Community Structures in Trade Flow

Pedro Garzon
Computer Science
Stanford University
Palo Alto, California 94306
Email: pgarzon@stanford.edu

Gerardo Rendon
Management Science and Engineering
Stanford University
Palo Alto, California 94306
Email: grendon@stanford.edu

Fidel Salgado
Computer Science
Stanford University
Palo Alto, California 94306
Email: fidels@stanford.edu

***Abstract*—Trade flow has been found to be correlated with war; that is, the more trade there is between nations, the less likely they are to go to war with each other. We explore further graph analysis on a data set of trade between countries originally compiled to analyze how trade influences war alliances and country stability. What was missing from this study was seeing if war trends could have been surfaced from the trade graph itself. Here we approach using cluster coefficients, shortest paths, spectral clustering, and graph evolution to study patterns in the trade dataset. Our end conclusion was that approaches that take into account weighted directed edges to separate the graph would give better insights to possible partitions in the data that might resemble alliances or communities of countries. At best, we were able to find a community of war-stable European countries using spectral clustering.**

***Keywords*—War, trade, clustering, spectral clustering, graphs**

I. INTRODUCTION

The effects of international wars can be devastating. From wars that span a small geographic region, such as the Iran-Iraq war, to wars involving armed conflicts across continents, such as WWI and WWII, international combat can leave behind thousands or even millions of casualties, irreparable damage, unstable government regimes, and other devastating consequences. Given these effects, the aim of this paper is to explore trade relationships between countries. It has historically been shown that stronger trade relationships between nations indicate a smaller chance of these nations going to war[1]. Thus, our attempt is to understand the global trade network to see what insights about military conflict can be derived from an analysis of this type. Having a sense of the factors that can lead to war can help avert it, either by leveraging trade negotiations between countries, through interventions by multilateral organizations, via new alliances, or by resorting to other similar means.

In order to better understand trade networks, we will look at historical imports and exports between countries. The data we'll be using spans back to 1870 and has some records all the way up to 2009. Mainly, we'll focus on doing two things with this data: analyzing the basic properties of the global trade network and how they've evolved over time, and using spectral clustering to split the data into significant trade networks that can lead to insights about military alliances.

II. RELATED WORK

Before diving into the two main objectives this paper aims to achieve, it is important to highlight some of the relevant literature that has explore topics similar to ours. We will discuss three papers that have reached important insights about related material.

A. Network of Military Alliances, Wars, and International Trade

The original paper that compiled the ward and trade dataset [1] uses historical data on wars and trade to show that there was a dramatic drop in interstate wars since 1950 because of the densification and stabilization of trading relationships and alliances. Furthermore, by examining specific relationships, the paper is able to find countries that traded the most with their allies and how these countries are less likely to be involved in wars with any other countries.

The model presented in the paper is defined as follows. $N = \{1, ..., N\}$ is a set of countries. Countries are linked through alliances, which is represented by a network of alliances, $g \subset N^2$ where if $i, j \in g$ countries i and j are allies. The paper then defines models for military strengths and what it means to win a war. It defines whats a vulnerable country in the network and what it means to have a war-stable network.

The value of this research comes as our baseline understanding of the full data set were using. This was the paper that aggregated the data. What our project will then aim to do, is to expand on the understanding of this paper by using tools out of the network analysis toolbox we have been exposed to in class. Our methods will discuss ways of expanding on this understanding of war-stable networks by first getting some baseline network statistics on the graph. Then from there well see if there are any other forms of structure in the graph and if certain relationships in a certain time-frame or different time-frames is possible

B. Network Analysis of International Relations

This report [2] explores how network analysis is currently used in the field of international relations – both its advantages and shortcomings, and then makes a case for how network analysis should be used to guide the study of international relations in the future. This paper is a more general approach to the analysis of relationships between nations than the other two works, for it is about the application of networks to the entire field of international relations, not specifically to trust or wars respectively. The main technical components of the paper are closely tied with the initial concepts presented in class. Hafner-Burton, Kahler, and Montgomery give a very comprehensive overview of network structure and theory in the initial parts of the paper, explaining some of these ideas thoroughly, and thus touching on concepts such as node degree, closeness, betweenness, and centrality, and how these concepts play crucial roles defining network structure in this interplay of nations.

This paper also gives critiques on how network analysis could be use to contrive numbers in order to fit some description of the data. Well take these critiques in mind to play devils advocate when we discuss our results and choice of algorithms and different properties to understand. These include the ones we went over in class such as degree, eigenvector centrality, betwenness, flow betweenness, closeness, and information. We hope to use these baseline measures as a way of understanding methods further custom analysis we will conduct on the dataset.

C. Network Clustering and Community Detection in Directed Networks

[3] comes as a general inspiration for us for trying to detect interesting patterns. It's our cookbook of ideas for finding communities in a graph. The paper goes through a throughout overview of typical algorithms for finding communities such as PageRank, spectral co-clustering, modularity, and density based methods for finding communities.

The paper weights pros/cons of algorithms along with suggestions on how to turn a directed and weighted graph into a usable form that the the mentioned algorithms could be implemented. However, this tends to have it's drawback as core properties of the original directed graph can be lost by changing directed edges to undirected ones. Some smarter approaches include turning the graph into a unipartite graph or a bipartite graph in order to maintain directionality. We bring this paper as relevant discussion to guide is on different methods we could apply for getting more information out of the trade flow network

III. Data

In this section, we'll take a deeper dive into explaining the data set that we'll be using for our analysis.

The Correlates of War (COW)[1] project is a research project aimed at collecting historical data about relationships between countries. Multiple data sets have stemmed from this initiative, some of which well use to conduct our research. The main data set well be using contains a list of country pairs at different years in history. Each country-country pair (dyad) has both, the amount of imports that country one receives from country two, and the amount of imports that country two receives from country one.

We can essentially think of the dyad trade data set as a collection of graphs – one graph for each that year data was collected, from the early 1900s to the 2000s. Since a node in the graph represents a certain country, it turns out that our graphs are fairly small in terms of nodes, but tend to have a high amount of edges. There's the consistent trend that the amount of nodes and edges increase over the years; however, there are clear order of magnitude changes over several decades. It seems we begin to have a view of the modern world trade flow around the 1960s when the amount of countries begins to slowly climb to the high 100s and the amount of edges climbs from the 1000s. For the sake of this project, we'll mostly look at the graphs post-WWII to see if trade flows shows any interesting characteristics of modern political structures between countries.

IV. Algorithms

This section explores the algorithms we implemented to reach the two goals mentioned in the introduction. The first part describes some findings about the graph and the evolution of the global trade network. The second part explores the spectral clustering algorithms.

A. Graph Basic Properties and Evolution

In order to get some baseline understanding of any interesting relations in the data, we first explored some structural properties of the graph. We chose to see how much the nodes are clustered together, the shortest paths throughout the years, and the closed and open triad counts. We used the built-in SNAP tools to get these initial findings.

The first item - how much the nodes are clustered - was measured by taking the clustering coefficient. Calculating the clustering coefficients of nodes gives us a measure of how tightly other nodes cluster those initial nodes. Intuitively, it is calculated as the proportion of edges between the nodes within the neighbourhood of a target node divided by the number of edges that could possibly exist between the neighboring nodes. Thus, clustering coefficients could tells us if there are certain countries in the trade graph that tend to encapsulate more of the network around it.

The shortest paths between any two given nodes is a measure of how difficult it is to get from one node on to the other, or how quickly it is for one country to reach another via its trade network. If shortest paths lengths are similar throughout the graph, this would give an indication that the graph is fairly uniform, with each node being relatively far from other nodes.

Looking at the proportion of closed versus open triads might also give us an insight if there is a skew in trade triangles. For instance, a node being in a large amount of closed triangles would indicate that it mutually trades with other countries. Being in a high amount of open triangles would indicate holes in the flow of trade. In other words, trade doesn't directly flow through the countries in an open triad. At least one country would be missing out of trade flowing in or out of it.

We also look into a baseline measure of modularity. Modularity is a measure of how compartmentalized a graph can be. To get a baseline measure of modularity over time, we opted to use the Clauset-Newman-Moore algorithm for computing modularity due to it's faster runtime over graphs with a high amount of edges. A modularity score closer to 1.0 means a graph we have a graph with a high modularity, which implies that each node belongs to a specific module/cluster of the graph.

Once we analyzed the basic graph properties, we also looked at how they've evolved over the years. The data is divided by year, so it was natural to study how the trading network evolved. We focused on the years 1948-2009 because that is when the Direction of Trade Statistics Historical Data started reporting trade. Data before then is not as accurate and is missing many statistics.

We first analyzed node ’arrival’ or ’integration’ into the network by plotting the number of nodes as a function of time (Figure 1). The result does not show a clear linear or exponential pattern; instead, the plot increases linearly for a period of time and then has jumpsïn particular years. For example, from 1960 to 1990, we see the number of nodes increasing linearly (about a new node per year); then from 1990 to 1992, we see a jump from 160 nodes to 182 nodes.

The are a couple potential reasons behind these trends. The jumps can most likely be attributed to historical events. The 1990 to 1992 jump is explained by the dissolution of the Soviet Union in 1991 that resulted in 15 post-Soviet states. The linearity of the graph can be explained by the collection of data. The data seems to start collecting trade information from new countries and it is starting to stabilize around 1995.

These properties gave us a good initial idea of the data set we would be working with. The results of these initial findings will be discussed in subsequent sections.

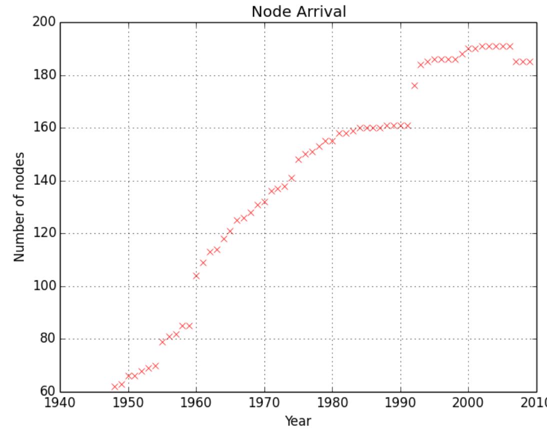

Fig. 1. Node arrival (1948-2009)

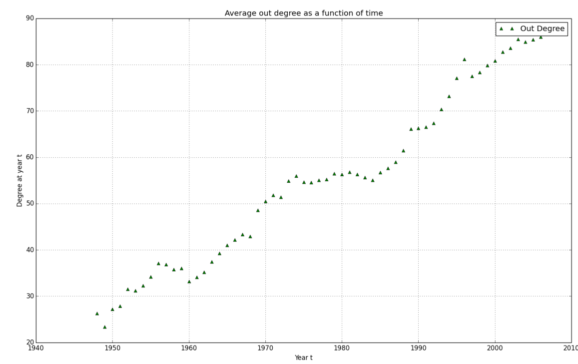

Fig. 2. Average Out-Degree over time (1948-2009)

B. Spectral Clustering

After studying the evolution of the trade network, we used using spectral clustering to analyze some of the most important global trade networks over the years. In brief, spectral clustering is a technique that leverages the eigenvalues of a matrix to split a given data set into coherent groups. In this case, we’re using spectral clustering to divide our data set into trade networks, and to then draw some insights about which networks have remained stable throughout the years and which have disintegrated. The goal here is to see if, at any point in time, spectral clustering could be used to divide the countries by different spheres of influence. For instance, perhaps spectral clustering could determine an interesting cluster of trade network that can be found through the weighted flow of trade in our data set.

Spectral clustering usually does not take into account weights since it uses a symmetric Laplacian matrix. This Laplacian matrix is used to map nodes according to eigenvector representations of the nodes. As a hack to account for our dyad structure, we make the entries of the adjacency matrix used to compute the Laplacian to be the average between the two flows corresponding to a pair of countries. We use the ready-made implementation of spectral co-clustering provided by Scikit-learn to then do the standard spectral clustering algorithm. Although this hack doesn’t guarantee the full properties of using spectral clustering, it at least does take into the account the weights and we can expect countries with a high amount of trade between them, even if it is skewed to one country importing more or less. A more detailed explanation of how spectral clustering works can be found in [3].

V. RESULTS

This section and the subsequent discussion talk about the results we obtained from implementing the algorithms mentioned above and what they mean in the context of trade networks. The results under each section here correspond to each of the two sections in the previous section.

A. Graph Basic Properties and Evolution

Due to the nature of being a trade network in dyadic form, it’s not too surprising to find out that there is a high amount of clustering throughout the graph. Most consistently we find that that the way nodes cluster together is fairly similar across the decades. Interestingly enough, the level of clustering is consistent throughout the decades and among node degrees. Below we can see how similar some baseline statistical data is between very far off years such as 1950 and 2009 as comparison. Results generalize over other years post-WWII. Here we notice that the clustering coefficient is fairly stable hovering just below 1.0 regardless of node degrees. Additionally, closed triads consistently seem to take about a 70% presence of all triads throughout the years. t There’s also a consistency with the the shortest path lengths never going above three and keeping a graph diameter around 1.5 throughout all years.

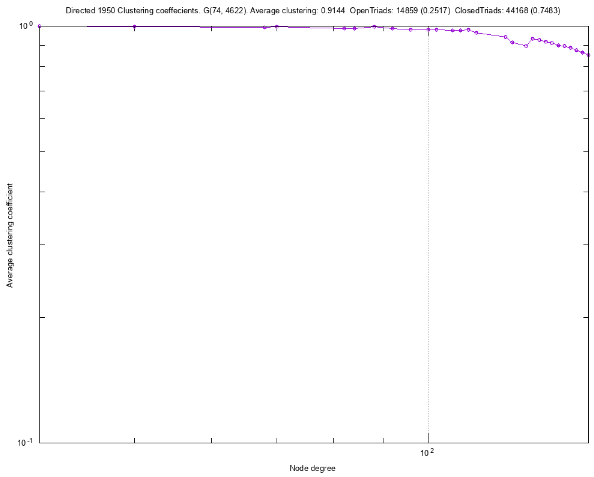

Fig. 3. Clustering Coefficients - 1950

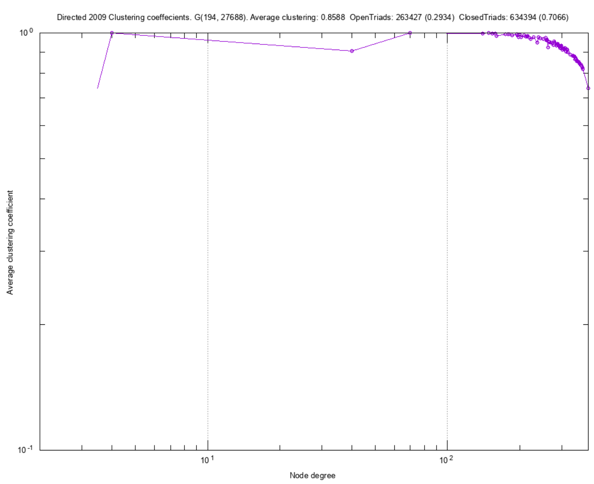

Fig. 4. Clustering Coefficients - 2009

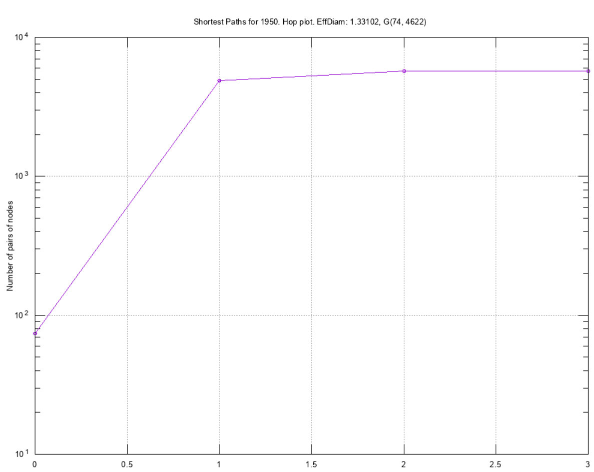

Fig. 5. Shortest Paths - 1950

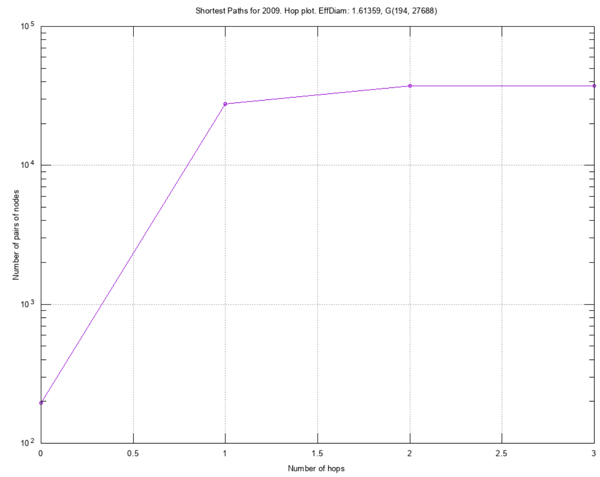

Fig. 6. Shortest Paths - 1950

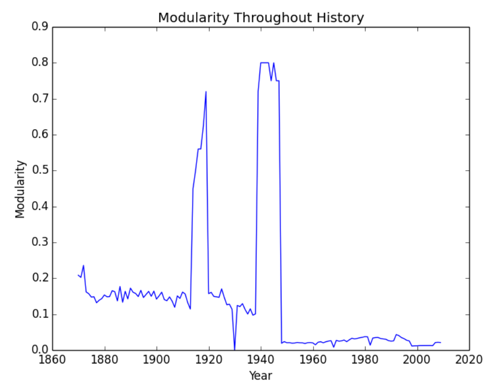

Fig. 7. Modularity History

Our yearly graph on modularity also shows some interesting details. Mainly, we see that the trade flow graph generally has a very low modularity; however, there are large spikes around 1917 and 1940 corresponding to WWI and WWII respectively. It should be noted though that these are dates when the data dropped significantly. These respective short time periods have less than 10 nodes and edges. Thus, we can discard these anomalies in the graph. What should be noted is that, overall, the modularity is decreasing.

We then analyzed the average out-degree as a function of time. We only analyzed the out-degree because the in-degree is equivalent (i.e. country A trades with country B if and only if country B trades with country A). Figure 2 shows the out-degree as function of time. It is evident that the out-degree increases linearly over time. In 1948 the average out-degree was barely around 25 (i.e. on average a country only traded with 25 other countries). In 2009, however, the average out-degree is almost 90. The linearly increase has been quite significant with globalization. This trend is likely still happening as more countries start trading.

B. Spectral Clustering

One of the strongest networks we found by using this technique is an important cluster of European nations. The United Kingdom, the Netherlands, Belgium, France, the German Federal Republic, Italy, and other nearby European countries have remained an important network over the years. This pattern has been evident from 1970 - which is when we started surveying data for spectral clustering - to 2005. This alliance makes sense given the proximity of the nations and the emergence of the European Union to strengthen these ties. Another important network, albeit not as significant as the first one, is one of Asian nations. This group seemed to appear later in the years. China and Japan showed a strong trade network since the 1970s, but other nations like Taiwan, South Korea, Singapore, and Malaysia joined the group later, around the year 2000. Lastly, two other nations that have also demonstrated a strong and consistent partnership throughout the years have been Canada and Mexico. This also makes sense, given their proximity and the emergence of the trade deal NAFTA in 1994, which further strengthened their trade relationships. It is important to note that as our clusters demonstrate, this trade relationship was more strongly evident after 1990.

The spectral clustering algorithm is successful because the networks it identifies as important are relevant in the context of actual global alliances. Furthermore, breaking down our data set into more clusters or less demonstrates which are the most relevant global trade networks today.

VI. Discussion

Our results of initial graph metrics gives the story of a graph structure that results in a large amount of homogeneity. Mainly we have a graph that is incredibly strongly connected and closed. What this means is that there are very few asymmetries in the graph and that it is really easy to move from any given node to the other following the edges. Moreover, the consistency of the baseline metrics such as clustering, shortest path, and triad structures throughout decades is somewhat surprising. We hoped to see some point where perhaps clustering increased steadily due to increased globalization of trade. However, this does not seem to be the case. This suggests that trade networks have already been fairly globalized post-WWII. What would probably change over time is not the trade network itself but rather the flow of goods throughout time. Additionally, the overall trend of decreasing modularity suggests that compartmentalizing the graph into communities is difficult to do using undirected edges alone. No noticeable trends relating back to wars could be found using these baseline methods.

Meanwhile, and as discussed previously, most of the trends derived through spectral clustering make sense given the proximity of nations and the emergence of international trade deals. Trade networks very often point to military alliances, so recognizing which trade networks are strongest is important to also identify which countries are at a lesser risk of going to war. Having said this, it isn't likely to see nations in the European Union, for instance, going to war with each other, nor is it likely to see countries in NAFTA fighting among themselves. Rather, the opposite is what normally happens; countries with strong trade alliances fight with each other, as is the case with NATO, for instance. Historically, wars tend to be between countries that are not significant trading partners, so having a better sense of who's trading with who gives us better insight into wars around the world.

The mentioned result hint that in order to find more interesting properties of the graph, it is suggested to depend on using the weighted properties of the trade flow as opposed to just directional edges for making further insights about the data.

As further work, we would recommend trying other methods of working with weighted directed graphs for clustering. For instance, perhaps a weighted PageRank or HITS approach to the graph could have been supplement to the our spectral clustering approach. We would also suggest elaborating more on the work of [4] who showed that spectral clustering could be applied to a weighted bipartite graph. First the trade flow dataset we have been using must be converted into a bipartite graph. Luckily [3] gives a few suggestions and background research on attempt for doing such a conversion in section 4.2. One technique could be seperating the graph into two sides by some threshold or in degree or out degree.

Another interesting approach would been using motif-based spectral clustering. Motif-based spectral clustering could have taken into account certain triad configuration and used those to separate the graph. Perhaps motif based spectral clustering could have separated the graph of trade into partitions around a certain flow direction such as a circular closed triad such that we get clusters with heavy inter-country trade circulation among partitions that match the given motif.

VII. Acknowledgements

Thanks to Jure Leskovec and to the CS 224W staff for putting in all the necessary work to make this class possible.

References

[1] Matthew Jackson and Stephen Nie, *Networks of military alliances, wars, and international trade*, 3rd ed. PENAS, 2015.

[2] Hafner-Burton, E., Kahler, M., Montgomery, A. (2009). Network Analysis for International Relations. International Organization, 63(3), 559-592.

[3] Fragkiskos D. Malliarosa, Michalis Vazirgiannisa (2009). Network Analysis for International Relations. International Organization, 63(3), 559-592.

[4] B. Coker, C. Plattner, A. Papaioannou. Clustering in Bipartite Graphs: State-Based Trade Networks. 2016.