Massachusetts Institute of Technology

Measuring Trade Profile with Granular Product-level Data*

In Song Kim[†]   Steven Liao[‡]   Kosuke Imai[§]

Author order: 1. In Song Kim, 2. Steven Liao, 3. Kosuke Imai

Short Title: Measuring Trade Profile

Key Words: clustering algorithm, trade competition, big data, comparative advantage, intra-industry trade, zero trade, dyadic data

Measuring Trade Profile with Granular Product-level Data

Abstract

The product composition of bilateral trade encapsulates complex relationships about comparative advantage, global production networks, and domestic politics. Despite the availability of product-level trade data, most researchers rely on either the total volume of trade or certain sets of aggregated products. In this paper, we develop a new dynamic clustering method to effectively summarize this massive amount of product-level information. The proposed method classifies a set of dyads into several clusters based on their similarities in *trade profile* — the product composition of imports and exports — and captures the evolution of the resulting clusters over time. We apply this method to two billion observations of product-level annual trade flows. We show how typical dyadic trade relationships evolve from sparse trade to inter-industry trade and then to intra-industry trade. Finally, we illustrate the critical roles of our trade profile measure in International Relations research on trade competition.

Replication Materials: The data, code, and any additional materials required to replicate all analyses in this article are available on the American Journal of Political Science Dataverse within the Harvard Dataverse Network, at: `https://doi.org/10.7910/DVN/SKPPLH`.

Word Count: 9,210 (abstract: 147)

1 Introduction

Since Ricardo, scholars have relied upon the concept of comparative advantage to explain why countries trade and to identify the winners and losers of trade (e.g., Rogowski, 1987; Scheve and Slaughter, 2001; Hiscox, 2002). While comparative advantage still plays a central role in explaining trade, consumer preferences for product variety and the use of global production chains by firms have dramatically altered patterns of international trade. The fast-growing political economy literature on product- and firm-level theories demonstrate the importance of examining bilateral trade at the granular level in understanding the distributional consequences of international trade (e.g., Antràs and Staiger, 2012; Grossman and Rossi-Hansberg, 2012; Jensen, Quinn, and Weymouth, 2015; Osgood, 2016; Kim, 2017).

Despite the substantive importance of products and the massive amount of product-level bilateral trade flow data that are becoming available, most studies still rely on the total volume of trade aggregated across products (e.g., Mansfield, Milner, and Rosendorff, 2000; Gartzke, 2007; Tomz, Goldstein, and Rivers, 2007; Carnegie, 2014) or certain sets of aggregated products (e.g., Dorussen, 2006; Elkins, Guzman, and Simmons, 2006; Goenner, 2010; Chatagnier and Kavaklı, 2017). For many, computational and methodological challenges prohibit effective summaries of the massive amount of product-level data and preclude insights based on product composition. For example, our data set, which is based on the United Nations Comtrade Database, covers more than 600 products and 59,000 directed dyads over 53 years (1962 – 2014). As illustrated in Figure 1, this yields approximately two billion observations of product-level bilateral trade flows. Identifying systematic patterns in such data is difficult for several reasons. The high-dimensionality of the data and a large number of meaningful comparisons can easily overwhelm researchers conduct-

1

ing simple descriptive analyses. Bigger data sets also may contain more noise, which can mask important systematic patterns. Regression models are also of limited use because they require researchers either to consider each product separately or to aggregate trade flows across multiple products, overlooking the composition of trade as a whole.

[Figure 1 about here.]

In this paper, we address this product-level trade data challenge by developing a new dynamic clustering method. Specifically, we group country-pairs into a fixed number of clusters based on the similarity of their *trade profile*, defined as the product composition of imports and exports. For example, U.S.–South Korea may be in the same cluster with U.S.–Japan because their current bilateral trades involve similar exchanges of chemical products and cars. However, the two dyads might have belonged to different clusters in the 1960s when they traded disparate products. This approach is different from Hidalgo and Hausmann (2009) who use product-level trade data to infer the relationships between products. In contrast, we model the dynamic patterns of trade between countries over time based on their trade portfolio.

We focus on dyadic trade relationships based on their product-level trade for two reasons. First, countries still play an important role in controlling the movements of goods as they set trade policies and negotiate international agreements. In addition, the growing number of bilateral trade agreements, in contrast to stagnant multilateral negotiations, attests to the significant and heterogeneous interests countries have vis-à-vis their partners. Second, the proposed dyadic clustering method allows researchers to distinguish bilateral trade relationships based on the *types* of products that countries exchange. This is in sharp contrast to the long-held approach where researchers consider the total *volume* of trade across certain sets of aggregated products or of

2

each separate product. The use of highly disaggregated products in clustering is also consistent with the recognition of firms as key political actors. That is, countries face different types of domestic and international political constraints as firms vary in their choice of entering foreign markets (Eaton, Kortum, and Kramarz, 2011) and their distinct global ties with partners across multiple production stages (Johns and Wellhausen, 2016). In sum, we consider the distribution of product-level bilateral trade in its entirety to characterize the nature and evolution of dyadic trade over time.

We overcome several methodological challenges that are unique in dealing with trade data. In particular, we model zero trade explicitly. In fact, many countries do not trade with each other, and the prevalence of zero trade becomes even more pronounced once we consider product-level trade. While there exist increasing concerns in the literature about systemic differences between dyads who trade versus those who do not (e.g., Silva and Tenreyro, 2006), most applied research still excludes non-trading dyads entirely from their analysis (e.g., Mansfield, Milner, and Rosendorff, 2000; Tomz, Goldstein, and Rivers, 2007). Furthermore, our dynamic clustering method, which is based on a hidden Markov model (Frühwirth-Schnatter, 2007; Park, 2012), allows researchers to effectively compare the composition of product-level trade not only across dyads (including non-trading pairs) but also across time given a dyad. Finally, we derive a fast Expectation-Maximization (EM) algorithm to address the computational challenges in modeling the evolution of bilateral trade relations over time (Dempster, Laird, and Rubin, 1977).

We find that there exists a path along which typical dyadic trade relationship evolves. Specifically, we show that most dyads engage in little trade with each other, but when they do they start by relying on comparative advantages especially in exporting crude materials and manufacturing goods. This relationship then evolves into intra-industry trade, in which two countries

3

simultaneously export and import products within the same manufacturing industry. Although many previous studies have identified comparative advantage, increasing returns to scale, and consumers' love of variety as distinct sources of gains from trade (e.g., Krugman, 1979), to the best of our knowledge, no study exists to identify dynamic changes and the sequence in their relative importance in characterizing dyadic trade relations at this level of disaggregation and scope. We also contribute to the literature that emphasizes the links between trade and development (Grossman and Helpman, 1990; Redding, 1999) by identifying the *timing* of structural transition for each dyad as well as the set of products that play distinct roles in the evolution of global trade.

Finally, while our cluster membership serves as a simple summary of complex bilateral trade patterns, we also demonstrate that this measure can be used to capture key variables of interest in International Relations research. In particular, we construct an improved measure of trade competition that encapsulates the degree to which two countries trade similar products with the same partners. Using our measure, applied researchers can effectively examine whether bilateral trade competition affects other state behaviors in international politics.[1]

The open-source software, `dynCluster: Dynamic Clustering Algorithm`, is available as an R package for implementing the proposed methods. All dyad-year cluster memberships, the measure of trade competition, and visualization tools used in this paper will also be made publicly available.

---

[1] In Appendix G, we provide an example of such application and show that trade competition has little effect on increasing the likelihood of signing Bilateral Investment Treaties (BITs) unlike previous studies (Elkins, Guzman, and Simmons, 2006).

4

2 Data and Methodology

2.1 Annual Product-Level Dyadic Trade Data

We analyze annual SITC (Standard International Trade Classification) four-digit product-level dyadic trade data from 1962 to 2014.[2] SITC is a widely used classification of internationally traded goods that is maintained by the United Nations. The classification reflects the materials used in production, the processing stage, uses of the products, and technological changes — facilitating economic analyses of long-term trends of international trade across various products.[3] Moreover, its hierarchical structure is useful for aggregating and disaggregating different sets of products and industries for analytic purposes where a four-digit classification gives the most detailed classification of products available for a large number of countries and periods. For example, the SITC commodity 6513 is "Cotton yarn & thread, grey, not mercerized," which belongs to Section 6 (Manufactured goods), Division 65 (Textile), and Group 651 (Textile yarn).[4]

To ensure that product classifications are comparable across the five decades, we use the list of all 625 SITC Revision-1 four-digit products consistently across the entire period. When countries report their trade statistics based on a different revision number, the United Nations Statistics Division maps them to the corresponding Revision-1 product using concordance tables.[5] We use the resulting data in our analysis. We then consider a total of 244 states based on the list of 289

---

[2] See Appendix D for a comparison with another widely used database.

[3] http://unstats.un.org/unsd/iiss/Print.aspx?Page=Standard-International-Trade-Classification

[4] We find that there exists significant variation at the 4-digit level. Specifically, almost 70% of the variation in trade volume can be explained by the variation across 4-digit products *within* 3-digit industry categories.

[5] The concordance table is available at https://unstats.un.org/unsd/trade/classifications/correspondence-tables.asp

country and region codes available from the United Nations Comtrade Database.[6] Out of the list, we include all countries and political entities that have existed for at least one year during the period while only excluding regional entities such as the European Union. For example, we include United Nations' non-member observer states such as Palestine. In addition, we consider newly independent countries (e.g., Belarus) as unique states after independence but record them as part of another distinct state (e.g., the Soviet Union) prior to independence. Likewise, we include three unique German states: the German Democratic Republic (East Germany) and the Federal Republic of Germany (West Germany) from 1962 to 1990, and unified Germany since 1991.

For each product and country pair, we record the volume of trade (measured as its value in US dollar). While the Comtrade database is one of the best sources available for trade data widely used in academic research, it is still possible that certain countries may fail to report their trade activities, especially for highly disaggregated commodity categories. Thus, we carefully check the availability of data for each product and partner: When reports on product-level trade are available from both importer and exporter, we use importer's valuation which generally includes the cost of transportation and insurance to the frontier of the importing country or territory, i.e., CIF (Cost insurance and freight) valuation. We use exporter's reports when no additional information from the importing country is available. When neither importer nor exporter reports a positive volume of trade, we consider the product as not being traded.[7] Although this choice might introduce some discrepancies due to the difference between CIF and FOB (Free on board) valuation, the reliance on reports from both direction ensures that products with some trade are

---

[6]The complete list of the country and region names along with their years of existence are available from http://unstats.un.org/unsd/tradekb/Knowledgebase/50377/Comtrade-Country-Code-and-Name

[7]One might view the zero trade cutoff arbitrary. For instance, the import statistics of the U.S. consist of goods valued at more than $2,000. We deal with this issue by modeling the selection probability explicitly in Section 2.2.

6

identified. Note that 72.4% of all dyads have a positive trade in at least one product for any given year. Since we also consider the absence of bilateral trade at the product level, we have a data set of approximately two billion observations ($\approx 244 \times 243 \times 625 \times 53$).[8] Table D.1 in Appendix D reports descriptive statistics for our dyadic trade data and shows the prevalence of sparse trade.

2.2 A New Dynamic Clustering Algorithm for Dyadic Data

We develop a new dynamic clustering algorithm to summarize the evolution of global trade patterns at the product level. Given the enormous size of our data, it is extremely difficult, if not entirely impossible, to discover systematic patterns by simply "looking at" the data. In this setting, a probabilistic model can provide useful summaries of this large data set. We develop a dynamic finite mixture model (Frühwirth-Schnatter, 2007; Imai and Tingley, 2012) and identify a pre-specified number of latent clusters, each of which represents a distinct pattern of bilateral trade.

The primary goal of the proposed method is to assign a cluster membership to each dyad so that a set of dyads with similar trade profiles, i.e., product compositions of exports and imports, are grouped together. We consider bilateral trade across all products in its entirety instead of focusing on either the total volume of trade or arbitrary sets of aggregated products separately from one another. In this regard, the proposed algorithm helps conduct systematic comparisons of trade composition across a large number of country-pairs given substantial noise in disaggregated product-level trade data (Mahutga, 2006). Furthermore, we allow the cluster membership of each dyad to evolve over time. In this way, the algorithm captures the dynamic patterns of global trade profile. Researchers specify the number of clusters based on the desired degree of summarization where a greater number of clusters implies a finer level of summary (we also offer a data-driven

---

[8]The raw data are downloaded from the UN Comtrade Database using its data extraction API.

method to choose the number of clusters below).

Methodology. Suppose we have a total of N countries over T years. The proposed algorithm requires researchers to choose the number of clusters, which is represented by M (though see below for a data-driven method to select the number of clusters). Let $Z_{ijt} \in \{1, 2, \ldots, M\}$ be a latent cluster membership for a dyad consisting of country i and country j in year t where $i, j \in \{1, \ldots, N\}$, $i < j$, and $1 \le \underline{T}_{ij} \le t \le \overline{T}_{ij} \le T$ with $\underline{T}_{ij}$ and $\overline{T}_{ij}$ representing the start and end years of the dyad, respectively. We allow different start and end years for each dyad because some countries do not exist for the entire period.

For the same dyad, $X_{ijtk} \in [0, \infty)$ represents the export of product k from country i to country j in year t. Similarly, X_{jitk} is the trade flow of opposite direction, representing the export of the same product k from country j to country i. We are interested in the trade profile or product composition of trade for each annual dyadic trade flow. To do this, we first compute the trade proportion for each product relative to the total volume of a given trade flow, $V_{ijtk} = X_{ijtk}/\sum_{k'=1}^{K} X_{ijtk'}$ such that $\sum_{k=1}^{K} V_{ijtk} = 1$ where K is the total number of products.[9] Then, a dyadic trade profile for country i and j in year t can be characterized by a $2K \times 1$ stacked vector $\mathbf{V}_{ijt} = (V_{ijt1}, \ldots, V_{ijtK}, V_{jit1}, \ldots, V_{jitK})^{\top}$ where $i < j$.

When clustering dyadic trade profile, the results should not depend on how the stacked vector of dyadic trade profile $\mathbf{V}_{ijt}$ is created. Specifically, although we defined $\mathbf{V}_{ijt}$ such that $i < j$ (so as to avoid double-counting the same dyad), this is an arbitrary constraint. Indeed, we can define $\mathbf{V}_{ijt}$ by flipping the order in which trade profiles of exports and imports are stacked. That is,

---

[9]Alternatively, if we wish to account for the relative trade volume of imports and exports, we can instead use the trade proportion relative to the total volume of both exports and imports rather than calculating it separately, i.e., $V_{ijtk} = X_{ijtk}/\sum_{k'=1}^{K}(X_{ijtk'} + X_{jitk'})$.

8

we can stack the trade profile for the exports from country j to country i on the top of that for the exports from country i to country j, i.e., $\mathbf{V}_{ijt}^* = (V_{jit1}, \ldots, V_{jitK}, V_{ijt1}, \ldots, V_{ijtK})^\top$. As a consequence, two dyad-year observations with similar dyadic trade profile, i.e., $\mathbf{V}_{ijt} \approx \mathbf{V}_{i'j't}$, may appear completely different if the order in which trade profiles are stacked is reversed, i.e., $\mathbf{V}_{ijt} \not\approx \mathbf{V}_{i'j't}^*$.

To address this "flipping problem," we create a total of $2M$ pseudo clusters so that each cluster corresponds to two pseudo clusters. This enables us to account for two possible ways in which one dyadic trade profile is similar to another (by flipping the order of stacking trade profiles for one of the dyad-year observations). We use $Z_{ijt}^* \in \{1, 2, \ldots, 2M\}$ to represent this pseudo cluster membership where for a dyad-year observation with $Z_{ijt} = z$ we have either $Z_{ijt}^* = z$ (the dyad-year belongs to cluster z without flipping) or $Z_{ijt}^* = z + M$ (the dyad-year belongs to cluster z once flipped). Since the model parameters stay identical for clusters z and $z + M$, these two pseudo clusters form one final cluster.

[Figure 2 about here.]

Figure 2 illustrates the proposed dyadic clustering methods when there are six distinct dyads consisting of four countries. Each country-pair exchanges four products. To capture the difference in trade profiles, we use three colors (blue, white, and red) to denote different levels of trade (low, medium, and high). It is clear from the figure that Dyad 3 and Dyad 5 exhibit similar patterns of trade whereby one country exports more of the first two products while the other exports more of the latter two. Likewise, Dyad 4 and Dyad 6 form Cluster 3 because countries 2 and 3 trade similarly compared to countries 3 and 4. The formation of Cluster 1 demonstrates the "flipping problem." Although Dyad 1 and Dyad 2 exhibit seemingly distinct patterns of trade, the trade

9

profile at the dyad-level resembles each other. Specifically, the stacked vector of $(\boldsymbol{V}_{12}^\top, \boldsymbol{V}_{21}^\top)^\top$ is similar to $(\boldsymbol{V}_{31}^\top, \boldsymbol{V}_{13}^\top)^\top$ once the order of data involving countries 1 and 3 is reversed. Thus, our algorithm groups the two dyads into one cluster given that countries 1 and 2, and 1 and 3 exhibit similar patterns of trade at the *dyadic* level.

As shown in Table D.1, a significant proportion of products have zero trade for many dyad years. Thus, we first model zero trade given a latent pseudo cluster membership,

$$D_{ijtk} \mid Z_{ijt}^* = z \quad \sim \quad \text{Bernoulli}(q_{kz}) \quad \text{for} \quad k = 1, \ldots, K \tag{1}$$

where $D_{ijtk} = \mathbf{1}\{V_{ijtk} = 0\}$. An important constraint here is $q_{kz} = q_{k,z+M}$ because two pseudo clusters, i.e., $Z_{ijt}^* = z$ and $Z_{ijt}^* = z + M$, imply the same cluster.

We then model the proportion of trade among non-zero trade products using the log normal distribution (Aitchison, 1982). This part of the model is defined as follows,

$$W_{ijtk} \mid D_{ijtk} = 0, Z_{ijt}^* = z \quad \sim \quad \mathcal{N}(\mu_{kz}, \ \sigma_{kz}^2) \quad \text{for} \quad k = 1, \ldots, K - 1 \tag{2}$$

where $W_{ijtk} = \log \frac{V_{ijtk}}{V_{ijtK} + c}$ with the baseline product K and c is a small constant used to avoid division by zero. We use a value of $c = 0.0001$ in our application. Again, we have important parameter constraints, i.e., $\mu_z = \mu_{z+M}$ and $\sigma_{kz}^2 = \sigma_{k,z+M}^2$, based on the relationship between pseudo clusters and clusters. While in theory one can allow for correlations across products, we assume independence given the computational challenge due to a large number of products in our data.[10] Next, we use the Hidden Markov Model so that cluster membership for a given dyad

---

[10]One possible approach is to incorporate the regularized estimation of a large covariance matrix into our dynamic clustering analysis (e.g., Bickel and Levina, 2008; Friedman, Hastie, and Tibshirani, 2008). We leave such an extension for future research.

10

changes over time (Frühwirth-Schnatter, 2007; Park, 2012).

$$Z_{ijt}^* \mid Z_{ij,t-1}^* = z \ \sim \ \text{Multinomial}(P_{z1}, \ldots, P_{z,2M}) \quad \text{for} \quad i < j \tag{3}$$

where $P_{zz'}$ is the transition probability from cluster z to cluster z'. Appendix A describes the details of this algorithm.

Quantities of Interest. To characterize the resulting clusters, we use the mean trade proportion for each product given a cluster. Note that the model parameter μ_{kz} is difficult to interpret because it is based on the log proportion scale relative to the arbitrary baseline product. Therefore, we estimate the average product proportion relative to the total *trade volume*, $\mathbb{E}(V_{ijtk})$, for product k given cluster z by Monte Carlo simulation. Specifically, we first sample W_{ijkz} from $\mathcal{N}(\hat{\mu}_{kz}, \hat{\sigma}_{kz}^2)$ for $k = 1, \ldots, K-1$ where $\hat{\mu}_{kz}$ and $\hat{\sigma}_{kz}^2$ are the maximum likelihood estimates of μ_{kz} and σ_{kz}^2. We then estimate the expected trade proportion $\mathbb{E}(V_{ijtk})$ by $\frac{1}{L}\sum_{l=1}^{L}\{\exp(w_{kl})/\sum_{k'=1}^{K}\exp(w_{k'l})\}$ where w_{kl} is the lth Monte Carlo draw of W_{ijk}, $w_{Kl} = 0$ for all l, and L is the total number of Monte Carlo draws. These estimates facilitate substantive interpretation of each cluster as we demonstrate below.

Choosing the Number of Clusters. We propose a data-driven approach to selecting the number of clusters based on the hold-out likelihood criteria. An advantage of this approach is that it avoids overfitting. We caution, however, that this type of data-driven approach, which measures the goodness-of-fit of the model, may not necessarily optimize the interpretability of the results. Thus, we suggest that researchers try different numbers of clusters and examine how sensitive their substantive conclusions are to the choice of clusters.

Since our model is dynamic, we set aside a certain number of last time periods as a validation data set while fitting our model with different numbers of clusters to the remaining data. We then

11

evaluate the log (observed-data) likelihood using the validation data. For example, if the last time period alone is set aside as the validation data set, then, the formal expression of the hold-out log-likelihood function to be evaluated is given by,

$$
\sum_{i=1}^{N}\sum_{j>i}\log\Bigg\{\sum_{(z_{\overline{T}_{ij}-1},z_{\overline{T}_{ij}})}\pi_{z_{\overline{T}_{ij}-1}}P_{z_{\overline{T}_{ij}-1},z_{\overline{T}_{ij}}}
$$
$$
\prod_{k=1}^{2K}q_{kz_{\overline{T}_{ij}}}^{D_{ij\overline{T}_{ij}k}}(1-q_{kz_{\overline{T}_{ij}}})^{(1-D_{ij\overline{T}_{ij}k})}\phi(W_{ij\overline{T}_{ij}k}\mid \mu_{kz_{\overline{T}_{ij}}},\sigma^{2}_{kz_{\overline{T}_{ij}}})^{\mathbf{1}\{k\neq K,k\neq 2K\}(1-D_{ij\overline{T}_{ij}k})}\Bigg\}
$$

where we have integrated out the latent group indicator variables. We then choose the number of clusters that maximize this hold-out log-likelihood.

3   Empirical Patterns of International Product-Level Trade

In this section, we first describe the characteristics of each cluster identified by the proposed dynamic clustering algorithm. We then show the evolution of dyadic trade relations with the changing cluster memberships over time. Our key finding is that typical dyadic trade relationships evolve from sparse trade to inter-industry trade and then to intra-industry trade while the specific timing for such transition varies significantly by dyads. The proposed algorithm enables us to examine these changes over time at any levels of aggregation including industries, countries, dyads, regions, and the whole world.

3.1   Characteristics of Dyadic Trade Profiles

We begin our analysis by setting the number of latent clusters to three in order to get a parsimonious summary of the massive product-level trade data. As we see later, the basic patterns consistently emerge in the analyses with greater numbers of clusters. Our hold-out log-likelihood calculation described above shows that the 7-cluster model is the most preferred and the 15-cluster

12

model has the second highest value (see Figure B.3 in Appendix B.3). As shown below, these models provide finer pictures of the patterns uncovered by the 3-cluster model. Therefore, throughout this paper, we supplement the results based on the 3-cluster model with those from the 7 and 15-cluster models.

As explained above, the proposed clustering algorithm assigns a cluster membership, $Z_{ijt} \in \{1, 2, 3\}$, to dyad-year observations with similar trade profiles. More precisely, the algorithm produces the estimated probability that a given dyad-year observation belongs to each cluster. Since the product-level trade data are high-dimensional, three clusters are well separated. Consequently, a vast majority of dyad-year observations belong to one cluster with a high probability, making it easy for us to classify observations.

[Figure 3 about here.]

To facilitate the characterization of each cluster as well as the comparisons across different levels of aggregation, we first consider the trade profiles of the resulting three clusters at the SITC one-digit industry level (see Table D.1). Figure 3 depicts the trade proportion for a given industry in each direction of trade flow, which is defined as the proportion of the relevant products relative to the total volume of exports from one country to another (see the discussion at the end of Section 2.2). We plot the trade proportion of each SITC one-digit industry in the exports from country A to country B (x-axis) against that in the exports from country B to country A (y-axis). If circles are located close to the 45-degree line, therefore, countries export similar products to each other. The size of a circle is proportional to the total volume of trade for the corresponding industry within each cluster.

13

We find three distinct clusters of international trade that we respectively denote as Sparse Trade, Inter-industry Trade, and Intra-industry Trade. First, as seen by the size of circles, dyads in the Sparse Trade cluster tend to trade very little (i.e., sparse) across various products relative to dyads in other clusters. That is, the membership in this cluster implies a sufficiently shallower bilateral trade relationship compared to other country-pairs, although it is still possible that there exist positive volumes of trade for some products between the trading partners. In fact, when these dyads do trade we find that in most cases one country exports crude materials to the other country in exchange for food/live animals.[11] Second, the trade profile of the Inter-industry Trade cluster shows that dyads in the cluster exchange dissimilar goods: one country exports crude materials while the other country tends to export manufactured goods. The force of comparative advantage is particularly pronounced in the machinery and transportation equipment industry (dark grey circle) as countries in this cluster tend to export such products only in one direction.

Finally, dyads in the Intra-industry Trade cluster tend to export and import products in the same industries and in similar proportions, as shown by the convergence of products toward the 45-degree line. For example, about 30% of exports and imports are from the manufacturing industry for both countries in a typical dyad-year of the cluster. To explore this pattern further at the product level, we calculate the extent to which dyads exchange similar products. The Intra-Industry Trade or IIT index at the top of each panel reports the mean product-level Grubel-Lloyd index for each cluster, measuring the degree to which two countries export products in similar proportions (averaging across all products).[12] The IIT Index equals 1 if for every product country

---

[11] We thank an anonymous reviewer for pointing out the possibility that countries may report zero trade for other strategic reasons, which might change the interpretation of the cluster.

[12] Formally, this index for cluster z is defined as, $1/K \sum_{k=1}^{K} \left\{ 1 - |m_{kz}^{\mathsf{AB}} - m_{kz}^{\mathsf{BA}}|/(m_{kz}^{\mathsf{AB}} + m_{kz}^{\mathsf{BA}}) \right\}$ where m_{kz}^{AB} denotes exports of product k from country A to country B in cluster z.

14

A exports the same amount as it imports from country B, while the index equals zero if for every product the trade occurs only in one direction. As expected, the Intra-industry Trade cluster has the highest score of 0.72, suggesting that dyads in this cluster tend to trade the same SITC four-digit products in similar amounts with each another. Similar patterns arise when we increase the number of clusters to seven and fifteen as shown in Appendix B.1.

We emphasize that the three types of cluster labels are general characterizations of dyadic trade patterns rather than referring to specific industries. For example, membership in the Intra-industry Trade cluster does not necessarily imply that in every industry there is intra-industry trade. To gain a better understanding of each cluster, we examine its trade profile at the product level. Figure 4 displays the product-level trade proportions for exports from country A to country B (left column of each cluster) and for exports from country B to country A (right column of each cluster) where each line segment corresponds to one of the 625 SITC four-digit products. We group the products by industry to facilitate the comparison. The color of a line segment indicates the extent to which the trade proportion of a product deviates from the mean proportion of trade across all clusters.[13] A darker red line segment represents a higher proportion while a darker blue line segment represents a lower proportion of the product's trade. In addition to the analysis based on three clusters (left plot), we also examine the results based on the seven cluster model (right plot).

[Figure 4 about here.]

---

[13]We calculate this quantity for each product k in cluster z for both directions. For example, given the product-level trade proportions m_{kz}^{AB} for exports of product k from country A to country B in cluster z, the deviation from the mean is given by, $m_{kz}^{AB} - \frac{1}{2K} \sum_{z'=1}^{K} (m_{kz'}^{AB} + m_{kz'}^{BA})$.

15

Several clear patterns emerge from the figure. First, dyads in Inter-industry Trade cluster tend to import and export different sets of products, as shown by the stark red-blue contrast between the two columns and across products. Specifically, bilateral trade in this cluster is characterized by one country exporting crude materials and food while its partner focuses on exporting industrial goods in chemical, manufacturing, and machinery industries. Second, the differences across the clusters are noticeable especially in these industrial goods as shown in the upper half part of each figure (see the differences especially in "Chemicals and related products" and above). Dyads in the Intra-industry Trade cluster tend to trade in higher proportions of industrial goods with each other (two red columns) and lower proportions of food, beverages, and crude materials (two blue columns). This suggests that exchanges of similar products occur mainly through industrial goods. Third, dyads in Sparse Trade tend to exchange little in industrial goods (two blue columns) while utilizing comparative advantages in food, beverages, and crude materials (red-blue contrast) if they do trade.

The right panel of Figure 4 shows that these patterns become even more conspicuous in the results based on the 7-cluster model. As we move from Sparse Trade to Inter-industry Trade and then to Intra-industry Trade, we observe an increase in exchanges of similar products indicated by the progressive change from two blue, blue-red, and two red columns for industrial goods. For food and crude materials, we also see the stark red-blue contrast as we move towards the Inter-industry Trade cluster in the middle. We emphasize that a finer degree of summary can be achieved by increasing the number of clusters. This can be seen from the maximum deviation of the mean proportion reported in the color-bar legend in each panel. For any given product, the mean proportion differs by up to 0.03 in the 3-cluster model whereas the 7-cluster model can distinguish the mean difference up to 0.08. Nevertheless, the basic patterns remain essentially identical. We

16

further illustrate this point with results from different numbers of clusters in Appendix B.2.

The three general trade profiles identified by our clustering algorithm shed light on some of the main theoretical underpinnings in the international political economy literature. First, the theory of comparative advantage has been a fundamental explanation for why countries have political cleavages across industries (Rogowski, 1987). However, most applications of the classical Stolper-Samuelson theorem conceptualize comparative advantages based on only a few factor endowments such as labor, land, and capital while distinctions across dyads and products are often ignored (Milner and Kubota, 2005). We leverage information on product-level trade to empirically identify the dyad-years with comparative advantage relationships and the products in which such forces are dominant.

Second, intra-industry trade has become an important factor in trade politics as most developed countries now exchange similar goods. We show that industries with differentiated products such as manufacturing are the primary venues for high intra-industry trade. It is important to emphasize that the co-occurrence of imports and exports within the same industry implies that import-competing domestic firms, importers, exporters, and even multinational firms may coexist within the same industry. Our analysis identifies a set of particular dyads and industries in which political cleavages within an industry might be particularly pronounced due to higher heterogeneity in firm preferences.[14] Finally, although zero trade-flows across pairs of countries are already well-known by researchers (e.g., Silva and Tenreyro, 2006; Helpman, Melitz, and Rubinstein, 2008), we show that there are significant variations in the levels of sparse trade at the product-level even among dyads with active trade relations. This raises concerns for most empirical studies that have neglected the product-level heterogeneity in the margins of trade. Our finding suggests that researchers should

---

[14]See also Kim (2017), which shows the importance of within-industry heterogeneity.

17

pay as much attention to the selection of trading partners (extensive margins) as to volumes of bilateral trade (intensive margins) at the product-level (Kim, Londregan, and Ratkovic, 2019).

3.2 Evolution of Dyadic Trade Relations

A vast literature on international political economy suggests that trade between two countries depend on many factors that change over time. These factors include barriers to market access (Bagwell, Mavroidis, and Staiger, 2002), improvements in information and communication technology (Baldwin, 2016), domestic politics (Grossman and Helpman, 1994), political institutions (Mansfield, Milner, and Rosendorff, 2000), alliances (Gowa and Mansfield, 1993), and state power (Krasner, 1976). An implication is that bilateral trade relations change as the trading environment and the global trading system evolve. However, few existing studies relate such factors to the changing composition of trade profiles. In contrast, as explained in Section 2.2, a key feature of our clustering algorithm is its ability to identify the dynamics of dyadic trade relations.

[Figure 5 about here.]

Figure 5 depicts the dynamic changes of cluster membership from 1962 to 2014. The left panel shows that the membership size of Sparse Trade has decreased continuously during this period while the number of dyads belonging to Inter-industry Trade and Intra-industry Trade increased, especially since the revolution in information and communication technology (ICT) in the 1990s accelerated the fragmentation of production processes (Baldwin, 2016, 79-105). Over the last several years, however, the growth in Inter-industry Trade appears to have slowed down while the growth in Intra-industry Trade has persisted. The growth in cluster membership, however, does not necessarily imply that more trade volumes are accounted for by the cluster. As it is seen from

18

the right panel of the figure, the overall trade volumes explained by the dyads in the Inter-industry Trade cluster have actually decreased over time. In fact, as of 2014, over 90% of global trade is due to the bilateral trade among the dyads that are in the Intra-industry Trade cluster even though only about 10% of dyads belong to the cluster.

[Figure 6 about here.]

Next, we shift our focus to monadic trade relations in order to investigate how individual countries underwent different dynamic changes in their trade relations over time. Figure 6 illustrates how each country's trade relationships with its partners have changed over the last 50 years. Each point in the triangles represents a country. The distance from each vertex corresponds to the proportion of dyads involving the country that belongs to each of the three clusters. For example, a point at the center of the triangle means the country is in each of three clusters with exactly one-third of its partners. The differences in the distribution of points across the three panels illustrate distinct landscapes of international trade in each period. We observe that most countries first increase their trade relationships based on comparative advantages (moving right), and then engage in intra-industry trade with more partners (moving up). To be sure, not all dyads follow the same path. This suggests that international specialization is a dynamic process that is determined endogenously by changes in comparative advantage (Proudman and Redding, 2000). As Redding (1999) argues, countries face a trade-off between specializing further based on existing comparative advantages, and investing in other sectors with no technological edges. The different trajectories followed by different countries are also illustrated by the movements of five countries from each continent highlighted in the figure. China has dramatically changed its trade relations

19

with its partners whereas the United Kingdom has maintained similar trade profiles with most countries.

[Figure 7 about here.]

Finally, this paper makes an important empirical contribution by characterizing the *evolution* of dyadic trade relations. Specifically, we identify the highly heterogeneous timing of any structural transition of bilateral trade patterns for *each* dyad across time. In Figure 7, we focus on five countries and their relationships with each partner country from 1962 to 2014. It shows that the U.S. engages in two-way intra-industry trade (red) with many of its partners. The list of such partners has grown steadily over time. This pattern is in contrast to those of China and the United Kingdom. In particular, China exhibits a strikingly steep trajectory of growing memberships in the Intra-industry Trade cluster while the United Kingdom maintains similar trade profiles with its partners as shown by the relatively flat color composition.[15]

In sum, the proposed dynamic clustering algorithm yields new insights about changes in global trade profile. Our analysis shows how typical dyadic trade relationships evolve and provides a simple summary of massive trade data. Our approach contrasts with existing empirical studies of international trade as we consider bilateral trade across numerous products in its entirety.

4 An Application: A New Measure of Trade Competition

Having detailed the value of our clustering algorithm in summarizing the dynamic evolution of bilateral trade, we now illustrate the use of our cluster membership in the analysis of trade

---

[15]Figure C.1 in Appendix C illustrates the evolutionary path of bilateral trade relations for all dyads that have existed between 1962 and 2014.

20

competition. Specifically, we show that our framework can incorporate the extent of competition that each country faces with all of its trading partners at the product level.

Trade competition has been one of the key theoretical concepts in international relations. For example, scholars argue that trade competition can affect how policies and institutions diffuse across borders (e.g., Jensen, 2003; Simmons and Elkins, 2004; Simmons, Dobbin, and Garrett, 2006). Despite its theoretical importance, surprisingly few measures are available to capture how countries compete for trade partners at the product level. In this section, we use our dyadic cluster membership to construct an improved measure of trade competition.

Simmons and Elkins (2004) define trade competition as "the degree to which nations compete in the same foreign markets" without reference to products that are traded (p. 178). In contrast, we exploit the availability of product-level trade data and define trade competition as the extent to which two countries trade *similar* products with the *same* partners. We argue that the degree of trade competition must be examined at the product level. This is because competition over price and quality tends to be specific to products that are sufficiently similar to each other. When two countries export different products to the same partners, they do not necessarily compete with each other even when the overall trade volumes are similar in the same time period.[16] Although it is possible that countries trading the same products with different partners have an "intention" to compete, we focus on directly measuring the existence of *observed* trade competition in each market.[17]

---

[16]See Cao and Prakash (2010) who criticize diffusion studies that do not distinguish between products or partners (p. 483).

[17] In other words, our measure does not capture latent strategic intentions for competition. We thank an anonymous reviewer for pointing this out.

21

The Proposed Measure. We use our dyadic trade cluster membership to measure trade competition between countries i and j in a given year t. We consider whether the dyadic trades of the two countries with the same trading partner h belong to the same cluster z. If dyads (i, h) and (j, h) belong to the same cluster, they trade similar products with the same trading partner, implying that the two countries are in competition with each other. Our measure captures trade competition in both exports and imports since our cluster membership is based on dyadic trades. This is a desirable feature because, for example, countries compete in importing raw materials as much as they complete in exporting manufacturing goods. In addition, we do not consider joint membership in the Sparse Trade cluster as evidence for trade competition because trade competition does not arise in the absence of trade.

Formally, we begin by defining an indicator variable for trade competition between countries i and j involving partner h in year t is defined as,

$$C_{ijt}^h \;=\; \mathbf{1}\{Z_{iht} = z, Z_{jht} = z \mid z \neq \text{Sparse Trade}\}. \tag{4}$$

Notice that we take into account the role of each country in defining this measure. That is, countries i and j are not in competition, even when they are in the same cluster with a common partner h, if the roles within the cluster involving h are reversed, i.e., $Z_{iht} = z$ and $Z_{hjt} = z$.

Next, we aggregate this competition indicator variable across all trading partners using Gower (1971)'s similarity metric with bilateral trade volumes as weights since the level of competition is likely to be higher in a larger market. That is, our measure of trade competition for countries i and j in year t is defined as,

$$\overline{C}_{ijt} \;\equiv\; \sum_{h \in \{1,\ldots,N\} \backslash \{i,j\}} \left(\frac{S_{iht} + S_{jht}}{\sum_{h' \in \{1,\ldots,N\} \backslash \{i,j\}} S_{ih't} + S_{jh't}} \right) C_{ijt}^h, \tag{5}$$

where S_{iht} denotes the share of country i's trade with partner h out of its total trade volume in

22

year t. Thus, our measure assigns the highest level of trade competition to two countries that belong to the same cluster for their trading relationships with all existing partners. Furthermore, the measure weights the importance of trade competition with specific partners by their dyadic trade volumes.

Comparison with the Existing Measures. The proposed measure of trade competition makes several improvements over existing measures. First, our measure is based on the similarity in trade profiles of all SITC four-digit products. This allows us to capture competition at a disaggregated level in a systematic fashion, yielding a more precise measure. In contrast, other measures are based on aggregate exports and imports (Lee and Strang, 2006), certain select industries (Simmons and Elkins, 2004; Elkins, Guzman, and Simmons, 2006), or only SITC first-digit products (Cao and Prakash, 2010, 2011).[18] Clustering based on the data at a finer product level improves the validity of our measure because substitution among different products can be easily justified at disaggregate levels. In other words, the elasticity of substitution decreases as the level of aggregation increases. For example, the degree of trade competition should be higher when two countries export oranges vs. mandarins (SITC 0571) than oranges vs. apples (SITC 0571 vs. SITC 0574), fruits vs. vegetables (SITC 057 vs. SITC 054), or food vs. manufactured products (SITC 0 vs. SITC 6). This is because it is easier to substitute between goods at more disaggregated levels.

Second, our measure discounts the level of trade competition when there exists little trade. Existing measures either ignore the importance of sparse trade entirely or deal with the problem by imposing strong constraints. For example, Chatagnier and Kavaklı (2017) calculate the simple correlation between two vectors of trade profiles to summarize the degree of trade competition

---

[18]We summarize and compare existing measures in Table E.1 in the appendix.

between two countries. Although their measure is also based on SITC four-digit product trade, the prevalence of sparse trade implies that two countries can be misleadingly considered to be in high competition when they do not trade most of the products with the same partners. This is because a high correlation between two trade profile vectors can result when most elements are close to zero. In contrast, our measure is based on the clustering algorithm that explicitly models zero trade. We consider trade competition to exist only when both countries engage in trade with its partners with sufficient amounts of product-level trade.

Third, our measure incorporates the levels of competition in each trading partner's market. While many existing measures of trade competition distinguish between different products at some level (Simmons and Elkins, 2004; Elkins, Guzman, and Simmons, 2006; Chatagnier and Kavaklı, 2017), few consider the levels of competition separately for each trade partner. Most measures are based on a simple correlation for two countries' product-level exports to the world. As a result, existing measures of trade competition may mask the fact that two countries could be exporting similar products to *different* partners, which also inflates the level of competition. In contrast, we build our measure explicitly on dyadic trade profiles and further weight the importance of competition in each partner countries by trade volume to capture such nuances (see Figure F.1 in Appendix F for this important distinction).

Finally, our measure considers both export and import competition. The rise of global supply chains implies that countries not only compete in their export markets but also compete in import markets for inputs to the products they produce (e.g. rare earth materials for the production of computer chips). However, to the best of our knowledge, all existing measures have focused exclusively on export competition. This leads to understated levels of trade competition for dyads

24

that compete mainly in imports and not exports.[19]

[Figure 8 about here.]

Figure 8 compares our measures of trade competition (labeled as 3-CL, 7-CL, and 15-CL for 3-cluster, 7-cluster, and 15-cluster models, respectively) against those based on the existing measures proposed by Elkins, Guzman, and Simmons (2006) (EGS) and Chatagnier and Kavaklı (2017) (CK). In particular, it shows the changing levels of trade competition between China, USA, and Japan and their top three competitors as of 2014.[20] In general, our measure based on our 3-cluster model (black solid line) changes smoothly over time as expected for bilateral trade competition. Yet, as shown in the top panel, it also captures the dramatic increase in China's trade competitiveness with others after its economic reforms in 1978, especially against Canada.

In contrast, EGS shows wide temporal fluctuations while CK exhibits little variation over time. In the middle row, our measure identifies the United Kingdom as the top and persistent trade competitor of the United States. In contrast, EGS would suggest that the United Kingdom or Switzerland's competition with the United States has actually decreased over time.[21] In the bottom panel, our measure identifies South Korea as Japan's top competitor, with increased competition

---

[19] Figure E.1 in Appendix E shows that the correlations between the proposed measure and the existing measures are centered around zero.

[20] The top three competitors are chosen according to our measure based on our 3-cluster model as of 2014. We focus on competitors that persist throughout the time period 1962–2014.

[21] Researchers' priors about top competitors may oftentimes be driven by patterns of aggregate industries, especially those that are conspicuous in public discourse. For example, one may think Canada is a top competitor of the U.S. because it also exports many food products or machinery. However, the level of competition might not be high if wheat is the primary agriculture products that Canada exports whereas soybeans and corns take up the most significant portion of the U.S. agricultural exports.

25

corresponding to South Korea's rapid economic growth based on export-oriented industrialization since the 1970s. This contrasts with EGS which shows that Japan experienced the highest level of trade competition with South Korea in the 1960s.

The measures based on our 7-cluster (grey solid line) or 15-cluster models (light-grey solid line) exhibits similar patterns over time. As expected, there exists more variation with a larger number of clusters as the underlying cluster membership captures finer differences in trade profiles. This can be useful to identify smaller changes in trade competition. For example, we find a big jump in the level of trade competition between Japan and South Korea in the mid-1970s. This period corresponds to the third Five-Year Plan (1972–1976) during which President Park Chung-hee transformed South Korea's economy by providing aggressive subsidies to heavy chemical industries. Note that measures based on different numbers of clusters should be interpreted in relative terms within each model, and hence smaller values from our 7-cluster or 15-cluster model do not suggest that trade competition is lower than what one finds from our 3-cluster model.[22]

In Appendix G, we apply our proposed measure of trade competition to the study of the diffusion of Bilateral Investment Treaties (BITs). In contrast to the original findings of Elkins, Guzman, and Simmons (2006), our reanalyses show that trade competition has no clear effect on increasing the likelihood of signing BITs. Instead, the results suggest that bilateral trade relationships built on vertical or horizontal production linkages play a more important role.

---

[22]Technically, finer gradations make it harder for two dyads to belong to the same cluster (see equation 4) and can thus shift trends downward. In Appendix E, we report a metric of discrepancy between the different measures. Appendix F provides further comparisons between our measure and other existing measures.

5   Concluding Remarks

In this paper, we characterized dyadic trade relations based on approximately two billion product-level trade data. We found that countries focus on different sources of gains from trade as their trade relationships evolve. In particular, a typical pair of countries starts trading based on their respective comparative advantage while variety gains from exchanging similar products within the same industry become more important later. This important sequential transition has been overlooked in most studies of international trade and development. Our findings suggest that the nature of bilateral trade relationships changes over time, and hence countries might have to deal with different domestic political conflicts depending on their trading partners at different points in time.

One important advantage of the proposed dynamic clustering methods is its ability to summarize a massive amount of highly disaggregated data with a simple cluster membership variable. Researchers can use our measure to account for the types of bilateral trade relationships over time without incurring enormous computational and methodological costs. We also demonstrate the use of this cluster membership to construct a measure of trade competition. Using this measure, we find that dyadic trade relationships can be more important than competitive economic pressures in explaining the likelihood of signing BITs (see Appendix G).

Dyadic clustering methods have broader applications in political science research. Indeed, measurements of social and economic interactions involving pairs of political actors are taken at increasingly disaggregated levels. For example, scholars in international relations observe highly specific dyadic exchanges of services (e.g., transportation, travel, communications, construction, insurance, financial, royalties, etc.), capital (in various forms of direct investment, portfolio in-

vestment, debt flows, aid, etc.), and people (with different skill sets and occupations). Outside of International Relations, dynamic clustering methods such as the one proposed here can be used to analyze various relationships between political actors that evolve over time (e.g., campaign contributions, lobbying activities, co-sponsorships among politicians, citations of court opinions). These methodologies provide an effective means to uncover systematic patterns underlying high-dimensional data.

References

Aitchison, J. 1982. "The Statistical Analysis of Compositional Data (with Discussion)." *Journal of the Royal Statistical Society, Series B, Methodological* 44 (2): 139–177.

Antràs, Pol, and Robert W. Staiger. 2012. "Offshoring and the Role of Trade Agreements." *American Economic Review* 102 (7): 3140–83.

Bagwell, Kyle, Petros C. Mavroidis, and Robert W. Staiger. 2002. "It's a Question of Market Access." *The American Journal of International Law* 96 (1): 56-76.

Baldwin, Richard. 2016. *The Great Convergence: Information Technology and the New Globalization.* Harvard University Press.

Bickel, Peter J., and Elizaveta Levina. 2008. "Regularized Estimation of Large Covariance Matrices." *Annals of Statistics* 36 (1): 199–227.

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning.* New York: Springer.

Cao, Xun, and Aseem Prakash. 2010. "Trade Competition and Domestic Pollution: A Panel Study, 19802003." *International Organization* 64 (jul): 481–503.

Cao, Xun, and Aseem Prakash. 2011. "Growing Exports by Signaling Product Quality: Trade Competition and the Cross-National Diffusion of ISO 9000 Quality Standards." *Journal of Policy Analysis and Management* 30 (dec): 111–135.

Carnegie, Allison. 2014. "States Held Hostage: Political Hold-up Problems and the Effects of International Institutions." *American Political Science Review* 108 (01): 54–70.

Chatagnier, J. Tyson, and Kerim Can Kavaklı. 2017. "From Economic Competition to Military Combat: Export Similarity and International Conflict." *Journal of Conflict Resolution* 61 (7): 1510-1536.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. "Maximum Likelihood from Incomplete Data Via the EM Algorithm (with Discussion)." *Journal of the Royal Statistical Society, Series B, Methodological* 39 (1): 1–37.

Dorussen, Han. 2006. "Heterogeneous Trade Interests and Conflict: What You Trade Matters." *Journal of Conflict Resolution* 50 (feb): 87–107.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz. 2011. "An Anatomy of International Trade: Evidence From French Firms." *Econometrica* 79 (5): 1453–1498.

Elkins, Zachary, Andrew T. Guzman, and Beth A. Simmons. 2006. "Competing for Capital: The Diffusion of Bilateral Investment Treaties, 1960-2000." *International Organization* 60 (oct): 811–846.

Feenstra, Robert C., Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo. 2005. "World Trade Flows, 1962-2000, with Production and Tariff Data." National Bureau of Economic Research (NBER) Working Paper Series available at `http://www.nber.org/papers/w11040`.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2008. "Sparse Inverse Covariance Estimation with the Graphical Lasso." *Biostatistics* 9 (3): 432–441.

Frühwirth-Schnatter, Sylvia. 2007. *Finite Mixture and Markov Switching Models.* New York: Springer.

30

Gartzke, Erik. 2007. "The Capitalist Peace." *American Journal of Political Science* 51 (1): 166–191.

Goenner, Cullen F. 2010. "From Toys to Warships: Interdependence and the Effects of Disaggregated Trade on Militarized Disputes." *Journal of Peace Research* 47 (sep): 547–559.

Gowa, Joanne, and Edward D. Mansfield. 1993. "Power Politics and International Trade." *American Political Science Review* 87 (02): 408–420.

Gower, J. C. 1971. "A General Coefficient of Similarity and Some of Its Properties." *Biometrics* 27 (4): 857–871.

Grossman, Gene M., and Elhanan Helpman. 1990. "Comparative Advantage and Long-Run Growth." *The American Economic Review* 80 (4): 796-815.

Grossman, Gene M., and Elhanan Helpman. 1994. "Protection for Sale." *The American Economic Review* 84 (4): 833–850.

Grossman, Gene M., and Esteban Rossi-Hansberg. 2012. "Task Trade Between Similar Countries." *Econometrica* 80 (2): 593–629.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein. 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes." *The Quarterly Journal of Economics* 123 (2): 441-487.

Hidalgo, César A., and Ricardo Hausmann. 2009. "The Building Blocks of Economic Complexity." *Proceedings of the National Academy of Sciences* 106 (26): 10570–10575.

Hiscox, Michael J. 2002. "Commerce, Coalitions, and Factor Mobility: Evidence from Congressional Votes on Trade Legislation." *American Political Science Review* 96 (sep): 593–608.

Imai, Kosuke, and Dustin Tingley. 2012. "A Statistical Method for Empirical Testing of Competing Theories." *American Journal of Political Science* 56 (January): 218–236.

Jensen, J. Bradford, Dennis P. Quinn, and Stephen Weymouth. 2015. "The Influence of Firm Global Supply Chains and Foreign Currency Undervaluations on US Trade Disputes." *International Organization* 69 (4): 913–947.

Jensen, Nathan M. 2003. "Democratic Governance and Multinational Corporations: Political Regimes and Inflows of Foreign Direct Investment." *International Organization* 57 (03): 587–616.

Johns, Leslie, and Rachel L. Wellhausen. 2016. "Under One Roof: Supply Chains and the Protection of Foreign Investment." *American Political Science Review* 110 (2): 31–51.

Kim, In Song. 2017. "Political Cleavages within Industry: Firm-level Lobbying for Trade Liberalization." *American Political Science Review* 111 (1): 1–20.

Kim, In Song, John Londregan, and Marc Ratkovic. 2019. "The Effects of Political Institutions on the Extensive and Intensive Margins of Trade." *International Organization*, Forthcoming.

Krasner, Stephen D. 1976. "State Power and the Structure of International Trade." *World Politics* 28 (3): 317–347.

Krugman, Paul R. 1979. "Increasing Returns, Monopolistic Competition, and International Trade." *Journal of International Economics* 9 (4): 469–479.

Lee, Chang Kil, and David Strang. 2006. "The International Diffusion of Public-Sector Downsizing: Network Emulation and Theory-Driven Learning." *International Organization* 60 (04): 883–909.

Mahutga, Matthew C. 2006. "The Persistence of Structural Inequality? A Network Analysis of International Trade, 1965–2000." *Social Forces* 84 (4): 1863–1889.

Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff. 2000. "Free to Trade: Democracies, Autocracies, and International Trade." *The American Political Science Review* 94 (2): 305–321.

Milner, Helen V., and Keiko Kubota. 2005. "Why the Move to Free Trade? Democracy and Trade Policy in the Developing Countries." *International Organization* 59 (1): 107–143.

Osgood, Iain. 2016. "Differentiated Products, Divided Industries: Firm Preferences over Trade Liberalization." *Economics and Politics* 28 (2): 161–180.

Park, Jong Hee. 2012. "A Unified Method for Dynamic and Cross-Sectional Heterogeneity: Introduci ng Hidden Markov Panel Models." *American Journal of Political Science* 56 (4): 1040–1054.

Proudman, James, and Stephen Redding. 2000. "Evolving Patterns of International Trade." *Review of International Economics* 8 (3): 373–396.

Redding, Stephen. 1999. "Dynamic Comparative Advantage and the Welfare Effects of Trade." *Oxford Economic Papers* 51 (1): 15–39.

Rogowski, Ronald. 1987. "Political Cleavages and Changing Exposure to Trade." *The American Political Science Review* 81 (4): 1121–1137.

Scheve, Kenneth F., and Matthew J. Slaughter. 2001. "What Determines Individual Trade-Policy Preferences?" *Journal of International Economics* 54 (aug): 267–292.

33

Silva, J. M. C. Santos, and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics* 88 (4): 641–658.

Simmons, Beth A., Frank Dobbin, and Geoffrey Garrett. 2006. "Introduction: The International Diffusion of Liberalism." *International Organization* 60 (04): 781–810.

Simmons, Beth A., and Zachary Elkins. 2004. "The Globalization of Liberalization: Policy Diffusion in the International Political Economy." *American Political Science Review* 98 (feb): 171–189.

Tomz, Michael, Judith L. Goldstein, and Douglas Rivers. 2007. "Do We Really Know That the WTO Increases Trade? Comment." *The American Economic Review* 97 (5): 2005–2018.
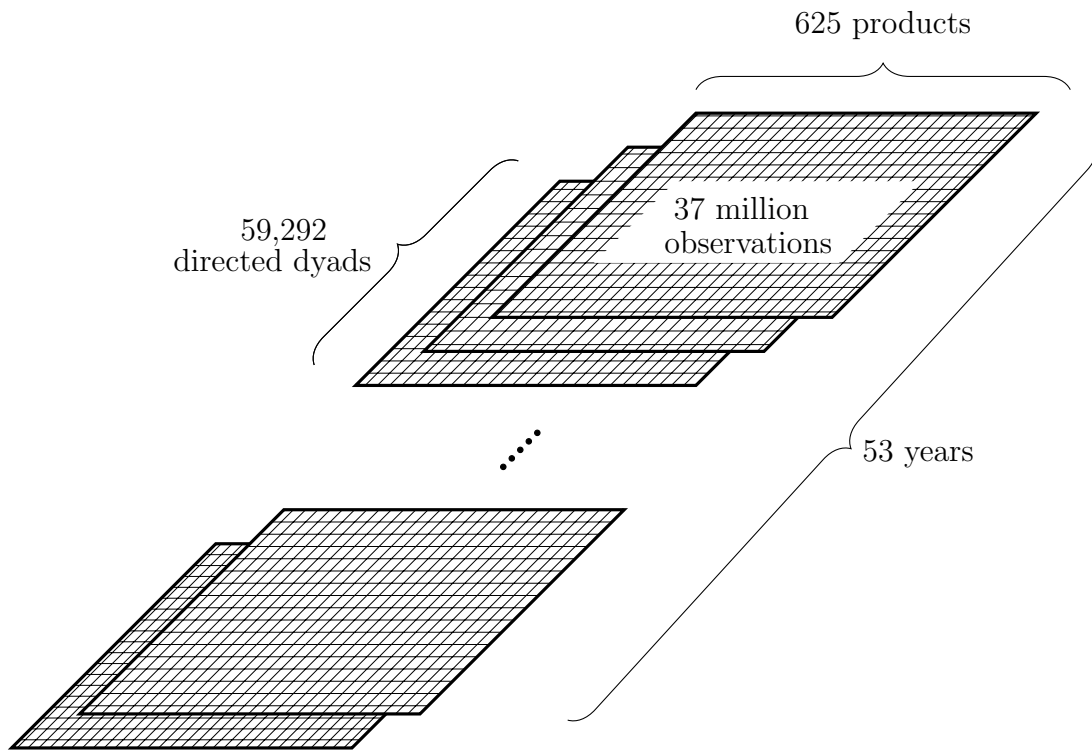
Figure 1: **Two Billion Observations of Product-level Trade Data**: We analyze bilateral trade of 625 products among 59,292 directed dyads over 53 years (1962 – 2014).
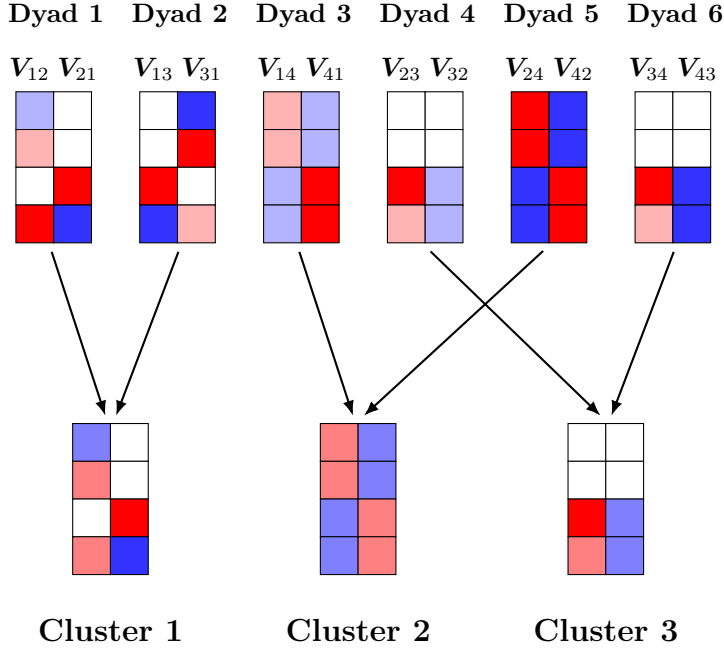
Figure 2: **An Illustration of the Clustering Methods:** This figure illustrates the proposed clustering method with four countries and four products. The first row represents the trade profiles of six dyads, while the second row describes the estimated mean trade proportions characterizing each cluster. Blue, white, and red colors represent general levels of trade (low, medium, and high) while the gradients of each color capture the differences across each dyad that researchers observe in the data. It shows that dyads with similar patterns of trade across products are grouped into a common cluster. Furthermore, the formation of Cluster 1 demonstrates the "flipping problem" in which the ordering of the stacked trade profile vector can be arbitrarily chosen by the researcher.
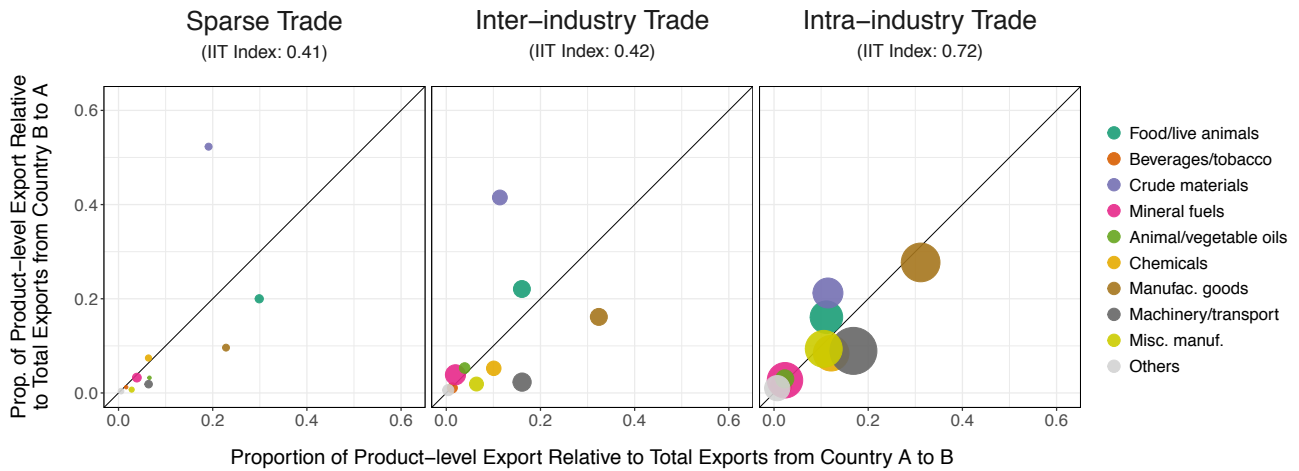
Figure 3: **Three Types of International Trade**. The location of each circle represents the mean industry level exports in each direction of bilateral trade. Circle sizes represent the magnitude of trade flows by dyads within the cluster. Circles on the 45-degree line indicate that dyads export and import products in the same industry in the same proportion. Dyads classified as Intra-industry trade tend to have most industries placed closer to the 45-degree line.
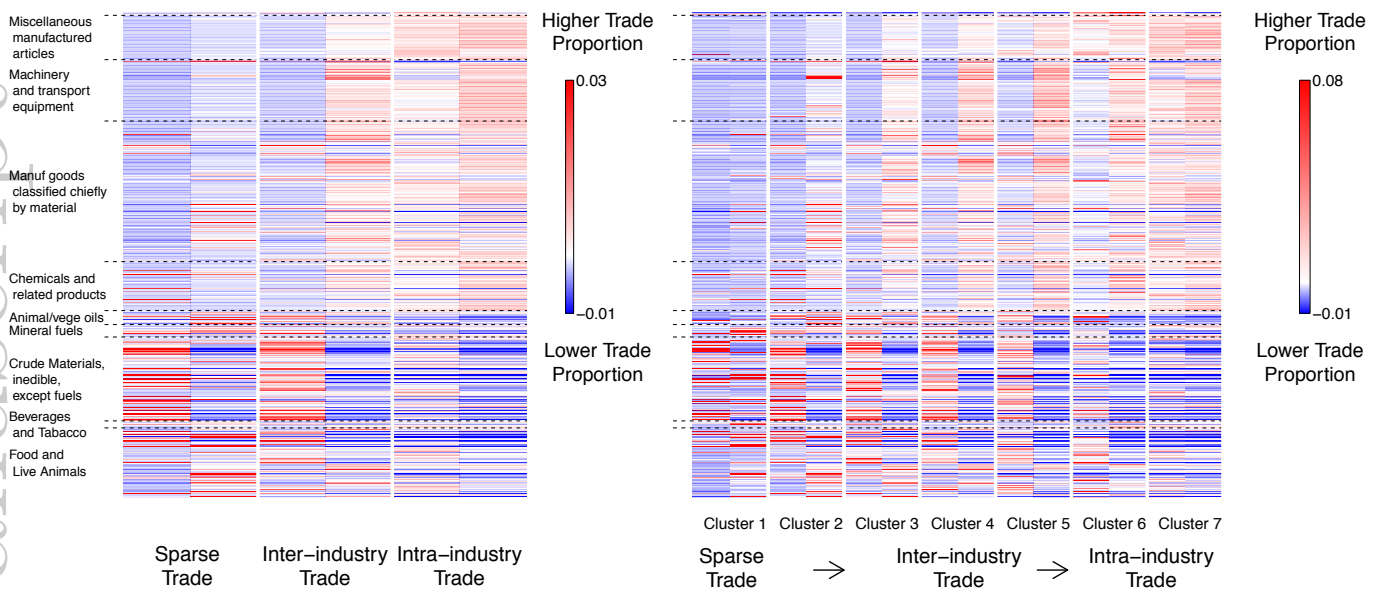
Figure 4: **Trade Profile at the Product Level**. The left panel shows results from our main 3-cluster model. Each row represents one product at the SITC four-digit level. Within each cluster, the left column plots exports from country A to country B while the right column plots exports from country B to country A. The color of each segment indicates the extent product-level trade deviates from the mean proportion of trade across clusters. Red (blue) represents higher (lower) proportions of trade. The right panel shows results from a more fine-grained 7-cluster model.

38
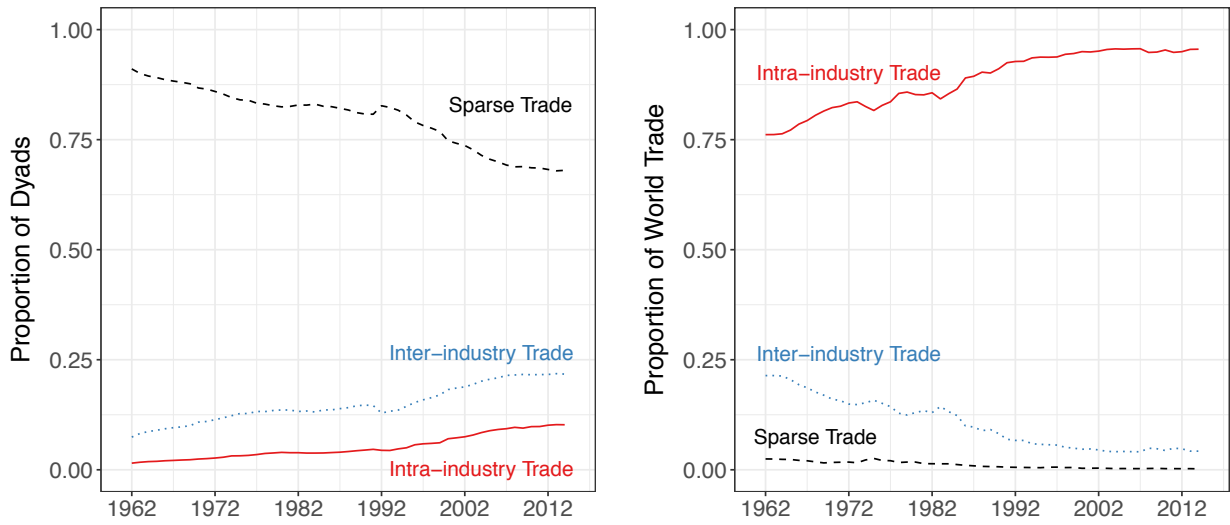
Figure 5: **Dynamic Changes in Cluster Membership from 1962 to 2014**: The left panel plots the proportion of dyad membership out of total dyads in each cluster. The right panel plots the proportion of world trade occupied by each cluster.
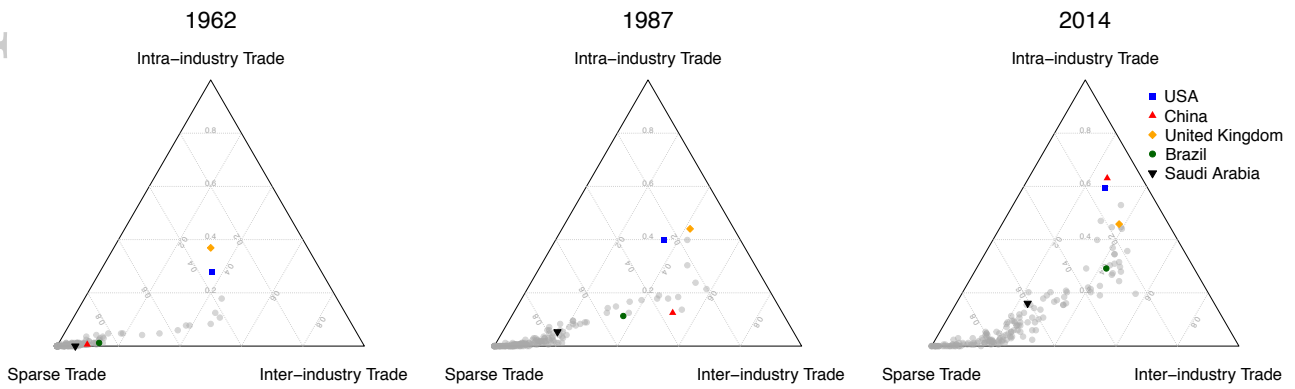
Figure 6: **A Path to Intra-industry Trade**: This plot describes a path along which dyadic trade relationships evolve from sparse trade to inter-industry trade and then to intra-industry trade. Each point represents a country. The location of each point corresponds to the proportion of dyads involving the country that belongs to each cluster. We highlight the location of five countries: USA, China, United Kingdom, Brazil, and Saudi Arabia.
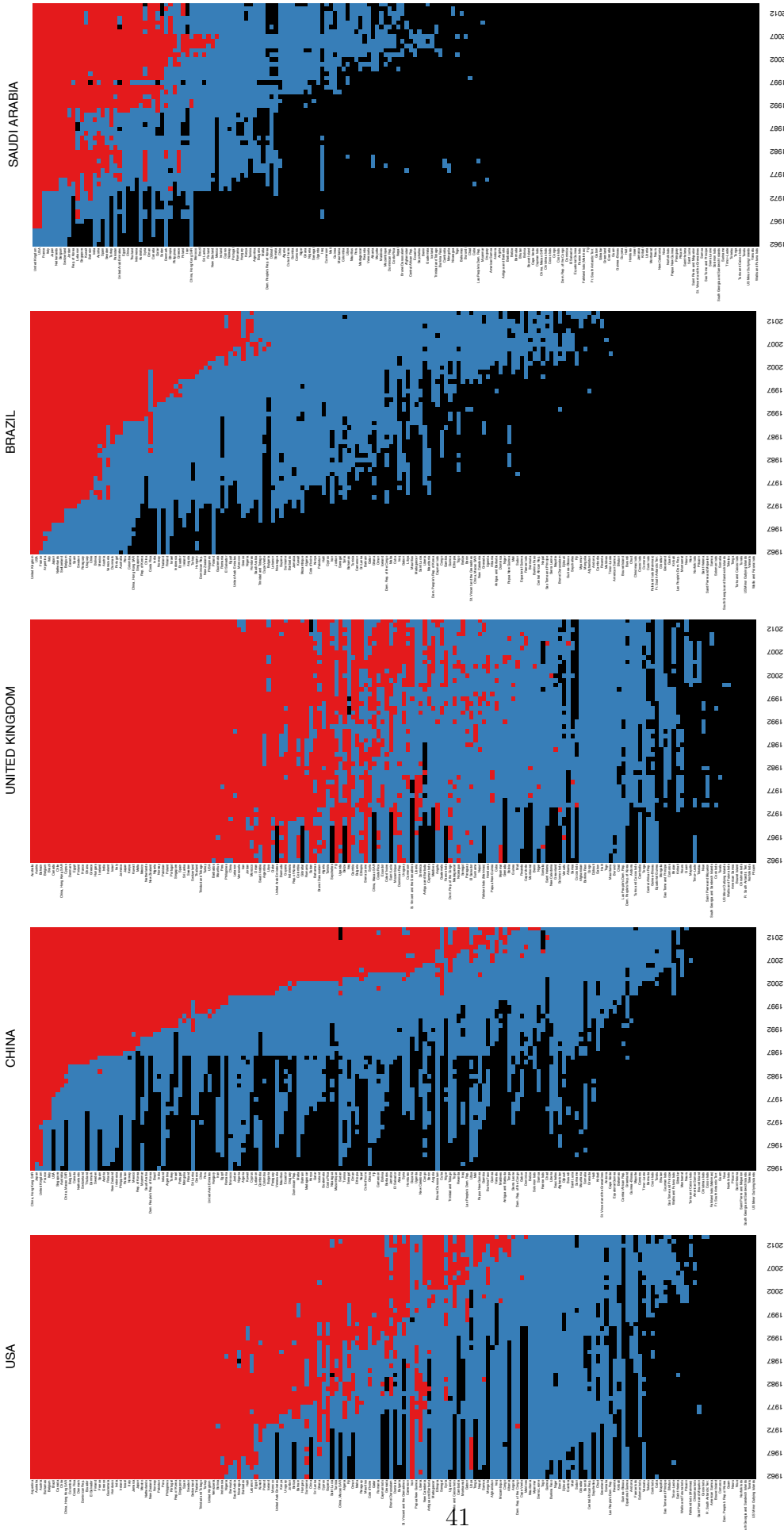
Figure 7: **Changing Trade Profiles with Partners.** This figure illustrates the dynamic changes of cluster membership for dyads involving five countries: USA, China, United Kingdom, Brazil, and Saudi Arabia. Each row represents one partner. We color each polygon according to its dyad-year trade profile: red (**Intra-industry Trade**), blue (**Inter-industry Trade**), and black (**Sparse Trade**). The panels exclude partners that do not persist throughout the time period 1962-2014.
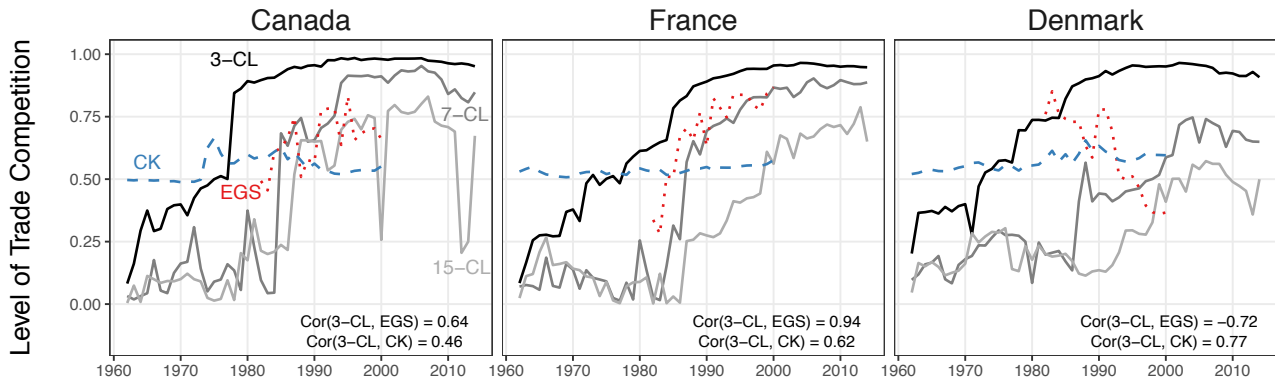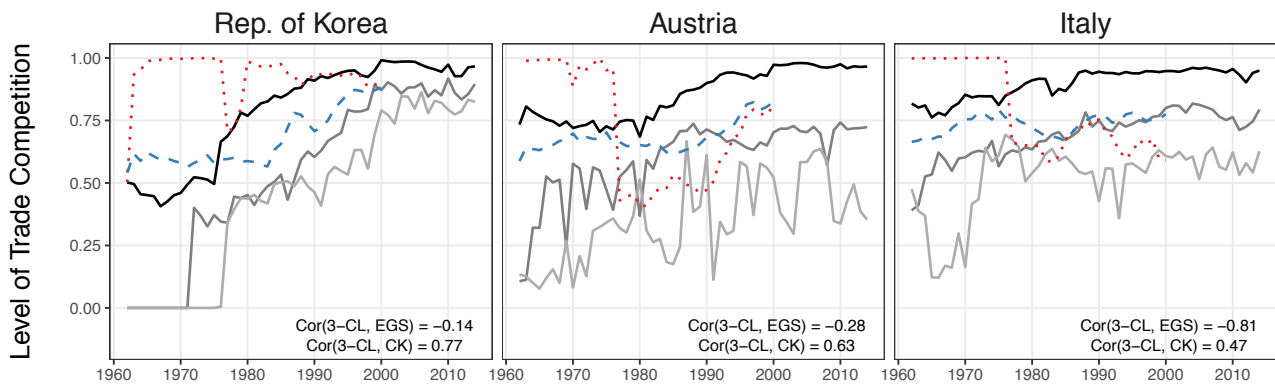
41
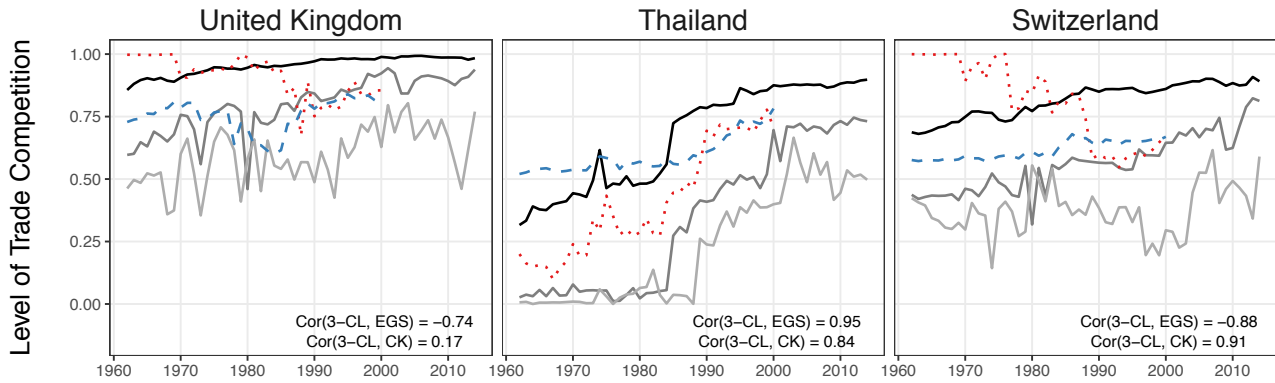
Figure 8: **A New Measure of Trade Competition:** This figure reveals the changing levels of trade competition between China, USA, and Japan and their respective top three competitors in 2014. The thick black and grey solid lines represent our proposed measures based on 3-cluster (3-CL), 7-cluster (7-CL), and 15-cluster models (15-CL), respectively. EGS (red dotted line) is based on Elkins, Guzman, and Simmons (2006) while CK (blue dashed line) uses the measure developed by Chatagnier and Kavaklı (2017). We present the correlations between our measure from the 3-cluster model and the other two measures at the bottom-right corner.