Modeling Trading Relationships Among the Top 50 World Economies with Exponential Random Graph Models

Peter Racioppo
December 19, 2020

***Abstract*—The nature of international trading relationships is of fundamental importance in the study of global political economy, and is bound up with patterns of investment, development, poverty and international competition over markets and resources. This paper investigates the use of exponential random graph models in modeling the international trade network, and attempts to identify major determinants of export and import activity. Bottlenecks in world trade, the influence of large economies in shaping network structure, and the role of China and the United States are also examined. Geographic region, national economy, and population are found to have significant effects on trading relationships, but economic growth does not have a significant effect. Furthermore, we find that the top 50 economies can be organized into two to three clusters on the basis of their export and import relationships.**

***Index Terms*—World trade, social network analysis, exponential random graph models**

I. Introduction

Much work has been devoted to modeling the world economy as a dynamical system, including recent research on the ongoing effects of the COVID-19 pandemic [1]. Recent work has extended these methods with the use of game-theoretic approaches [2], differential games [3], time-series regression [4], and reinforcement learning [5], but time series and regression models may not adequately capture the relational character of much economic data. Social networks are a natural means of representing the complex interactions of production and exchange in the world economy. The international trade network, the network of trading ties between the world's economies, has been widely studied, including in the work of [6], who found that world trade networks have scale invariant properties, similar to networks generated by the preferential attachment mechanism studied by [7]. Research on more sophisticated generative models of world trade include a 2011 study which attempted to reproduce the statistical properties of the international trade network using a gravity model [8], but found that the ability of the model to predict links was poor, and a 2019 study modeling dynamic transport networks with matrix factor models [9]. Other studies have attempted to leverage network structure for inference, including a 2014 study on the rise of China [10], which studied community structure and found a correlation between regional and global dynamics.

Exponential-family random graph models (ERGMs) are generative models for networks which attempt to fit an exponential family, using sufficient statistics which describe nodal and edge covariates. These models can be fit with maximum likelihood estimation and Markov Chain Monte Carlo techniques [11], [12], [13], and are implemented in the ergm package in R [14].

The purpose of this paper is to study the international trade network of the fifty largest national economies in the world. In particular, this paper explores the role of large economies in shaping network structure, attempts to identify clusters of countries, and seeks to identify the role of geographic region, national economy, population, and economic growth in determining trading relationships. We also examine the rivalry between the United States and China, and the distinctions in the roles played by the two largest economies in the global trading network.

II. Constructing the Social Networks

I built a social network of the top 50 world economies using data from Worldometer, a reference website for real-time statistics, and the Observatory of Economic Complexity (OEC), a research project at MIT's Collective Learning group, which is now operated by Datawheel [15]. The OEC records the top five exporting and importing partners of every country during the previous year (2019). I combined this data with Worldometer data for GDP, population, annual GDP growth, GDP per capita, and share of world GDP for 2017 [16]. The latter data is two years out of date, but not readily available for 2019. I defined 7 geographical regions: North America, South America, Europe, Asia, Sub-Saharan Africa, North Africa and the Middle East, and Australia and the Pacific. The 50 nations in this data set together account for some 93% of world GDP and about 78% of the world's population, as of 2017.

I created a weighted, directed social network in which the top 50 countries are nodes, with out-edges denoting exports to top exporting partners and in-edges denoting imports from top importing partners. I created five such networks, recording the top 1, 2, 3, 4 and 5 trading partners. There are four node covariates in total: GDP, population, GDP growth, and region (GDP per capita and share of world GPD are redundant).

III. Social Network Analysis

Visualization:

The top trading partner network is shown in Fig. 1. In the top part of the figure, the edge thickness is proportional to trade volume; the bottom part of the figure shows countries with node areas proportional to GDP and geographical regions denoted by color. Of the top 50 economies, 21 are in Europe (yellow), 11 in Asia (red), 6 in South America (orange), 6 in the Middle East and North Africa (dark green), 3 in North America (blue), 2 in sub-Saharan Africa (light green), and 1 in the Pacific (purple). There are no isolates in the network. Three large clusters are apparent in the figure: an Asian cluster centered on China, a North American cluster centered on the US, and a European cluster centered on Germany. In each of these clusters, smaller economies are arranged around larger economies in star-graph layouts. Most edges in the graph appear to be between larger economies and from larger economies to smaller economies.

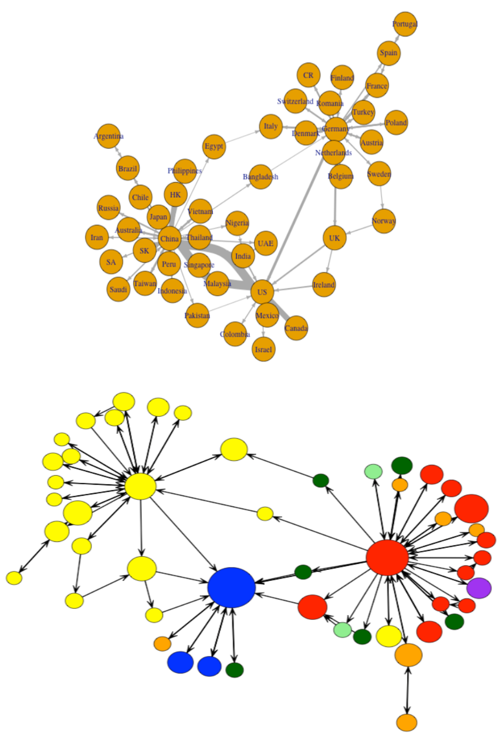

Figure 1: 1-partner trading relationships among the top 50 economies. Top: Edges are proportional to trade volume. Bottom: Node area is proportional to GDP and colors denote regions.

As we consider social networks with more trading partners, the network density increases from 4% (1 partner) to 17% (5 partners) and the three clusters become less distinct, with the US and China moving into almost overlapping positions at the center of the network, using the Fruchterman-Reingold layout, as shown in Fig. 2. However, we can continue to observe strong homophily based on geographic region. In the five-partner network shown in Fig. 2, a European cluster can be observed on the right of the network and an Asian cluster centered on Japan on the left. All but one of the nodes in the Sub-Saharan Africa and the Northern Africa & Middle East regions are clustered in the bottom section of the network, while the North American and South American countries are clustered in the top of the graph. Australia has strong ties with the countries in the Asian cluster. Germany is the closest of the European countries to the center of the network, but the UK and France (right of Germany) are also close to the center of the European cluster. Japan, which is the third largest economy, does not appear to play as central a role in the world economy as Germany, the fourth largest economy. This is likely due to the influence of China and the fact that there are fewer Asian countries than European countries in the top 50 economies. India (the large red circle in the bottom left) is positioned on the border separating the Asian economies and the African/Middle East economies, which reflects its geographic location.

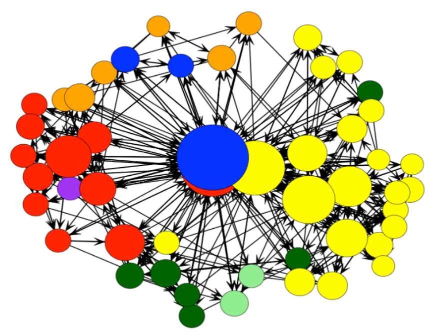

Figure 2: Top five trading relationships among the top 50 economies. The area of each node is proportional to its degree.

Centrality Measures:

To get a better idea of the economic influence of the countries in the network, I computed the in-degrees, out-degrees, and total-degrees of each of the nodes. In all five networks, China has the highest out-degree, Germany has the second highest, and the US has the third highest. The US has the highest-in degree in all but the 1-partner network, followed by China and

Germany. This reflects the position of the United States as the world's top importer and China as the world's top exporter. The US trails behind China and Germany in the 1-partner networks, but reaches parity with China in the 2-partner networks, and is the highest-degree node in the 3, 4, and 5 partner networks. China is the most important trading partner of the most nations, but the US seems to have a larger and more spread-out web of trading relationships. The top-5 trading network, with node areas scaled by degree, is shown in Fig. 2.

The degree correlation was negative for all 5 networks, indicating that the networks are disassortative, meaning that majority of trading relationships are between countries with high degree and countries with low degree. This tendency was most pronounced among the 1-partner network, with a degree correlation of -0.64, with the magnitude steadily decreasing with the number of trading relationships, falling to -0.30 for the 5-partner network. The out-degrees and in-degrees are correlated in all five networks, with 92% correlation in the top-1 partner graph and 66% correlation in the top-5 partners graph.

The out-degree, in-degree, and total-degree distributions all appear to follow power laws. Plotting the degree distribution on a log-log plot, as in Fig. 3, shows a good linear fit on a log-log scale, with higher-variance for nodes with lower degrees. This is consistent with the hypothesis that the generative model of the network is approximately a preferential attachment process, also known as a Yule process, in which nodes with higher degree more readily form new attachments. This "cumulative advantage" mechanism is a natural way to model relationships in a market economy.

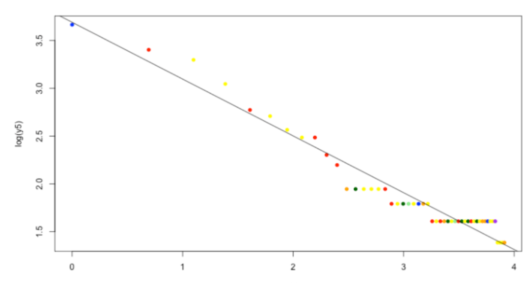

Figure 3: Log-log plot of the degree distribution.

I fit Yule, Waring, Poisson, and Conway-Maxwell-Poisson models to the degree distributions for all five networks. The AIC and BIC values of the models were minimized by the Waring distribution for the 1-partner network, the Poisson distribution for the 2-partner network, and the Yule distribution for the 3, 4, and 5-partner distributions. In the latter three cases, the Pareto distribution was the second most optimal fit, while the Waring, Poisson, and CMP distributions were far worse.

This is again consistent with a preferential attachment process.

I also computed the eigenvector-centrality, the closeness-centrality, and the betweenness-centrality of the nodes. The eigenvector-centrality, displayed for the 3-partner network in Fig. 4, measures how often each node is visited in an infinitely-long random walk through the network. In particular, consider the graph $G = \langle V, E \rangle$. Given a vertex x, denote the set of vertices with a directed edge out of x as out(x) = $\{w \in V | x \to w\}$, and the set of vertices with a directed edge into x as in(x) $= \{y \in V | y \to x\}$. Let $\pi(x) = \sum_{y \in \text{in}(x)} \pi(y)/|\text{out}(y)|$, which can be thought of as a measure of the total flow into a node. We can sample from this distribution by defining a Markov chain transition probability $K(y, x) = 1/|\text{out}(y)|$, since in this case $\sum_{y \in V} \pi(y) K(y, x) = \pi(x)$. The stationary distribution of this Markov chain can be cheaply computed via power iteration [17].

Eigenvector centrality is a more global measure of centrality than degree. It appears to increase the prominence of the exporting countries outside Europe and North America. China is dominant in the smaller networks, but roughly equal to the US in the larger networks.

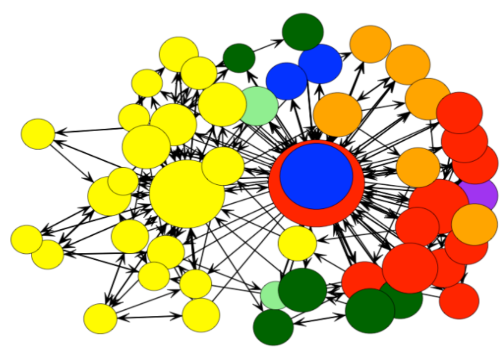

Figure 4: Top-3 trading partners network with node areas scaled by their eigenvector-centrality.

The closeness-centrality scores – the reciprocal of the sum of the distances from a node to every other node – are quite uniform throughout the network, due its centralized structure: any node is able to quickly reach any other node through its connections with the central countries (especially the US, China, and Germany).

The betweenness-centrality – the number of geodesics passing through a node – exaggerates the importance of the most central nodes. China, Germany, and the US dominate the network. China and Germany play the largest roles in the top 1-4 trading networks but are surpassed by the US in the top-5 trading network.

To get a better idea of the structure of the networks, I constructed a second set of networks in which the US, China, and Germany are removed. The result is displayed in Fig. 5. The network appears far less clustered, but it still possible to observe clustering of the European countries around the three large EU economies of France, the UK, and Italy, and some clustering of the Asian countries around Japan. The network appears to still be quite well-connected, which can be taken as an indication of the robustness of trading routes.

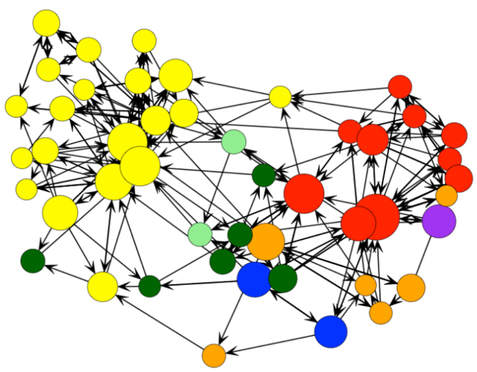

Figure 5: The 5-partner trading network with the US, China, and Germany removed.

Bottleneck and Conductance of a Simplified Network:

The bottleneck of a graph G, the edge in the graph with maximum stochastic flow, is defined as $\kappa = \max_{e \in E} \sum_{\Gamma_{xy} \ni e} \gamma_{xy} \pi(x)\pi(y)$, where $\gamma_{xy} = \sum_{<s,t> \in \Gamma_{xy}} \frac{1}{\pi(s)K(s,t)}$. The conductance, which is a measure of how "well-knit" the graph is, is defined as follows. Suppose the vertex state space is partitioned into two disjoint subspaces, $\Omega = S \cup S^c$. Let $\pi(S) = \sum_{x \in S} \pi(x)$, $K(S, S^c) = \sum_{x \in S} \sum_{y \in S^c} K(x, y)$, and $Q(S, S^c) = \sum_{x \in S, y \in S^c} \pi(x)K(x, y)$. The conductance is then $h = \min_{S:\pi(S) \leq \frac{1}{2}} \frac{Q(S,S^c)}{\pi(S)}$. Both of these quantities can be used to bound λ_{slem}, the the second largest eigenvalue modulus of a stochastic matrix. The convergence of an irreducible, aperiodic Markov chain is in turn bounded by λ_{slem} [18].

I built a 5-node weighted, directed network of the top four economies, with the fifth node representing all other economies combined, and the edge weights denoting percent of world imports and exports. The stationary probability distribution of the Markov chain defined by this stochastic matrix is 0.26 (US), 0.20 (China), 0.09 (Japan), 0.10 (Germany), 0.34 (Other economies), which appears to be quite close to the shares of world GDP, 0.24 (US), 0.16 (China), 0.06 (Japan), 0.04 (Germany), 0.40 (Other economies).

The bottleneck edge in this reduced network is the edge which represents exports from Germany to the US. The minimum conductance edges are from Germany to the world market and from China to Germany. The conductance of the network, which is the lowest edge conductance, is 0.69.

Additional Metrics:

I computed the edge-betweenness – the number of geodesics passing through an edge – as a measure of the importance of trading relationships. The four edges with the highest edge betweenness in all five networks involved either Germany or China, with the exception of only one edge.

I also computed the common neighbors of the three most central economies, the US, China, and Germany, in order to measure overlaps in their trading relations. Let us define the number of common neighbors of nodes i and j as $CN(i, j) = |S_i \cap S_j|$, where S_i is the neighbor set of node i and S_j is the neighbor set of node j, and the Jaccard measure, a scaled version of the common neighbors measure, as $Jaccard(i, j) = \frac{|S_i \cap S_j|}{|S_i \cup S_j|}$. The Jaccard measure of the US-China relationship is higher than the US-Germany and the China-Germany relationships in all five of the partnership networks. This is a characterization of the fact that the US and China both occupy a more central position in the networks than does Germany, and their trading partners thus tend to overlap to a greater degree.

IV. Community Detection

I implemented the Markov Clustering (MCL) algorithm, a fast, unsupervised clustering algorithm which simulates stochastic flow through iterative "expansion" and "inflation" steps. The expansion step simulates random walks through the graph with matrix multipications of the graph adjacency matrix. In the inflation step, the elements of the matrix are exponentiated and the columns are then renormalized. The effect of this operation is to weaken weaker connections and strengthen strong connections, which tends to break the graph into separate star graphs [19].

I constructed undirected versions of each trade network, in which an edge exists between two nodes if there is a directed edge between them in the directed networks, and tested my implementation of the MCL algorithm on the undirected 1-partner network. With an expansion parameter of 3 and inflation parameter of 4, the network, pictured at the top of Fig. 6, clustered into two connected communities: a purely European community, centered on Germany, and all other countries, centered on China. As shown at the bottom of Fig. 6, with an expansion parameter of 2 and inflation parameter of 6, the clusters broke into additional communities: a European community, again centered on Germany, a North American community, centered on the United States, two small clusters of European countries, and a large cluster of all other countries, centered on China.

I cross-referenced these results with those produced by the fast greedy algorithm and the edge-betweenness

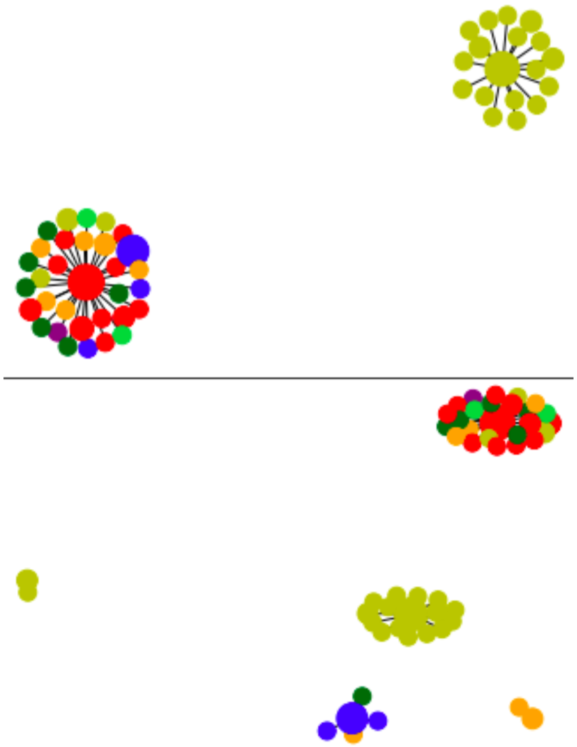

Figure 6: Markov clustering applied to the 1-partner trading network. Top: expansion parameter of 3 and inflation parameter of 4. Bottom: expansion parameter of 2 and inflation parameter of 6.

algorithm. The fast greedy algorithm is an efficient clustering algorithm which starts with a subnetwork of the most highly-connected nodes, and then randomly samples edges, checking whether the new edges improve the modularity scores of the subnetworks, until convergence. The edge-betweenness algorithm is also a greedy algorithm, but it operates in the opposite direction, starting with the original network, and progressively removing the edges that have the lowest edge-betweenness scores. The clusters produced in the 1-partner network are shown for the fast greedy algorithm in Fig. 7 and for the edge-betweenness algorithm in Fig. 8. These algorithms, with the default parameters, produced clusters similar to those produced by the second MCL. Namely, the majority of nodes are clustered around China, with secondary North American and European clusters, clustered around the United States and Germany, respectively.

V. Fitting Exponential Random Graph Models

I fit a series of ERGMs to the trading partner networks, beginning with a homogeneous Bernoulli model and models for homophily on region, GDP, population, and GDP growth (using the nodematch and absdiff terms). There appears to be strong homophily on GDP, population, and region in all five networks, but GDP

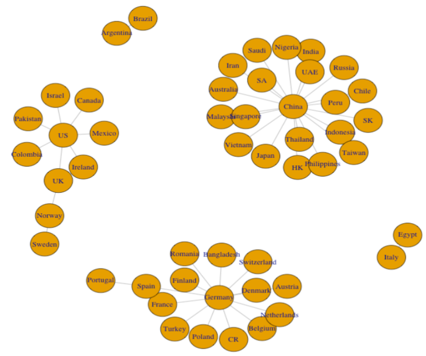

Figure 7: Fast-greedy algorithm for community detection.

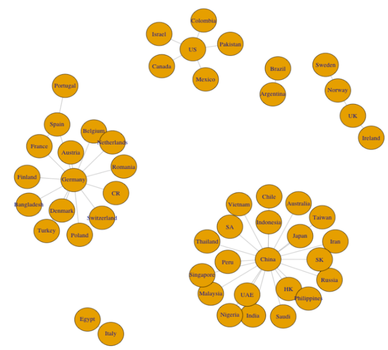

Figure 8: Edge-betweenness algorithm for community detection.

growth was not a statistically significant covariate, when controlling for the other covariates. GDP appears to have the largest effect of all the covariates in all five networks (in that its coefficient has the largest magnitude in the models, and models independently fit on each covariate had the lowest AIC and BIC values when fit on GDP).

Fitting for differential homophily on region, I found that only Asia, Europe, and South America had statistically significant homophily. However, the addition of differential homophily slightly increased the AIC and BIC scores of the fits.

Comparing AIC and BIC values, I found that, for both the 1-partner and the 5-partner networks, the fits improved after adding homophily on GDP to the homogeneous Bernoulli model, and further improved after adding homophily on population and region. Adding a term for mutual ties also improved the fit. I found (using the ttriple term) that the triads in the networks also have a statistically significant tendency to be transitive.

I also tried adding terms for sender and receiver nodes. In the 1-partner network, China is the only statistically significant sender and the US is the only statistically significant receiver. In the 5-partner network, the statistically-significant sender nodes are China, the US, and Germany and the statistically-significant receiver nodes are China, the US, Germany, Japan, the UK, India, and France. Including terms for these nodes removed the statistical significance of population in both the 1-partner and 5-partner networks and decreased the AIC and BIC values of the fits.

When controlling for homophily on GDP, region, and population and for mutual ties, the statistically significant triads were 012 (one uni-directional edge per triad), 021C (one uni-directional two-path per triad), and 300 (a triad with three bi-directional edges). The predominance of the 012 and 021C triads is a reflection of the sparsity of the network and the fact that many relationships are uni-directional. The presence of the 300 triad indicates that the transitive relationships tend to be bidirectional. Evidently, the cliques of countries that heavily trade with each other prefer to both export and import to each other.

Fitting ERGMs to the 5-partner network with the US, China, and Germany removed again produced fits with strong homophily on GDP and population, dependence on mutual ties, and no dependence on population or GDP growth rate.

To assess the quality of the best fit models, I simulated 1-partner and 5-partner trading networks using these models, which included terms for the number of edges, homophily on GDP and population, sender/receiver and mutual ties for select countries, and, in the former case, transitivity. These results are displayed in Fig. 9. Both networks display clustering by region and a preponderance of ties from large economies to small economies. The simulated 1-partner network reproduces the star-graph appearance and the approximate number of connections of the clusters around China and the United States, but overestimates the connectivity of the European countries, and incorrectly includes three isolates. The simulated 5-partner network reproduces the clustering of the larger economies at the center of the network (using the Fruchterman-Reingold layout). As in the actual graph, the United States and China occupy almost exactly the same position in the graph. The simulated graph also captures the high degree of connectivity among the European states.

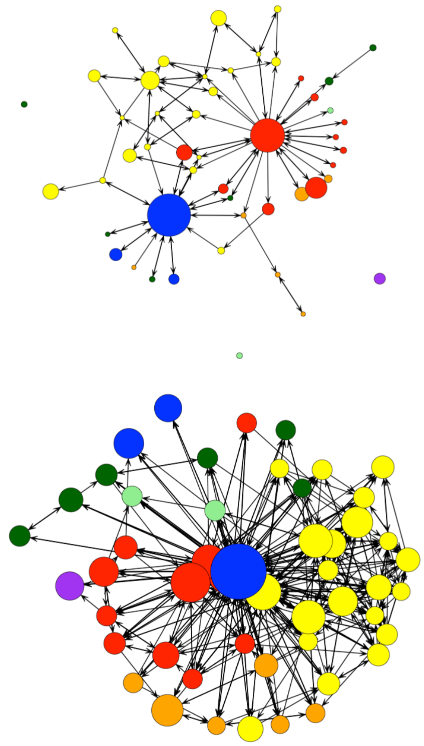

Figure 9: Simulated trading networks using fits on the number of edges, homophily on GDP and region, and sender/receiver and mutual ties. Top: simualted 1-partner network. Bottom: Simulated 5-partner network.

VI. Conclusion

The results in this paper demonstrate that the main features of the international-trade network — the dominant role of a few large economies in shaping network structure, clustering by geographical region, and the star-graph like character of the top national trading relationships — can be captured by exponential random graph models. It is somewhat surprising that ERGMs are capable of learning fairly accurate models of international trade using only the geographical region and gross domestic product of each country, and the presence of major trading relationships, without access to trade volume data. The next logical step to improve the quality of the ERGM models is to include these data (using e.g. ergm.count). A network with many more nodes and edges, including all 189 or so countries in the world and perhaps splitting up the larger countries into smaller regions, would provide a much richer representation of trading relationships. Latent variable models and other modeling techniques should be considered, and other node covariates should be included in the analysis, such as membership in trading agreements, military alliances, the goods which are exported and imported, and other economic indicators like debt. Trading relationships change over time, and more sophisticated models should be able to capture the dynamical behavior of the international trade network, and not only its structure. These relationships are further complicated by the existence of transnational corporations, which produce goods in multiple countries for sale on the world market. Capturing these relationships in greater detail is an important and challenging task.

References

[1] I. M. Trotter, L. A. C. Schmidt, B. C. M. Pinto, A. L. Batista, J. Pellenz, M. Isidro, A. Rodrigues, A. G. S. Suela, and L. Rodrigues, "Covid-19 and global economic growth: Policy simulations with a pandemic-enabled neoclassical growth model," 2020.

[2] Y. Ikeda, "An interacting agent model of economic crisis," 2020.

[3] C. K. Leong and W. Huang, "A stochastic differential game of capitalism," pp. 552 – 561, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304406810000297
[4] J. Shi, "Comprehensive time-series regression models using gretl u.s. gdp and government consumption expenditures gross investment from 1980 to 2013," 2014.
[5] A. Charpentier, R. Elie, and C. Remlinger, "Reinforcement learning in economics and finance," 2020.
[6] K. Bhattacharya, G. Mukherjee, and S. S. Manna, "The international trade network," p. 139–147. [Online]. Available: http://dx.doi.org/10.1007/978-88-470-0665-$2_1$0
[7] G. U. Yule, "A mathematical theory of evolution, based on the conclusions of dr. j. c. willis, f.r.s."
[8] M. Duenas and G. Fagiolo, "Modeling the international-trade network: A gravity approach," 2011.
[9] E. Y. Chen and R. Chen, "Modeling dynamic transport network with matrix factor models: with an application to international trade flow," 2019.
[10] Z. Zhu, F. Cerina, A. Chessa, G. Caldarelli, and M. Riccaboni, "The rise of china in the international trade network: A community core detection approach," p. e105496, Aug 2014. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0105496
[11] M. S. Handcock, "Assessing degeneracy in statistical models of social networks," 2003.
[12] T. A. B. Snijders, "Markov chain monte carlo estimation of exponential random graph models," 2002.
[13] G. L. R. Tom A. B. Snijders, Philippa E. Pattison and M. S. Handcock, "New specifications for exponential random graph models," 2002.
[14] C. T. B. S. M. G. P. N. K. Mark S. Handcock, David R. Hunter and M. Morris, "ergm: A package to fit, simulate and diagnose exponential-family models for networks," 2010.
[15] "Exports, imports and economic complexity data at subnational level," 2020.
[16] "Gdp by country. worldometer; https://www.worldometers.info/gdp/gdp-by-country," 2020.
[17] A. Barbu and S. Zhu, "Monte carlo methods," 2020.
[18] Y. Wu, "A note on monte carlo methods (ucla statistics)," 2017.
[19] "Mcl - a cluster algorithm for graphs."