Clustering in Bipartite Graphs: State-Based Trade Networks

Berk Çoker
bcoker@stanford.edu

Charissa Sonder Plattner
chariss@stanford.edu

Aristidis Papaioannou
papaioan@stanford.edu

December 11, 2016

Abstract

International trade relationships are complex networks whose creation and evolution are influenced by geography, history and ever-changing agreements. Existing research focuses on trade between countries, modeling their relationships as graphs of nodes of the same type. We focus on trade relationships between states and territories of the Unites States of America and other countries, modeled as a directed, weighted bipartite graph. We propose and evaluate two null models: (1) the Coupled Erdos-Renyi model and (2) the Bipartite Configuration Model. We identify highly unbalanced relationships using normalized weights, and extend two existing clustering algorithms (BRIM and Spectral Co-Clustering) to handle directed, weighted bipartite graphs. Early results from our spectral analysis provide some insightful clusters that seem to capture latent cultural and historical properties of the corresponding nodes.

1 Introduction

Trade partnerships have been at the center of economic discussions, eliciting reactions and proposals from think tanks, researchers, and even presidential candidates. Globalization has increased the interdependence among countries in complex and oftentimes unexpected ways.

Using graph theory, researchers have investigated the effects of trade by creating models that connect partners in the world according to their trading volume. The bulk of this research has treated the United States as a single node in these networks, analyzing export trade between all fifty states combined and various other countries.

Similarly, the problem of detecting community structure in networks has recently received a great deal of attention in the scientific community. Many different kinds of algorithms have been proposed to solve this problem, ranging from hierarchical clustering to modularity-based methods. The majority of these methods make few assumptions about the underlying structure of the network in order to be as general as possible, but in so doing often overlook properties specific to the particular graph.

In order to further explore the nature of trade partnerships, we model the import and export volume data between the 50 states and other countries around the world as a bipartite network, such that all edges were drawn between individual states in the US and foreign countries. Unlike past contributions, we analyze trade between various countries and each individual state in a bipartite framework.

The motivation for this research stems from the authors' perception that state-of-the-art research on bipartite graphs was missing a key element in network analysis: a strong null model. We launched an investigation into null models for bipartite graphs, specifically for the import-export weighted, directed bipartite graph of world trade. A fundamental contribution of this work is the creation and evaluation of bipartite-specific null models that captures the weighted trade volumes in this world trade network, and can even be further extended to other directed, weighted bipartite networks exhibiting similar properties.

2 Related Work

2.1 Network Analysis on World Trade

Previous research on trade networks has traditionally taken a country-based approach to capture the properties of import and export relationships. Studies such as [1] have made inferences on the topological properties of the world trade web using network models. In the scope of this paper, we consider the most recent discoveries regarding the trade network in [2] and [3] and evaluate the extent to which similar results hold when studying trade between states and countries.

In [3], Fagilio et al. study different network models to identify the best model that would capture the properties of trade relationships between countries. The paper evaluates the structure and intensity of trading among 159 countries over a period of 20 years (from 1981 to 2000). The network is first modeled by a directed unweighted network. This graph is massively connected, dense and reciprocated: almost all countries import from partners to whom they export. In a weighted model, where node degree is weighted by the trade intensity of each edge, the resulting distribution is very close to a power law: most countries have relatively weak connections, while a few have very intense connections. The weight of edges between partners is almost perfectly symmetrical.

Viewed as a weighted graph, the network displays statistical properties that are significantly different from the unweighted model: (i) the majority of existing links are associated to weak trade relationships; (ii) the weighted web is only weakly disassortative; (iii) countries holding more intense trade relationships are more clustered.

Results highlight that countries with stronger trade relationships are mostly clustered together. Furthermore, the researchers label nodes weak or strong nodes based on their actual trade capacities. They develop a binary model that used thresholding to remove "unimportant" edges and then transformed the graph into an unweighted model. In this model, a weak node had a high probability of connecting to a strong node. However, this property is not found in the weighted model, which shows that stronger nodes cluster primarily together [4].

2.2 Clustering in Bipartite Networks

In the interest of our problem, we studied community detection methods that were specifically tailored for bipartite graphs. Two algorithms emerged in literature: Bipartite Recursively Induced Modules and Spectral Recursive Embedding:

Bipartite Recursively Induced Modules (BRIM), introduced by Michael Barber, is a direct extension of the standard modularity maximization algorithm to bipartite networks [5]. It is an iterative algorithm that employs a refined modularity matrix and null model to accommodate for the bipartite structure. At each iteration, the algorithm fixes a partition on one side and maximizes the modularity with respect to the other. The results show significant improvements, demonstrating the importance of leveraging the bipartite structure in identifying clusters. A pitfall of BRIM, as acknowledged by Barber, is that it only handles unweighted and undirected bipartite networks. Further work can generalize the algorithm to all bipartite graphs. Moreover, BRIM has been evaluated only on one null model so far. More complex null models for bipartite graphs can improve the performance of the algorithm.

Spectral Recursive Embedding (SRE), introduced by Zha, is an adaptation of the standard spectral clustering algorithm to bipartite graphs [6]. It is an iterative algorithm that finds partitions, where partitions are constructed to minimize the normalized sum of edge weights between unmatched pairs. The algorithm finds an approximate solution to this minimization problem by employing a partial singular value decomposition on the weight matrix of the bipartite graph. As Zha demonstrates, normalized cut in this case can be minimized with left and right eigenvectors that correspond to the second largest eigenvalues. Zha et all acknowledge that further work is needed on developing the SRE algorithm. Two suggestions are finding a more principled approach on identifying cut points and on expanding the algorithm to cover partitions with overlaps.

3 Data

Data points represent the average import and export volumes between United States and foreign countries, between the years 2008 and 2015 [7]. When projected onto a bipartite graph, the data has 291 nodes, with 54 United States and territories on one side, and 237 foreign countries on the other. Edges are weighted (in US dollars) and directed (import/export).

The graph is almost complete: 86% of all possible edges are defined, joining 79% of all possible node combinations that preserve the bipartite nature of the network. It consists of 4 strongly connected components, the largest of which contains 99% of all nodes. All nodes belong to the same weakly connected component. The degree distribution does not follow a power law: most states trade with most countries, and vice-versa (see figure 1). As seen in 2, bi-directional edges are mostly balanced, confirming the observation in [3].

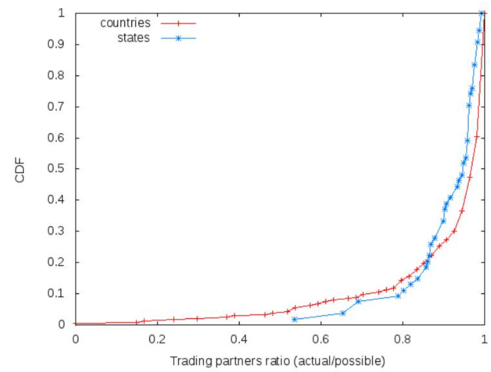

Figure 1: Trading partner (i.e., degree) distribution

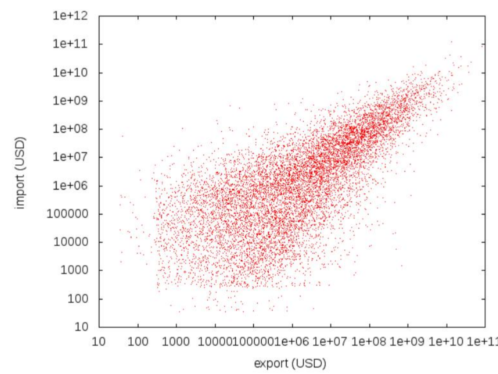

Figure 2: Edge weights of all bi-directional relationships

4 Models

4.1 Normalization

In order to capture the relative significance of trading partners regardless of their absolute size, edge weights are normalized such that the sum of all exports (respectively, imports) for each state node sums up to one. Let S, C be the set of state nodes and country nodes respectively, and let w_{s_i,c_j} be the weight of the outgoing edge (i.e., exports) from state $s_i \in S$ to country $c_j \in C$, and w_{c_j,s_i} the weight of the corresponding incoming edge (i.e., imports). The normalization process replaces those weights with:

$$w'_{s_i,c_j} = \frac{w_{s_i,c_j}}{\sum_{c_k \in C} w_{s_i,c_k}}$$

4.2 Bipartite Random Model

This random model is the unweighted, undirected industry standard for bipartite graphs described in [5]. In this null model, the probability of an edge existing between two nodes is proportional to the product of the degree of the first node and the degree of the second node. If the two nodes are in the same independent set, then the probability of that edge is

0. Thus, the expected degree of any node is equal to that node's corresponding node in the real graph.

Let m be equal to the number of edges in the original graph. Let k_i be the degree of node i and d_j be the degree of node j. The probability of an edge existing between nodes i and j in the null model, where i and j are in different independent sets, is equal to $P_{ij} = \frac{k_i d_j}{m}$.

4.3 Coupled Erdos-Renyi Bipartite Random Model

This model extends the previously described random model to produce bipartite graphs that are weighted and directed. It builds upon the observations obtained from our data to produce a distribution of edge weights that matches real networks.

1. We create undirected edges between any (state, country) pair with probability p. The weight of the created edge is sampled from a distribution with parameters tuned from our bipartite network (see below).

2. We calculate the *strength* of each node in the graph by summing the weights of all incoming edges to that node.

3. For each undirected edge, we reapportion the weights of that edge into two directed edges, such that the weight on each new edge is proportional to the "strength" of its start node. This effectively distributes the constant amount of trade volume between partners proportionally to their individual total imports.

The edge weight probability function used to generate this model is obtained by observing that the empirical CDF of the edge weight distribution in the original dataset strongly resembles a sigmoid function in lin-log space. We arbitrarily decide to fit a logistic function to it, of the form:

$$P(W \leq w) = \frac{1}{1 + \exp(\alpha(\log w - \mu)}$$

Maximum likelihood estimation of parameters α and μ results in a close fit to the empirical CDF, as seen in figure 3. Furthermore, parameter μ has a natural interpretation: it corresponds to the median trade volume across all states and countries. This interpretation is confirmed by our evaluation, as seen in the following sections.

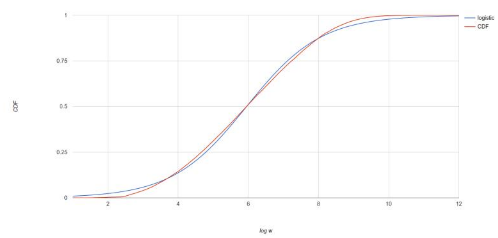

Figure 3: Logistic fit of empirical weight distribution

4.4 Bipartite Configuration Model

For this final iteration of null model, we use the real weight data in the graph rather than drawing weights from the function described in 4.3. We make the assumption that the trade volume of a node in our graph is representative of the intrinsic trading capacity of a country or state, and randomly distribute that weight between edges in the following manner.

The input to the function that produces this null model is the set of desired node strengths (both import strength and export strength) for each state and country.

1. Let our real graph be called G. Initialize a new unweighted multigraph N with the same nodes as G but with no edges yet.

2. Replace each weighted edge in G of weight w with w/u edges in N, where u is a resolution unit. We used $u = \$1000$.

3. We then create edges from states into countries in the following way:

   (a) For each state node, count how many outgoing edges that state has, and initialize that many outgoing "stubs" for that state.

(b) For each country node, count how many incoming edges that country has, and initialize that many incoming "stubs" for that country.

(c) Randomly match in-stubs with out-stubs.

4. Lastly, we create edges from countries into states in the same way, except we count the number of incoming edges to states and outgoing edges from countries.

At the end of this process, we have a null model N with weighted, directed edges whose nodes have the same node strength as the nodes in G. This model matches another intuition in Fagliolo [2] that strong nodes are likely to have strong connections with other strong nodes. We saw in our tests of this null model that pairs of strong nodes almost always have edges between them with high weights.

5 Algorithms

5.1 BRIM

The BRIM algorithm recursively identifies bipartite modules and was first developed in [5].

BRIM depends on a few matrices: A, P, B and S. For the purposes of this experiment, we have a graph with ns state nodes and nc country nodes, where $ns + nc = n$, or total nodes.

A is an $n \times n$ adjacency matrix of the graph of n nodes. P is an $n \times n$ probability matrix, where P_{ij} is the probability that an edge exists between i and j in the null model. B is the "modularity matrix", given by $B = A - P$. S is the assignment matrix, wherein $S_{nc} = 1$ if node n is in cluster c and $S_{nc} = 0$ if node n is not in cluster c.

We can partition S in the following way: Let R be the $ns \times c$ matrix that contains assignments for only state nodes and let T be the $nc \times c$ matrix that contains assignments for only country nodes.

The modularity is given by

$$Q = \frac{1}{m} Tr R^T \widetilde{B} T$$

.

$\widetilde{B}$ is a $ns \times nc$ matrix that represents the upper-right corner of the B matrix. Because we only have edges between state and country nodes, we need only to examine that particular portion of the B matrix.

To find the best cluster given the number of clusters c, they implement an iterative maximization scheme. To start, they randomly assign all the nodes to clusters and build the corresponding assignment matrix S, from which we can extract the R and T state and country assignment matrices.

They want to maximize the modularity, so they want to maximize the equation $Q = \frac{1}{m} Tr R^T \widetilde{B} T$.

They then repeat the following steps until modularity stops increasing:

(1) They fix the assignment for the countries, thereby fixing the matrix T. Because T is fixed, let $\widetilde{T} = \widetilde{B} T$. Then choose the R that maximizes $Q = \frac{1}{m} Tr R^T \widetilde{T}$.

(2) They then fix the assignment for the states, thereby fixing the matrix R. Because R is fixed, let $\widetilde{R} = R^T \widetilde{B}$. Then choose the T that maximizes $Q = \frac{1}{m} Tr \widetilde{R} T$.

Once the iteration stops, S contains a complete cluster assignment with possibly the best modularity score for c clusters. As mentioned previously, the algorithm can get stuck in a local maxima.

In the paper, they find the optimal c using a simple binomial search in the range of possible c values: from 1 to n, where n is the number of nodes.

However, one thing we must note is that this algorithm can get stuck in a local maxima and is not guaranteed to find the globally optimal solution every time. We also see that the modularity calculation is depending on B, which is dependent on P, our null model probability matrix.

5.2 Spectral Recursive Embedding

The SRE algorithm recursively identifies partitions that minimize the normalized sum of edge weights between unmatched pairs. At each iteration, the bipartite graph is split into two clusters and are treated independently thereon.

Let's first denote the bipartite graph as $G(X, Y, E)$ where X and Y represent the two sides of the bipartite graph. Then, $V = X \cup Y$ and $X \cap Y = \emptyset$.

The algorithm will be computing a vertex partition of $G(X, Y, E)$ denoted by $\Pi(A, B)$ where $X = A \cup A^c$ and $Y = B \cup B^c$. The algorithm is based on the following proposed variant of the normalized cut equation [6]:

$$NCUT(A, B) = \frac{cut(A, B)}{W(A, Y) + W(X, B)} + \frac{cut(A^c, B^c)}{W(A^c, Y) + W(X, B^c)} \quad (1)$$

Intuitively, this refined NCUT not only captures a partition with small edge cut, but also forms two subgraphs that are as dense as possible when minimized.

To approximate this minimization, the paper then delves into proving a series a mathematical statements (which we have chosen to exclude for the sake of space), showing :

$$\min_{\Pi(A,B)} NCUT(A, B) = 1 - \max_{x \neq 0, y \neq 0} \frac{2x^T W y}{x^T D_X x + y^T D_Y y} \quad (2)$$

where $x^T D_X e + y^T D_Y e = 0$ [6]. Relaxing this minimization by doing away with the constraint, Zha et al arrives at Algorithm 1.

Our implementation improves on this model in the following ways: instead of using the simple strategy of setting $c_x = 0$ and $c_y = 0$ which divides the network into two subgraphs, we find a cluster that we finalize at each step and then recurse on the remainder graph. To do this, we order the left and right eigenvectors so that the values that are closest to -1 and 1 are at the two ends. Without the relaxation, the values would have been -1 and 1, and so intuitively, those that are closest must denote a a stronger relevance. Then, starting at the positive end, we iterate over the sorted eigenvectors in order to find the minimum NCUT. We do the same thing for the negative end, and pick the smallest value. Doing this let's us identify a cluster where the next edge would have increased the sum. Notice that after this partition, one of the subgraphs is finalized as a cluster.

Algorithm 1 Spectral Recursive Embedding (SRE)

Input: Given a weighted directed bipartite network $G = (X, Y, E)$ with its edge weight matrix W:
Output: Clusters $(A_1, B_1), ..., (A_n, B_n)$

1) Compute D_x and D_Y and form the scaled weight matrix $\hat{W} = D_X^{-1/2} W D_Y^{-1/2}$.

2) Compute the left and right singular eigenvectors that correspond to the second largest eigenvalue of $\hat{W}$, $\hat{x}$ and $\hat{y}$.

3) Find cut points c_x and c_y for $x = D_X^{-1/2}\hat{x}$ and $y = D_Y^{-1/2}\hat{y}$.

4) Form partitions $A = \{i | x_i \geq c_x\}$ and $A^c = \{i | x_i < c_x\}$ for vertex set X, and $B = \{j | y_j \geq c_y\}$ and $B^c = \{j | y_j \geq c_y\}$ for vertex set Y.

5) Recursively partition the sub-graphs $G(A, B)$ and $G(A^c, B^c)$ if necessary

6 Results

6.1 Normalization

The normalized model further confirms that trade relationships are also balanced in relative terms: over 90% of normalized bi-directional edges have weights that are within 1% of each other (see Figure 4).

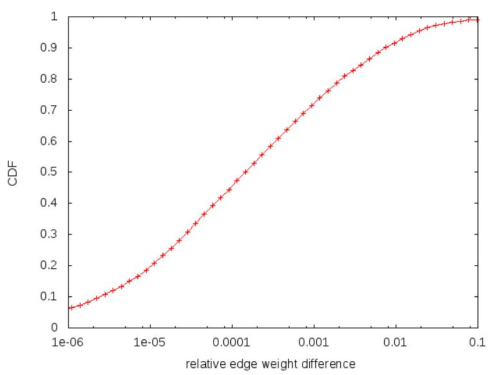

Figure 4: Trade imbalance in normalized partnerships

Country	State
Canada	Wyoming
United Arab Emirates	District of Columbia
Canada	Montana
Canada	New Hampshire
Venezuela	Virgin Islands
Switzerland	Nevada
Australia	Hawaii
Ireland	Puerto Rico
China	Nevada
Canada	Vermont
Mexico	Utah
China	Tennessee
United Kingdom	Utah
China	California
China	Minnesota
Germany	Rhode Island
Chine	Arkansas
China	New Mexico
Canada	Alaska
Israel	New Mexico

Table 1: Most unbalanced (normalized) relationships

What is even more interesting however is that the normalized graph allow us to detect *outliers*, that identify relationships in which one of the partners (most typically, the state) imports good from the other in a much proportion than it exports to it (see Table 1). In addition to highlighting possible economical dependencies, those cases oftentimes highlight geographical or historical partnerships. Many of them surface again in the results of the spectral clustering results below.

6.2 Null Models

6.2.1 Coupled Erdos-Renyi Model

As seen in Figure 5 and Figure 6, the CER null model produces graphs with similar edge weight distributions, but with radically different degree distributions. This is to be expected, because the edges were randomly assigned and created bidirectionally.

We also see that the edge weight distribution is almost identical to our raw data, which means that our empirical weight distribution logistic function was very well-fitted to the data, which means that our reapportion step did not noticeably affect the edge weight distributions. It is interesting that the trade volume between countries and states can be so accurately fitted with a logistic function. Our original decision to fit the data with a logistic function was arbitrary, but upon seeing this new distribution, we believe there must be an underlying reason why this is true and this is definitely something that we want to look into further.

We see in Figure 7 that the node strength of the nodes in the CER model was generally a bit lower than in the real graph. This means that, in the null model, there are fewer nodes model that are extremely strong.

6.2.2 Configuration Model

Comparatively, the CM null model has a very different edge weight distribution, which comes from the

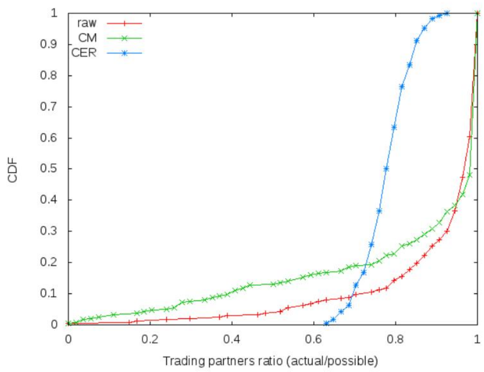

Figure 5: Null Model Evaluation: Degree Distribution

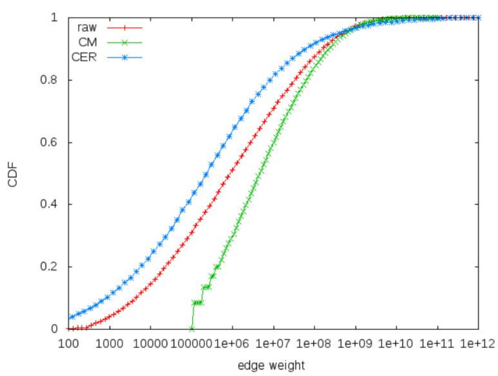

Figure 6: Null Model Evaluation: Edge Weight Distribution

algorithm that paired the generated stubs randomly rather than preferentially attaching them to stronger nodes.

Unsurprisingly, we do see that the CM's node strength distribution matches the raw data exactly. This is an artifact of the algorithm, which enforces that an node in the real graph retains its node strength in the null model.

However, what's more important is that the degree distribution in this null model matches the real graph even closer than the CER model. We see from these results that this null model is a very good candidate to replace BRIM's current null model. This is also something we want to look into further, to see if we

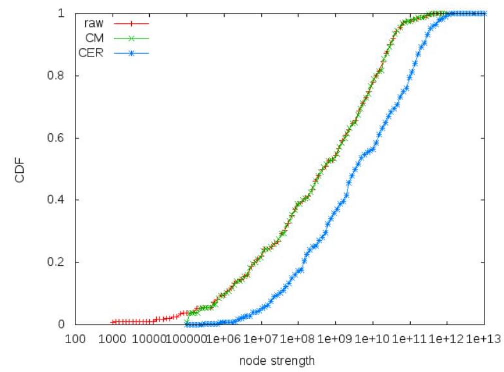

Figure 7: Null Model Evaluation: Node Strength Distribution

can improve the clustering results from BRIM even further with a new null model.

6.2.3 Reflection

We learn from these comparative distributions that the two developed null models can play two different roles in the analysis of a network. Should a researcher want to analyze the significance of the edge weight distribution of a directed, weighted graph, they should compare their graph with a CM null model. If they want to analyze the significance of their degree distributions, they should compare their graph with a CER null model.

6.3 BRIM

We expected that running BRIM [5] on our raw graph would produce uninteresting results because the graph was complete. This was confirmed when we saw a 0.06 modularity score when we ran BRIM on the raw graph.

BRIM requires an unweighted, undirected graph. When we transformed our raw graph into this format, the nuanced information about partners that were mainly encoded in the weighted, directed edges, was lost. Thus, we set out to improve the undirected, unweighted representation of our graph by developing normalization and thresholding techniques that we hoped would encode the nuances of those weighted edges into the remaining undirected edges.

We first attempted to normalize the edges, described under the Model section. This demonstrated mild improvement in our modularity scores, increasing the score from 0.06 to 0.1. However, with the two versions of thresholding, once with the parameter set to 0.005 and second with the set to 0.01, we were able to double our modularity score. Though these are marginal gains, it showed that our graph transformations were able to remove edges that were insignificant in the creation of communities.

As seen in Table 2, in which we report the modularity scores of the various graph models, our normalization schemes were able to remove the "unimportant" edges from the complete graph and thus retain the most important partners of the nodes in the undirected graph. We were able to iteratively improve the modularity scores given by BRIM on our graph.

Graph Transformation	Modularity
Raw Graph	0.0610
Normalized	0.1062
Normalized with 0.5% Thresholding	0.2033
Normalized with 0.01% Thresholding	0.2075

Table 2: BRIM performance

6.4 Spectral Clustering

We apply Algorithm 1 to the directed bipartite graph with normalized weights in order to detect clusters of related states and countries based on trading patterns alone. The goal of this experiment is two-fold: (i) since finding the global optimum of the normalized cut value is impractical, validating that our heuristic-based search for a local maximum yields relevant results; (ii) empirically evaluate the relevance of the returned clusters.

We confirm that the improvements to the SRE algorithm discussed previously allow us to obtain variable-size clusters. By only reaching a local optimum, we have indeed seen combinations of states and countries that would have yielded higher normalized cut values. A closer inspection of the returned clusters however already provides interesting insights, and reveal common properties that are not captured in numbers (such as shared history, culture or language). In fact, we have a cluster made up entirely of islands, as seen in Table 3! The SRE algorithm was able to unearth this unexpected cluster base solely on trading volumes, which was quite incredible. We also observed in the examples of obtained clusters (Table 3) that this technique is somewhat resilient to a possible Matthew effect, which would result in simply clustering most nodes with their strongest partners (i.e., China). We evaluated our clustering algorithm explicitly by looking at the clusters:

States	Countries
Virgin Island	Martinique
	Guadeloupe
	Sint Maarten
	Curacao
	Antigua and Barbuda
Hawaii	Kiribati
	Australia
	Marshall Islands
	Micronesia (Federated States of)
	Palau
	Christmas Island
	Cook Islands
New Mexico	Mexico
	Israel
Nevada	Switzerland
	India

Table 3: Top 4 clusters obtained by Modified SRE

It is worth highlighting the interesting overlap between some of the clusters presented below and the "outliers" in the normalized trade relationships listed in table 1. Notice that New Mexico and Israel; Hawaii and Australia; Nevada and Switzerland are all clustered together (In fact, Israel is New Mexico's third largest trading partner).

7 Conclusion

We achieved our goal of utilizing our data to create smart bipartite-specific null models for bipartite graphs that can be used for analysis of other bipartite networks. We employed variations of network models—weighted and unweighted, directed and undirected—to find the most optimal way to (i) capture the relations between states and countries, and (ii) find communities among states and countries based on import and export relationships. Our study expands on previous work, demonstrating that imports and exports are mostly symmetrical in state-country trade relationships as well. In our analysis, however, we have also identified some outliers, which in fact surfaced within the clusters we obtained from our algorithms. This showed that, even though import and export data were mostly symmetrical, considering the directed graph had value.

Some suggestions for future work are: (i) to incorporate our new null models into BRIM and (ii) implement SRE with overlapping partitions. Unfortunately, we were not able to obtain confident results when we used our null models on BRIM. Further mathematical analysis is required to modify BRIM accept more complex null models. The case is similar for SRE.

References

[1] Serrano, Ma Angeles, and Marián Boguná. *"Topology of the world trade web."* Physical Review E 68, no. 1 (2003): 015101.

[2] Fagiolo, Giorgio, Javier Reyes, and Stefano Schiavo. *"On the topological properties of the world trade web: A weighted network analysis."* Physica A: Statistical Mechanics and its Applications 387, no. 15 (2008): 3868-3873.

[3] Fagiolo, Giorgio, Javier Reyes, and Stefano Schiavo. *The Evolution of the World Trade Web. A Weighted-Network Analysis.* 49. ma Riunione Scientifica Annuale. Società Italiana degli Economisti. No. info: hdl: 2441/9795. Sciences Po, 2008.

[4] Garlaschelli, Diego, and Maria I. Loffredo. *"Fitness-dependent topological properties of the world trade web."* Physical review letters 93, no. 18 (2004): 188701.

[5] Barber, Michael J. *"Modularity and community detection in bipartite networks."* Physical Review E 76, no. 6 (2007): 066102.

[6] Zha, Hongyuan, et al. *"Bipartite graph partitioning and data clustering."* Proceedings of the tenth international conference on Information and knowledge management. ACM, 2001.

[7] https://catalog.data.gov/dataset/state-imports