



Extracting geography from trade data



Yuke Li^a, Tianhao Wu^b, Nicholas Marshall^c, Stefan Steinerberger^{d,*}

^a Department of Political Science, Yale University, New Haven, CT, USA

^b Department of Statistics, Yale University, New Haven, CT, USA

^c Department of Mathematics, Program in Applied Mathematics, Yale University, New Haven, CT, USA

^d Department of Mathematics, Yale University, New Haven, CT, USA

HIGHLIGHTS

- We derive geopolitical information using only the trade volume between countries.
- The implemented method uses nonlinear dimensionality reduction (diffusion maps).
- We discover different historic, linguistic and geographic factors as relevant.

ARTICLE INFO

Article history:

Received 23 July 2016

Received in revised form 20 December 2016

Available online 5 January 2017

Keywords:

Trade distance

Geopolitics

Dimensionality reduction

Spectral embedding

ABSTRACT

Understanding international trade is a fundamental problem in economics—one standard approach is via what is commonly called the “gravity equation”, which predicts the total amount of trade F_{ij} between two countries i and j as

$$F_{ij} = G \frac{M_i M_j}{D_{ij}},$$

where G is a constant, M_i , M_j denote the “economic mass” (often simply the gross domestic product) and D_{ij} is the “distance” between countries i and j . Here “distance” is a complex notion that includes geographical, historical, linguistic and sociological components. We take the *inverse* route and ask ourselves to what extent it is possible to reconstruct meaningful information about countries simply from knowing the bilateral trade volumes F_{ij} : indeed, we show that a remarkable amount of geopolitical information can be extracted. The main ingredient is a spectral decomposition of the Graph Laplacian as a tool to perform nonlinear dimensionality reduction.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

1.1. The gravity equation

Understanding the geometry and structural properties of world trade is a problem of obvious significance, great appeal and has a long history. It has become challenging to summarize the existing literature outside of the framework of a survey article, a starting point is given by the (non-exhaustive) list [1–9] and references therein. One of the dominant paradigms

* Corresponding author.

E-mail addresses: yuke.li@yale.edu (Y. Li), tianhao.wu@yale.edu (T. Wu), nicholas.marshall@yale.edu (N. Marshall), stefan.steinerberger@yale.edu (S. Steinerberger).

was first formalized in [10]: given a set of n countries, the “gravity equation” predicts the bilateral trade flow F_{ij} between countries i and j as

$$F_{ij} = G \frac{M_i M_j}{D_{ij}},$$

where G is a constant, M_i , M_j denote the “economic mass” (often simply the gross domestic product) and D_{ij} is the “distance” between countries i and j . Here “Distance” is a complex notion for which various models have been proposed –these include factors such as common colonizers [11], cultural proximity [12], linguistic ties [13], mutual trust [14], past conflicts [15], shared borders [16], shared currency [11] and others.

1.2. Our approach

Motivated by the complex notions of “distance” in the literature, we were interested in the question of whether it is possible to understand what type of underlying factors dominate multinational trade *using trade volume alone*: given the amount of trade F_{ij} between different countries, how much information can be extracted about the trade-distances between the countries? We believe that any such minimal approach, if successful, is most suited in furthering a precise understanding of the underlying processes because no preconceived notions are ever part of the computation –the method only uses trade volume itself. Our approach is as follows: we encode trade structure as complete, weighed graph. This graph is fairly complicated and not well-suited for a direct analysis; however, we can ask whether it is possible to embed the graph in \mathbb{R}^2 without introducing a large distortion of distances. Our approach is motivated by the *diffusion maps* framework of Coifman and Lafon [17]: by performing a spectral analysis of a suitable diffusion operator related to the Graph Laplacian, we can use its ground state and first excited state as a nonlinear coordinate system. Since the technique can also be applied to sub-graphs, this provides a map

$$\phi : \{\text{any collection of countries}\} \rightarrow \mathbb{R}^2$$

using only the total amount of trade done between any two countries in the collection within a given year. We demonstrate that the map faithfully represents a complex notion of “distance” comprised of geographical factors (different continents being easily identifiable), shared history, common language, colonization history and others.

The contribution of this paper is to construct an embedding into the Euclidean plane \mathbb{R}^2 that

1. automatically separates countries into different continents and meaningful subregions,
2. provides an accurate insight into what “trade distance” might be *since the computation does not use any information about cultural, historical, linguistic or sociological matters*,
3. shows that trade is driven by a few underlying factors that dominate the dynamics
4. and, finally, it demonstrates that tools from nonlinear dimensionality reduction may be highly useful in the analysis of international trade.

1.3. Difference to existing approaches

The difference between our approach and existing studies is as follows: first, to the best of our knowledge, our work is the first that applies nonlinear dimensionality reduction techniques in the studies of trade; secondly, in doing so, we obtain detailed and informative empirical regularities about selected countries’ distances in trade *without* any external information regarding culture, geography, language or sociology: *all information is solely derived from the trading behavior*. This is in stark contrast to the more classical approach of using models for “distance” that incorporate various factors which are considered to be of potential interest and then fitting parameters—our method is completely agnostic about any of these underlying structures; in particular, any arising structures are intrinsically contained in the data. One of the contributions of our paper is to show that this agnostic method does indeed recover various factors (geographical, sociological, linguistic, historical) which we believe further substantiates their use.

1.3.1. Principal component analysis

The classical method for dimensionality reduction is Principal Component Analysis (PCA). However, PCA is most effective on data with linear structures—when data has nonlinear structures, nonlinear dimensionality reduction techniques can be significantly more effective. We give a simple (and classical) example by taking a closed loop with helix structure and applying both methods to reduce the dimensions of the data points. In Fig. 2, diffusion embedding obviously outperforms PCA, as it accurately recovers the geometry of the original data as a closed loop.

1.3.2. Community detection

The existing approach that might be closest in spirit to our method comes from the subfield dealing with problems of community detection. We wish to especially emphasize Barigozzi et al. [18], who studied the community structure of the commodity-specific trade relations among countries over the 1992–2003 period and compared them to the exogenous community structures induced by geography and regional trade agreements. Another related result is [19] which studied dynamical aspects of the international trade network over time (in particular, detecting with a switch of the economic leadership of Asia-Oceanic region from Japan to China in the 1995–2011 period). The main difference is that our approach recovers clusters as well as an underlying geometry that encapsulates pairwise distances between the detected clusters.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات