Modeling and Detecting Shifts in International Community Structure

Ayush Kanodia[1], **Raghav Gupta**[1], **Pulkit Agrawal**[2]
[1]Department of Computer Science
[2]Department of Electrical Engineering
Stanford University
Stanford, CA 94025
{kanodiaayush,rgupta93,pulkital}@stanford.edu

1 Introduction

The advent of network theory as a well-studied area within computer science, coupled with the emergence of huge social network datasets has led to the development of techniques to analyze network effects on user behavior. This theory has been successfully applied to various domains ranging from biology to social science to describe complex network phenomena. These recent advances beg the question of how network effects work in the sphere of international trade. Econometric theory uses statistical modeling techniques to model the interaction between trade and endogenous economic variables of a country and these simple statistical models fail to incorporate complex network effects arising from inter-dependencies between international trade partners due to diplomatic alliances.

Motivation and Practical Relevance: In this work, we use network models to describe the evolution of community structure in international trade network. We are intrigued by the idea of shifts in community structure across the network and stability of communities across time. In particular, we are interested in the behavior of certain nodes which change communities entirely, for example the shift of North Korean diplomacy from Russia to China. Our goals for network modeling here are two fold. The first goal is to qualitatively test if we can capture such major international events as the breakdown of Soviet Union, Indo-China war, etc. just by observing the community structure in the international trade network across years. Our second goal is to explain the network effects that characterize/accompany node shifts in the community structure by developing a network evolution model for the trade network which is successful in modeling evolution in trade network across years.

We also realize that there is a strong link between international alliances and international trade, and that trade indicators present a strong picture of a nation's economic standpoint. This leads us to question whether community formation in trade networks is a strong (possibly latent) determinant of a nation's economic policy, affecting its endogenous variables. To this end, we use regression analysis to model the effect of trade and community structure in trade networks on national variables such as GDP.

Paper Organization: We begin by presenting a brief overview of prior work that has been done in this area. This is followed by a description of the dataset that we use and the graph we work with. We then discuss the three types of analyses we perform on our graph and discuss the results of our experiments. We end our discussion with proposition for future directions of research in the area.

2 Related Work

The prior work closest to what we aim to do is by Zhu et al. (2014), which characterizes the growth and evolution of China in the international trade network. Notably, it looks at all forms of trade. Its key technical contributions are manifold. First, it analyses the important members within a community, apart from the community formation itself, which they call *community core detection*.

They present a detailed analysis of the evolution, collapse and reformation of communities in the Asian region. They try to empirically explain this with respect to economic changes during this time frame. They follow this up with a variant of the preferential attachment model presented as a generative model for the emergence of such network structures. Taking inspiration from the model proposed here, we develop our own version of preferential attachment model which attempts to explain node shifts across communities.
However, one of the weaknesses of their analysis is that they do not demonstrate the application of their model on the actual international trade network and they do not provide any comparison between results generated by their model and the ground truth. We address this weakness by applying our model to the full trade network and comparing the model parameters obtained by fitting Kronecker graphs on the ground truth graph and the graph generated by our model.

Another fairly similar piece of work is by Zhong et al. (2014), a novel study on the formation of communities in international trade networks. The goal and technical contribution of this paper are key insights into network effects in the international trade community since 2001. They apply a state of the art community detection algorithm from Blondel et al. (2008) to trade network datasets, and present the evolution of such communities over time. They look at community stability and community evolution over time, and further describe network effects of large externalities in the market, such as the Iraq War or the Housing Crisis. Their analysis, however, comes across as primarily qualitative, with very dispersed conclusions of one off cases. We perceive the merit of this work as more of establishing how clear communities are in world trade, than in making conclusions or reasoning about these communities.

Jackson and Nei (2015) try to establish whether trade networks have a strong effect in determining military pursuits of nations. Their main contribution is a formal model of alliances between nations, based on mutual incentives to form alliances. They present an analysis of the stability of such alliances based on their model, followed by a proof that no such stability is possible within their model without trade between participants. They then present an augmentation of their model to allow trade effects, and present a theoretical proof describing when stable military alliances form in the presence of trade. They, however, keep their results to a qualitative graph model and not actually test these hypothesis on an actual quantitative network. While their methods are not directly relevant, this work served as an inspiration for us to formulate our questions for this project.

3 Data and Network Model

3.1 Dataset

We leverage our data from the Correlates Of War (COW) portal Maoz (2016), a portal maintained by multiple research groups providing quantitative data in international relations. Some significant datasets available from this source that we use are as follows (the following list includes attributes of links between nations):

- **Bilateral Trade:** Data recording bilateral trade volumes across all commodities between nations, over the period 1870-2009
- **Conflicts:** Data corresponding to approx. 3000 inter-state wars and militarized disputes (all instances of when one state threatened, displayed, or used force against another) and approx. 5000 intra-state conflicts, with locations, over the period 1817-2007, approx every 5 years.
- **Diplomatic Representation:** The degree of diplomatic representation at the level of chargé d'affaires, minister, and ambassador between nations, every 5 years from 1817-2007
- **Formal Alliances:** Details on formal alliances between nations, including mutual defense pacts, non-aggression treaties, and ententes, over the period 1816-2012
- **Territorial Change:** Records of all peaceful and violent changes of territory from 1816-2014
- **Territorial Contiguity:** Records of all changes in territorial contiguity (direct sharing of international borders between nations).

In addition, for GDP data for countries across time, we use the World DataBank Kim (2016) to obtain GDP data for all countries starting 1960.

3.2 Graph definition

We represent our network of nations as an **undirected weighed multigraph** with nodes as nations and various types of international links as edges. There are multiple types of edges in our graph. *For each year, there is a different graph*, but these can be linked over time since most nations are present across all years.

From the dataset described above, our network model needs to account for the following entities and events-

- Nations as nodes, with economic and military indicators of their own changing with time. These node attributes will not a play a role in community detection, but will however play a role in linking communities across time. National attributes of interest include
  - Military expenditure and personnel
  - Energy consumption
  - GDP per-capita in that year
  - Total and urban population
- Diplomatic and trade links as edges, changing weights with time. We describe the various sets of edges that are part of our graph -
  - **Trade edge** - The weight of a trade edge from nation A to B will be the value of the exports from nation A to B. While this graph is directed in nature i.e. the $A \rightarrow B$ edge is different from the $B \rightarrow A$ edge, we will take the $A \rightarrow B$ edge weight to equal the sum of the bilateral trade volumes in either direction.
  - **Alliance edge** - The weight of an alliance edge between nations A and B is the level of treaty between them i.e. $w_1 \times \delta_{defence} + w_2 \times \delta_{neutrality} + w_3 \times \delta_{nonaggression} + w_4 \times \delta_{entente}$, where $\delta_i = 1$ when there is a treaty of type i between the two nations in that year, 0 otherwise. These weights are set to $1.0, 0.5, 0.25$ and 0.2 respectively, for modeling purposes.
  - **Diplomatic edge** - Diplomatic links will be weighted by the level of diplomatic representation between the nations. Specifically, the edge weight equals $w_1 \times \delta_{ambassador} + w_2 \times \delta_{minister} + w_3 \times \delta_{charged'affaires} + w_4 \times \delta_{other}$, where $\delta_i = 1$ if the nation A has diplomatic representation at level i in nation B, else 0. The weights, for modeling purposes, are set to $1.0, 0.5, 0.25, 0.2$ respectively.
  - **War 'edge'** - Wars are modeled as edge strength disruptions. Specifically, in the event of a war between nations A and B, their edge weight is decreased by 25% during the years of the war, and this penalty of 25% is successively halved for the next 5 years.

4 Experiments, Evaluations and Results

As part of this work, we have performed four basic evaluations for our analysis of the international trade network. We start out with **community detection**, to discover community structure in the international trade network. We then try to model **community stability** using mutual information. Once we have this information, we try to build a **preferential attachment model** which tries to capture the evolution of community structure in the international trade network. Finally, we try to evaluate community effects on endogenous economic variables of a nation. For this, we use **regression analysis.**

This section is organized as follows. We first present the analytical method for each of the above analyses, and then detail our evaluations as well as contributions alongside.

4.1 Community Detection and Global Community Stability

In this section we describe our method for community detection, how it extends to a directed network, and how we model community stability for communities discovered over time.

4.1.1 Method

We detect communities by heuristically minimizing modularity, well known in the related literature, as defined below. Heuristic optimization is required because in general this optimization is NP-hard, and we cannot solve this problem for our graph in reasonable time.

$$Q = \frac{1}{2m} \sum_{i,j} \left[w_{i,j} - \frac{A_i \times A_j}{2m} \right] \delta(c_i, c_j)$$

where where $w_{i,j}$ is the weight of the edge $i \to j$. $A_i = \Sigma_j w_{i,j}$ is the sum of the weights of the edges attached to node i . c_i is the community where node i is assigned. $\delta(c_i, c_j)$ is 1 if $c_i = c_j$ and 0 otherwise. $m = \sum_{i,j} wi, j$. We use the algorithm described in Blondel et al. (2008). Roughly, this algorithm performs the following two iterations until convergence is reached.

- Merge a node with another node which causes the most increase in modularity of the graph. Repeat till no more nodes can be merged.
- Construct a new graph with groups of nodes as a single node, edges across these *supernodes* as sum of edges of individual nodes in supernodes.

This definition of modularity, however, works with undirected graphs. A simple extension to directed graphs is proposed in Nicosia et al. (2009) as

$$Q = \frac{1}{2m} \sum_{i,j} \left[w_{i,j} - \frac{A_i^{out} \times A_j^{in}}{m} \right] \delta(c_i, c_j)$$

where A_i^{out} is the sum of weights of the edges going out of node i, and A_j^{in} is the sum of weights of the edges coming into node j, everything else remaining the same from the previous definition.
Contribution: We note that we use an extension of the technique that we mentioned for community detection in directed graphs to perform community detection in undirected graphs, although we did not find techniques like this in the literature. Upon experimentation, we find that we detect similar (almost identical) communities with directed and undirected graphs.

Importantly, we also note that our model can be altered to weigh each edge by the GDP of each notion. We note that this was the case with Zhong et al. (2014), while it was not so for Zhu et al. (2014). We go with unweighted edges, although community detection on weighted edges is similar.

We further calculate how stable communities that are formed in the graph are. We do this using the notion of NMI (Normalized Mutual Information)Zhong et al. (2014). This is an important metric which we expect economic behaviour to depend on. This is defined as:

$$NMI_{y^t, y^{t+1}} = \frac{\sum_{h=1}^{k^t} \sum_{l=1}^{k^{(t+1)}} n_{h,l} \log \frac{n.n_{h,l}}{n_h^{(t)} n_l^{(t+1)}}}{\sqrt{\left(\sum_{h=1}^{k^t} n_h^t \log \frac{n_h^t}{n} \right) \left(\sum_{l=1}^{k^{t+1}} n_l^{t+1} \log \frac{n_l^{t+1}}{n} \right)}}$$

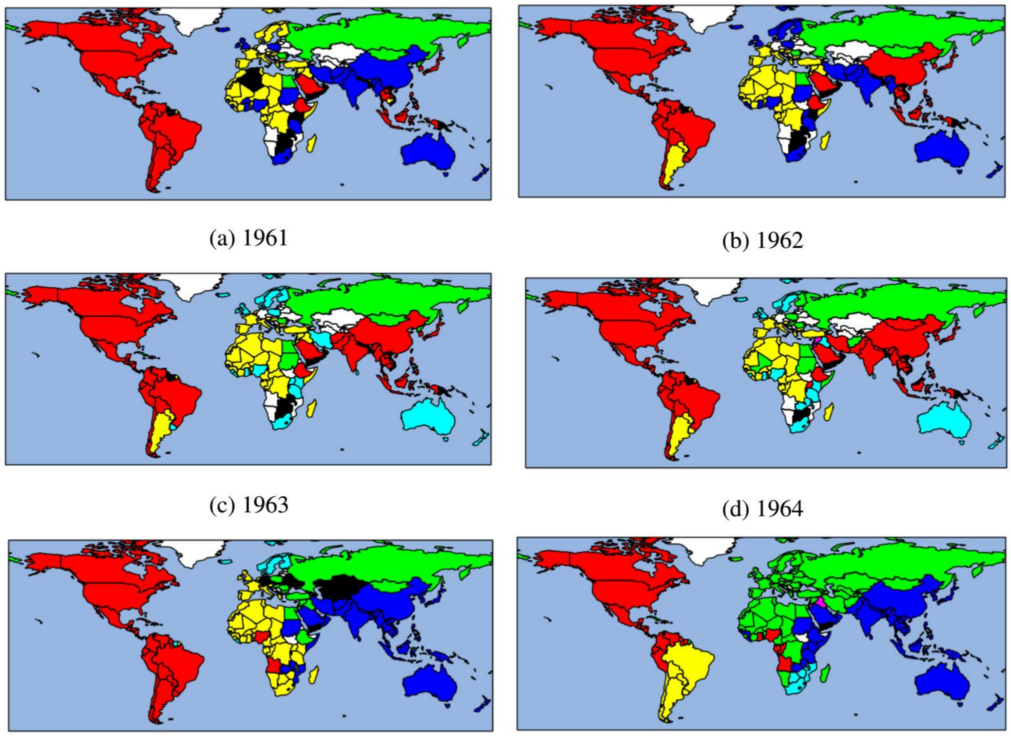
where y^t is year, t, n_h^t is the number of nodes in cluster g in year t, $n_l^(t+1)$ is the number of nodes in cluster l in year $t+1$, and $n_{h,l}$ denotes the number of nodes both in cluster h in year t and in cluster l in year $t+1$.

4.1.2 Evaluation and Analysis

For our experiments, we focus on the international trade network. We have also run our experiments for military alliances, and have modeled territorial contiguity via land and sea. However, we present our most significant results which are with the trade network and omit the rest.

Community Detection: We implemented the community detection algorithm described in 4.1.1 for the trade-only network for each year, and plot the communities on the world map using `carto.py`, a Python library. In the community plot for years 1961-1964 and 1990-1993 (Figure 1), we see some interesting examples of real world events being reflected in our communities-

1. In the communities discovered in the trade graph, North Korea, starting 1950 until 1963, belongs to the community containing Russia and centered around it. However, starting 1964, it moves to the community containing China, and stays there since. This is concordant with the 1961 Mutual Aid and Cooperation Friendship Treaty between the two nations.



(a) 1961 (b) 1962

(c) 1963 (d) 1964

(e) 1990 (f) 1993

Figure 1: Communities returned by our algorithm over the bilateral trade network

2. India and China are mostly in the same communities until 1962, but they break that in this year, presumably due to the Indo China war and the ensuing sanctions during that year, which are not present in the 1963 and 1964 graphs. Since then, China and India have generally belonged to the same community over the years.
3. After the USSR split, most of the new nations are in a community of their own, until they merge with that of Russia in a few years as in the 1993 plot.

Global Community Stability: We present a plot of Community Stability versus year. Here, community stability for the trade network for a year is its Mutual Information with the previous year's trade network. We note that the plot of NMI in Fig 2 affords us the following interesting insights. For one, we see that stability falls during the 1960s, which was the most intense period of the cold war between the United States and the Soviet Union. This period saw a lot of strong new trade ties being forged. Further, we see that stability sharply falls during the period of the collapse of the Soviet Union as well as the unification of Germany during the early 1990s. This is because many of the new nations born out of this even went to different communities in this period, leading to reduced stability. Both these observations show that international political events have significant impact on community structure in the trade network.

4.2 Network Evolution

We develop a network evolution model similar to the preferential attachment model where we define the notion of community stability of a node, which gives us a measure of how strong the node's presence is in this community and, more importantly, how likely it is to leave its community and join another one.

Motivation: We reason that even within a community, there are nodes which may not be strongly connected within their own community and may also possess strong links with nodes in other communities and may attempt to leave. We characterize this by the probability

Figure 2: Plot of Mutual Information vs Years for International Trade Network

of a node wishing to change community by the relative volume of its total trade within the community.

As has been seen historically, nodes changing communities (especially common among developing countries changing blocs from the US to communist, for example) are enticed largely by higher aid/trade volumes from the central nodes in the new community, but also increased trade with their partners. This is also reflected in our model.

For our model, we start out by constructing the countries graph with the trade edges in a given year. Now, for each node, define

$$P_{\text{reachout}}(i) = \frac{\sum_{j \in V, c_i \neq c_j} w_{i,j}}{\sum_{j \in V} w_{i,j}}$$

i.e. the proportion of edge weights the node has within the community versus overall, as the probability the node will attempt to leave the community. We use this distribution to sample nodes which will attempt to leave its community.

If in an iteration a node attempts to leave the community it is in, we model this by increasing the edge weight of the node with another node in a different community by a fraction β of the total outgoing trade of the node. We set $\beta = 0.05$ in our experiments. An equivalent amount of trade is reduced uniformly from all other edges of the node in the network. The node with which the weight is increased is selected in two steps - we first pick a community according to the unigram distribution

$$P_{\text{community}}(i,k) = \frac{\sum_{j \in V, c_j = k} w_{i,j}}{\sum_{j \in V, c_i \neq c_j} w_{i,j}}$$

giving us the probability that node i will now attempt to connect to community k. In the final step, we select a node in this new community to which we will increase the weight of the edge to node i. We model the probability of choosing node j in the new community to be proportional to a linear combination of the edge weight $w_{i,j}$, the GDP of the new node GDP_j and the community centrality of node j, CC_j which we define as

$$C_{\text{community}}(j) = \min_{\{k \in K - C[j]\}} |Q(g,C) - Q(g,C_k^j)|$$

where K is the set of all communities in the graph, C is a mapping from node to community ID and C_k^j is same as C for all nodes except j for which the community is changed to one of the communities it is not already a part of. Thus, the centrality of a node is the minimum absolute change change in modularity obtained by removing node j from its community and assigning it to a different community. It is thus a measure of how important a node is to its community. More precisely, we choose our new node to increase the weight to with a probability distribution (weights w_1, w_2 below are modeling parameters)

$$P_{\text{change}}(i,j) \propto w_1 \times w_{i,j} + w_2 \times \frac{C_{\text{community}}(j)}{\sum_j C_{\text{community}}(j)} \times \sum_k w_{j,k}$$

We set $w_1 = w_2 = 0.5$ in our experiments. We find these values to lead to stable results across different simulation runs, and are hence sound choices for these hyperparameters.

Contribution: Our ultimate goal is to observe the evolution of communities with our modified preferential attachment model, and examine whether this evolution is similar to the actual evolution of communities in the international trade network as discovered by our algorithm in 4.1. Specifically, we fit the 2x2 **Stochastic Kronecker graph (SKG)** model as described in Leskovec et al. (2010) to our ground truth graphs and the graphs generated by our network evolution algorithm and measure the similarity of the 4 model parameters. We note that in this, we provide a solid analytical comparison between our model and the ground truth, which is missing in the prior literature such as Zhu et al. (2014) entirely and is a limitation of the same as we pointed out.

Figure 3a shows the Kronfit parameter plots for the actual trade networks (1990-1994) and for the graph generated by our PA model (once every 50 iterations). Note that the model iterations update only a small fraction of edge weights at a time, and thus a large number of them collectively equal one year in the ground truth. However, our model parameters have trends similar to those in the actual networks and asymptotically mimic them. Further, the community structure in the actual networks and the generated networks are strikingly similar in nature as regards the community size distribution and the number of communities changing nodes across years/iterations.

4.3 Regression Analysis

As another aspect of our analysis, we train regression models to predict endogenous variables *per country* based on certain independent variables. Some of these variables are reflective of trade network effects. More concretely, we perform regression analysis using the following linear model (GDP is simply Gross Domestic Product). Let C be the community of our nation of interest $y = \sum_i a_i x_i + b\epsilon$, where ϵ is our error and y is δGDP in a year and change in total trade volumes in a year. We use the following variables to regress on δGDP for each country:

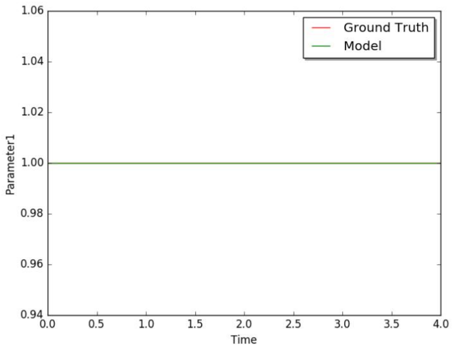
- x_1 = incoming trade volumes
- x_2 = outgoing trade volumes
- x_3 = current GDP of the country
- x_4 = mutual information between community structure in that year and the next
- x_5 = average GDP of all countries in the community of that country

Contribution: We perform the regression on all countries present in our dataset, from 1960 through 2008. We note that this is the first attempt at quantifying the correlation between network structure and endogenous economic variables of a nation precisely, therefore we do not have baseline comparisons to compare with. We obtain an $R^2 = 0.63$ for this regression, which indicates a reasonable fit. Further, the individual correlation coefficients for x_1, x_2, x_3, x_4 and $x_5$are respectively 0.67, 0.62, 0.66, -0.0068, 0.17. This indicates that the effect of average community strength (for which average GDP of all nations in the nation's community is a proxy) is weakly correlated with a nation's GDP. However, we note that global network stability across years, as measured by mutual information change in GDP is non existent.

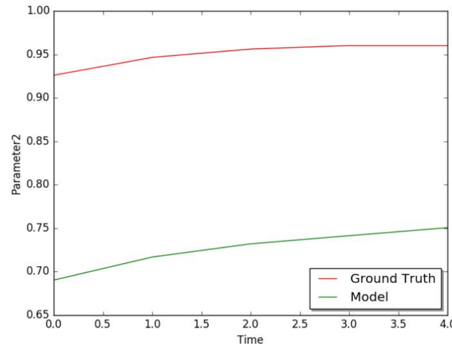
We note that we also experiment with network effects on other variables, but these do not yield strong statistical significance. We use the following variables to regress on change in total trade volumes:

- x_1 = mutual information
- x_2 = $\sum_{community(i)=C} GDP(i)$

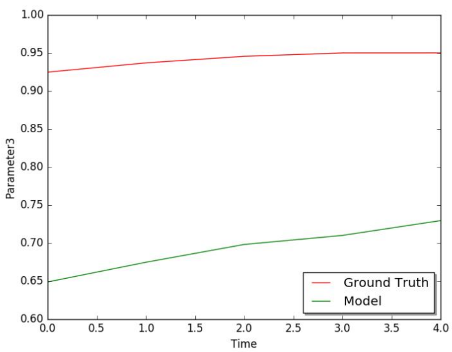
for which we obtain an $R^2 = 0.05$, which is very low.



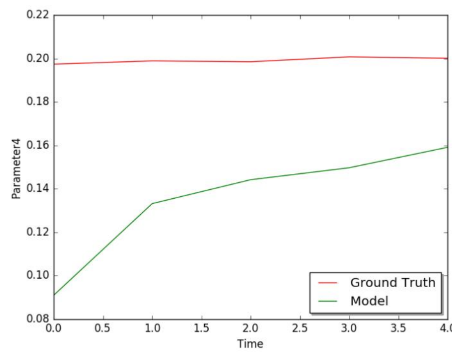
(a) Kronfit parameter α



(b) Kronfit parameter β_1



(c) Kronfit parameter β_2



(d) Kronfit parameter γ

Figure 3: Plots of Kronfit parameters for actual graph and our preferential attachment model across iterations/years

5 Conclusions and Future Work

We were able to see promising results on some of the goals we set out to accomplish. We were able to recover communities in international trade network which roughly correspond to the various international trade alliances formed thorough the years. We were also able to observe the effects of international events like the cold war, the Indo-China war, and dissolution of the Soviet Union in the communities formed in the trade network graph. We also presented a modified version of the preferential attachment model to explain temporal node shifts in community structure and a quantitative comparison of the model against the ground truth graph evolution data. Our model shows promising trends and seems to roughly mimic the true evolution in the network over the years. We have tried to present an exact quantitative comparison of our model with the ground truth, using the Kronfit Leskovec et al. (2010) algorithm.

Geography and Trade: Are they the same thing? There is a lot of scope for several interesting analyses on our dataset. One interesting question is how geographical factors like land and maritime distance affect trade and war between nations. We were able to perform community detection on multiple layers of the graph – trade, military alliances and territorial contiguity. In our research for this study, we found that the gravity theory Nello (2009) in economics is losing relevance as geographic distances start to reduce. We observed that there is strong correlation between territorial contiguity and trade ties to begin with, but we note that this reduces over time. We believe that correlating trade and territorial communities is a very promising direction of future work, and has the potential of uprooting the well agreed upon gravity model of international trade.

Exports vs Imports From an economic perspective, modeling communities in the directed network is a gray area, since this leaves a void for unmodeled, significant economic debt. When a nation has more exports vs imports (or the other way round), the directed network is an incomplete picture of international economic dependencies. If we can reliably estimate the value of international borrowings (such as the USA, which as a huge trade deficit), then we can evaluate community effects on directed trade networks, as well as model the same.

References

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Jackson, M. O. and Nei, S. (2015). Networks of military alliances, wars, and international trade. *Proceedings of the National Academy of Sciences*, 112(50):15277–15284.

Kim, J. Y. (2016). World databank. `http://databank.worldbank.org`.

Leskovec, J., Chakrabarti, D., Kleinberg, J., Faloutsos, C., and Ghahramani, Z. (2010). Kronecker graphs: An approach to modeling networks. *Journal of Machine Learning Research*, 11(Feb):985–1042.

Maoz, Z. (2016). Correlates of war project. `http://www.correlatesofwar.org/data-sets`.

Nello, S. (2009). *The European Union: Economics, Policies and History*. McGraw-Hill.

Nicosia, V., Mangioni, G., Carchiolo, V., and Malgeri, M. (2009). Extending the definition of modularity to directed graphs with overlapping communities. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03024.

Zhong, W., An, H., Gao, X., and Sun, X. (2014). The evolution of communities in the international oil trade network. *Physica A: Statistical Mechanics and its Applications*, 413:42 – 52.

Zhu, Z., Cerina, F., Chessa, A., Caldarelli, G., and Riccaboni, M. (2014). The rise of china in the international trade network: a community core detection approach. *PloS one*, 9(8):e105496.