USE OF EIGENVALUES AND EIGENVECTORS TO ANALYZE BIPARTIVITY OF NETWORK GRAPHS

Natarajan Meghanathan

Jackson State University, 1400 Lynch St, Jackson, MS, USA
natarajan.meghanathan@jsums.edu

ABSTRACT
This paper presents the applications of Eigenvalues and Eigenvectors (as part of spectral decomposition) to analyze the bipartivity index of graphs as well as to predict the set of vertices that will constitute the two partitions of graphs that are truly bipartite and those that are close to being bipartite. Though the largest eigenvalue and the corresponding eigenvector (called the principal eigenvalue and principal eigenvector) are typically used in the spectral analysis of network graphs, we show that the smallest eigenvalue and the smallest eigenvector (called the bipartite eigenvalue and the bipartite eigenvector) could be used to predict the bipartite partitions of network graphs. For each of the predictions, we hypothesize an expected partition for the input graph and compare that with the predicted partitions. We also analyze the impact of the number of frustrated edges (edges connecting the vertices within a partition) and their location across the two partitions on the bipartivity index. We observe that for a given number of frustrated edges, if the frustrated edges are located in the larger of the two partitions of the bipartite graph (rather than the smaller of the two partitions or equally distributed across the two partitions), the bipartivity index is likely to be relatively larger.

KEYWORDS
Eigenvalue, Eigenvector, Network Graph, Bipartivity Index, Partitions

1. INTRODUCTION

Network analysis is the study of complex relational data that capture the relationships among the members of the system. The typical goals of network analysis are to characterize the structure of the system, identify patterns of relationships, rank the constituent members based on the connections that they are part of as well as to detect communities of the members of the system. Network analysis has applications in many disciplines such as Social networks, Biological networks, Citation networks, Co-author networks, World Wide Web, Internet, Particle Physics, Electrical networks, and etc. Accordingly, the members of the system could be anything - from individual users, user groups and organizations, molecular complexes (like proteins), scholarly publications, computers and routers, websites, electrical grids and etc. The power of network analysis is to abstract the complex relationships among the constituent members simply in the form of a graph with vertices (a.k.a. nodes) and edges (a.k.a. links), which could be directed or undirected or a combination of both as well as be weighted or unit-weight edges, depending on the nature of the interactions among the members.

We model any given complex network as a graph of vertices and edges: a vertex represents an individual component of the system being modeled (e.g., users, computers, actors, protein complexes, etc) and an edge captures the interactions between them. The adjacency matrix A(G) of the network graph G essentially captures the presence of edges between any two vertices. For any two vertices i and j in graph G, the entry in the i^{th} row and j^{th} column of A(G) = 1 is 1 if there is an edge from vertex i to vertex j and 0 otherwise. Depending on the nature of the interactions, the edges could be undirected (symmetric adjacency matrix) or directed (non-
symmetric adjacency matrix). We encounter undirected edges when we model networks where two-way interaction is the de facto standard for any association between two nodes in a network (for example: Protein-protein interaction networks, Power grid, Science collaboration networks, Internet, Actor network, etc). Directed edges are encountered when the interactions need not be in both directions of the link between any two nodes (for example: Phone call network, Email network, World wide web, etc). Sometimes, there could be both directed and undirected edges in a network graph (e.g., metabolic networks where certain chemical reactions are reversible while certain reactions proceed in only one direction).

In this paper, we present spectral decomposition (eigenvalues and eigenvectors of the adjacency matrix of the graph) based analysis of network graphs to detect whether they are bipartite or close-to being bipartite or not and if found to be either of the two cases, we show how to predict the two partitions of the "true" or "close-to" bipartite graph. A bipartite graph is a graph wherein the set of vertices in the graph could be partitioned to two disjoint partitions and the edges of the graph connect the vertices across the partitions. In a "true" bipartite graph, there are no edges connecting the vertices within a partition. In a "close-to" bipartite graph, there may be one or few edges (called the frustrated edges) connecting the vertices within a partition.

Spectral decomposition consists of generating a continuous multi-dimensional representation (a set of eigenvalues and the corresponding eigenvectors) of the adjacency matrix of the network graph. An eigenvector is referred to as the vector of coordinates of the points along each axis of the multi-dimensional space and the corresponding eigenvalue is the length of the projection on the particular dimension. Depending on the underlying network characteristic that is to be studied, we identify the set of axes (eigenvectors) that essentially capture the variability in the data (in this paper, we make use of the smallest eigenvalue and its corresponding eigenvector): the first axis corresponds to the direction of greatest variability in the data; the second axis (orthogonal to the first axis) captures the direction of the greatest remaining variability and etc. Though the number of dimensions in the spectrum is the number of vertices in the graph, most of the variations could be captured in the first few dimensions of the coordinate system represented by the eigenvalues and the eigenvectors.

**Related Work:** Most of the work in the literature has focused on developing algorithms that minimize the number of edges (i.e., the frustrated edges) that need to be deleted from a graph to extract a bipartite spanning graph. Though this is an NP-hard problem for general graphs [1], for fullerene graphs (cubic 3-connected planar graph with exactly 12 pentagonal faces and an optional number of hexagonal faces), it has been found that there exists a polynomial-time algorithm [2] to determine the minimum set of edges that could be removed from fullerene graphs to extract a bipartite spanning graph. In [3], the authors developed a mathematical programming model and a genetic algorithm to determine the minimum number of frustrated edges to be removed from fullerene to extract a bipartite subgraph. In [4], the authors compute the bipartite edge frustration of a polybuckyball (a fullerene polymer) by extending the splice and link operations on the two partitions of the graph. In [5], the authors derive theoretical bounds on the maximum frustration index of a complete graph with a set of $l$ and $r$ vertices constituting the two partitions. Thus, most of the work in the literature is focused on minimizing the number of frustrated edges that need to be removed from selected graphs (mostly chemical compounds) to obtain a bipartite spanning graph. Our contributions in this paper are to predict the two bipartite partitions of any given graph that is hypothesized to be "true" bipartite or "close-to" bipartite as well as to analyze the impact of the distribution of the frustrated edges on the bipartivity index.

**Roadmap of the Paper:** The rest of the paper is organized as follows: Section 2 explains the eigenvalues and eigenvectors in greater detail and illustrates their computation with a simple example. Section 3 presents the calculation of bipartivity index on undirected graphs. Section 4
illustrates the prediction of the partitions for undirected "true" and "close-to" bipartite graphs. Section 5 illustrates the prediction of the partitions for directed "true" and "close-to" bipartite graphs. Section 6 concludes the paper.

2. **Eigenvectors and Eigenvectors**

Spectral decomposition is a standard method to handle multivariate data in statistics and identify the directions of maximum variability [6]. The directions are called the eigenvectors and the relative importance to be given for each direction is represented by the eigenvalues. The spectrum is the collection of all the (eigenvalues, eigenvector) pairs of the multivariate data represented as a matrix. In this paper, we show that spectral decomposition of a unit-weight adjacency matrix (where the entries are either 0 or 1) of a network graph could be conducted to extract information on the extent of bipartivity in an underlying network as well as to predict the two partitions constituting the network graph.

We now illustrate an example to determine the calculation of eigenvalues and eigenvectors. Figure 1 illustrates the computation of the characteristic polynomial of an adjacency matrix for the network graph shown. The roots of the characteristic polynomial (i.e., roots of the equation $\lambda^4 - 4\lambda^2 - 2\lambda + 1 = 0$) are the eigenvalues. Accordingly, we solve the characteristic polynomial $\lambda^4 - 4\lambda^2 - 2\lambda + 1 = 0$; the roots are $\lambda = \{2.17; 0.31; -1; -1.48\}$. The eigenvector $X$ for an eigenvalue $\lambda$ is the one that satisfies $(A - \lambda I) X = 0$ [7]. Note that $X$ is a column vector with $n$ rows where $n$ is the dimension of the adjacency matrix $A$.

We illustrate the computation of the eigenvector for eigenvalue 2.17 in Figure 2. Note that 2.17 is the largest eigenvalue for the adjacency matrix and it is called the principal eigenvalue and its corresponding eigenvector is called the principal eigenvector. For the calculation of the eigenvalues and eigenvectors of adjacency matrices used in this paper (including Figure 2), we...
use the website: http://www.arndt-bruenner.de/mathe/scripts/engl_eigenwert.htm. A screenshot of the results obtained for the adjacency matrix of Figure 1 is shown below in Figure 3.

$$A - \lambda I = \begin{bmatrix} -\lambda & 1 & 0 & 0 \\ 1 & -\lambda & 1 & 1 \\ 0 & 1 & -\lambda & 1 \\ 0 & 1 & 1 & -\lambda \end{bmatrix}$$

Let $\lambda = 2.17$

\[-2.17x1 + x2 = 0 \quad \ldots (1)\]
\[x1 - 2.17x2 + x3 + x4 = 0 \quad \ldots (2)\]
\[x2 - 2.17x3 + x4 = 0 \quad \ldots (3)\]
\[x2 + x3 - 2.17x4 = 0 \quad \ldots (4)\]

Solve $\ (A - \lambda I) \ X = 0$

\[
\begin{bmatrix} -\lambda & 1 & 0 & 0 \\ 1 & -\lambda & 1 & 1 \\ 0 & 1 & -\lambda & 1 \\ 0 & 1 & 1 & -\lambda \end{bmatrix} \begin{bmatrix} X1 \\ X2 \\ X3 \\ X4 \end{bmatrix} = 0
\]

Solving the above 4 equations, we get the Eigenvector corresponding to eigenvalue 2.17:

\[
X = \begin{bmatrix} 0.281 \\ 0.611 \\ 0.522 \\ 0.522 \end{bmatrix}
\]

Figure 2. Calculation of the Principal Eigenvector for the Network Graph in Figure 1

3. **Bipartivity Index**

A graph $G = (V, E)$ is said to be bipartite if the vertex set $V$ could be divided into two disjoint sets $V1$ and $V2$ such that there are no edges connecting vertices within the two subsets and every edge in $E$ only connects a vertex in $V1$ to a vertex in $V2$ or vice-versa (if the graph is directed) [8]. More formally, $G = (V, E)$ is said to be bi-partite if the two partitions $V1$ and $V2$ of the vertex set $V$ and the edge set $E$ are related as follows:

(i) $V1 \cup V2 = V$ and $V1 \cap V2 = \Phi$ (empty set)

(ii) $\forall (i, j) \in E$, either $i \in V1$ and $j \in V2$ or $i \in V2$ and $j \in V1$
Figure 4.1 illustrates a bipartite graph that has no edges within its two vertex set partitions. In reality, it may not be possible to find network graphs that are truly bipartite. There may be few edges between the vertices within the same partition. Such edges are called frustrated edges. Figure 4.2 illustrates a graph that is close to being bipartite, with the majority of the edges connecting the vertices across the two partitions but there are two frustrated edges. The eigenvalues of the adjacency matrix can be used to determine the extent of bipartitivity of a network graph $G$ in the form of a metric called the bipartitivity index, $b_3(G)$, calculated as follows. Let $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ be the eigenvalues of the adjacency matrix of $G$.

\[
b_3(G) = \frac{\sum_{j=1}^{n} \cosh(\lambda_j)}{\sum_{j=1}^{n} \cosh(\lambda_j) + \sum_{j=1}^{n} \sinh(\lambda_j)}
\]

The calculation of the bipartivity index for a "true" bipartite graph and for a "close-to" bipartite graph are shown in Figures 5 and 6 respectively. We can notice that for a "true" bipartite graph, the $\sinh(\lambda_j)$ values in the formula for the bipartivity index add to 0, resulting in the bipartivity index being 1 for such graphs. On the other hand, for a non-bipartite graph, the sum of the $\sinh(\lambda_j)$ values contribute a positive value - leading to an increase in the value of the denominator compared to the numerator in the formula for bipartivity index. Thus, the bipartivity index for a non-bipartite graph is always less than 1; if the $b_3(G)$ values of graph $G$ is closer to 1, we call such graphs as "close-to" bipartite, as the one in Figure 6 (where edge 2 - 4 is removed from the graph in Figure 5 and edge 1 - 4 is added as a frustrated edge).

Figure 4. Examples of "True" and "Close-to" Bipartite Graphs

The calculation of the bipartivity index for a "true" bipartite graph and for a "close-to" bipartite graph are shown in Figures 5 and 6 respectively. We can notice that for a "true" bipartite graph, the $\sinh(\lambda_j)$ values in the formula for the bipartivity index add to 0, resulting in the bipartivity index being 1 for such graphs. On the other hand, for a non-bipartite graph, the sum of the $\sinh(\lambda_j)$ values contribute a positive value - leading to an increase in the value of the denominator compared to the numerator in the formula for bipartivity index. Thus, the bipartivity index for a non-bipartite graph is always less than 1; if the $b_3(G)$ values of graph $G$ is closer to 1, we call such graphs as "close-to" bipartite, as the one in Figure 6 (where edge 2 - 4 is removed from the graph in Figure 5 and edge 1 - 4 is added as a frustrated edge).

Figure 5. Bipartivity Index Calculations for a "truly" Bipartite Graph

Figure 7 illustrates the impact of the number of frustrated edges and their location on the bipartivity index values for several sample network graphs. The bipartivity index of the graphs
decreases with increase in the number of frustrated edges that connect vertices within the same partition. We can observe that for a given number of frustrated edges, a larger value of the bipartivity index is observed for graphs that have a relatively larger number of frustrated edges in the larger partition vis-à-vis the smaller partition.

The Bipartivity Index Calculations for a "close-to" Bipartite Graph

| Eigenvalue, \( \lambda \) | \( \cosh(\lambda) \) | \( \sinh(\lambda) \) |
|--------------------------|------------------|------------------|
| -2                      | 3.762            | -3.627           |
| -1.562                  | 2.489            | -2.279           |
| 0                       | 1                | 0                |
| 0                       | 1                | 0                |
| 1                       | 1.543            | 1.175            |
| 2.562                   | 6.519            | 6.442            |
| Total                   | 16.313           | 1.711            |

\[ b_s(G) = \frac{\sum_{j=1}^{n} \cosh(\lambda_j)}{\sum_{j=1}^{n} \cosh(\lambda_j) + \sum_{j=1}^{n} \sinh(\lambda_j)} \]

\[ b_s(G) = \frac{16.313}{16.313 + 1.711} = 0.905 \]

Figure 6. Bipartivity Index Calculations for a "close-to" Bipartite Graph

Figure 7. Impact of the Number of Frustrated Edges and their Location on the Bipartivity Index of Network Graphs

4. Predictions of the Partitions in a Undirected Bipartite Graph

We now illustrate how to predict the two partitions of a "true" or "close-to" bipartite graph. For this purpose, we will make use of the smallest of the eigenvalues and its corresponding eigenvector, hereafter referred to as the bipartite eigenvalue and the bipartite eigenvector respectively. The bipartite eigenvalue is most likely a negative value in "true" or "close-to" bipartite graphs. The bipartite eigenvector is likely to comprise of both positive and negative entries. The node IDs whose entries in the bipartite eigenvector are of the positive sign constitute one of the two partitions and those of the negative sign constitute the other partition. The above approach has been found to accurately predict the two partitions of a "true" bipartite graph, as shown in Figure 8. However, for "close-to" bipartite graphs, the partitions predicted (using the smallest eigenvalue and its corresponding eigenvector) may not be the same as the partitions expected (hypothetical partitions) of the input graph whose adjacency matrix had been used to determine the eigenvalue and the eigenvector. Nevertheless, the predicted partitions of
the "close-to" bipartite graphs and the hypothetical partitions of the original input graph contribute to the same bipartivity index value. This shows that two "close-to" bipartite graphs that physically look similar (i.e., same set of vertices and edges connecting the vertices), but are logically different (i.e., differ in the partitions) would still have the same bipartivity index; the difference gets compensated in the number of vertices that form the two partitions and/or the distribution of the frustrated edges across the two partitions. Note that the predictions of the partitions get less accurate as the bipartivity index gets far lower than 1.

Figure 8. "True" Bipartite Graph: Predicted Partitions Match with the Hypothetical Partitions of the Input Graph

Figure 9. "Close-to" Bipartite Graph: Predicted Partitions Appear to be the Same as that of the Input Graph

Figure 9 illustrates the prediction of the partitions of a "close-to" bipartite graph that looks complex enough for one to initially hypothesize the two partitions; hence, we assume the predicted partitions are the same as what is expected of the original input graph. However, Figure 10 illustrates an example where the predicted partitions of a "close-to" bipartite graph are of unequal sizes with two frustrated edges (whereas the input graph is hypothesized to have two equal-sized partitions with one frustrated edge, as shown); but, both the graphs have the same bipartivity index. The predicted "close-to" bipartite graph consists of a larger partition with four vertices and a smaller partition with two vertices; there are two frustrated edges in the larger partition and none in the smaller partition. The input "close-to" bipartite graph for this illustration has three vertices in each of the two partitions, with a frustrated edge in one of the two partitions. This example reiterates our earlier assertion that for two "close-to" bipartite graphs that physically look the same and have the same bipartivity index, there could be logically different sets of partitions: a topology with equal number of vertices in both the partitions and fewer frustrated edges could offset for a topology with a larger partition containing relatively larger number of frustrated edges.
5. **Prediction of the Partitions in a Directed Bipartite Graph**

To predict the two partitions of a directed bipartite graph, we can still use the same approach as explained above. We need to transform the directed graph to an undirected graph (replace all the directed edges with undirected edges), determine the bipartite eigenvalue and bipartite eigenvector of the undirected graph, and predict the constituent vertices of the two partitions based on the sign of the entries (corresponding to these vertices) in the bipartite eigenvector. Finally, we restore the directions of the edges. If the input directed graph is "truly" bipartite, the predicted partitions will be the same as that hypothesized for the input graph (refer the example shown in Figure 11). However, for "close-to" bipartite directed graphs, the predicted partitions need not be the same as that of the hypothetical partitions of the input graph; but, as long as the
set of vertices and edges (including the directions of the edges) remain unaltered, the bipartivity index of the "close-to" bipartite graph will remain the same with both the hypothetical expected partitions and the predicted partitions. Figure 12 illustrates an example wherein the set of vertices constituting the predicted partitions is observed to be different from the hypothetical partitions expected of the input graph. The hypothetical partitions contributed to two frustrated edges, whereas the predicted partitions contributed to only one frustrated edge. Nevertheless, since the set of vertices and the set of edges are the same for both the input graph and the graph based on the predicted partitions, the bipartivity index value remains the same.

6. CONCLUSIONS

The paper demonstrates the application of the eigenvalues and eigenvectors to analyze the bipartivity of both undirected and directed graphs. We observe that for a given number of frustrated edges, the bipartivity index is more likely to be larger if more of these edges are located in the larger of the two partitions of the bipartite graph. For "close-to" bipartite graphs, we observe the predicted partitions of the vertices to be different from that of the hypothetical partitions of the input graph; but nevertheless, since the set of vertices and set of edges constituting the bipartite graphs do not change, the bipartivity index remains the same for both the input and predicted graphs. In other words, for a given number of vertices and edges, there could be more than one instance of a bipartite graph (i.e., there could exist one or more combinations of the two partitions) that could have the same bipartivity index value. The above argument holds good for both directed and undirected bipartite graphs.

REFERENCES

[1] Z. Yarahmadi, T. Doslic and A. R. Ashrafi, "The Bipartite Edge Frustration of Composite Graphs," Discrete Applied Mathematics, vol. 158, no. 14, pp. 1551-1558, July 2010.
[2] T. Doslic and D. Vukicevic, "Computing the Bipartite Edge Frustration of Fullerene Graphs," Discrete Applied Mathematics, vol. 155, no. 10, pp. 1294-1301, May 2007.
[3] Z. Seif and M. B. Ahmadi, "Computing Frustration Index using Genetic Algorithm," Communications in Mathematical and in Computer Chemistry, vol. 71, pp. 437-443, 2014.
[4] Z. Yarahmadi, "The Bipartite Edge Frustration of Extension of Splice and Link Graphs," Applied Mathematics Letters, vol. 23, pp. 1077-1081, 2010.
[5] G. Bowlin, "Maximum Frustration in Bipartite Signed Graphs," The Electronic Journal of Combinatorics, vol. 19, no. 4, #P10, 2012.
[6] C. Yang, B. A. Poser, W. Deng and V. A. Stenger, "Spectral Decomposition of Susceptibility Artifacts for Spectral-Spatial Radiofrequency Pulse Design," Magnetic Resonance in Medicine, vol. 68, no. 6, pp. 1905-1910, December 2012.
[7] D. C. Lay, Linear Algebra and its Applications, Pearson, 4th edition, May 2011.