International Comparisons of Prices Using Shortest Path Spatial Chaining

Presenter: Robert Hill
University of Graz

Joint work with
Prasada Rao (University of Queensland)
Sriram Shankar (University of Western Sydney)
Reza Hajargasht (University of Melbourne)

Society for Economic Measurement Conference – OECD Paris
July 22-24, 2015

Bilateral Comparisons

A number of bilateral price index formulas exist in the literature. Here we focus on Fisher defined as follows:

$$P_{jk}^{F} = \sqrt{P_{jk}^{P} \times P_{jk}^{L}} = \sqrt{\frac{\sum_{n=1}^{N} p_{kn} q_{kn}}{\sum_{n=1}^{N} p_{jn} q_{kn}} \frac{\sum_{n=1}^{N} p_{kn} q_{jn}}{\sum_{n=1}^{N} p_{jn} q_{jn}}},$$

where P_{jk}^{P} and P_{jk}^{L} denote Paasche and Laspeyres price indexes, and $n = 1, \ldots, N$ indexes the headings over which the comparison is being made.

As soon as the comparison is extended to three or more countries, we have a whole matrix of Fisher price indexes:

$$F = \begin{pmatrix} 1 & P^F_{1,2} & \cdots & P^F_{1,I} \\ P^F_{2,1} & 1 & \cdots & P^F_{2,I} \\ \vdots & \vdots & & \vdots \\ P^F_{I,1} & P^F_{I,2} & \cdots & 1 \end{pmatrix}, \quad \text{where } P^F_{j,k} = 1/P^F_{k,j}.$$

In general, Fisher price indexes are intransitive. This means that

$$P^F_{j,I} \neq P^F_{j,k} \times P^F_{k,I}.$$

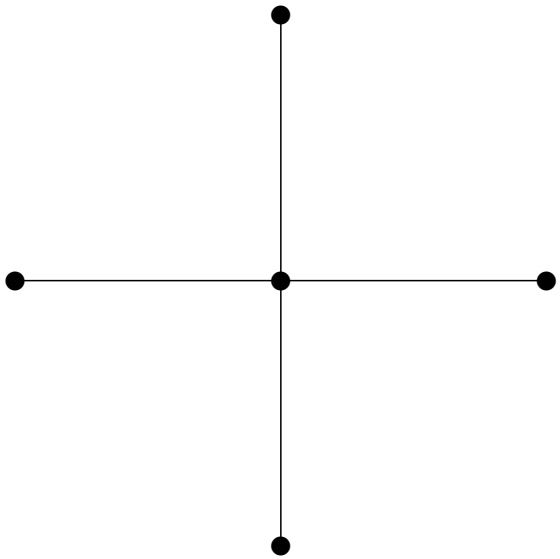
International Comparisons with Bilateral Building Blocks: A Graph Theory Perspective

Let each country be denoted by a vertex and a Fisher price index comparison between a pair of countries by an edge connecting the respective vertices of these countries.

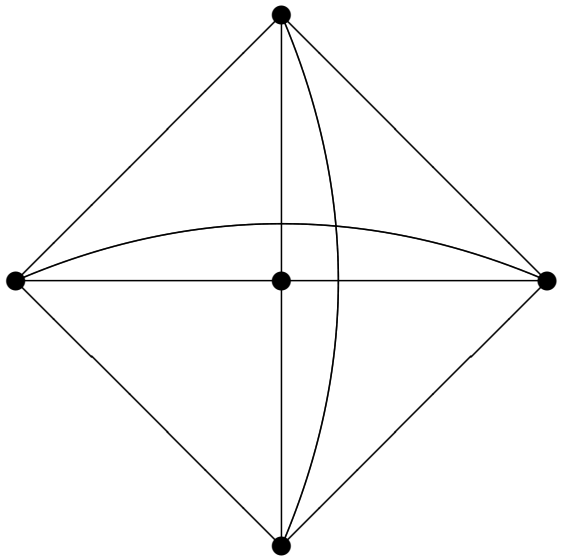
A spanning tree is a graph without any cycles that connects all the vertices.

Since Fishers are intransitive it follows that:
(i) When Fisher indexes are chained across a graph that includes cycles, the results will be internally inconsistent.
(ii) When Fisher indexes are chained across a spanning tree, each spanning tree will generate a different set of results.

FIGURE 1. — EXAMPLES OF GRAPHS



Star Graph



Complete Graph

The Gini-Eltetö-Köves-Szulc (GEKS) Method

The GEKS method imposes transitivity on the Fisher price indexes as follows:

$$\frac{P_k^{GEKS}}{P_j^{GEKS}} = \prod_{i=1}^{I} \left(\frac{P_{i,k}^F}{P_{i,j}^F} \right)^{1/I} .$$

From a graph theory perspective, the GEKS method constructs I star spanning trees (each with a different country at the centre), and then takes the geometric mean of these I sets of results.

Hence the GEKS method uses all the Fishers and gives them equal weight.

Problems with GEKS

(i) Some countries may have poor quality data, which can then contaminate the overall GEKS comparison.
(ii) Some pairs of countries are very different in terms of the baskets of goods and services they consume. As a result a bilateral comparison between these countries will be of poor quality.
(iii) It should be possible to improve on GEKS by giving more weight to bilateral comparisons that are deemed more reliable.

The Minimum Spanning Tree (MST) Method

The MST method proposed by Hill (1999) uses the following distance metric:

$$D^1_{jk} = \left| \ln \left(\frac{P^L_{jk}}{P^P_{jk}} \right) \right| .$$

It is then argued that the Fishers of bilateral comparisons with a smaller distance D_{jk} are more reliable.

The motivation comes from the Hicks and Leontief aggregation theorems.

The spanning tree is constructed using Kruskal's algorithm.

The edge with the smallest D_{jk} is selected subject to the constraint that its inclusion does not create a cycle. Repeat for the edge with the next smallest D_{jk}. Once $I - 1$ edges have been selected the algorithm terminates.

The resulting graph is the minimum spanning tree (i.e., the spanning tree that minimizes the sum of the weights on the edges).

Problem: The MST is not robust to slight changes in the data or the choice of metric for measuring the reliability of bilateral comparisons.

Something in between GEKS and MST may be better.

Some Alternative Distance Metrics

The following distance metrics are obtained from Diewert (2002).

$$D_{jk}^2 = \sum_{n=1}^{N} \left[\left(\frac{s_{j,n} + s_{k,n}}{2} \right) \left(\frac{p_{k,n}}{P_{jk}^F \times p_{j,n}} + \frac{P_{jk}^F \times p_{j,n}}{p_{k,n}} - 2 \right) \right]$$

$$D_{jk}^3 = \sum_{n=1}^{N} \left\{ \left(\frac{s_{j,n} + s_{k,n}}{2} \right) \left[\left(\frac{p_{k,n}}{P_{jk}^F \times p_{j,n}} - 1 \right)^2 + \left(\frac{P_{jk}^F \times p_{j,n}}{p_{k,n}} - 1 \right)^2 \right] \right\}$$

$s_{j,n}$ in D_{jk}^2 and D_{jk}^3 denotes the expenditure share of heading n in country j .

The Weighted-GEKS Method

Weighted-GEKS proposed by Rao (1999) specifies a weights matrix as follows:

$$W = \begin{pmatrix} 0 & w_{12} & \cdots & w_{1K} \\ w_{21} & 0 & \cdots & w_{2K} \\ \vdots & \vdots & & \vdots \\ w_{K1} & w_{K2} & \cdots & 0 \end{pmatrix}.$$

The weighted-GEKS price indexes are now derived as follows:

$$\begin{pmatrix} \ln P_2 \\ \ln P_3 \\ \vdots \\ \ln P_K \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{K} w_{2j} & -w_{23} & \cdots & -w_{2K} \\ -w_{32} & \sum_{j=1}^{K} w_{3j} & \cdots & -w_{3K} \\ \vdots & \vdots & & \vdots \\ -w_{K2} & -w_{K3} & \cdots & \sum_{j=1}^{K} w_{Kj} \end{pmatrix}^{-1} \begin{pmatrix} -\sum_{j=1}^{K} w_{2j} \ln P_{2j}^{F} \\ -\sum_{j=1}^{K} w_{3j} \ln P_{3j}^{F} \\ \vdots \\ -\sum_{j=1}^{K} w_{Kj} \ln P_{Kj}^{F} \end{pmatrix}.$$

The price index for country 1, P_1, is normalized to 1. In the case where $w_{j,k} = w$ for all j, k, the weighted-GEKS method reduces to GEKS.

Choosing a Weights Metric

Any of the distance metrics discussed above can be converted into a weights metric for use in the weighted-GEKS formula as follows:

$$w_{jk} = \frac{1}{1 + D_{jk}}.$$

A weights metric w_{jk} can be generalized as follows:

$$w^*_{jk} = \frac{(w_{jk})^x}{\max^I_{a,b=1}[(w_{ab})^x]}, \quad \text{where } x \geq 0.$$

MST as a Special Case of Weighted-GEKS

Theorem: In the limit as x tends to infinity, the weighted-GEKS method converges on the MST method.

Proof: Weighted-GEKS starts from a complete graph. Take the edge with the smallest weight. As x gets large the impact of this edge on all the cycles of which it is part will tend to zero in the weighted-GEKS formula. Hence for large x, this edge can be deleted as long as the resulting graph is still connected. Repeat for the next smallest edge. This algorithm continues until the graph is reduced to a spanning tree. This is the reverse delete algorithm, and the resulting spanning tree is the minimum spanning tree (see Kruskal 1956).

The Problem with Weighted-GEKS

Weighted-GEKS implies judgments of the form: the bilateral comparison between countries j and k should be given 50 percent more weight than the bilateral comparison between countries a and b.

It is not clear how the weights should be chosen in weighted-GEKS.

We consider in what follows two approaches based on shortest path chaining in which the weighting structure emerges naturally rather than needing to be imposed ex ante.

Also, with shortest path methods the focus is on finding the best comparison between say countries j and k, rather than on saying how much more reliable one bilateral comparison is than another.

Shortest Path Chaining

Let $SP(jzk)$ denote the shortest path containing z edges between countries j and k. For example,

$$SP(j1k) = D_{jk},$$

$$SP(j2k) = \min_a(D_{ja} + D_{ak}),$$

$$SP(j3k) = \min_a(D_{ja} + D_{ab} + D_{bk}),$$

where D again denotes a distance metric.

The overall shortest path between countries j and k is now calculated as follows:

$$SP(jk) = \min_z[SP(jzk)].$$

This shortest path can be calculated using Dijkstra's algorithm.

With Shortest Path chaining it must makes sense to compare D_{jk} and $D_{ja} + D_{ak}$. It is not clear that such a comparison is meaningful for D^2_{jk}. However, it is for D^1_{jk}.

$$D^1_{ja} + D^1_{ak} = \left| \ln \left(\frac{P^L_{ja} \times P^L_{ak}}{P^P_{ja} \times P^P_{ak}} \right) \right| .$$

This is a chained Paasche-Laspeyres spread.

It is reasonable to say that we prefer a chained comparison to its direct counterpart when the chained comparison has a smaller Paasche-Laspeyres spread.

Note: A complication arises when for some bilateral comparisons Paasche is greater than Laspeyres. In this case, the shortest path definitions above need to be altered slightly.

In a shortest path context it would be better to define the Paasche-Laspeyres spread distance metric as follows:

$$D^1_{jk} = \ln\left(\frac{P^L_{jk}}{P^P_{jk}}\right).$$

Defined this way, the distance can be negative.

The shortest paths $SP(jzk)$ now need to be defined as follows:

$$SP(j1k) = |D_{jk}|,$$

$$SP(j2k) = \min_a |D_{ja} + D_{ak}|,$$

$$SP(j3k) = \min_a |D_{ja} + D_{ab} + D_{bk}|.$$

The overall shortest path between countries j and k is now again calculated as follows:

$$SP(jk) = \min_z [SP(jzk)].$$

When some of the distances are negative, Dijkstra's algorithm does not solve this problem.

These methods are illustrated for a sample of 14 countries using ICP 2005 data for consumption.

The countries in the sample are:

OECD-Eurostat: Germany, USA, Australia, Japan
CIS: Russia, Tajikistan
Africa: Morocco, Nigeria, Tanzania
Latin America: Brazil, Peru
Asia-Pacific: India; Thailand
Western Asia: Saudi Arabia

Note: D^2_{jk} and D^3_{jk} generate the same minimum-spanning tree for this data sample.

Figure 1: Minimum Spanning Tree D^1_{jk}

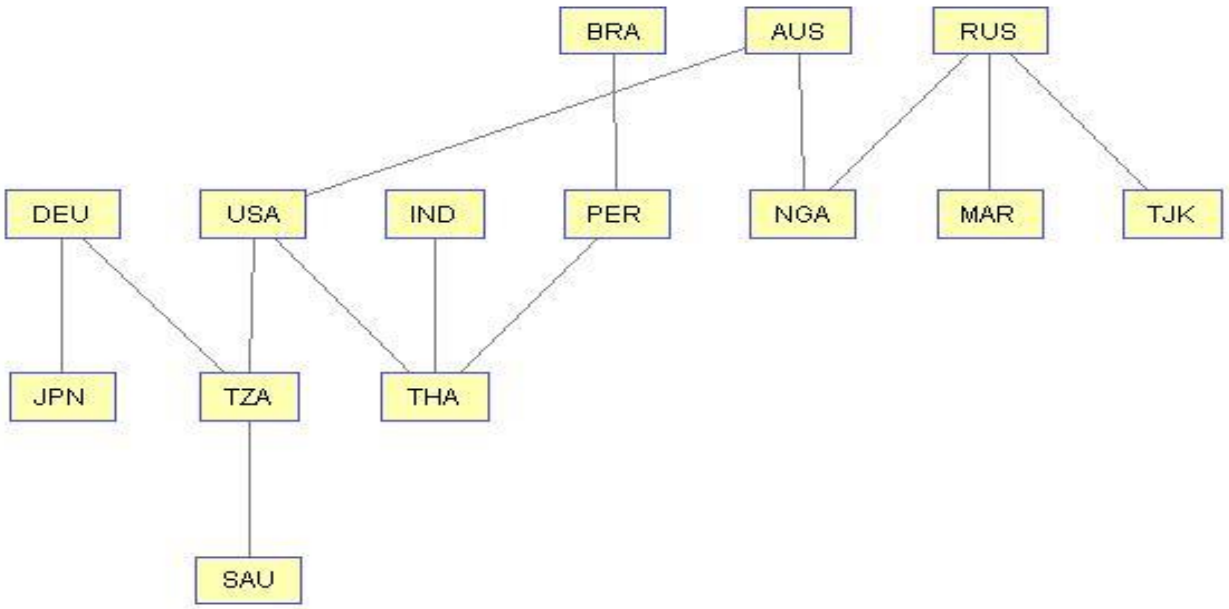


Figure 2: Minimum Spanning Tree D^2_{jk}

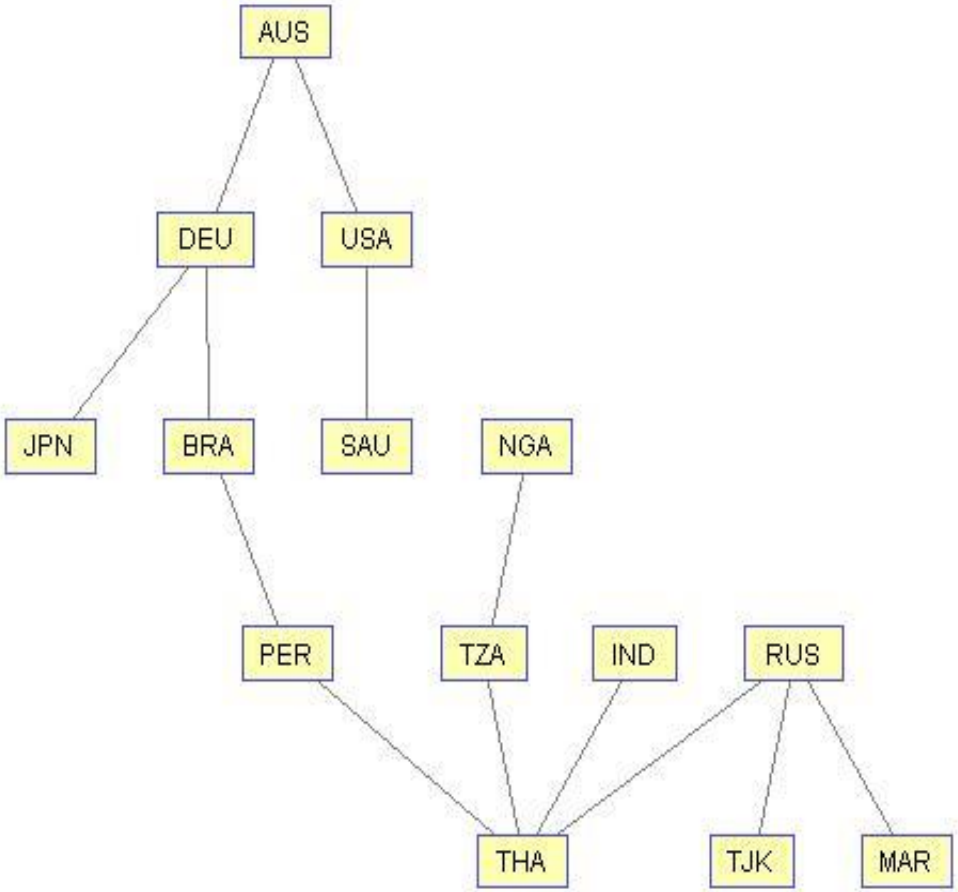


Figure 3: Shortest Path Spanning Tree for the USA D^1_{jk}

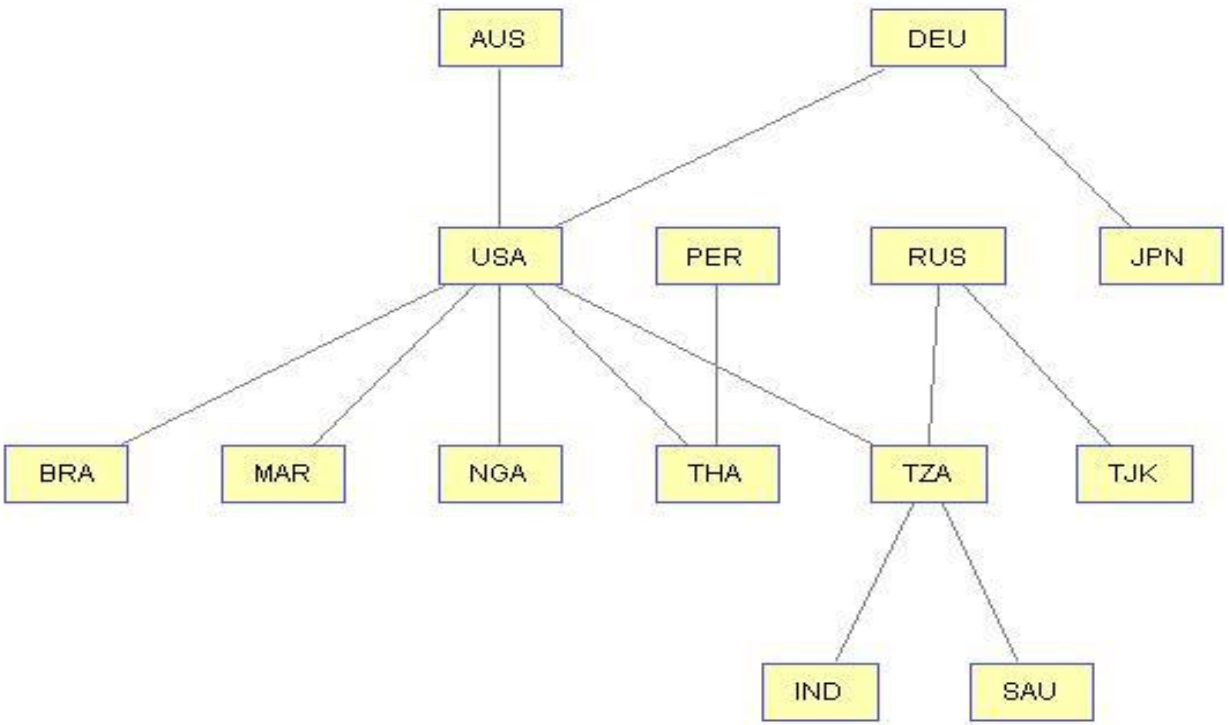
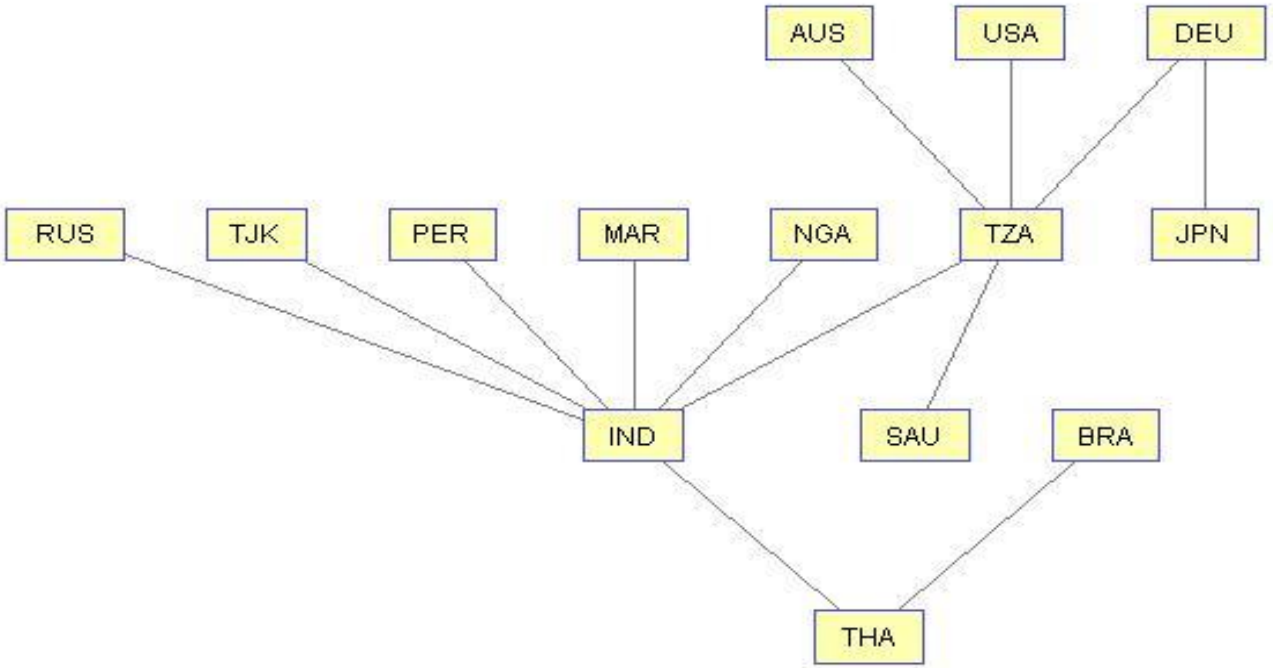


Figure 4: Shortest Path Spanning Tree for India D^1_{jk}



Shortest Path GEKS

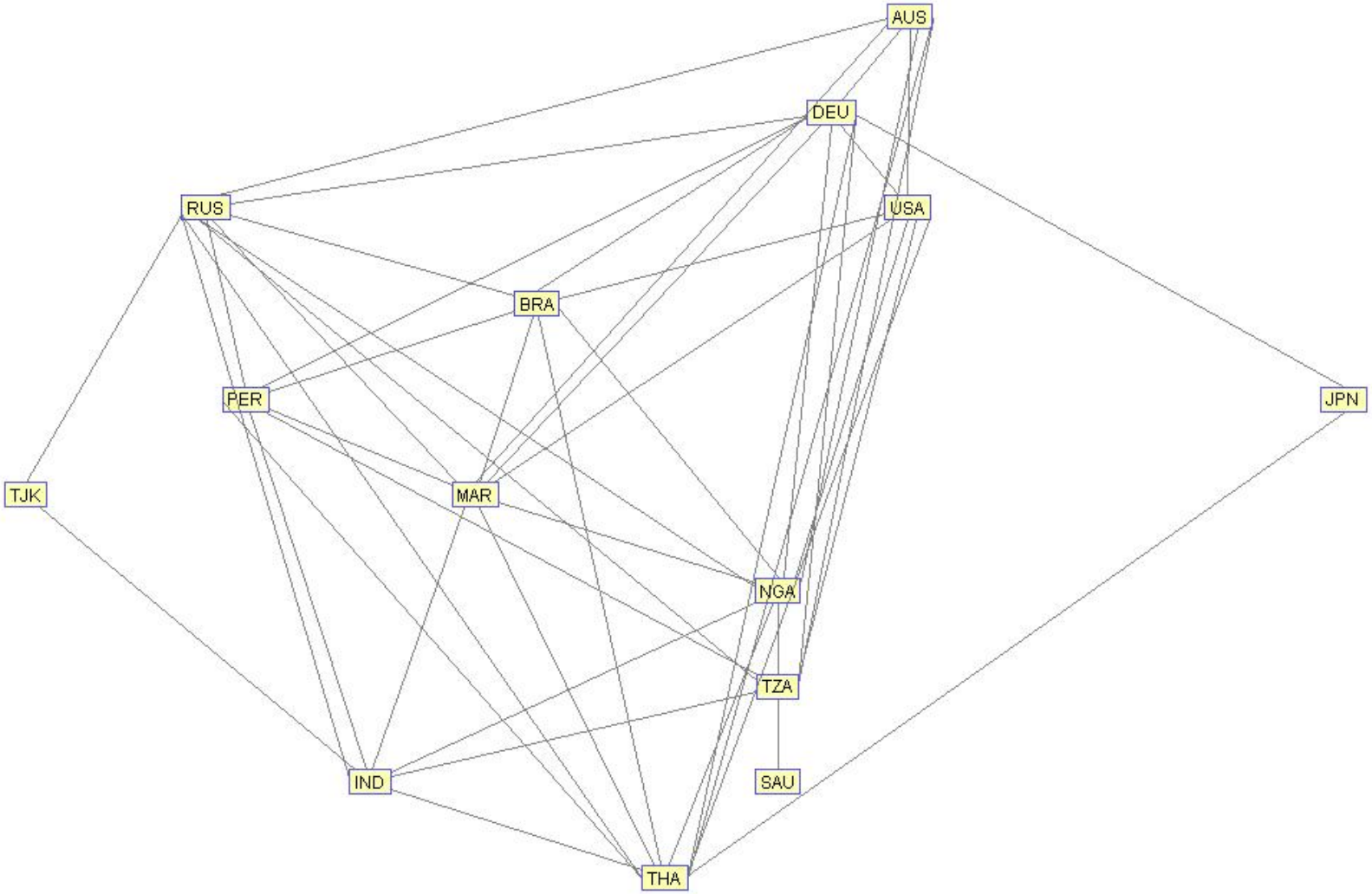
Each element of the Fisher price matrix is replaced by its shortest path equivalent.

For example, suppose the shortest path between countries j and k is via countries a and b. Then P^F_{jk} is replaced by $P^F_{ja} \times P^F_{ab} \times P^F_{ak}$.

The GEKS transitivization formula is now applied to the shortest path Fisher matrix.

This method has the advantage over GEKS that the most egregious bilaterals are removed prior to the application of the GEKS transitivization formula.

Figure 5: The Union of Shortest Paths Graph D^1_{jk}



Weighted-GEKS Applied to the Union of Shortest Paths Graph

The weights matrix W can be constructed as follows:

$w_{jk} = 1$ if the edge connecting countries j and k appears in the union of shortest paths graph.

$w_{jk} = 0$ otherwise.

This method like the shortest-path GEKS method ensures that the most egregious bilaterals are removed prior to the application of the transitivization formula.

Also, like the shortest-path GEKS method, this method does not require differential weights ot be applied to the bilaterals included in the comparison.

Within-Region versus Between-Region Bilateral Comparisons

ICP is divided into six regions.

91 bilateral comparisons can be made between the 14 countries in our data set.

12 of these comparisons are within region and 79 are between regions.

In the union of shortest paths graph there are 9 within-region bilaterals out of 48.

The proportion of within-region bilaterals rises from 12/79=0.13 in the complete graph to 9/48=0.19 in the union of shortest paths graph.

Our preliminary analysis indicates that over the full set of countries the increase in the share of within-region bilaterals in the union of shortest paths graph is much more dramatic.

Conclusions

Some bilaterals are more equal than others.

Scenarios that lie in between the GEKS and MST methods deserve closer scrutiny.

Methods that combine GEKS and shortest paths look quite promising.