$ where $\boldsymbol{C}$ is a $\alpha \times \alpha$ matrix with all its components being α. Since $\boldsymbol{B} - \boldsymbol{C}$ is a diagonal matrix with all its diagonals being β, we have $\boldsymbol{Bv} = (\boldsymbol{B} - \boldsymbol{C})\boldsymbol{v} + \boldsymbol{Cv} = (\boldsymbol{B} - \boldsymbol{C})\boldsymbol{v} = \beta\boldsymbol{v}$. Thus, the eigenvalues of $\boldsymbol{B}$ are $\lambda_j(\boldsymbol{B}) = \beta$ for $j = 1,\dots,\alpha-1$ and $\lambda_\alpha(\boldsymbol{B}) = \alpha^2 + \beta$.

Theorem 4. Suppose that the hub nodes are connected as a ring. Then, for $j = 1,\dots,\alpha$,

$$\beta \le \lambda_j(\boldsymbol{B}) \le \beta + 4$$

Proof: When the hub nodes are connected as a ring, we have

$$\boldsymbol{B} = \begin{bmatrix} 2+\beta & 1 & & & & 1 \\ 1 & 2+\beta & 1 & & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & 2+\beta & 1 \\ 1 & & & & 1 & 2+\beta \end{bmatrix}$$

By the Gershgorin Disc Theorem, eigenvalues of $\boldsymbol{B}$ are bounded below by $(2+\beta) - 2$ and above by $(2+\beta) + 2$.

From Theorems 3 and 4, we note that when the clusters increase their size β, $\lambda_1(\boldsymbol{B})$ will increase and thus, by Corollary 2, also the eigenvalue separation. More precisely, we can establish the corollaries below. .

Corollary 4. Suppose that the hub nodes in the system are connected as a clique or ring. Then

$$\frac{\lambda_{m-k+1}(\boldsymbol{A}^T\boldsymbol{A})}{\lambda_{m-k}(\boldsymbol{A}^T\boldsymbol{A})} \ge \frac{\beta}{\mu}$$

Proof: The result follows from Corollary 3 and Theorems 3 and 4, while noting that $d^{(\alpha\beta)} = \mu$.

Corollary 5. Suppose that the hub nodes in the system are connected as a clique. Then

$$\frac{\lambda_m(\boldsymbol{A}^T\boldsymbol{A})}{\lambda_{m-1}(\boldsymbol{A}^T\boldsymbol{A})} \ge \frac{\alpha^2 + \beta}{\beta + \mu}$$

Proof: From Theorem 2, we note that, for $j = 0,\dots,k-1$,

$$\lambda_{k-j}(\boldsymbol{B}) \le \lambda_{m-j}(\boldsymbol{A}^T\boldsymbol{A}) \le \lambda_{k-j}(\boldsymbol{B}) + \lambda_{m-k}(\boldsymbol{D})$$

For $j = 0$, we have

$$\lambda_k(\boldsymbol{B}) \le \lambda_m(\boldsymbol{A}^T\boldsymbol{A})$$

For $j = 1$, we have

$$\lambda_{m-1}(\boldsymbol{A}^T\boldsymbol{A}) \le \lambda_{k-1}(\boldsymbol{B}) + \lambda_{m-k}(\boldsymbol{D})$$

By Theorem 3, we have

$$\lambda_k(\boldsymbol{B}) = \alpha^2 + \beta \text{ and } \lambda_{k-1}(\boldsymbol{B}) = \beta$$

It follows that

$$\lambda_m(\boldsymbol{A}^T\boldsymbol{A}) \ge \lambda_k(\boldsymbol{B}) = \alpha^2 + \beta$$

and

$$\lambda_{m-1}(\boldsymbol{A}^T\boldsymbol{A}) \le \lambda_{k-1}(\boldsymbol{B}) + \lambda_{m-k}(\boldsymbol{D}) = \beta + \mu$$

by noting that $\lambda_{m-k}(\boldsymbol{D}) = \mu$. The Corollary follows from the above two inequalities.

Corollary 6. Suppose that the hub nodes in the system are connected as a ring. Then

$$\frac{\lambda_m(A^T A)}{\lambda_{m-k+1}(A^T A)} \leq \frac{\beta + 4 + \mu}{\beta}$$

Proof: From Corollary 1, we note that, for $j = 0, \ldots, k-1$,

$$\lambda_{k-j}(B) \leq \lambda_{m-j}(A^T A) \leq \lambda_{k-j}(B) + \lambda_{m-k}(D)$$

For $j = 0$, we have

$$\lambda_m(A^T A) \leq \lambda_k(B) + \lambda_{m-k}(D)$$

For $j = k-1$, we have

$$\lambda_1(B) \leq \lambda_{m-k+1}(A^T A)$$

By Theorem 4, we have

$$\lambda_k(B) \leq \beta + 4 \text{ and } \lambda_1(B) \geq \beta$$

It follows that

$$\lambda_m(A^T A) \leq \lambda_k(B) + \lambda_{m-k}(D) \leq \beta + 4 + \mu$$

and

$$\lambda_{m-k+1}(A^T A) \geq \lambda_1(B) \geq \beta$$

The Corollary follows from the above two inequalities.

VII. SYSTEM IMPLICATION

We can design a cluster ensemble system which will lead to a desired eigenvalue separation ratio of the leading eigenvalues and the rest of the non-zero eigenvalues. For example, we can form an alpha-beta cluster ensemble with a relatively large β, so that β/μ will be large, where μ is the largest in-degree of a non-hub node in a cluster. Then, when the hub nodes are connected as either a clique or ring, Corollary 4 implies that the resulting system will exhibit strong dominance by the α leading eigenvalues. This means that the overall system behavior will essentially be governed by the corresponding α eigenvectors. For example, the values or signs of the components in these eigenvectors will reveal clustering structures of the system [7]. The resulting clusters, for example, can help detect malfunctioning or malicious sensors in a sensor network application [5].

Furthermore, suppose that values for α and μ are fixed and the hub nodes are connected as a ring. Corollary 6 implies that separation ratios among the largest α eigenvalues will be bounded above by a quantity which will approach 1 when β increases. Thus, for large β, there will be large separations between the α leading eigenvalues and the rest of eigenvalues (Corollary 4), but there are no significant separations among the α leading eigenvalues themselves (Corollary 6). This means that when the hub nodes are connected as ring all the corresponding α eigenvectors will be of equal significance in characterizing the system.

As a contrast, consider the case when hub nodes are connected as a clique. Assume that the values for β and μ are fixed. Then, by Corollary 5, when $(\alpha^2 + \beta)/(\beta + \mu)$ is large (e.g., when α is large relatively to β and μ), the principal eigenvalue will exhibit strong dominance. Thus, for example, component values of the principal eigenvector alone will basically characterize the clustering structure of the system [7].

From the above discussion, we see that by choosing proper values for α, β and μ in the design of a cluster ensemble, we can tune the dominance properties of its leading eigenvalues.

VIII. SUMMARY AND CONCLUDING REMARKS

We have motivated the problem of designing distributed systems which exhibit strong dominance of the leading eigenvalues, in terms of underlying network topologies. We have introduced a special class of networks called cluster ensembles, and established some fundamental mathematical results of ensuring such dominance for this class of networks. As future work, we plan to generalize these characterizations by allowing more general topology than the alpha-beta ensemble and develop applications in system design based on spectral properties established in this paper.

REFERENCES

[1] R. Bhatia, Matrix Analysis, Springer-Verlag, New York, 1997.

[2] F. R. K. Chung, Spectral Graph Theory, American Mathematical Society, 1997.

[3] R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, UK, 1985.

[4] H. T. Kung and B. W. Suter, "Hub Matrix Theory and Applications to Wireless Communications," EURASIP Journal on Advances in Signal Processing, Vol. 2007, Issue 1, Article ID 13659, 2007.

[5] H. T. Kung and D. Vlah, "Validating Sensors in the Field via Spectral Clustering Based on Their Measurement Data," Military Communications Conference (MILCOM 2009), October 2009.

[6] B. N. Parlett, The Symmetric Eigenvalue Problem. Prentice-Hall, Englewood Cliffs, NJ, 1980.

[7] H. T. Kung and D. Vlah, "Sign-based spectral clustering," 25th Biennial Symposium on Communications, May 2010.