Statistical toolbox for flow and network analysis

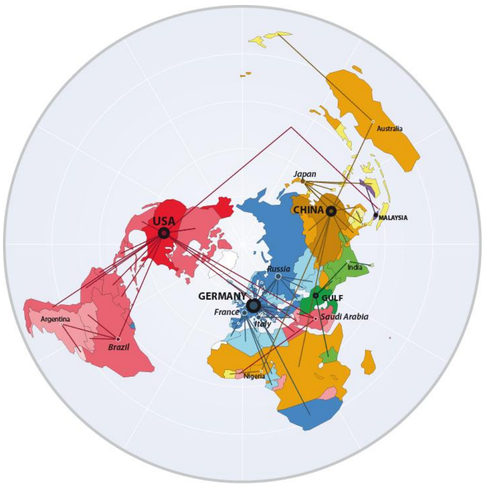
Gilles van Hamme, Claude Grasland

Toolbox

Work Package 5: Flows and Networks

From deliverable 5.1

July 2009 - revised November 2011



Statistical toolbox for flow and network analysis

Gilles VAN HAMME (ULB-IGEAT, Belgium) and Claude GRASLAND (CNRS-UMR Géographie-cités, France) (coord.)

Edited by Laurent Beauguitte

Abstract: This EuroBroadMap working paper proposes a review of selected statistical tools and theoretical concepts that can be used for the elaboration of division of the world in regions, on the basis of matrix of flows or connexions (network). In order to be able to compare the strengths and weaknesses of each family of tools or concepts, and to benchmark their results, each family of tools and concepts has been applied to the same data set: the CHELEM (CEPII) matrix of trade between countries of the world for the periods 1994-96 and 2004-06, eventually disaggregated by type of goods (72 types). In some cases, we have also used a more detailed matrix of trade between 168*168 countries (1996-2000) that was aggregated into CHELEM units (93*93) in order to evaluate the effects of aggregation of countries on results.

Key-words: Trade flow, centre/periphery, Intramax, statistical model, social network analysis, gravity model

The building of the toolbox involved many persons and we have to thanks especially Geoffrey PION (ULB-IGEAT, Belgium) Clarisse DIDELON and Sophie de RUFFRAY (CNRS-UMR IDEES), Françoise BAHOKEN, Laurent BEAUGUITTE and Maude SAINTEVILLE (CNRS-UMR Géographie-cités), Nuno MARQUES da COSTA and Diego Do ABREU (IGOT, Portugal).

Introduction

Many methodological tools can be applied to a matrix of trade flows in order to obtain a division of the world into economic regions. But each method relies on implicit theoretical considerations that are not always clearly expressed by researcher more interested in the application of new quantitative tools than in the empirical discoveries on the organization of world economy. On the other hands, researchers that focus on thematic results can be subject to the temptation to select the methodological tools that support their epistemological or ideological point of view.

The purpose of this paper is to combine theoretical and methodological points of view through an empirical application on the evolution of trade flows at world scale during the period 1994-96 and 2004-06. This period is indeed characterised by a major reorganization of the geographical distribution of trade flows at world scale.

1 Nodal regions based on dominant flows

Section written by Claude Grasland

The idea of nodal regions was firstly proposed by geographers and spatial planners in the 1960s in order to delineate functional regions based on flows between cities. The objective was to establish a hierarchy of central places based on the intensity of flows. The aim of this section is to examine if, and how, it is possible to apply this method to world division.

1.1 Concepts and bibliography

The delimitation of nodal regions by the method of dominant flows was proposed for the first time by Nyusten and Dacey in 1968 but developed and popularized by Taaffe and Gauthier in their reference book, *Geography of Transportation*, published in 1973. Many variations and improvement of the initial method has been proposed since this pioneer work, but the solution proposed by these authors remains actually widely used by geographers and spatial planners, because of the great simplicity of the solution and the general efficiency of the results obtained. This method was initially applied to bilateral information flows (telephone calls) between cities and it is not necessary obvious to transpose directly this method to asymmetric trade flows between countries. (1) The asymmetry of trade flows implies indeed that the relation of domination are not necessary the same for import and export. (2) States are areal units and not punctual object as cities, which mean that the notion of polarization is considered here as more political than geographical. An example of recent paper using this method for the analysis of trade can be found in Serano *et al.* (2007). The authors try to evaluation the propagation of economic crisis between countries that are linked by a network of trade. But the method they propose is based on the asymmetry of flows $(F_{ij} - F_{ij})$ considered as a factor of potential perturbation.

1.2 Methodology: selected tools for world division

DOM_FLOW: The original method of dominant flows (Nyusten and Dacey)

In the original formulation proposed by Nyusten and Dacey, a spatial unit i is dominated by as spatial unit j if and only if two conditions are fulfilled: (a) the maximum flow send by i is directed toward j and (b) the total sum of in-flows from j is greater than the total sum of in-flows from i.

$$(a) \quad Max_k(F_{ik}) = F_{ij}$$

$$(b) \quad \sum_k F_{ki} < \sum_k F_{kj}$$

Let us take some example to illustrate the application of the rule.

If we consider the matrix of total trade in 2004-06, the maximum export flow of Italy is directed toward Germany (47 685 M$) and the total import of Germany (767 524 M$) is greater than the total import of Italy (376 023 M$). The rules (a) and (b) are therefore fulfilled and we will consider that Italy is dominated by Germany for export trade.

In the case of Germany, the maximum export flow is directed toward France (88 409 M$) but the total import of France (492 856 M$) is lower than the total import of Germany. Therefore, Germany is not dominated and is considered as a terminal node or dominant country.

The result of the analysis is the building of an oriented graph that produces a strict hierarchy of countries. For example, Bosnia is dominated by Croatia, which is dominated by Italy, which is dominated by Germany. We can derive from this graph a typology of countries:

- Dominant countries (Germany) are the top level of hierarchy as their main flow is directed toward a country of lower level.
- Intermediate countries (Italy, Croatia) are at the same time dominated and dominant.
- Dominated countries (Bosnia) are dominated by a country but are not dominating others.

Concerning our objective to establish a partition of the world, a first obvious solution is to consider that the number of regions of the world is derived from the area of influence of dominant countries. In mathematical terms, each dominant country is the head of a tree (or water basin) that defines the area that is directly or indirectly under its control.

DOM_IMP/DOM_EXP: An oriented version of the method

This method was originally applied to bilateral flows of telephone calls between cities, which means that the direction of flows was not important ($F_{ij} = F_{ji}$). But in the case of trade, it is obviously not the case and we can produce two different graphs of domination according to the direction of flows.

- DOM_EXP is related to the domination of export countries by import countries and reveals the power of the client that can decide to restrict its import by means of different tools (protection of internal market, external tariff, rules and norms, etc.).
- DOM_IMP is related to the domination of import countries by export countries and reveals the power of the supplier that can decide to restrict its export, for example in the case of high level technological products able to produce weapons.

A good example of this difference is provided by Gabon. In terms of export, this country is dominated by USA because the majority of exports are based on oil that is bought by American companies. But in terms of import, the dominant country for Gabon is France that has inherited from the colonial period a strong position in the provision of manufacturing goods.

Possible variations

Basically, we can point two families of criticism that introduce different variants regarding rule (a).

Variation on the number of rank taken into account: They are many situations where the difference of intensity between the first flow and the following ones is very narrow, that introduce the risk of bifurcation in the structure of the graph for minor changes that are not statistically significant. For example, the first export flow of Italy is directed toward Germany (12.6%) but the second one directed to France is very near (12.3%) and as the sum of import of France is greater than the one of Italy, we can consider that Italy is dominated in a nearly equivalent way by France and Germany. It is not the case for Canada where the first flow of toward USA represents 83.1% of exports and the second one toward Japan represents only 2.5% of exports. In this case, the consideration of the second flow is clearly not justified.

An interesting variation to Nyusten and Dacey initial formulation is to replace the rule (a) by a rule based on a percentage of export. For example, we can decide that a country is dominated if the % of export is equal to a minimum value of 15%[1]. Following our previous examples, we can say that Italy is not dominated because the threshold of export of 15% is not reached and Italy will be therefore a dominant country. On the contrary, Belgium will be dominated by both Germany (17.4% of export) and France (16.5% of export). And Canada will be dominated only by USA (83.1%) but not by Japan. The domination graph will become an oriented graph but not a tree because one country can be dominated by more than one. The world partition can be for example based on dominant countries as before, but some dominated countries can belong to different clusters, introducing fuzzy situations.

It is also possible to replace the criteria of total import (derived from the flow matrix) by another criteria like GDP, if we consider that all economies are not equally opened and that the size of import or export do not necessary reflect the economic power of a country.

They are many possible variant around the initial method proposed by Nyusten and Dacey (1968) and it is not possible to define *a priori* the best mathematical solution. It is rather the comparison of results that does matter, and not the research of an "ideal" solution.

1.3 Application to trade data

The method of Nyusten and Dacey provides very interesting results when applied to the CHELEM matrix of total trade in 1994-96 and 2004-06, especially if export and import are analysed separately (Figures 1 and 2).

As it was expected, the regionalization of the World appears different when the domination is based on the "power of buyer" (dominant export flows) or the "power of supplier" (dominant import flows). The results are not limited to a division of the world but introduce also a hierarchy of states (dominant, relay, dominated) with possibility of definition of sub-regions, according to the influence area of relay countries.

Comparison through time appears of particular interest and the example of dominant import flows in 1994-96 and 2004-06 reveals very clearly the emerging power of China as "World factory". More precisely, the comparison reveals how the world has become "triadic" very recently and how Japan has been overcome by China as dominant supplier in Easter Asia, Middle East and part of Africa.

Comparison of different products appears also very promising and reveals clearly that the pattern of domination can change very strongly in case of

[1]We can also modify the rule (b) in the same way, and consider that the dominant country should have a sum of import not simply greater but 1.5 times greater than the sum of import of the dominated one.

segmentation of the model by products. For example, the situation of Russia and Germany is completely reversed when dominant flows of import are built for (1) Energy or (2) Machine. One difficulty appears with the CHELEM database that is not fully disaggregated at state level and involves artificial aggregates of countries. In the case of Energy, it appears to be a real problem in the area of Persian Gulf where all countries exporting oil (except Saudi Arabia) are clustered in one single unit (with an over-estimation of its domination as it is not a state with a coordinated policy). To evaluate the biases introduced by this aggregate, we have elaborated a benchmarking of results for a complete matrix of trade 1996-2000 that can be available at state level (168x168) or aggregated according to CHELEM divisions (92x92). The benchmarking of results (Figure 3) demonstrates that problems are generally limited (the main regions of the World remains equivalent) except in the case of Subsaharian Africa and Middle East. For these two areas, the diversity of domination is removed by the aggregation of CHELEM data.

1.4 Interest and limits

The preliminary experiments that has been realized on trade flows 1994-96 and 2004-06 lead to the conclusion that the method of nodal region can present a great interest for the objective of regionalization of EuroBroadMap:

- Simplicity: the method easy to understand and reproduce on various datasets.
- Efficiency: the results of the method are consistent with major studies on recent evolution of world trade (growing role of China, the relative decline of Japan and USA).
- Sensitivity: the distinction of export and import is of particular interest, as well as the distinction of domination by type of products.

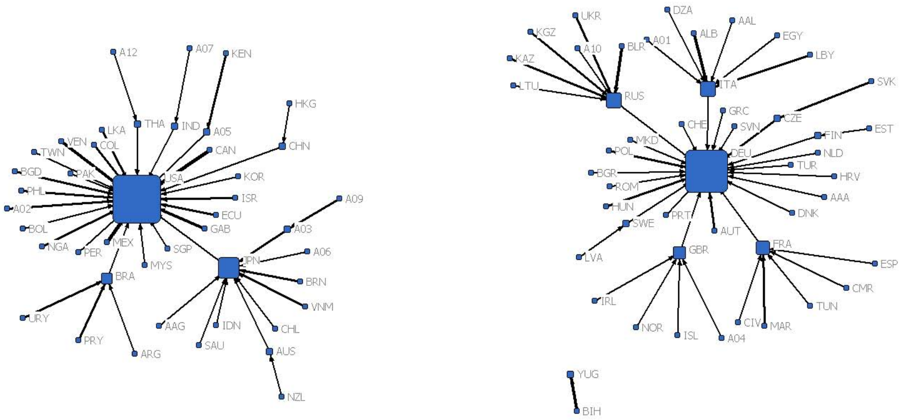
But these advantages have to be balanced by some problems of the method from both theoretical and empirical point of view.

Theoretical problems: the initial method is based on an assumption of single domination based on first rank flows. This position is not obvious and we can imagine variant of the methods where multiple domination are allowed. This would fit better with strategies of minimization of risk that are chosen by countries (that avoid having one dominant partner).

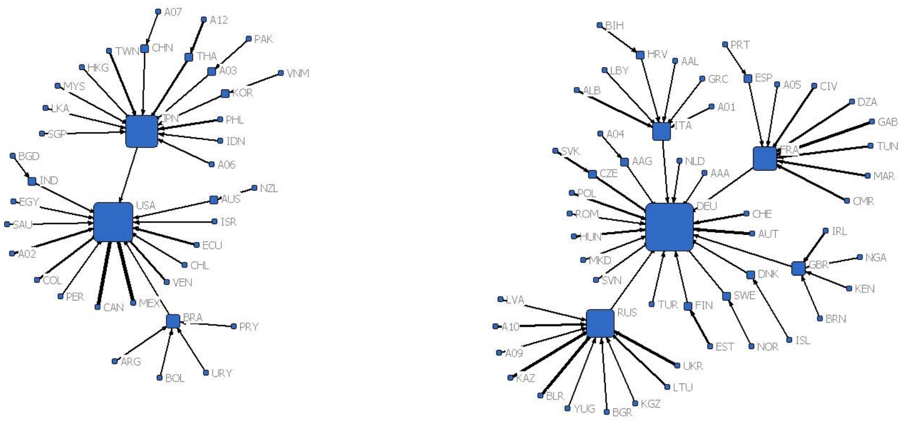
Empirical problems: the fact to choose states as building block of the analysis is not a problem if we assume that trade is strongly related to national strategies. We can easily introduce commercial blocks in the model (like EU or MERCOSUR) and measure their effect on the graph of domination. But what is more troubleshooting is the fact that the data are not always available at state level and can be based on artificial aggregates.

Figure 1: Graph of dominant trade flows - 1994-96 - All products - Export (top) & Import (down)

All products - Export 1994-96



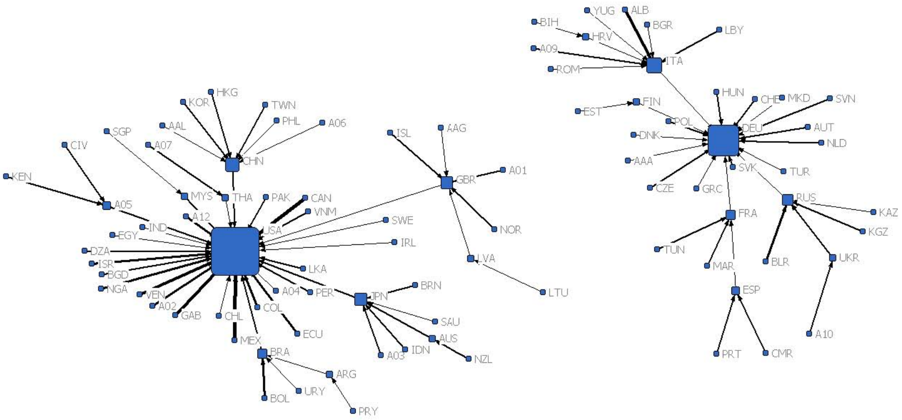
All products - Import 1994-96



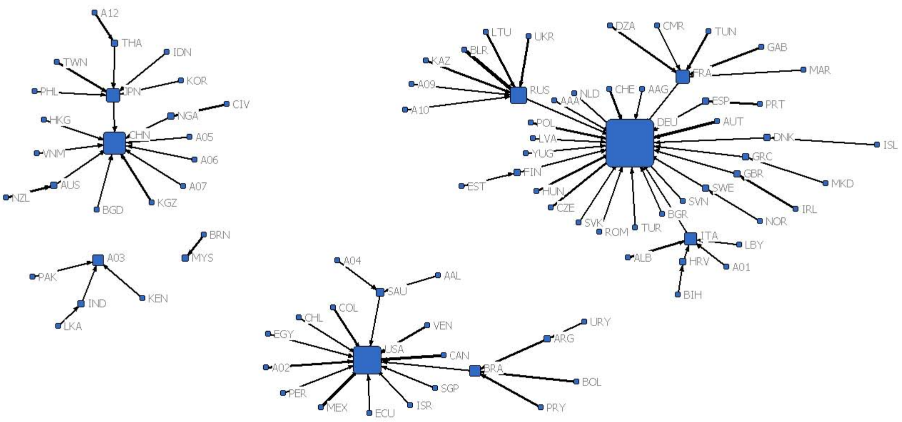
In terms of export domination as in terms of import domination, the world of 1994-96 was clearly divided in two influence area. The Pacific region is dominated by USA with a major relay in Japan and secondary dominant center in Brazil. The Euromed region is dominated by Germany with major relays in Italy, France, Russia and Great Britain. The influence of the former soviet empire remained very strong and all former soviet republics remained strongly connected to Russia, especially for importations. China appeared as a minor actor of the world trade system. Colonial links appeared always strong, especially from African countries toward France and UK.

Figure 2: Graph of dominant trade flows - 2004-06 - All products - Export (top) & Import (down)

All products - Export 2004-06



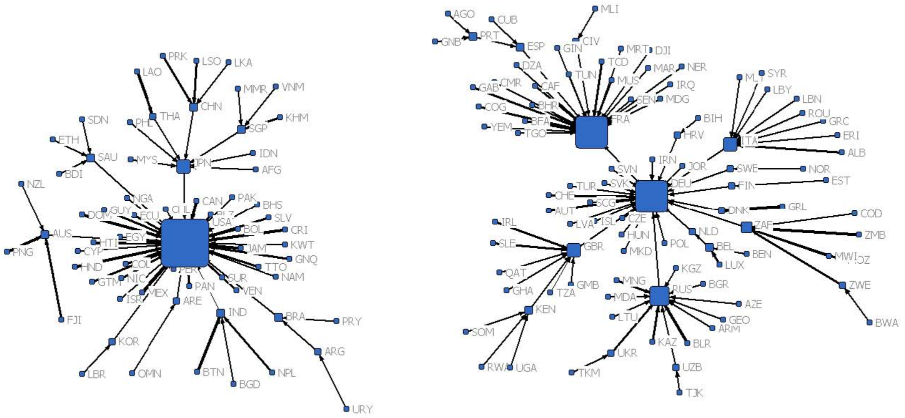
All products - Import 2004-06



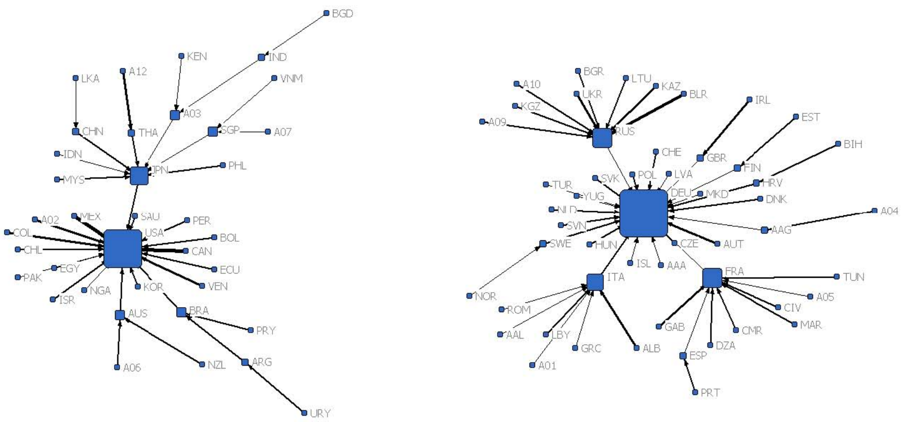
Very important changes can be observed in 10 years. In terms of export the world of 2004-06 is always divided in two influence area controlled by Germany and USA. But important changes can be noticed, especially with the transfer of UK and other northern European countries toward USA instead of Germany. In terms of import domination the modification are more dramatic: the former USA regions is now limited to Americas and part of Middle East. A third major area appears, dominated by China and Japan. A fourth cluster appears in southern Asia, linking India and Pakistan to the region A3 which is in fact a cluster of Iran and Iraq.

Figure 3: Comparison of dominant import trade flows 1996-00 at different aggregation levels

168 States



92 CHELEM units



The impact of aggregation from 168 states to 92 Chelem units on the results appears more important for import domination as compared to export domination. The global divisions of the world are more or less the same (two regions dominated respectively by USA and Germany) but many differences appear in the detail. The best example is provided here by United Kingdom (GBR): at state level, they are 7 countries directly dominated by UK and 10 if we take into account indirect domination through Kenya; but after aggregation to CHELEM units, it remains only one country (Ireland) dominated by UK, instead of 10.

Figure 4: Dominant export trade flows in 2004-06

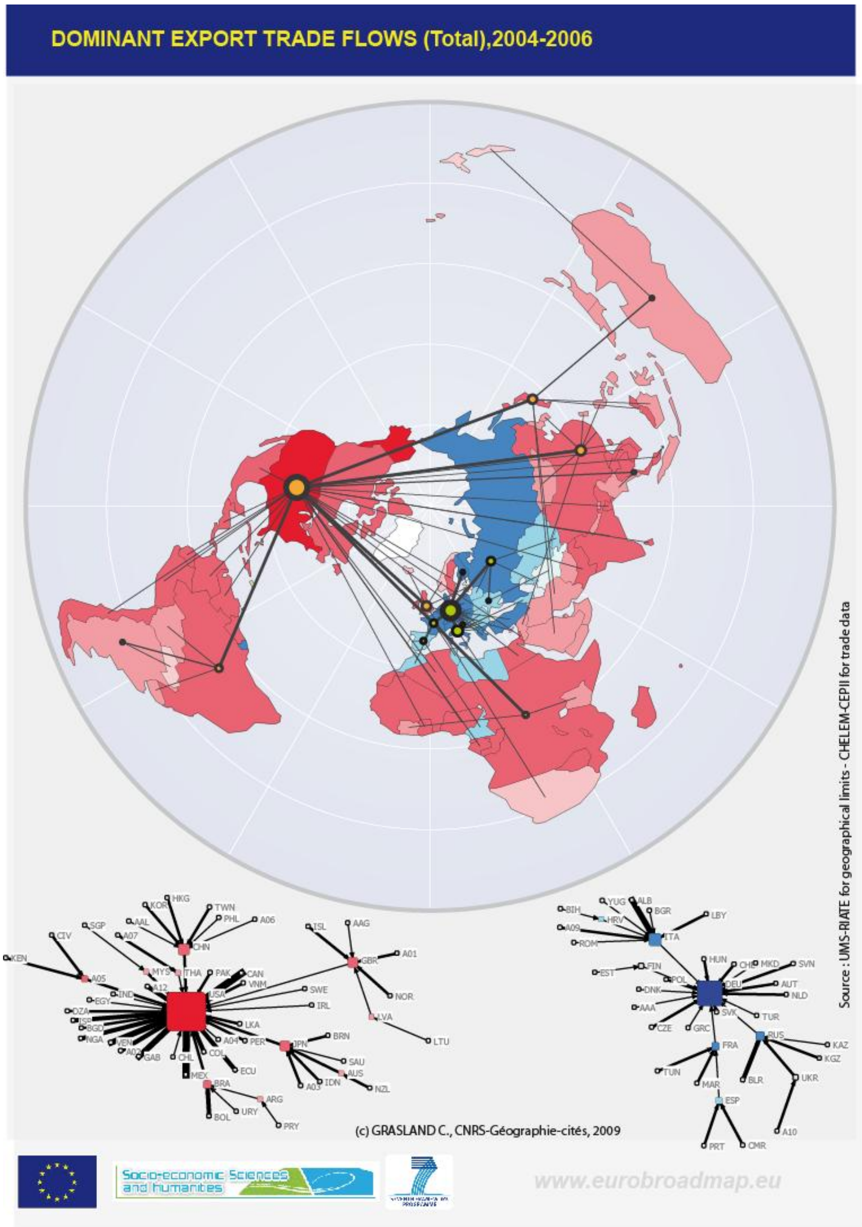
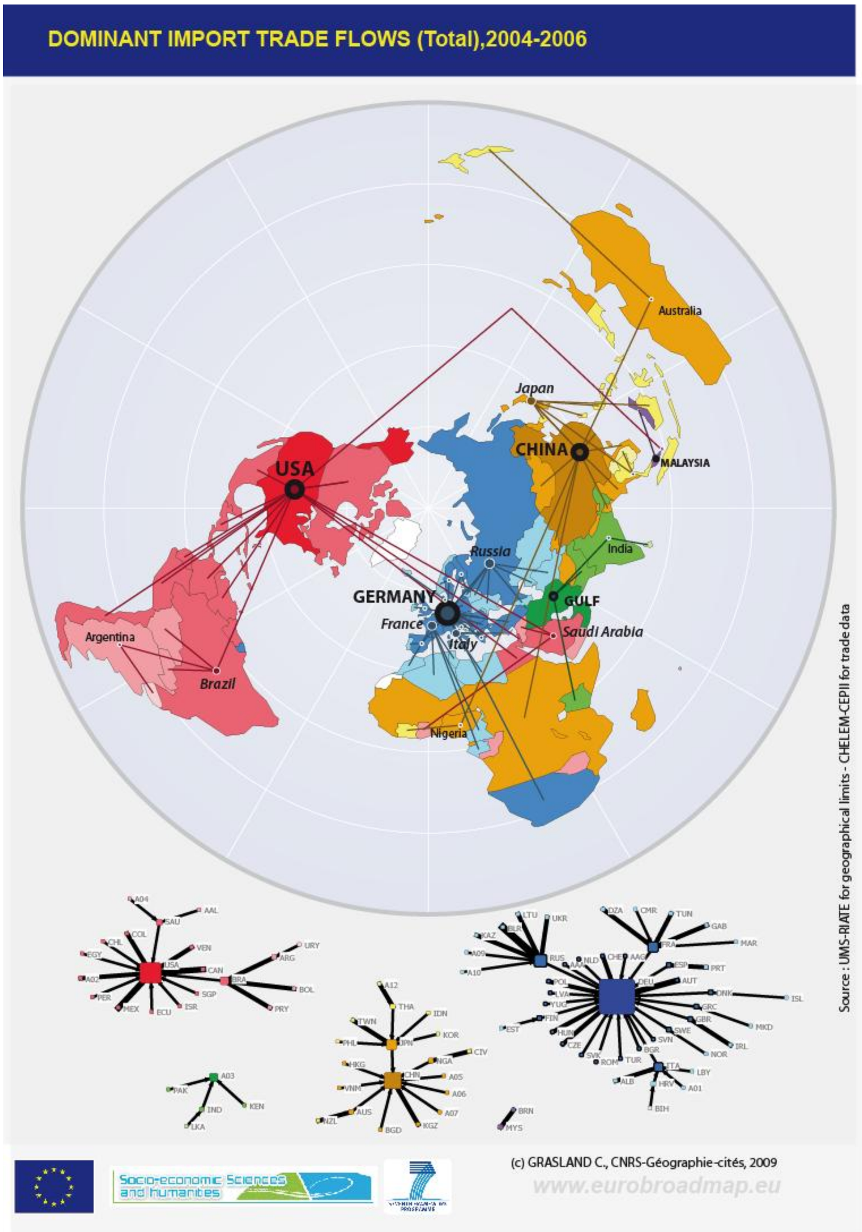


Figure 5: Dominant export trade flows in 2004-06



2 Functional regions based on Intramax procedure

Section written by Françoise Bahoken and Nuno Marques da Costa

The clustering procedures in statistics have two main objectives. The first one aims to summarize spatial attributes in order to differentiate spatial types and create means profiles. The second one aims to create new zoning of the space based on the key concept in spatial organization which are the "Functional Areas". The main idea of such a zoning system is to define a multi-level space typically based on flows data, included in some case a contiguity criteria.

The aim of this section is to examine if, and how, it is possible to apply a method of functional clustering to divide the world based on Chelem's international trades data, and if the spatial results are valid. To create the functional areas on Chelem's world division, we carried out an analysis using a hierarchical clustering algorithm called Intramax procedure, incorporated in Flowmap© (Van der Zwan *et al.* 2003).

2.1 Concepts and bibliography

The concept of Functional Areas is clearly derived from the gravity model. The more interactions between spatial units, the shorter the distance. Interactions are then be observed as a measure of functional distances between spatial units: high level of interactions indicates short level of functional distance. The spatial units close in terms of functional distance will be grouped together in order to create functional regions.

In the context of flows, the zoning design problem is consider in one hand as an alternative of administrative zoning systems depending on the Modifiable Area Unit Problem (MAUP); and in the other one, as Openshaw (1977) said, as provided flows data that better match to theoretical requirement for gravity modelling.

The zoning design problem have four main objectives (Alvanides *et al.* 2000):

- A data reduction in order to create a convenient level for spatial management and data reporting purposes;
- An optimization of the spatial representation performance so that the partitioning minimize distortions and bias;
- A tool to make the flows visible in an understanding map form as a useful description of spatial organization and flow structure;
- To enhance models performance by tuning the spatial aggregation to generate more model friendly data.

A main issue consists then on the choice of the objective function of the zone design and their effects on the scale and the aggregation. This choice has to be made according to the observed data and the problem investigated. Several objectives functions (Alvanides *et al.* 2000) have been defined in the 1970s in the relevant literature in order to maximize or to minimize the intra or the inter-zonal flows.

The first objective function suggested by Masser *et al.* was fully implemented by Masser and Brown (1975) to study movement data for London and Liverpool. That well-known method is the so-called Continuous Intramax Analysis developed by Masser and Scheurwater (1977). The Intramax objective function aims to:

$$Maxf(Z) = \sum i \sum j(a_{ij} - a * ij) \quad \text{for all} \quad i \neq j$$

with $a_{ij} = \frac{F*ij}{S}$ and $a * ij = \sum a_{kj} \sum a_{ik}$

F_{ij}: the original matrix of flows between the origin i and the destination j,

S: the sum of the observed flows in the original matrix F_{ij},

$F * ij$: the aggregated matrix $(N * N)$ of the N spatial units,

$a * ij$: the variation of the size of the row and column total.

The function merges together the N units step-by-step by maximizing the proportion of the total interactions.

The second objective function is based on the suggestion of Hirst (1977) to handle variations in the row and columns total on the residual values. The modification aims to divide the previous function f by $a * ij$.

$$Maxf(Z) = \frac{\sum i \sum j(a_{ij} - a * ij)}{a * ij} \quad \text{for all} \quad i \neq j$$

This function is considered as the original Intramax model.

Broadbent (1969) noted that a basic objective of a good zoning system is to ensure that it will produce more interaction between the spatial units than within them (i.e. the intra-zonal interactions are smaller compared to the inter-zonal interactions).

Masser and Brown (1978) noted that the Broadbent criteria is achieve when 85-90% of the total number of flows cross a boundary and then reformulated their Intramin objective function as minimizing intra-zonal flows:

$$Minf(Z) = \frac{\sum iF * ij}{S}$$

with $F * ij$ the intra-zonal flows for the zoning system and S the sum of the observed flows in the original matrix F_{ij}.

This function had been commonly used with the Intramax one in order to create zoning systems for gravity modeling where the requirement was to maximize the inter-zonal flows (intra-zonal flows are less represented in gravity model).

The function which maximizes inter-zonal flows is expressed as:

$$Maxf(Z) = \frac{\sum iF * ij}{S}$$

Another methods of zoning systems from flows include contiguity constraints. One of them is the so-called Mirabelle, developed by the French institute of statistics (INSEE). The Mirabelle function aims to design zone system according to flows ($F_{ij} + F_{ji}$) and the mass (or weight) of flows (M_i and M_j) which are for example the inhabitants of the zone or the margins of the matrix.

2.2 The Intramax procedure

The original Intramax model has been implemented in the GIS program Flowmap© (http://flowmap.geog.uu.nl). The software is specialized in the treatment and the analysis of flow data like migration and commuting flows, network analysis, interaction analysis and gravity modelling (Flowmap© uses three types of data: flow data, distance tables and maps. All tables can be done into Flowmap© or import from *DBFIII format or/and other current GIS software.

The Intramax analysis is a step-by-step procedure with no contiguity criteria which carries out a regionalization of an interaction matrix.

At the first step, two spatial units are grouped together in order to create another spatial unit. "The objective of the Intramax procedure is to maximize the proportion within the group interaction at each stage of the grouping process, while taking account the variations in the row and column totals of the matrix" (Van Der Zwan *et al.*, 2005). This implies that the aggregation process maximized the function:

$$I = \frac{F_{ij}}{O_i * D_j} + \frac{F_{ji}}{O_j * D_i}$$

with F_{ij} the interaction between the origin area i and the destination area j, O_i and D_j the margins of the matrix (respectively $\sum jF_{ij}$ and $\sum iF_{ij}$).

The function can be calculated in Flowmap© for all $D_j > 0$ and for all $O_i > 0$. A zone which has no in-flow or no out-flow has to be ignored and suppressed from the matrix.

At the second step, the interaction between the two areas becomes intrazonal interaction for the new resulting areas. These two areas take the place of the parent area at the next step of the procedure and so on. At the end of the procedure, all the units are grouped and create a single spatial unit. In the case of N areas, after $N-1$ steps, all areas are grouped and all interactions are intrazonal.

The result of the Intramax analysis is a:

- A dendogram of the clustering;
- The different steps of the procedure list in a Fusion Report (FR);
- A new column in the structural table with the result of the spatial aggregation: in order to make it, we have to choose the number of the fusion steps to be calculated regarding the dendogram, the FR and our knowledge of the data and the subject.

After theses results, we can calculate outside Flowmap© the aggregated flows and create the new zoning system by merging previous zones. We can also make the complementary analysis after the Intramax procedure called "Trip-end-ranking". This analysis is useful to determine the role of the distance to the end destinations. For example, most countries do not have current trade relations with countries far away, but have some with their neighbours...

2.3 Application to trade data 1994-96 and 2004-06

We used the conventional Chelem's boundaries as the starting point of the aggregations. The flows data are the total in 1994-96 and in 2004-06.

In 1994-96, the initial intra-zonal interaction is 0.13%. The cumulative intra-zonal interaction is increasing with the steps. After 92 steps, all areas are merged and the cumulative intrazonal interaction percentage becomes 100. In order to map the new zoning system at the two dates and compare it with the previous one, we decided to stop the aggregation process at two different steps.

In order to determine some valid levels of aggregations between previous zones (i.e. the moment to stop the procedure) and according to Poon (1997), we examined the fusion report and found in one hand, where the increment in intrazonal interaction was the most important and in the other hand, where the cumulative intrazonal percentage was in the average of 50%.

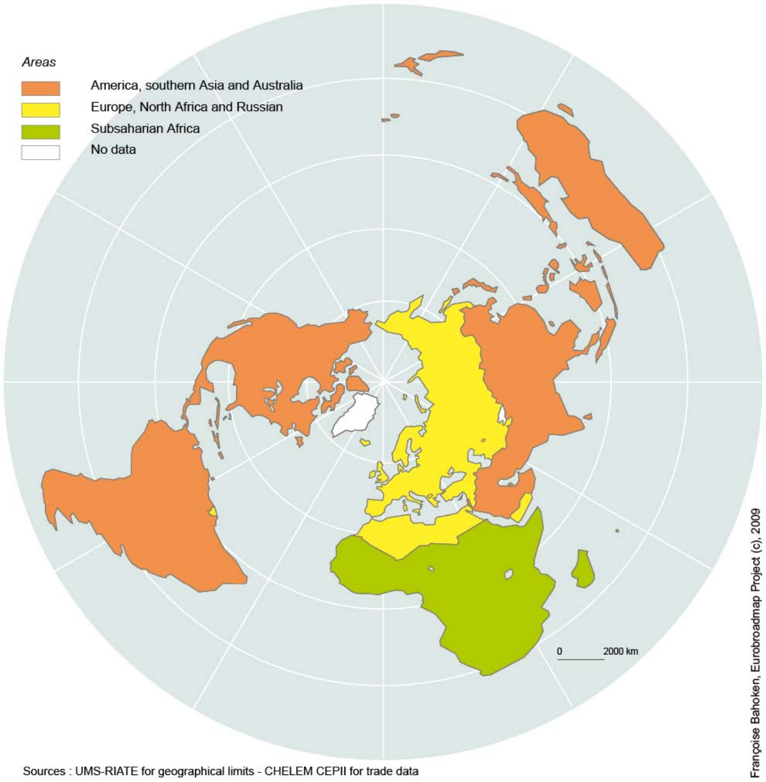
For 1994-96, the most important increment occurred at step 90 corresponding to + 77.04% intrazonal flows: the world is then divide into three blocks (America, Europe, Russian Federation, Africa and Yemen, Asia and Australia) (Figure 6).

At step 86, the increment is + 10.11% and corresponds to 51.45% of cumulative intrazonal percentage: the world is divided into 6 blocks + Israel. At the next step (87), the increment is + 9.17% and the cumulative intrazonal percentage is 60.62. The main difference with the previous clustering is that Northern Europe is merged with Occidental Europe (Figure 7).

This map (Figure 7) has to be compared with Poon's spatial structure of world trade regions in 1990, but one should take into account the fact that Chelem's world boundaries are already partially aggregated at the beginning of the clustering procedure. The fact that Sub-Saharan Africa is composed

Figure 6: Intramax clustering on Chelem's international trade flows (1994-96) - 1

World divided at 77.04% intrazonal flows



by few countries creates differences: where in our maps, Africa is a whole region, in Poon's one: the whole Africa is divided into two majors blocks (Germany and UK regions), Egypt being merged with the US region and Ethiopia with the Japan one. Regardless that point, the single difference between Poon's map and our maps is Israel case: in our maps, Israel is a whole region at the opposite of Poon's one which is merged with Germany Region.

For 2004-06, the most important increment occurred at step 88 with +12.29% intrazonal flows corresponding to 57.48% of cumulative intrazonal percentage (Figure 9).

At step 91, the increment is +12.20% intrazonal flows and the cumulative intrazonal percentage is 78.46%: the world is then divided into two blocks: Russian Federation, Europe and North Africa on one hand, and the rest of the world on the other (Figure 8).

In the maps focused on 2004-06, the mains observations are firstly: the

Figure 7: Intramax clustering on Chelem's international trade flows (1994-96) - 2

region which merged Occidental Europe and North Africa. In both case, Algeria is merged with the Russian Federation areas. Secondly, we can note that Africa is in the same region than emerging countries like India or petroleum's one (i.e. Persian Gulf's countries).

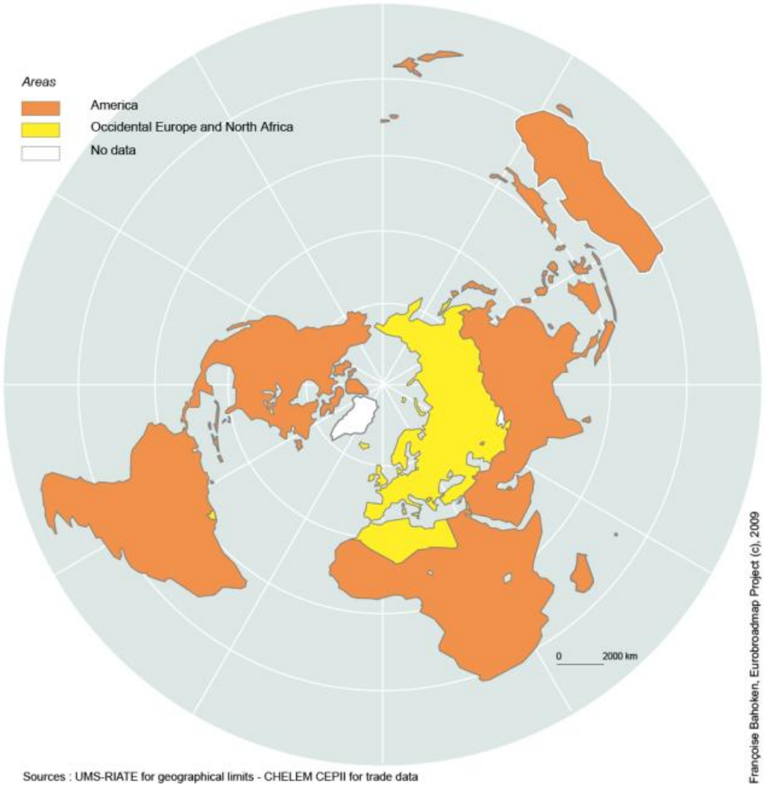
2.4 Interest and limits for EuroBroadMap project

The Intramax zoning system based on flows data is providing an aggregate spatial representation of the initials flows. The off diagonal elements are therefore the intrazonal flows: the small areas area aggregate into larger one and the inter-flows and the intra-flows are both recalculated at each step. Simultaneously, the inter-flows between small areas become intra-flows.

The zoning system seems to be relatively stable at the two dates for aggregations at steps 86 (1994-96 with 51.45% intrazonal flows) and 87 (2004-06 with 45.20% intrazonal flows), but there is only one difference which have

Figure 8: Intramax clustering on Chelem's international trade flows (2004-06) - 1

World divided at 78.46% intrazonal flows



to be mentioned regarding the Region Africa. In 2004-06, the Region Africa is merged with petroleum countries (Yemen, Oman and Saudi Arabia) and Algeria is included in the Russian federation Region while in 1994-96, Africa it a whole region.

This analysis does not provide the correspondence between the aggregation steps (in the Fusion Report) and the number of zones out; the justification of the choice of the fusion step: the choice must depends on another characteristics, nor the appropriate zoning system for data modelling.

Figure 9: Intramax clustering on Chelem's international trade flows (2004-06) - 2

World divided at 57.48% intrazonal flows

3 Natural/Residual regions based on spatial interaction models

Section written by Claude Grasland

3.1 Concepts and bibliography

The use of spatial interaction model for the description and explanation of flows is based on a theoretical framework that has been elaborated in various fields of social sciences since the end of the 19th century. Demographers (Ravenstein, Courgeau), geographers (Stewart, Tobler, Hägerstrand), economists (Reilly), and sociologists (Stouffer, Zipf) have all contributed at different period and with different points of view to the elaboration of this large family of models called spatial interaction models or gravity models. The common point of all this contribution is (1) the direct inspiration of the Newtown law of gravity as initial starting point, with a particular role of assumptions on the effect of "masses and distances" - even if this analogy was further criticized and (2) the quantitative and mathematical approach of flows.

The different social sciences use more or less the same family of mathematical equation for the description of flows, but with very different theoretical assumptions on the meaning of the parameter of this equation and, more important, on the interpretation of the residuals of the models of spatial interaction. For example, Stouffer has elaborated in the 1960s a specific variant of the gravity model (intervening opportunities) in order to demonstrate that distance was NOT a factor that influences directly human behaviour. According to Stouffer's view, the decrease of migration with distance is related to the fact that social opportunities of interaction are located between the origin of the migrant and the destination. It is therefore a fallacy to attribute to the geographical factor of distance what is in fact a social effect that can be explained by human behaviour. The case of economist, especially the one interested in the analysis of trade, is different from the one of sociologist: when sociologist denied the effect of distance, economist tried rather to find a theoretical basis to the gravity model, in order to integrate it in a wider perspective.

The relation of economists with gravity model in particular - and geography in general - are also characterized by a strong ambiguity, that is perfectly illustrated by the different contributions of the book *The Regionalization of the World Economy* published by Frankel (1998). As quoted by Bergstrand, there is a "frustration fascination of trade economist with the gravity equation" because they recognized its very high explanatory power, but they have many difficulties to explain this explanatory power by "real" economic theory like Heckscher-Ohlin model of equilibrium. In a review of thirty years of use of gravity model by trade economists, A.V. Deardorff (in Frankel 1998) observe that since the pioneer work of Tinbergen (1962) or Linneman (1966), the economist has encountered many difficulties to link the empirical model of gravity with relevant theoretical explanations on why it works. But with further development of research (Linneman 1966; Leamer and Stern 1970; Leamer, 1974; Anderson 1979; Bergstrand 1985, 1989 and 1990) another problem appeared as many economic theories of trade appeared likely to provide alternative explanations on why gravity model worked... More important, Grossman suggest that none of the theories proposed by economist are sufficient to explain the power of distance decay effect in a globalised world where the cost of transportation as become relatively low. Krugman (2004) expressed the same surprise that distance and borders remained more important than expected according to classical economic framework, suggesting as Grossman that something was still missing in the economic models of world trade: "What seems to have emerged from the empirical work of the past dozen years is a compromise vision. Distance matters a lot, though possibly less than it did before modern telecommunications. Borders also matter a lot, though possibly less than they did before free trade agreements. The spaceless, borderless world is still a Platonic ideal, a long way from coming into existence."

Natural region and free trade areas

The interest of economist for the elaboration of theoretical foundation of gravity model is directly related to a major scientific - but also ideological - challenge which is to evaluate the respective effect of the regionalization and globalization of world economy after the fall of the socialist block (1989) and the growth of protectionist trade area in European Union (1992) and further in other parts of the world (NAFTA, ASEAN, MERCOSUR...). Krugman played a major role in the development of this research area when he proposed at the beginning of the 1990s two opposite interpretation of the effect of Regional Trade Agreement (RTA) on the liberalization of world economy. A good summary of the problem can be found in a paper by Frankel *et al.* (1995) entitled "Trading blocs and the Americas: The natural, the unnatural, and the super-natural". Briefly summarized, Frankel start from the neoclassical assumption that free trade between states of the world is a factor of progress of economic welfare. The problem is that the perfect solution of integral free trade between all countries of the world is actually impossible to establish and it is therefore of interest to analyse what could be the "second best solution" of partial free trade. On the one hand, it is a good thing as it develops international free trade between countries belonging to the same bloc; on the other hand, it is a bad thing as it produces external trade barriers between countries belonging to different blocs. What is therefore the balance between positive and negative effects? Based on the econometrics analysis of economic welfare, the answer is ambiguous. We have

Table 1: Trading Blocs and Natural Regions: The Krugman *vs* Krugman Case

In his first contribution (Krugman 1991a), he focused on the idea that when individual countries form larger groupings, they are liable to become more protectionist, and thus to move farther from the ideal of world free trade. The reasoning was that as a group they would set higher tariff levels *vis-à-vis* the rest of the world, since they would have more monopoly power to exploit. Units were assumed to set tariffs at a self-maximizing optimal levels. He showed that world welfare is lower with a few trading blocs than with the extremes of one or many, and that for specific plausible parameter values, three turned out to be the worst possible number of blocs to have!
His second contribution, Krugman (1991b), provided a useful review of the whole array of issues and factors involved. But it also included a very simple argument that leads to the diametrically opposite conclusion from the first one, that trading blocs are good. It is observed that even without the formation of regional free trade areas or preferential trading arrangements of any sort, countries trade more with their neighbors than with countries from which they are far removed, in part because of transportation costs.
Source: Frankel *et al.*, 1995.

therefore a Krugman (1991a) *vs* Krugman (1991b) case that is summarized in table 1.

The Geography Strikes back: Natural regions *versus* Residual regions

It is important to turn back to geographers' point of view on (1) theoretical justification of gravity model, (2) status of distance and (3) delineation of world region.

The basic point of debate between geographers and economist is related to the interpretation of the role of mass (GDP, Population) and proximity (distance, contiguity, common language) in the gravity model.

For neoclassical trade economist, this factor is considered as "natural" or, more precisely, as exogenous parameter that is independent from the configuration of trade flows. As a typical example, Frankel *et al.* (1995) indicate that the measure of the effect of RTA is possible only when this exogenous factors are controlled: "First, we shall measure the extent, by looking at the magnitude of bilateral trade flows after one adjusts, by means of the gravity model, for such natural determinants of bilateral trade as

GNPs and proximity". The implicit assumption is therefore the existence of a universality of this factor that produces the same effect on trade all around the world. More precisely, it implies that (1) A given amount of GDP will generate the same amount of export or import all around the world (with eventual differences related to size effects but with the same elasticity) and (2) that a given transport cost will reduce the trade by the same amount, according to Samuelson's iceberg hypothesis.

For geographers working on advanced spatial interaction model the development of gravity models has followed a completely different way during the last 40 years. The initial formulation of the gravity equation in bi-logarithmic form has been replaced very early by more convenient models, taking into account the problems of error measurement and solving the question of zero-flows (Fotheringham and O'Kelly 1989; Sen and Smith 1995). New forms of gravity model have been proposed with double constraint on origin and destination, either in multiplicative (Wilson, 1967) or additive form (Dorigo and Tobler 1983). This family of double constraint model is particularly useful for the evaluation of barriers and preferences under the assumption of an equilibrium model of trade between countries of the world (i.e. under the assumption that all exports and imports of countries are given - margin of the matrix - this model provides an exact solution for trade allocation between countries). Despite its theoretical interest, this family of model was very few applied to world trade flows, with the notable exception of Bröcker and Rohweder (1990). But the most crucial difference between geographer and economist point of view is related to the question of distance considered as a central parameter of the analysis. The classical assumption of the gravity models that flows depend from distance can be reversed and transform into the reverse assumption that distance can be revealed by the observation of flows if we reverse the gravity model (Tobler 1983).

The same is true for "regions" that are not necessary considered as predefined for geographers. Of course, it is possible to adopt a deductive approach and to test the effect of a given division of the world that is supposed to have an influence on flows. Geographers can introduce variables that try to capture preferences and barriers according to different partition of space that are established a priori: effect of RTA on trade (Bröcker and Rohweder 1990), effect of linguistic barriers on telephone calls (Klaassen *et al.* 1972; MacKay 1953), effect of political and historical divisions on internal migratory flows (Cattan and Grasland 1992), etc. But it is also possible to adopt an inductive approach and to try to reveal unknown division of space in region characterized by internal preferences and external barriers.

3.2 Selected tools for world division

We propose first a review of the family of models of spatial interaction that can be used for the description of flows, and then we derive from one of the proposed family a set of solution for the elaboration of regional division of the world. We conclude with a reflection on distance.

The classical family of gravity model takes the form of a linear regression after bi-logarithmic transformation. The explanatory variables are divided in two group of factors: (1) the so-called natural factors that are supposed to have the same effect in all countries and (2) the barriers that are related to RTA. Some authors suggest to add (3) lagged trade flows (i.e. trade flows in a previous period) in order to evaluate the historical dimension of the phenomena and potential effects of hysteresis (Eichengreen and Irwin 1996). One classical example of this family of model is:

$$F_{ij} = k.(Y_iY_j)^{\beta1}.(P_iP_j)^{\beta2}.(DIST_{ij})^{\alpha}.(CONT_{ij})^{\lambda1}.(LANG_{ij})^{\lambda2}.(COL_{ij})^{\lambda3}$$

with
F_{ij}: bilateral flows between countries i and j
Y_iY_j: GDP of countries i and j
P_iP_j: population of i and j
$DIST_{ij}$: distance between i and j
$CONT_{ij}$: common border between i and j (dummy variable, 1 = yes, 0 = no)
$LANG_{ij}$: common langage netween i and j (dummy)
COL_{ij}: colonial link between i and j (dummy)
$k \quad \beta1 \quad \beta2 \quad \alpha \quad \lambda1 \quad \lambda2 \quad \lambda3$: parameters to estimate.

The evaluation of parameter is ordinary made after log linear transformation of the equation:

$$\begin{aligned} log(F_{ij}) = a_0 + a_1log(Y_iY_j) + a_2log(P_iP_j) + a_3log(DIST_{ij}) \\ + a_4(CONT_{ij}) + a_5(LANG_{ij}) + a_6(COL_{ij}) + \epsilon_{ij} \end{aligned}$$

with
$k = exp(a_0)$, $\beta_1 = a_1$, $\beta_2 = a_2$, $\alpha = a_3$, $\lambda_1 = a_4$, $\lambda_2 = a_5$, $\lambda_3 = a_6$ and ϵ_{ij} = error of the residuals.

The linear transformation provide easiest statistical solution but introduce many problems in the estimation of the model, especially when equation is solved by Ordinary Less Square (OLS):

- Zero flows are removed or fixed to an arbitrary value;
- Gaussian assumption of residuals is not fulfilled;
- Real uncertainty of flows (that is ordinary proportional to the square root of F_{ij}) is not properly taken into account.

Table 2: Modelisation of bilateral trade with unconstrained gravity model in 2004-06

Model	Param.	Sign.	Dev.	Adj1	Adj2
Intercept	-5.878	<.0001	92323745	0.0%	-
$\log(Y_iY_j)$	0.742	<.0001	24436846	73.5%	-
$\log(P_iP_j)$	0.129	<.0001	24192717	73.8%	0.0%
Log(dist)	-0.771	<.0001	11103387	88.0%	54.1%
Contiguity	0.365	<.0001	10306775	88.8%	57.4%
Language	0.484	<.0001	9755970	89.4%	59.7%
Colonization	-0.217	<.0001	9696475	89.5%	59.9%
Scale	41.391				

A more convenient solution from statistical and thematic point of view is offered by the family of Poisson regression models that use a variant of Maximum Likelihood criteria on flows without logarithmic transformation, making possible to keep zero flows in the analysis and insure a better representation of each flow as regard to the uncertainty of measure. The only important point for the use of Poisson regression model (Calzada *et al.* 2000) is to introduce a scale parameter (internal to the model) that allows a stability of the results, independently from the unit of measurement of trade flows ($, thousands of $, billions of $ ...). Accordingly, the model to be solved can be written as:

$$F_{ij} = SCALE.exp[a_0 + a_1 log(Y_iY_j) + a_2 log(P_iP_j) + a_3 log(DIST_{ij}) + a_4(CONT_{ij}) + a_5(LANG_{ij}) + a_6(COL_{ij})] + \epsilon_{ij}$$

The application of this model to bilateral trade flows 1994-96 and 2004-06 of the Chelem database give the following results (see tables 2 and 3).

With Poisson Regression Model, the explanatory power of the model is not measured by r^2 value (as it is the case in log-linear regression with OLS) but by the reduction of the deviance which can be used as a measure of the information explained by the model. We can see in the example of trade 2004-06 that the initial value of deviance given on the line “Intercept” is equal to 92323745, that will define the maximum error (Adj1=0%). We can now try to evaluate the effect of each explanatory variable on the model through the examination of the reduction of deviance. The introduction of economic size effect (GDP) is clearly the most important. In 2004-06, it reduces the deviance from 92323745 to 24436846 which means that 73.5% of initial deviance of trade flows is explained by the difference of economic size of the states. The effect of population (or GDP/capita[2]) appears sta-

[2] As the product of GDP is firstly introduced in the model, adding the product of

Table 3: Modelisation of bilateral trade with unconstrained gravity model in 1994-96

Model	Param.	Sign.	Dev.	Adj1	Adj2
Intercept	-7.601	$<.0001$	46502872	0.0%	-
$\log(Y_i Y_j)$	0.802	$<.0001$	11474330	75.3%	-
$\log(P_i P_j)$	0.023	0.002	10934157	76.5%	0.0%
Log(dist)	-0.658	$<.0001$	5381558	88.4%	50.8%
Contiguity	0.424	$<.0001$	4890187	89.5%	55.3%
Language	0.556	$<.0001$	4500808	90.3%	58.8%
Colonization	-0.053	0.138	4499011	90.3%	58.9%
Scale	28.404				

tistically significant but less important and produces only a small increase of explanatory power to 73.8%. The distance introduces another important step as this parameter reduce the deviance to 7125424 which means that 88% of the initial deviance has been removed (Adj1).

The other parameters (contiguity, language, colonial relations) are clearly less important in terms of explanatory power, but what is important is the measure of their significance, that can be evaluated by a χ^2 test, either in a sequential way (the effect being measured as compared to variables introduced before) or in a simultaneous way (the effect being measured all thins being equal with the effect of the other variables). The contiguity parameter ($\log(\lambda 1) = 0.365$, so λ_1=1.44) reveals for example an increase of trade of +44% between countries sharing a common border in 2004-06. The existence of a common language shared by 20% of inhabitants of each country produce an increase of trade of + 62%. But the existence of colonial relations (in the past) produces an unexpected negative effect of reductions of trade flows by -20%. It is important to keep in mind that this effects are understood "all things being equal with the other explanatory variables". The comparison of the results of the same model applied to bilateral trade data 10 years before reveals some interesting evolutions of the effect of the explanatory variables (see table 3).

The effect of economic size on trade is less important in 2004-06 than in 1994-96. But the effect of population size is more important. Contrary to a general idea, the effect of distance is rather increasing as the parameter of distance effect (α) is growing from -0.66 to -0.77. When distance is doubled (x 2), the flows were reduced by -37% in 1994-96 and by -42% in 2004-06. At the same time, we can observe a reduction of the effect of contiguity (+52%

population is equivalent to adding the product of GDP per capita as the model use a multiplicative form. The aim is to control a possible effect of increasing trade between more or less developed countries.

to +44%) or common language (+74% to +62%) and the emergence of a reverse effect of colonial relation (-6% to -20%).

The application of gravity model to trade flows reveals interesting evolution of globalization but does not provide a direct answer to EuroBroadMap's problem of definition of world region without a priori on the results. Of course, we could introduce in our previous models some variables related to existing Regional Trade Agreement (EU, NAFTA, MERCOSUR...) and obtain an evaluation of their effects. But in this case our approach would be clearly deductive and the only thing that we could obtain could be a confirmation or invalidation of the effects of this existing division on world trade.

Gravity Model 1: Control of size effect

In this first model, we assume that the only explanatory parameter that can be considered as global is the size of the countries. We will therefore eliminate all other parameters from the model and build a gravity model 1 that is based only on GDP.

Hypothesis: the volume of bilateral trade between two countries is proportional to the product of their GDP.

$$F_{ij} = SCALE.exp[a_0 + a_1 log(Y_i Y_j)] + \epsilon_{ij}$$

Gravity model 2: Control of economic size and geographical proximity

In this second model, we assume that flows depend not only on economic size but also on geographical proximity. We will therefore add distance and contiguity to GDP in order to obtain a gravity model 2.

Hypothesis: The volume of bilateral trade between two countries depends on their economic size and their geographical proximity.

$$F_{ij} = SCALE.exp[a_0 + a_1 log(Y_i Y_j) + a_2 log(DIST_{ij}) + a_3(CONT_{ij})] + \epsilon_{ij}$$

Problems of interpretation of residuals of unconstrained gravity models

The interest of the comparison of residuals of gravity models 1 and 2 is to demonstrate that regional divisions that will be obtained by a classification procedure are heavily dependant on the assumption made on what are the "natural" determinants of trade. If we control only the size effects (Model 1), the factor of proximity is included in the residual and the clustering methods will probably revealed groups of states that are near each other. On

Table 4: Gravity model 1

i	j	F_{ij}	$F*_{ij}$	R_{ij}	X_{ij}
Most significant positive residuals					
USA	Canada	465180	96025	369155	11.60
Belgium	Netherlands	93669	4316	89353	8.52
Belgium	Germany	124470	13740	110731	7.73
USA	Mexico	279985	76544	203441	7.66
Germany	Pays-Bas	126843	19287	107556	6.92
France	Belgium	96633	11158	85476	6.70
Germany	Austria	91231	10777	80454	6.46
Singapore	Malaysia	32561	498	32063	6.16
S. Korea	China	105229	19158	86072	5.83
Taiwan	China	76869	10185	66684	5.69
Most significant negative residuals					
i	j	F_{ij}	$F*_{ij}$	R_{ij}	X_{ij}
USA	Spain	16014	96094	-80080	-4.35
USA	Italy	41382	137052	-95669	-4.14
USA	France	55150	158903	-103753	-4.12
Italy	Japan	11687	62202	-50515	-3.37
France	Japan	17126	72120	-54994	-3.34
USA	Russia	18820	71720	-52900	-3.19
UK	Japan	21241	75864	-54623	-3.18
USA	Turkey	9246	48051	-38805	-2.94
USA	Poland	3201	33831	-30630	-2.90
USA	UK	88627	167152	-78525	-2.89

F_{ij}: observed trade flow (millions of US$); $F*_{ij}$: predicted trade flow (millions of US$); $R_{ij} = F_{ij} - F*_{ij}$: raw residual (millions of US$); and X_{ij}: standardised residual (according to deviance).

The residuals of this model are clearly correlated with distance which is related to the fact that this "natural" parameter was not introduced in the model. The pair of countries with the most significant positive residuals is therefore big and rich neighbouring countries located inside the different cores of the world economy. The countries with significant negative residuals are logically couple of rich countries located in different cores of the world economy. A classification method trying to maximize positive residuals inside group and negative residuals between groups will therefore logically produce a strong regionalization of the world economy around the main economical cores. This solution is very near from the Intramax method that has been discussed in the previous section.

Table 5: Gravity model 2

		Most significant positive residuals			
i	j	F_{ij}	$F*_{ij}$	R_{ij}	X_{ij}
China	USA	274284	89676	184608	11.35
China	Taiwan	76869	18051	58818	7.38
Canada	USA	465180	288137	177043	7.18
Malaysia	Singapore	32561	3088	29473	6.99
S. Korea	Other Gulf	35484	4027	31456	6.88
Japan	Other Gulf	65045	15419	49627	6.75
Singapore	Hong Kong	14912	505	14408	6.11
Chine	S. Korea	105229	40619	64611	6.08
Malaysia	USA	41084	7737	33348	6.04
China	Malaysia	28578	4624	23953	5.39
		Most significant negative residuals			
i	j	F_{ij}	$F*_{ij}$	R_{ij}	X_{ij}
Germany	France	150296	275537	-125242	-6.14
Italy	France	86065	176232	-125242	-5.53
Spain	USA	16014	66342	-90166	-5.36
Switzerland	France	24595	74189	-90166	-4.86
France	USA	55150	111521	-50329	-4.30
Italy	USA	41382	89444	-50329	-4.12
USA	Poland	3201	21276	-49593	-3.53
Denmark	Germany	28961	62896	-49593	-3.47
UK	France	74820	123498	-56371	-3.47
Japan	Italy	11687	35595	-56371	-3.36

F_{ij}: observed trade flow (millions of US$); $F*_{ij}$: predicted trade flow (millions of US$); $R_{ij} = F_{ij} - F*_{ij}$: raw residual (millions of US$); and X_{ij}: standardised residual (according to deviance).

The residuals of this model 2 are clearly less correlated with distance but we observe that some positive residuals remain important between countries located in the same geographical area (especially in Eastern Asia). Long distance trade flows can now be revealed as preferential relations, in particular the cross-pacific linkages (China-USA, Malaysia-USA) and the relation between Eastern Asia and Persian Gulf for oil. We observe negative residuals between many countries of Western Europe, but it is clearly related to the introduction of the variable contiguity in the model. Many negative residuals can be observed between European Union countries and Northern America, despite the fact that geographical proximity has been taken into account.

the contrary, if we control for both size effects and geographical proximity (Model 2), we will obtain cluster of countries that will be generally separated, especially if contiguity effect is controlled. We can also imagine to have built a Model 3 with language effects and colonial links, producing more and more complex groups of states located at long distance.

Despite this general problem (how to produce objective results ?) we can observe that the unconstrained gravity model present some obvious weaknesses related to the fact that it does not take into account (1) the effect of internal trade of countries and (2) the systemic effect of international trade as a whole.

Concerning the first problem, Wei (1996) or Head and Mayer (2002) suggest that the residuals of classical gravity models are biased when they do not take into account intra-national flows. For this author, it is not possible to evaluate the effect of RTA without considering two levels of barriers/preferences: (1) Home bias, which is the preference of firm for the development of relation inside their own country; (2) Regional bias, which is the preference of firm of one country for a given group of country (region) as compared to countries located in other regions. Head and Mayer (2002) suggest that these two effects are not independent and that the evaluation of barriers related to RTA can be overestimated or underestimated according to the level of home bias. They propose a solution which is to introduce intra-national trade flows based on the difference between the GDP of the country minus its exports. By computing an estimation of internal distance inside each country, it is possible to apply the gravity model to a full matrix, including the diagonal (i.e. the flow from a country to itself).

Concerning the second problem, the major difficulty with classical gravity model is related to the fact that only the total sum of flows ($\sum\sum F_{ij} = \sum\sum F{*}_{ij}$) is introduced as systemic constraint through the intercept parameter (k). Accordingly, the some of residual flows is not necessary equal to 0 for a given country in terms of exports ($\sum_i (F_{ij} \neq \sum_i F{*}_{ij} \Leftrightarrow O_i \neq O{*}_i$) as in terms of imports ($\sum_j (F_{ij}) \neq \sum_j F{*}_{ij} \Leftrightarrow D_j \neq D{*}_i$). As a consequence, it is difficult to analyse the residuals as one country eventually obtains systematically positive or negative residuals. This situation is difficult to interpret as the residuals can be related to the number and size of countries located in the neighbourhood and, more generally, to the accessibility of a country to the other countries of the world. The fact that negative residuals are observed between Western Europe countries is for example certainly related to the small size of states in this part of the World.

The uncertainty on the meaning of residuals of gravity model is therefore a real problem for the elaboration of world division.

Double constraint model as equilibrium model

Double constraint model has been very few applied to trade flows analysis despite their methodological and theoretical advantages. Bröcker and Rohweder (1990) has published their famous paper "Barriers to International trade" in the *Annals of Regional Science* rather than *Journal of Development Economics* or *Economic Journal*. On the other hand, geographers that has the most contributed to the theoretical reformulation of gravity models with double constraint like Wilson (1971) or Tobler (1983) has been more interested by application to migration of population, travel to work or sales at intra-national or intra-urban levels. We propose to examine how the development of spatial interaction models achieved by geographers can contribute to provide innovative answers to the unsolved problems revealed by economist concerning the meaning of gravity model applied to trade flows at world scale.

We consider firstly a double constraint model applied to a flow matrix between all countries of the world where intra-national flows are unknown. We assume the existence of a global market with zero-cost of transportation and zero barrier effects. We assume also that intra-national demand is satisfied by specific national firms and that inter-national trade is realized by specific firms that are fully independent from the first one. Each country proposes to the other countries an amount of products for exportation (O_i) and asks for an amount of product for importation (D_j). This is a situation of short term equilibrium that does not necessarily means symmetry of trade between countries. For this reason, we do not analyse bilateral trade blows ($F_{ij} + F_{ji}$) but oriented trade flows (F_{ij}). And we take into account the fact that the sum of exports and imports can be different ($O_i \neq D_i$), which means that some countries can have positive or negative trade balances.

Under this general assumption, what is the most likely distribution of trade flows $F*_{ij}$ between countries of the world in a perfect market without transport costs and without any type of barriers?

If intra-national trade was involved in the model (diagonal not empty), the answer would be very simple and the estimated trade flow would be simply $F*_{ij} = F_i.F_j/F..$ as in the Intramax method presented earlier. But the solution is complicated here by the fact that we have excluded intra-national trade flows and we are therefore obliged to build a more sophistical model that can only be solved by iteration:

Hypothesis: Export and Import of countries are randomly distributed.

$$F_{ij} = a_i O_i b_j D_j$$

with $a_i = \frac{1}{\sum_i b_j D_j}$ and $b_j = \frac{1}{\sum_j a_i O_i}$.

We have adopted here the solution proposed by Wilson (1971) which is based on a multiplicative model. But it is also possible to solve the equation through the additive model ($F_{ij} = A_i + B_j$) proposed by Tobler (Tobler 1983; Dorigo and Tobler 1984). For the model of perfect equilibrium without distance effect, both models produce the same results. But some differences appear between the two families of model when we introduce further parameters (Sen and Smith, 1995). We adopted the multiplicative formulation which is easier to solve with usual statistical tools, in particular through a variant of Poisson Regression Model where the parameters a_i and b_j are introduces as sets of dummy variables ($I_1 \dots I_n$ and $J_1 \dots J_n$) in the equation and can be combined with a scale parameter that measure the size of independent "packages" of trade. In statistical terms, the equation that is solved by maximum likelihood is the following one:

$$F_{ij} = SCALE.exp[a_0 + a_i I_i + \cdots + a_n I_n + b_1 J_1 + \cdots + b_n J_n] + \epsilon_{ij}$$

In practical term a set of $(n-1)$ parameters is sufficient for a_i and b_j because the export and import of the n^{th} country satisfy necessarily the constraint when the conditions are satisfied for the $(n-1)$ other countries. An important consequence is the fact that it should normally not be applied on sample of trade flows but only on complete matrix of trade at world scale. We can eventually apply this model on a subset of countries (e.g. trade between EU countries) but in this case we are obliged to assume that trade inside this subset of countries forms and independent system of exchange. With CHELEM database, we do not have to face this problem as all trade flows has been estimated between all countries or group of countries of the World. Zero flows are not excluded but consider as information on the state of trade relation.

For comparison with classical gravity model 1, we have first applied the double constraint model to bilateral flows in 2004-06. The explanatory power jumps from 73.5% to 80.3% of the total deviance. This is obviously due to the fact that we use better predictor of trade (total of export and import of each country rather than single GDP). The double constraint is less useful for prediction purpose but is much more interesting for the understanding of choices of trade destinations and factors that are underlying this choice (distance cost, barriers. . .). With double constraint model, the margin of the matrix (O_i and D_j) is given and the some of residuals is necessary equal to zero for each line or column. We can therefore better observe what the revealed preferences in terms of exports and imports are (see table 6).

Looking at the 10 most significant positive and negative residuals we can observe significant differences as compared to gravity model 1. For example, the flow between Russia and Belarus appears now is the top ten of the most significant residua, which was not the case when GDP was used as predictor. The most important significant negative residual according to double constraint model is USA-Germany, which was not present in the top-10 of the gravity model.

Table 6: Residuals of double constraint model

		Most significant positive residuals			
i	j	F_{ij}	$F*_{ij}$	R_{ij}	X_{ij}
Canada	USA	465180	93402	371778	15.41
Mexico	USA	279985	55593	224391	11.91
Japan	China	201149	80267	120882	6.19
Netherlands	Belgium	93669	21648	72021	6.04
Germany	Austria	91231	21660	69571	6.00
Malaysia	Singapore	32561	3157	29405	5.00
Belgium	France	96633	31981	64652	4.90
Russia	Belarus	17606	550	17056	4.84
Portugal	Spain	27990	2387	25603	4.83
Spain	France	79427	23579	55848	4.78
		Most significant negative residuals			
i	j	F_{ij}	$F*_{ij}$	R_{ij}	X_{ij}
USA	Germany	123678	261548	-137870	-5.58
USA	France	55150	134191	-79041	-4.40
Japan	Germany	38587	105645	-67058	-4.17
Italy	USA	41382	104614	-63231	-3.98
China	Germany	73422	133514	-60092	-3.19
USA	Belgium	35867	92775	-56908	-3.80
USA	Netherlands	37436	90835	-53399	-3.58
USA	Spain	16014	68403	-52389	-4.27
Canada	Germany	11749	62755	-51005	-4.33
Chine	Belgium	12645	47359	-34715	-3.24

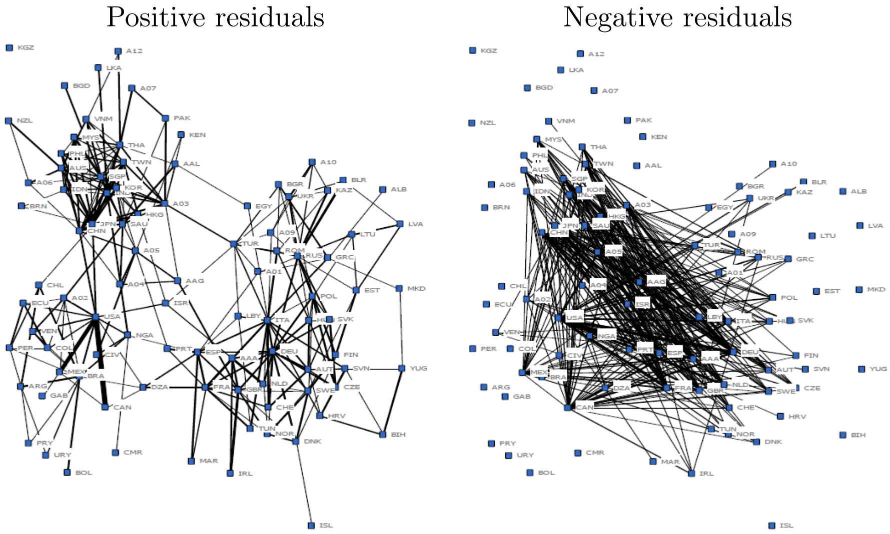
F_{ij}: observed trade flow (millions of US\$); $F*_{ij}$: predicted trade flow (millions of US\$); $R_{ij} = F_{ij} - F*_{ij}$: raw residual (millions of US\$); and X_{ij}: standardised residual (according to deviance).

The advantages of the double constraint model are much more important when we use the oriented version of the gravity model. A typical example is provided by trade relation between China and USA. According to double constraint model, the residual is significantly positive from China to USA (227 instead of 152 billions of US$) and negative but less significant from USA to China (47 instead of 59 billions of US$). This crucial difference is not visible with bilateral trade flows.

Analysis of residuals of double constraint model

The analysis of the graph of positive residuals (Figure 10, left) reveals the existence of preferential relations trade relations between countries. The

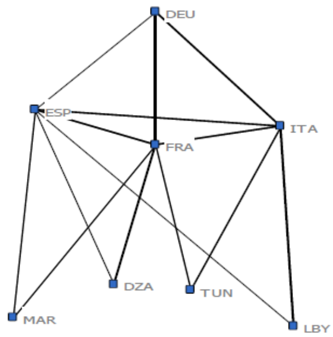
Figure 10: Graph of most significant residuals for Trade 2004-06



explanation of this preferential relation is generally related to transportation costs and it is easy to see geographical clusters on Figure 10 (Eastern Asia, Americas, Europe and Mediterranean). But some preferential links are also related to specific linguistic or historical ties (France with Maghreb; former Soviet Union, Former Yugoslavia) or to geopolitical alliances (USA-Israel). Specialization in specific products and complementaries can also produce some specific linkages or specific linkages established by transnational firms (Nigeria-USA). The analysis of the graph of negative residuals (Figure 10, right) reveals the existence of barriers effect between countries. As we have used a double constraint model, there is an obvious relation between the two type of residuals. If a cluster of country develops strong preferential internal relations (e.g. Western Europe), this will necessary produce as counterpart negative residuals with the countries located outside this cluster. In this perspective, trade relations appear as a zero-sum game between countries of the world.

The clusters of countries that can be derived from this analysis are not exactly natural regions but rather residual regions. Their definition involves all factors that can contribute to increase or limit the level of trade between countries (distance, language, regional agreement, embargo. . .). The objective of residual regions is not to distinguish the role of each factor but to analyse the global result.

Figure 11: The interest of correlation of residuals for regionalization



Construction of correlation matrix between residuals

Different methods can be applied to the matrix of residual for the elaboration of a classification of countries. One obvious solution is to try to maximize the amount of positive residual inside the clusters and the amount of negative residuals between the clusters. This is the Intramax solution discussed in the previous section. But it is possible to develop another approach that takes better into account the structure of direct and indirect relation between countries.

In the example presented on figure 11, we have selected 8 countries of northern and southern shore of Mediterranean Sea. Northern countries are all connected by positive residual and form clearly a clique. But the situation is different for southern shore countries because they are not connected each other by direct relation, but all of them has preferential relation with countries of the northern shore. We find here a typical situation of structural equivalence that can not be directly capture by Intramax procedure but that can be easily revealed if we examine the correlation of export or import of the countries. All countries from Maghreb are clearly correlated in terms of trade, even if they do not trade directly. In this sense, they belong to the same region.

We propose therefore to use the correlation between lines or arrows of the matrix of of residuals in order to define residual regions that are characterized by the proximity of their directions of exports or import. The figure 12 indicates that a clear structure of world is revealed by the correlation matrix, defining three main clusters of countries with high level of correlation that is obviously related to the influence area of the three main economic poles of the World, the so-called Triad: Northern America, Eastern Asia, and Western Europe. The countries that are attracted by the same pole of the Triad are logically correlated. But the graphic reveals also the existence of countries that are more exceptional and that are not really correlated with the other. This is for example the case of Israel which presents exceptional positive residual with USA but also with some European countries like Belgium. It is also the case of many African countries or oil produces that are polarized at the same time by different economic poles.

Elaboration of clusters through Hierarchical Cluster Analysis and cartography of results

The last step of the analysis is the derivation of clusters from the correlation matrix of residuals. We suggest using a method of complete linkage that avoids the creation of cluster defined by exceptional linkage. The comparison of results for different matrix can be obtained either by the definition of number of clusters or by the choice of the same level of correlation as end of the clustering procedure. For matrix of different size, it is also possible to use the significance of the correlation parameter. Finally, it is also possible to examine the discontinuities of the tree of hierarchical classification in order to cut at a relevant level, but this solution is mainly useful for the examination of an isolated matrix as it does not provide guarantee of objective comparison.

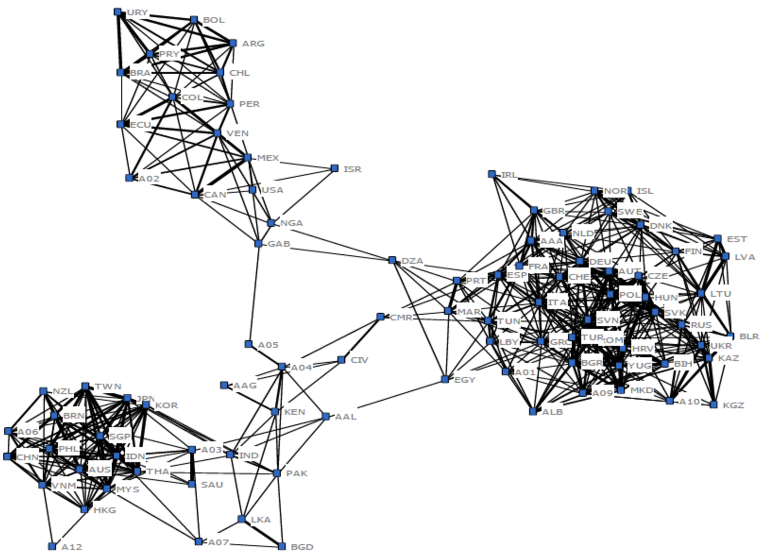
In our example of trade flows 2004-06, the tree reveals a clear division of countries in 2 main clusters that can be further divided in 4 (if we choose the threshold of R=0) or 7 (if we prefer to build group of countries with significant level of internal correlation). Once we have decided of the best level, we can propose cartography of the resulting regions where the color of clusters reflects their proximity in the hierarchical tree. In our example, they should be a main opposition between two groups (dark and cold colours) but also between 4 main clusters (blue, green, orange, and red) and finally variations in order to identify the last internal variations of the 7 sub-clusters (light/medium blue; green, light/medium orange; light/medium red). Last but not least, the name of regions is given in an abstract form (A, B, C, D) reflecting the hierarchy of the classification tree (A.1, A2...) but without giving immediately geographical names ("Europe", "America", etc.). The EuroBroadMap project tries indeed to produce world regions without a priori on the results. And it is therefore only in the interpretation of results that clusters can be compared to existing division.

3.3 Application: trade data 1994-96 and 2004-06

We have applied the residual method defined in previous section to both period 1994-96 and 2004-06, but with distinction of correlation of exports and import flows. We produce therefore 4 divisions of the world in residual regions. Each partition has been realized in 6 clusters because it was in every

Figure 12: Correlation of residual (exports and import) trade flows 2004-06

Graph for R > +0.4



Graph for R > +0.6

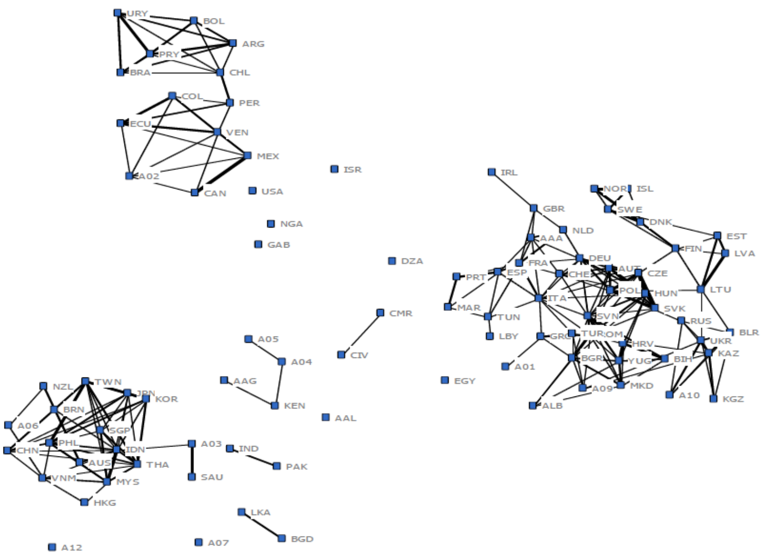
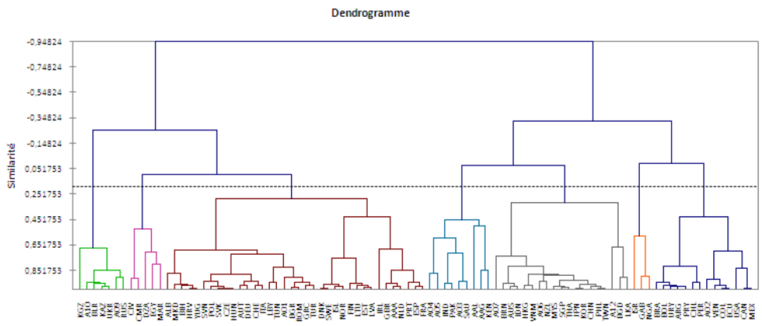


Figure 13: Hierarchical cluster derived from correlation of residual trade flows 2004-06



case an obvious discontinuity related to the threshold R=0.00.

At both period of time, the world is characterized by a strong division in two clusters of country: one related to countries of EU and its eastern and southern peripheries (A+B); another one related to the rest of the world (C+D). At this level of division in two clusters, there is a relative stability of the situation and it is only in Africa and Middle East that some reorganizations can be pointed.

At a more detailed level (4 to 6 regions), some clusters appears relatively stable like Americas, former USSR, Eastern and Southern Asia. In the specific case of the so-called “Europe”, we can observe that the West-East Division that was visible in 1994-96 is removed in 2004-06. We can notice that this area of integration (cluster B.1) includes both shores of Mediterranean Sea and is clearly determined by polarization of export toward EU.

In terms of importation, we obtain very similar results. The most obvious one is the fact that the world appears one more time divided in two main groups, one centred on EU and its eastern and southern peripheries. But the structure of the regions is a bit different, especially in the case of Africa. The cluster B.2 (countries with few internal trade but strong polarization toward EU) covered the majority of Africa in 1994-96. It is reduced to northern Africa and some isolated countries in 2004-06.

The East/West division inside EU is not removed between 1994-96 and 2004-06, as it was the case for exports. Moreover, some countries that was included is the “western cluster” (B.1) are joining the “eastern cluster” (A), in particular Finland and Turkey. The exports of energy and raw material from Russia and other republics from former Soviet Union is probably responsible from this modification. New member states (Poland, Hungary, Romania. . .) has been able to modify their exportations of manufacturing products, but not their importations of energy.

Figure 14: World trade regions according to the residual method in 2004-06

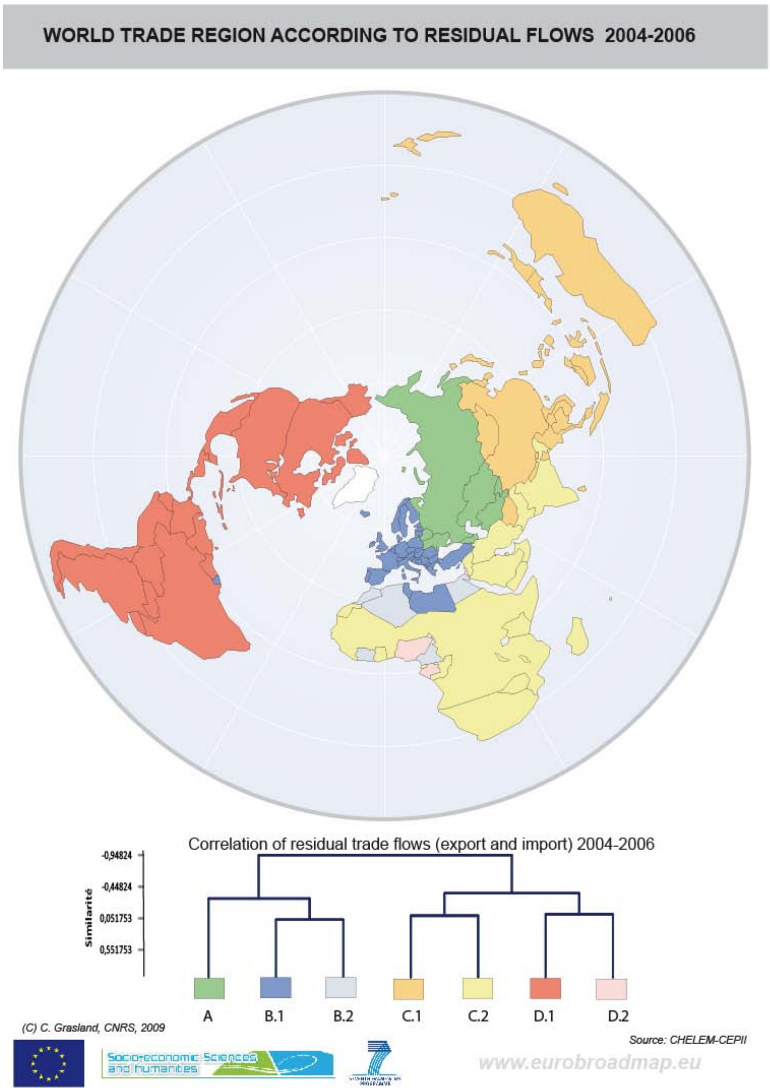


Figure 15: Division of the world according to export flows in 1994-96 and 2004-06

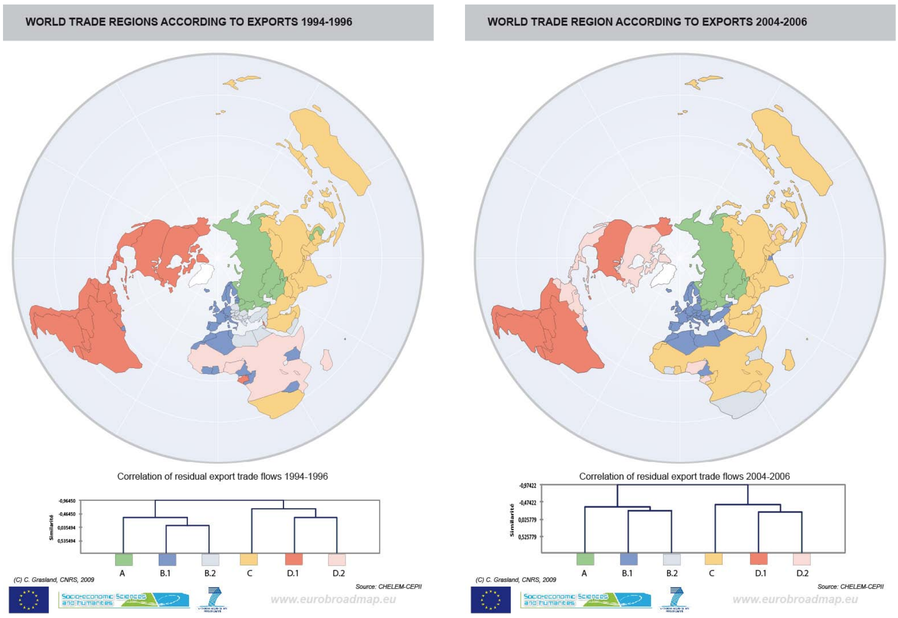
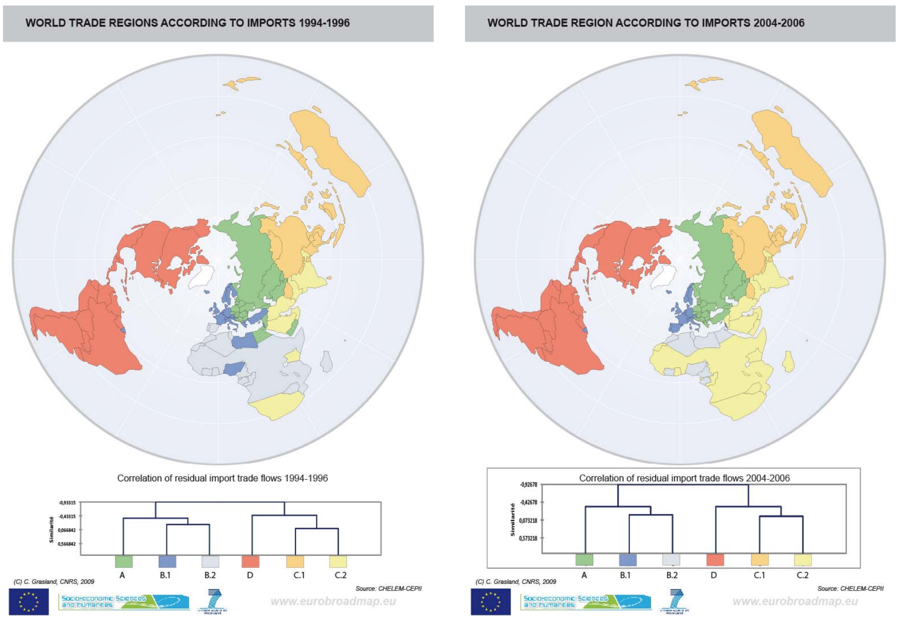


Figure 16: Division of the world according to import flows in 1994-96 and 2004-06



3.4 Interest and limits for EuroBroadMap project

We focused our analysis on the theoretical foundation of so-called natural regions and try to demonstrate that it was better to speak from residual regions, according to different assumptions. We have less develop the empirical results and not examined fully the consequences of the method of partition that is derived from double constraint model. We will therefore have to complete this chapter by further in depth analysis in many directions:

- Comparison of residual method with Intramax. Apparently, both methods are similar (as they are based on estimation of flows by margin) but they are some important differences, in particular related to the fact that the cluster of countries are based on concentration of residuals with Intramax method, as they are only based on correlation with Residual method.

- Interpretation of cluster of countries obtained with residual method is therefore less obvious as some countries are in fact related by relation with countries located in another cluster. It is therefore necessary to add a graph of intensity of flows between clusters for a better understanding of results.

- Distance can be introduced in the gravity model as an unknown parameter to be estimated. We can indeed derive from the residuals of the double constraint model a measure of functional proximity (T_{ij}) which is not necessarily equal to geographical distance (D_{ij}). The clusters of country that are obtained by the residual method are in fact derived from this functional proximity. One alternative method could be therefore to extract this functional distance from an inverse gravity model and to build regions based on this functional proximity with statistical methods like Multi Dimensional Scaling.

As a provisional conclusion, the gravity model and the related methods defines a large field of investigation for EuroBroadMap but more in depth analysis is needed before to decide on the best theoretical and methodological solutions to be applied.

4 Centre-periphery divisions based on world system theory

Section written by Geoffrey Pion and Gilles Van Hamme

4.1 Concepts and bibliography

The center/periphery paradigm was mainly developed by two historians, Braudel and Wallerstein. The central concept of Braudel (1985) is the notion of "economy-world": a large and coherent economic unit dominated by one and only one core. For example, in the 15th century, Europe is an economy-world as well as China, the Arabic-Muslim world and India. As for Wallerstein (1985), he studied the way Europe progressively made the whole world a unique economy-world for its own interests.

Along the same lines, Amin (1976) built a history of worldwide accumulation in order to study the unequal exchange between a developed center and underdeveloped peripheries. This unequal exchange is the consequence of the connection between developed capitalist economies and peripheral social formations which do not constitute coherent capitalist economies. The centre dominates this relation and imposes successive specializations to peripheral economies benefiting of the possibility to maintain low salaries, notably because of large reserve of labour force and the possibility to get cheap products from precapitalist segments of the economy. This process is the base of the unequal exchange in terms of working hours between centre and peripheries (see also Emmanuel 1969).

The domination of the centre toward peripheries has taken different aspects over time: direct political domination, unequal trade, capital exports, world institutions dominated by the centre (Vandermotten 2004). Of course, the spatial configuration of the centre(s) and peripheries has also evolved. For example, USA became the major centre of the world economy during the 20th century.

As far as trade is concerned, the domination of the centre toward peripheries includes different dimensions:

- The trade specialization of the centre and peripheries are different and the latter tend to be specialized in primary products or low technological segments of manufacturing industries;
- Some authors have insisted on the deterioration in the terms of trade for primary products and for peripheral countries (Singer 1950 and 1984; Sarkar 2001). The price of products sold by peripheral countries tend to decrease in comparison to those sold by the centre;
- The centre(s) constitute integrated economies, whose trade is mainly with the centre, while peripheries mainly trade with the centre. This

is related to the duality of the peripheral economies, with a modern segment which is integrated in the world economy and precapitalist segments which are not (Vandermotten 2004);

- Trade is based on unequal exchange in terms of working hours (Emmanuel 1969; Amin 1976).

The methodology we will use to regionalize the world according to trade is in accordance with this centre/periphery approach. First, we will identify centre(s) and peripheries through the type of products they sell to the rest of the world, that is to say the way they insert in the world economy. Second, since the EuroBroadMap project is about the intensity of relations between territories of the world, we divide the world according to preferential trade links between periphery(ies) and the centre(s).

Of course, it will give only a partial vision of the centre/periphery division of the world since it does not take into account other major aspects such as capital exports or political power relations.

4.2 Selected tools for world division

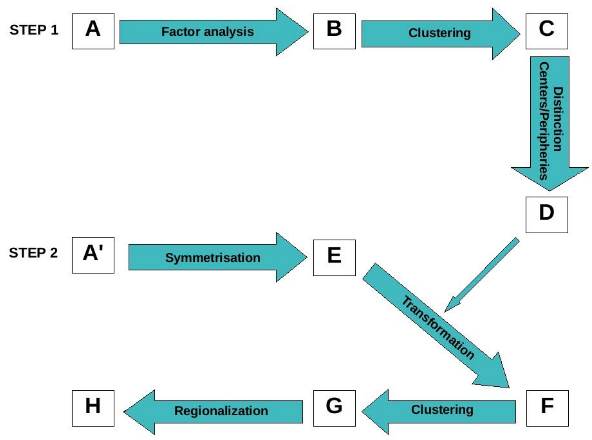
In order to divide the world according to the core/periphery theory, we need to take into consideration the nature of the trade (which countries are selling which types of products) and the direction of the trade. Indeed, the centre(s) of the world is characterized by the high share of high technological manufacturing products in its exports as well as trade flows which mainly go to other parts of the centre. On the other hand, peripheries should be characterized by primary products or labour intensive manufacturing products and flows mainly directed to the centre.

In order to regionalize the World according to a World system or center/periphery approach, we opted for a two-step analysis (Figure 17).

In the first step, we use a country/merchandises matrix (A). It enables us to define the countries belonging to the centers according to the type of products they are selling. To achieve this classification, we first ran a factor analysis on the 70 products for each country or block (92 blocks) (B). It gives a synthetic view of the export structure of the countries. Using this synthetic picture - the components of the factor analysis - we then run a cluster analysis with the Ward's method (C) which gives a synthetic map of regionalization of World trade according to export structure. This step enables us to identify the core countries (D).

In the second step, we classify the world according the direction of their trade flows. To achieve this objective we use an origin/destination matrix of trade flows (A'). First, we used it to see whether it is relevant to divide the core countries into different centre(s). Second, on the base of the centre(s) identified, we transform the initial matrix in a country/centre(s) origin-destination matrix (F). This matrix enables us to determine the sphere of

Figure 17: Scheme of the regionalization of the world according to core/periphery theory



influence of the different parts of the centre. By a cluster analysis, we group together the different countries according to the direction of their trade flows (G). It gives a regionalization of the world according to the direction of the flows toward the core countries (H).

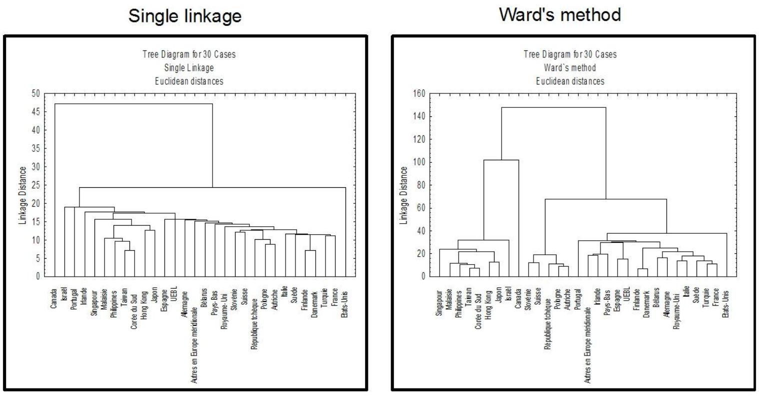
We will now justify the use of a factor analysis like PCA or cluster analysis like HAC with Ward's method in our approach.

Principal component analysis (PCA), which was formulated for the first time by K. Pearson in 1901, is a mathematical algorithm that transforms several correlated variables into a smaller number of factors called principal components. The first principal component accounts for as much of the variability in the data as possible, and each following component accounts for as much of the remaining variability as possible. At each step, the PCA tries to maximize the sum of correlation square with the original variables. The process ends when all the factors have been extracted (Beguin 1979).

A PCA has been used in order to synthesize a lot of variables into few factors. Basically, the aim was to divide several dozens of merchandises into a small number of different sectors structuring the World economy. Thus, it allows defining group of products without any *a priori*. The use of the first components rather than the 70 merchandises to run the cluster analysis gives much more coherent results since it has allowed to group similar products together.

Cluster analysis with Ward's method (HCT) is a hierarchical ascendant classification designed to optimize the minimum variance within clusters (Ward 1963). The purpose is to minimize the distance between individual members and group's centroid. Individuals or groups are merging when

Figure 18: Classification trees



they minimize the variance growth of distances within the group (Beguin 1979).

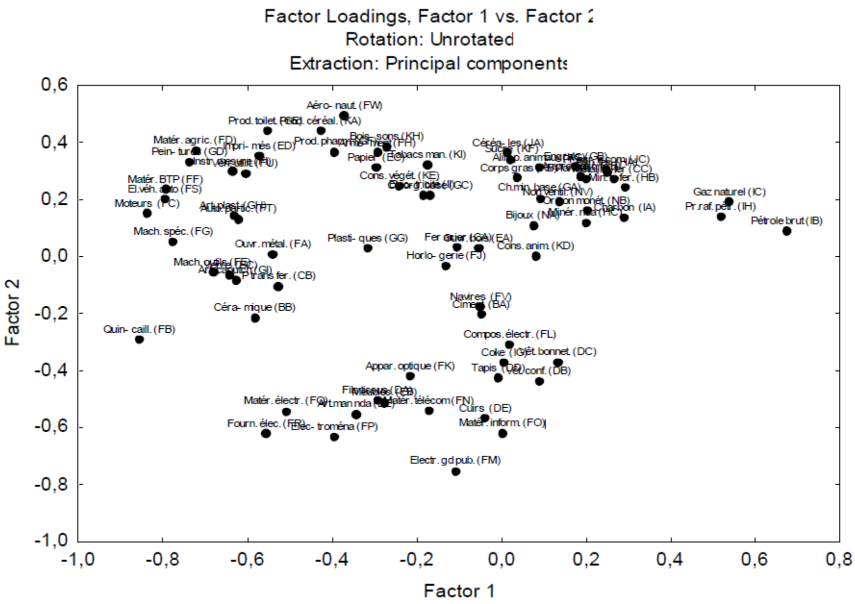
The major advantage of the Ward's method, in comparison to others hierarchical ascendant classification like single, complete or average linkage, is clearly to minimize intraclass variance. As a result, strong differences can be observed between a primary classification tree and a classification tree with Ward's method as illustrated on figure 18.

4.3 Application: Trade data 1994-96 and 2004-06

The first step of our method consists in regionalizing the World according to the export structure. In order to achieve this first objective, we will begin by a PCA to synthesize the structure of merchandises exports. Then, we will make a cluster analysis in order to classify the blocks according to the first components of the PCA analysis.

The PCA with the matrix A has enabled us to extract 35.8 per cent of the total variance for the period 1994-96 (37.4 per cent in 2004-06). Only the three first factors were kept because of the gap with the fourth factor in terms of the variance which is accounted for. The first factor explains 17.5% (17.2 per cent in 2004-06) of the inertia and opposes the raw materials (especially petroleum and gas) on the positive side and industrial goods like machine tools, motors or car manufacturing on the other side. The

Figure 19: Scheme of Factor Loadings - 2004-06 - Factor 1 (17.2%) *vs* Factor 2 (11%) - Export

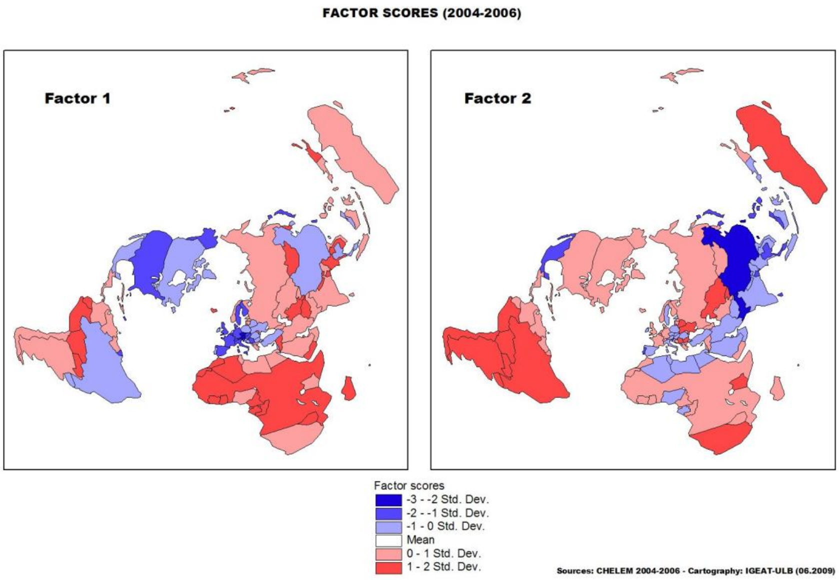


second factor, which explains 9.8% of the total variance (11% in 2004-06), opposes light industry (textile, electronics) below in the scheme (Figure 19) to miscellaneous goods (chemistry) and some raw material (food especially) at the top. The third factor explains between 8.5% (1994-96) and 9.5% (2004-06) of the total inertia. It groups together some high technological merchandise such as chemical products.

On the figure 20, we have mapped the first and second factors. The first factor put in evidence the opposition between European and North American countries in blue which mainly export industrial goods to African, Latin American and Asian (except those on the Pacific Coast) countries in red which specifically export raw materials. The second factor puts into the fore the assembling countries (Southern Asia, Mexico, Maghreb and to a lesser extent Mediterranean Europe).

On the base of the scores of each country on the three first components of the analysis, we ran a cluster analysis. We can note that the hierarchical classification tree is basically dividing between two parts: countries which are mainly raw materials producers on the left side and those which are mainly industrial goods producers on the right side. We have decided to stop the aggregation process at a linkage distance of 5. So we have clustered the World between ten regions; four raw materials regions and six industrial regions.

Figure 20: Cartography of factor scores 1&2 - 2004-06 - Export



The classification allows us to create a map of world regions according of export structure for the two periods. In 1994-96, the cluster analysis clearly shows the opposition between raw material producers (in blue and grey tones) and industrial good producers (in yellow and red tones). A red core in Northern America and Western Europe can be identified as well as a dark orange core with Mexico and Eastern Asia's countries (Figure 21). The first one gathers countries which produce especially high technology goods (chemistry, machine tools) whereas the second one groups together countries with diversified manufacturing industries, including labour intensive industries. On the other side, there are several types of countries which constitute the peripheries. Whereas some of them are hydrocarbons and ores producers (Middle East, Africa, Southern America), others are specialized in light industry with high intensity of labour force (Southern Asia, Eastern Europe).

In the 2004-06 period, a strong contrast still exists between high technology good and hydrocarbons producers (Figure 22). Despite the fact that the red core and the dark orange core are approximately the same than in 1994-96, countries like Turkey or Poland have been integrated into it (in the same time, Mexico and Thailand are no more in the dark orange class). Two main evolutions in the peripheries should be noticed. Firstly, the dark blue class has gathered much more countries than 10 years before. It has to be linked with the rise of petroleum prices which make that much more countries belong to the Petroleum and gas group. Secondly, most of Southern and Eastern Asia are now in orange tones which mean that they reinforce their assembling functions in the world division of labour.

The second step aims at regionalizing the World according to trade flows: firstly, between countries of the cores and secondly with periphery/core countries flows. An HAC with Ward's method on country/country trade flows between core countries enables us to divide the countries of center between three classes: one in Northern America in red (with or without Mexico), an other one in Europe in blue (including some Central European countries and even countries like Belarus or Turkey) and a final one in green in the Asian Pacific Coast (which includes Israel because of the similarity of external trade between Israel and Japan toward USA).

Now that intra-cores regionalization is achieved, we can study trade flows between all the peripheral countries and the three cores. We should mention that we added a fourth class, rest of the World.

We used again an HAC with Ward's method to classify peripheral countries according to the direction of their exports. According to the classification tree, we've divided the World between six regions. Three of them are cores with their areas of influence: Europe in dark blue, Northern America in red and Eastern Asia in green. The Eastern Asia group gathers most of the Eastern Asia countries as well as Middle East countries which provide hydrocarbons to Japan and NIP. European region includes European countries of the core, Central Europe, Mediterranean countries and some African countries. As for North America, it gathers ALENA's countries, Caribbean and Latin American countries as well as African hydrocarbons producers like Nigeria or Gabon. An attenuated blue class groups some countries which export toward Europe as well as countries of the rest of the World (notably Russia). The yellow class gathers countries which mostly make their exportations to Brazil (Argentina, Uruguay, Bolivia, and Paraguay) or Russia (Ukraine, Kyrgyzstan, Other CEI countries). Finally, the grey region gathers countries which export towards one of the three cores and the rest of the World; it's a by default class. There is not any major switching between the two periods. However, we can notice the change of PMA in Africa which pass through from American area of influence to the grey class. Furthermore, Balkanic and ex-USSR countries are in the European sphere of influence in 2004-06 whereas they were oriented toward Russia ten years before.

Some unexpected result should be pointed out. One of the most significant example is the block PMA in Africa which makes part as the Northern American group despite the fact that Europe is the main destination of their exports. It could be explained by the fact that PMA in Africa and several other countries are gathered together early in the tree because of their trade orientation's similarities. Each of the four blocks (PMA in Africa, Bangladesh, Nigeria and Sri Lanka) make between a third and a half of their exports toward Europe and Northern America.

The results we obtain seem to be promising and consistent with the

Figure 21: Regionalization of World trade based on exportation's structure (Step D) - 1994-96 - All products - Export

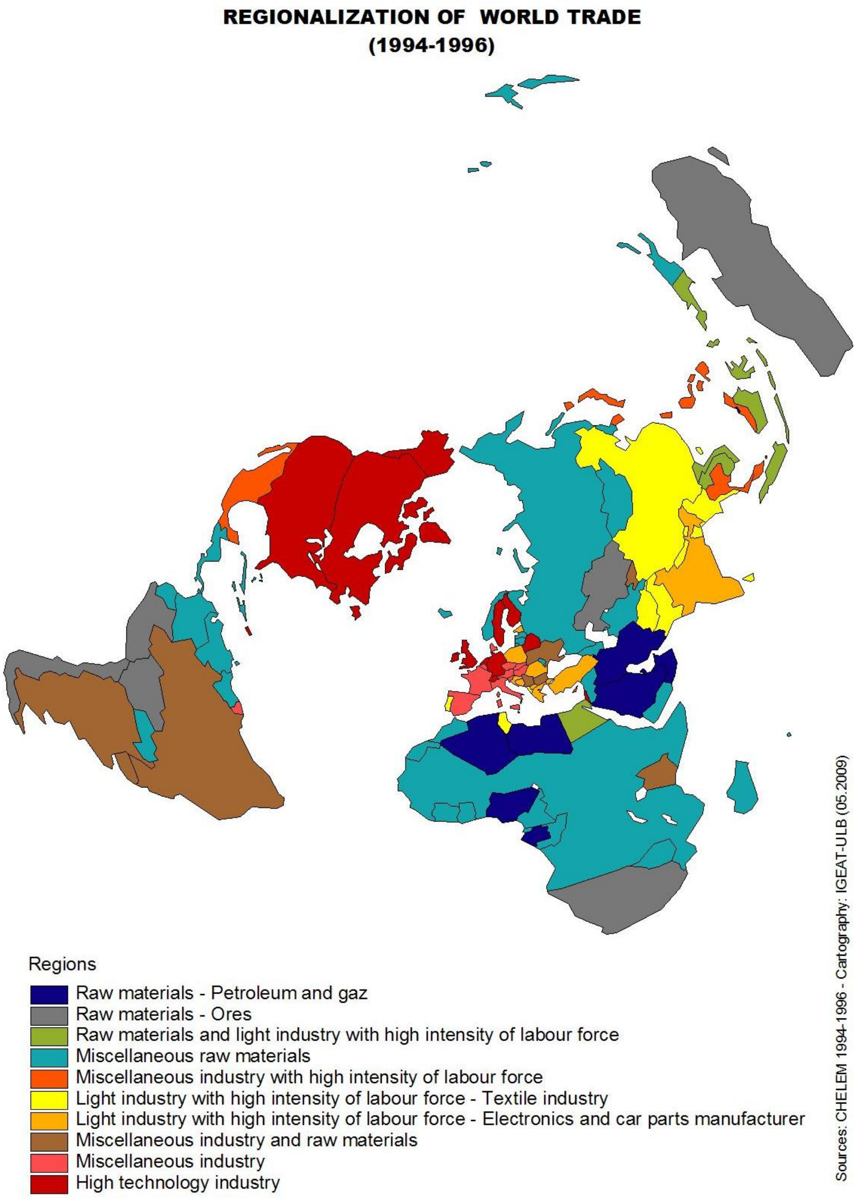


Figure 22: Regionalization of World trade based on exportation's structure (Step D) - 2004-06 - All products - Export

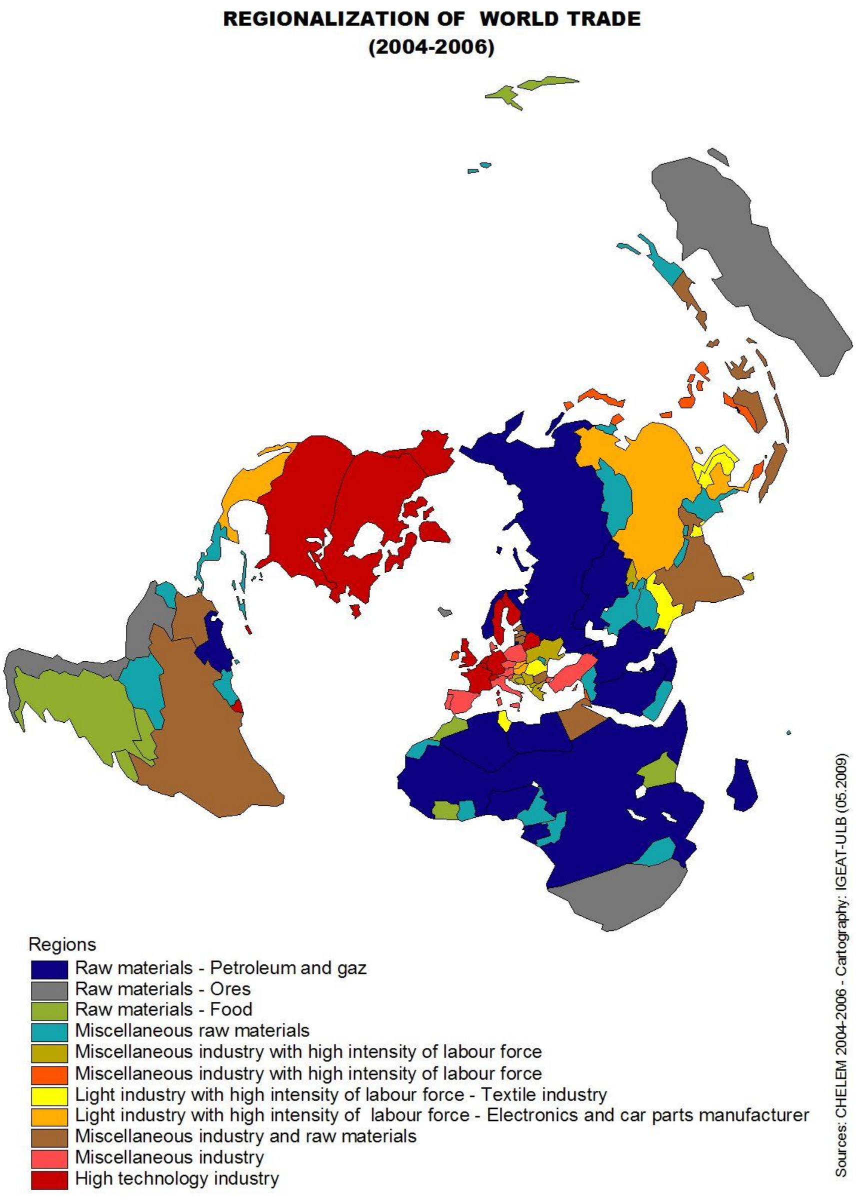
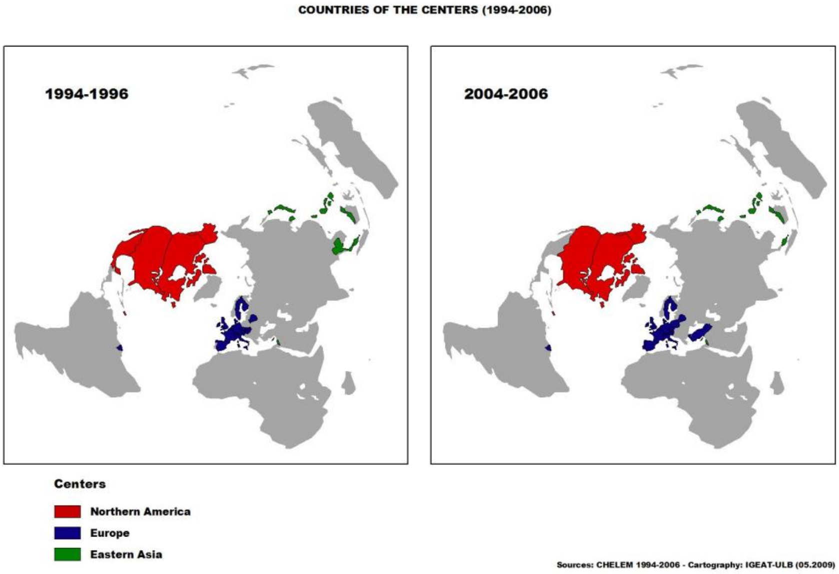


Figure 23: Clustering of core's countries (Steps E and F) - 1994-96 - All products - Export

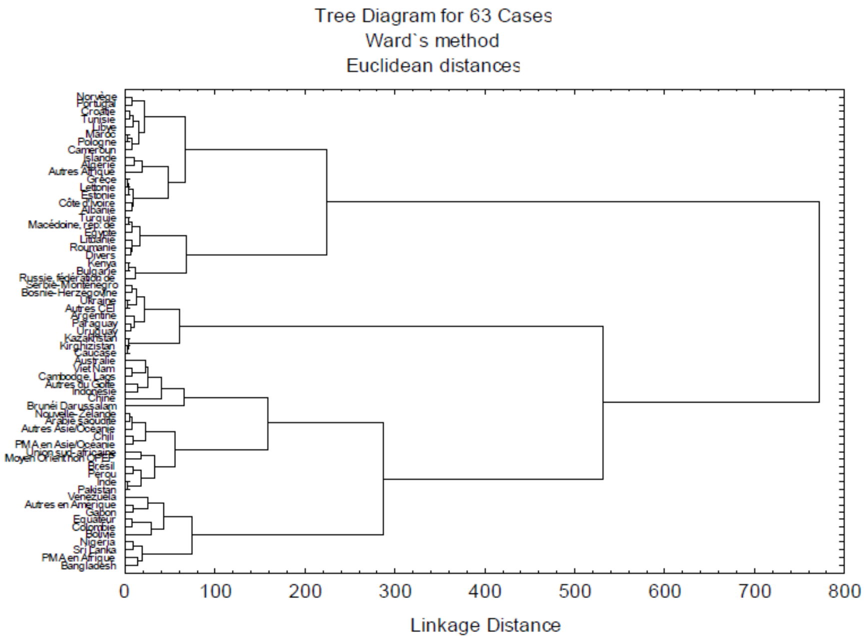


existing literature.

Scientific literature provides a lot of studies which try to divide the world between cores and several types of peripheries. The geographers have developed since a long time some methods of regionalization in a center/periphery approach. French planners developed the MIRABEL algorithm with tries to determinate employment cores and their area of influence (Techer 1994). Such a method would allow dividing the world in a core periphery approach only according to the direction of the flows without taking into account the export structure. Unfortunately this method is not reliable regarding our purpose because of the necessity to calculate a degree of autonomy. Considering that we are working on a national level without taking into account the intra-national relations, we could not calculate it.

Reynaud in his study, *Société, espace et justice* explains the concept of socio-spatial justice and analysed the relation between cores and peripherals areas (Reynaud 1981). He tried to discriminate several types of cores and peripheries. At the end of his book, Reynaud had a go at regionalizing the World according to his concept of socio-spatial justice (Figure 26). But this regionalization is not based on a clear statistical methodology. Reynaud used some variables like the population, the capital loan, the direct investment, the exports, the imports but does not explain clearly the building of his typology.

Figure 24: Example of HCT (Ward's method) in order to regionalize the peripheries considering their relations toward the cores (Step F) - 1994-96 - All products - Export



As for Snyder and Kick, they tried to regionalize the World on the base of four economic and political criterion (trade flows, military interventions, diplomats flows and bilateral treaties) thanks to a block-model analysis (Snyder and Kick, 1979). Even if their method is a little bit obscure, their regionalization of the World in the sixties distinguish a core (blocks C and C') and several types of peripheries or semi-peripheries (Figure 27).

Their method could be useful to regionalize the World according different criteria since it allows taking into account other aspects of core/periphery relations than trade. But their method imposes to work with binary matrix, which would oblige to impoverish our initial database. When working only on one dimension, it is better to keep as much as possible the richness of the initial information. But of course, working only on trade is only dealing with one aspect of core/periphery imbalances.

4.4 Interest and limits for EuroBroadMap project

One of the objectives of Eurobroadmap is to confront subjective perception of the world, and specifically of Europe, with objective measures of flows.

In this perspective the center/periphery approach is particularly relevant.

Figure 25: Regionalization of World trade based on exportation's destination (Step H) - 1994-96 & 2004-06 - All products - Export

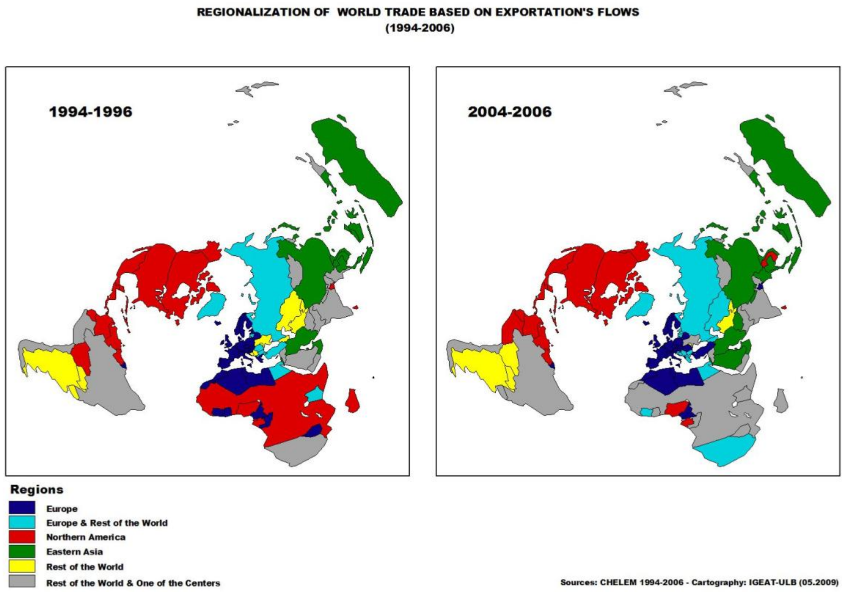


Figure 26: Regionalization of the World according to Reynaud (1981)

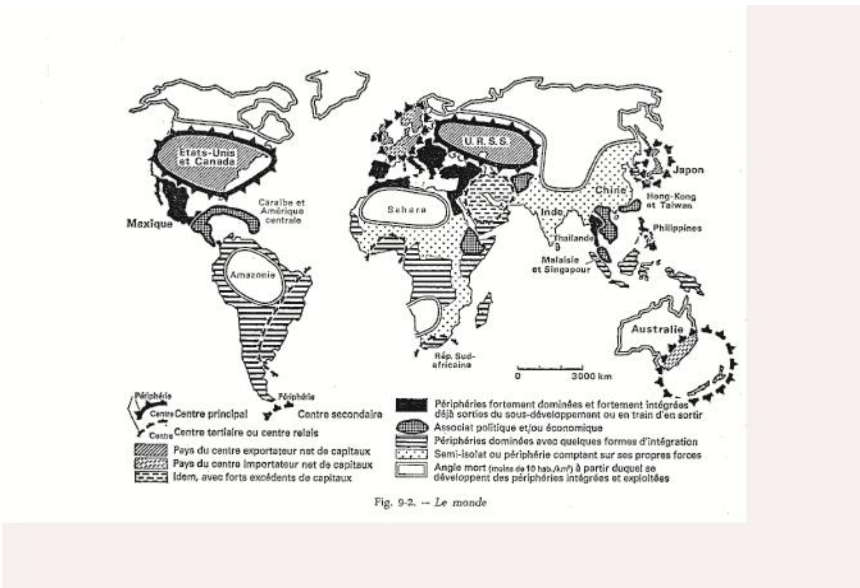
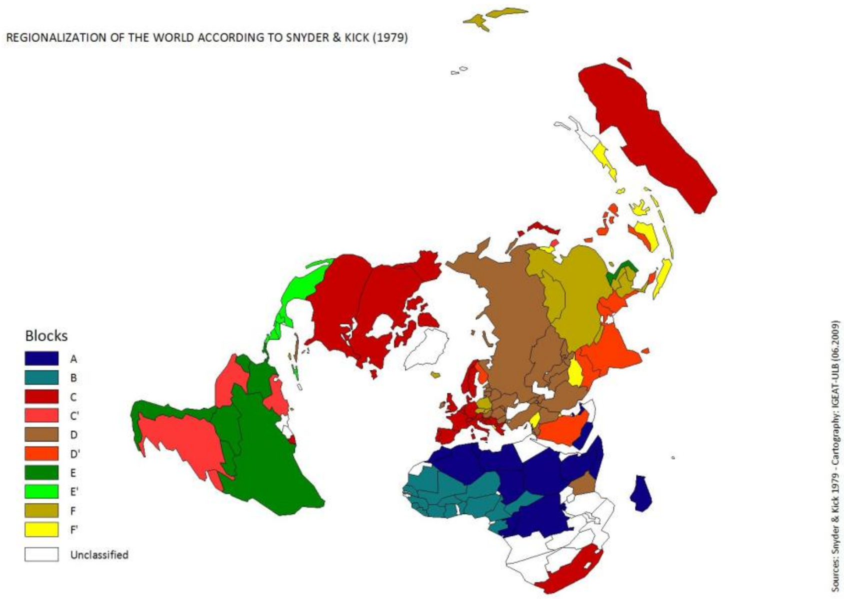


Figure 27: Regionalization of the World according to Snyder and Kick (1979)



This approach supposes unequal relation between center(s) and periphery, notably between Europe and its peripheries. In this context we proposed divisions of the world which take into account these imbalances in the world trade flows. It is of particular interest to cross these world classifications with subjective perceptions of Europe.

This could be done by answering the two following questions:

- Is Europe, and the other centres, more important (better known, more attractive, more repulsive…) for residents of its influence area than for the rest of the periphery (or the other center(s))?
- Which relation can be observed between the subjective perception of Europe and the nature and intensity of the flows between the different parts of the world and Europe? More precisely, does the domination of Europe on its peripheries make it more attractive (as a land of "opportunities") or more repulsive (as a land of "exploiter")? In this perspective, it would also be very important to take into account migratory flows.

Two kinds of empirical limits can be raised. Firstly, imperfect geographical breakdown could strongly influence some results. In addition, intra-blocks trade could be very important and introduce some biases into our analyses. Secondly, insufficient merchandise breakdown obliges to work with heterogeneous sectors. For example, Japan and Philippines were always in the same class because of their specialization in electronic goods but it hides the fact that the segments of production in which they are specialized are dissimilar.

So, improvements can be achieved by using more disaggregated data on both geographical and products fields. However, it would probably not solve all the problems. To refine our approach, it would be necessary to take into consideration at the same time the direction of the flows and the nature of it rather than using the two steps analysis we opted for here. To take again the Japan/Philippines example, to take into consideration the merchandises structure of their bilateral trade would undoubtedly emphasizes the technological superiority of the Japanese economy.

Statistical analyses should also consider imports rather than only focusing on exports. Finally, a more relevant centre/periphery perspective would integrate different types of flows and not only focus on trade. It would oblige to use other methods than the one which is well adapted to trade data.

We can summarize the future improvements as follow:

- Taking into consideration imports for example by working with trade balance by products rather than exports;
- Taking into consideration more refined range of products;
- Taking into consideration direction and nature of flows together rather than successively;
- Crossing trade data analyses with other types of data to better fit the centre/periphery approach.

It is worthy noticing that PCA has some limits, especially the fact that some spatially coherent variables (such as gas and petroleum) - because of their similar geography - are well taken into account by PCA and to a certain extent determines the way other variables are taken into consideration. For example, the first component isolates gas and petroleum as well as other extraction products on the first component and opposes it to a diversified range (but not homogeneous) of manufactured products which you do not find in these primary producing countries.

As far as hierarchical cluster analyses are concerned, they raise problems when trying to assess evolutions. By definition, cluster analyses provide discrete classifications. Changes in the classification do not provide a good picture of structural evolutions because they might be due to threshold effects. It means that adapted methods should be used to evaluate evolutions.

5 Graph theory and network analysis

Section written by Laurent Beauguitte

The aim of this part is to present a couple of methods used on social network analysis, to test it on two matrices (world trade 1994-96 and 2004-06) and to examine if they can be useful to achieve the aim of the Work Package Flows and networks which is to find "objective" divisions of the world.

5.1 Concepts and bibliography

The social network approach is a method used mainly on sociology which is based on mathematical theory of graph. The main idea is to focus on relations between actors; an actor can be a person, a group, a state, etc. In a same way, nature of the relation can be various from friendship relations until world trade relations.

One as the main obstacle to its use is the heterogeneity of the vocabulary used by different authors from different fields[3]. More important, some common tools of social network analysis can hardly be applied to anything else than social relations between individuals (i.e. betweenness).

Anyway, its application to world trade is not new and many articles used it to create world divisions (Snyder and Kick, 1979, Smith and White 1992, Kim and Shin 2002). If most of those articles use binary matrices, some recent articles show that some partitioning methods can be used on signed (Doreian 2008) or on valued networks (Nordlun 2007).

Centrality indicators can be very useful to describe relative positions of actors in a graph. If many indicators exist, we'll use only a couple of them because they can be applied on relations between states, even if they have been created to study relations between persons (Freeman 1977, Bonacich 2007).

One of the most relevant concepts of social network analysis for the project EuroBroadMap are those of subgroups and of equivalence. A subgroup is a set of actors (here states) which share most properties between them than with the rest of the network (White *et al.* 1976). Studying the equivalence of actors means that we look for the relational properties; it helps creating sets of actors who have the same type of relations with the same type of actors (Wasserman and Faust 1994).

All analysis was performed on Ucinet (Borgatti *et al.* 2005). The free book from Hanneman and Ridddle (2005) is both a general introduction to social network analysis and a tutorial for Ucinet.

[3]Sociologists look for cliques, physicists for communities and computer scientists look for clusters. But these three words design the same object.

5.2 Selected tools for world division

This presentation is made of three parts. In the first one, basic indicators used to describe graph are presented. In a second part, I'll explain how to create world divisions with directed graphs (i.e.: the matrices are asymmetrical and F_{ij} may be different from F_{ji}). I'll end with methods suitable for undirected graphs (symmetrical matrices and $F_{ij} = F_{ji}$).

Even if they cannot provide a world division, some indicators allow presenting the general structure of the graph and the main properties of nodes involved.

Starting from the most simple, the order of a network is given by the number of nodes and its size is the number of links. The density is equal to the number of ties present divided by the number of possible ties. The degree of a node is the number of lines incident with a node. The geodesic distance between two nodes is the shortest distance, generally expressed in number of links, between two nodes.

The main way to divide a directed matrix is to look for equivalence of nodes. Groups created will join nodes that share common relational properties. By definition, it's only suitable for directed graphs. The CONCOR method was one of the first used, it detects structural equivalence. The principal is quite simple: the input is a binary matrix and iterated correlations will create division in subgroups. Each iteration divides the matrix in two sets. The advantage is that you get a hierarchical division of the world. But there are two problems:

- It always splits a group in 2 subgroups, even if a division in 3 subgroups would be more relevant
- Even if results are often readable, mathematical properties of this algorithm are "not well understood" and "remain obscure" (Wasserman and Faust 1994, p.380).

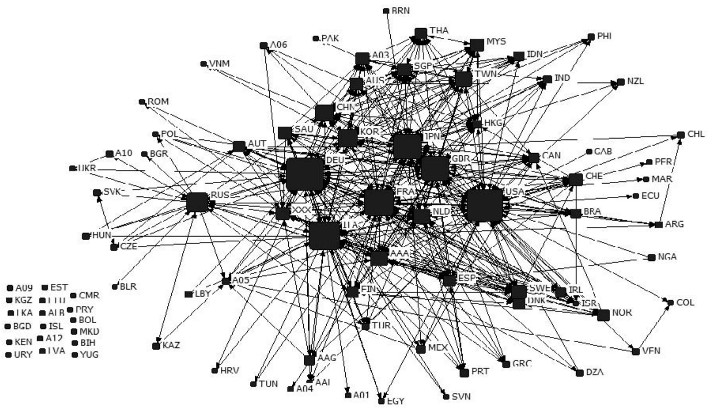
The REGE algorithm is more recent. Instead of looking for structural, it looks for regular equivalence. Theoretically, it should perfectly fit with our purpose.

If we deal with undirected graphs, we won't be able to divide completely the world. But we'll easily be able to reveal the centre of the system and its evolution. Looking for cliques[4] will reveal the most connected part of the network. It can be useful to complete it with an examination of k-plex and k-cores[5] because the definition of a clique is really strict and we could miss possible highly connected subsets in the matrix.

[4]A clique is a maximal complete subgraph of at least three nodes. In other words, it is the biggest group of nodes where all possible links are present.

[5]These wo notions relax the clique definition, allowing a given amount of links to be absent.

Figure 28: Export flows in 1995. Each link represents a value greater or equal to 0.025% of world trade



5.3 Application to trade data

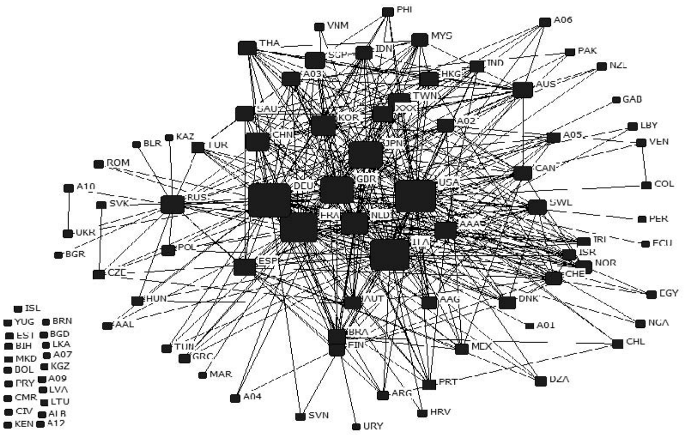
Results presented here concern four types of matrices, three of them derived from the original ones. The input is valued directed matrices from database CHELEM. But if some methods are convenient with directed networks, some are not, so we'll also use symmetrical matrices. More important, many methods need binary matrix (directed or not).

We made many tests on different matrices to check methods available. We first studied networks taking into account the size of the export and trade flows. As most of these methods fit better with binary matrices, we had to choose a cut point value. For the matrices of total exports in 1994-96 and 2004-06, we put a zero when the value of the export was below 0.025% of the world trade, we put a 1 elsewhere. For the matrices of flows (export + import) in 1994-96 and 2004-06, we took into account the 10% more important values. Tests were made with different cut point values and results were quite similar.

Figures 28 and 29 show these two types of networks. On each figure, the size of the node depends on its degree (out-degree for directed graph). A quick look allows to see which countries are important (big in the centre) and which are not (isolate). But taking into account the size of the trade flows causes a problem; graphs are disconnected, which prevent some measures; and the number of isolate states is too important (nearly 20%). So methods available did not gave great results with these graphs. We also noticed some problems with the CHELEM database, especially with the XX ensemble.

The first graph is directed, the second is none; but they look quite similar. On both, 3 kinds of countries appear; first, countries belonging to the interconnected core of the graph; second countries related with only a cou-

Figure 29: Trade flows in 1995. Each link represents a flow greater than 1500M US$



ple of other countries which suppose dependence; and third group, isolate countries. The fact to take symmetrical flows does not change the general structure of the network, and the number of isolate groups of countries remains high.

To suppress size effect and to minimize the number of isolates, we used another way to dichotomize the trade matrices; we put a 1 in cell when the export flows was greater or equal to 5% of the total national export flows. For example, Canada sends nearly 80% of its exports to USA so we put a 1; on contrary, its exports to Germany represent only 1.3% of its total exports so we put a 0.

To get a symmetrical matrix, with is useful to make other kinds of measurement, we used the geometric mean of export and import between two countries (see box 7).

For the symmetrical matrix, the cut-point chosen is lower (greater or equal to 3% of total flow) which allows to get only a few isolates (less than 10% of total). Besides, the entity XX was deleted from the original matrices.

Figures 30 to 33 shows the network studied. Size of nodes is related with the degree (in-degree for the first two networks)

Basic measures on networks and nodes

Dividing the world, looking for equivalence

When we study equivalence, we need directed graphs. Using methods to partition graphs is useful because finding structures with 93 nodes and more than 400 ties is quite difficult (Figures 30 and 31). We first used one of the oldest methods available, the CONCOR one. As expected, the CONCOR

Table 7: From directed to indirected matrix

Let's take the export flows between Iceland and USA. Iceland sends nearly 11% of its export to USA, while USA sends 0.03% of its export to Iceland. One main flow for a country might be insignificant for the other. Some easiest ways to symmetrise it is to take the minimum or the maximum, in both case, it alters the significance of the flow. Using the arithmetic mean will give a value of $(11+0.03)/2 = 5.5\%$ which over estimates the importance of this trade flow. A convenient way to symmetrise the matrix is to use the geometric mean. In this case, we take the square root of their product $M = \sqrt{11} * 0.03 = 0.6\%$.

Figure 30: Dominant export flows in 1994-96

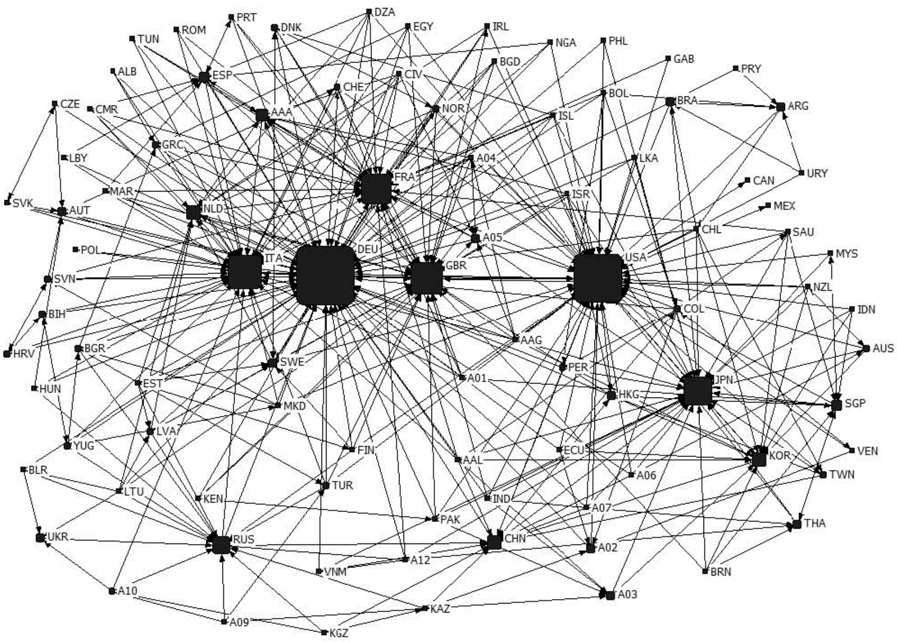
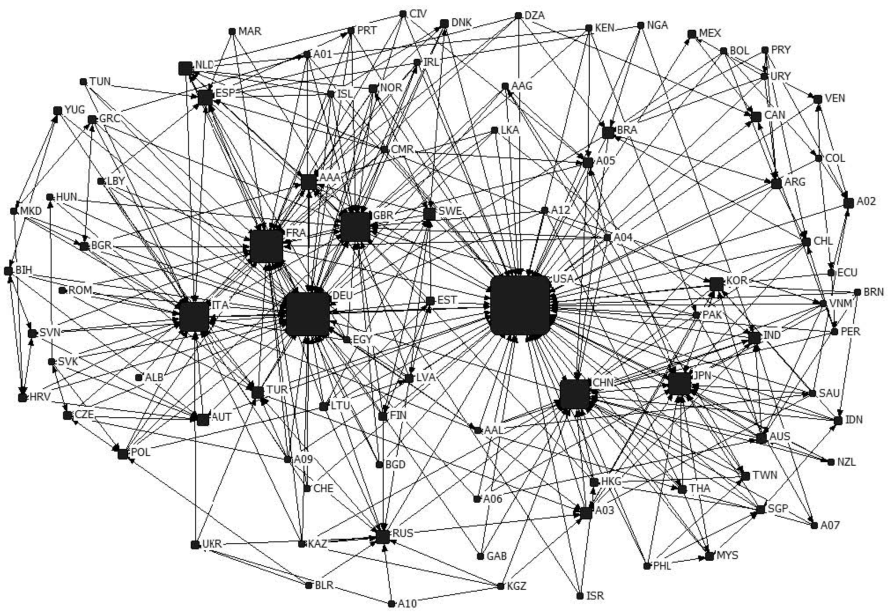


Figure 31: Dominant export flows in 2004-06



The hierarchy of world trade clearly appears on these graphs. A couple of countries are dominating it; but some evolution is visible, like the rise of China. Three kinds of countries appear; core of the system with high in degree, a couple of medium size countries (Russia, Spain) and some countries which trade only with few other countries.

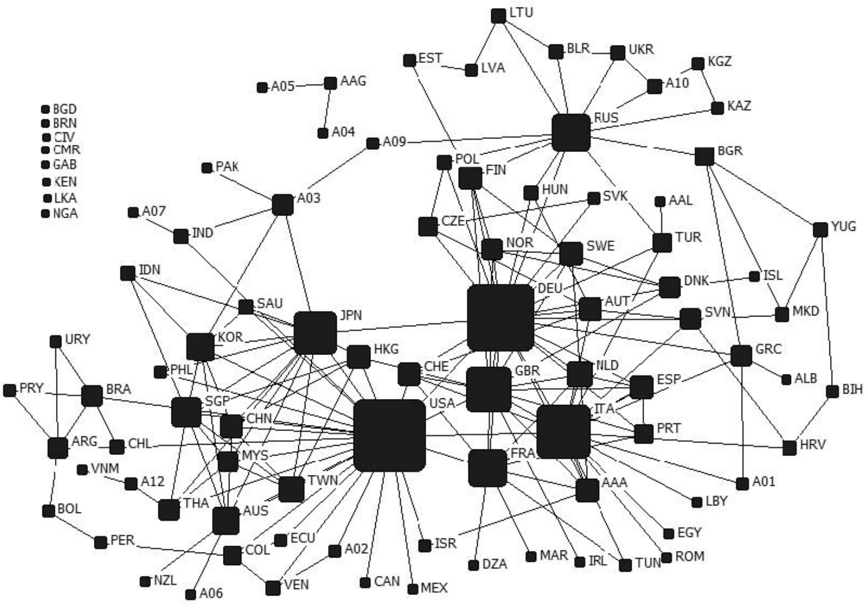
Table 8: Basic indicators on the 4 matrices used

	Export95	Export05	Import95	Import05
Order	93	93	93	93
Nb isolates	0	0	8	7
Size	417	438	185	217
Density	0.049	0.051	0.043	0.051
Av. dist.*	2.34	2.06	3.12	3.05

Av. dist.: average distance between reachable pairs

Getting 4 graphs of same size allows comparisons. The density is slightly increasing, and, logically, the average distance decreases. But the general structure does not move a lot; the number of isolates remains stable, which could illustrate that a growing trade does not necessarily means a decreasing of inequalities.

Figure 32: Dominant trade flows in 1994-96



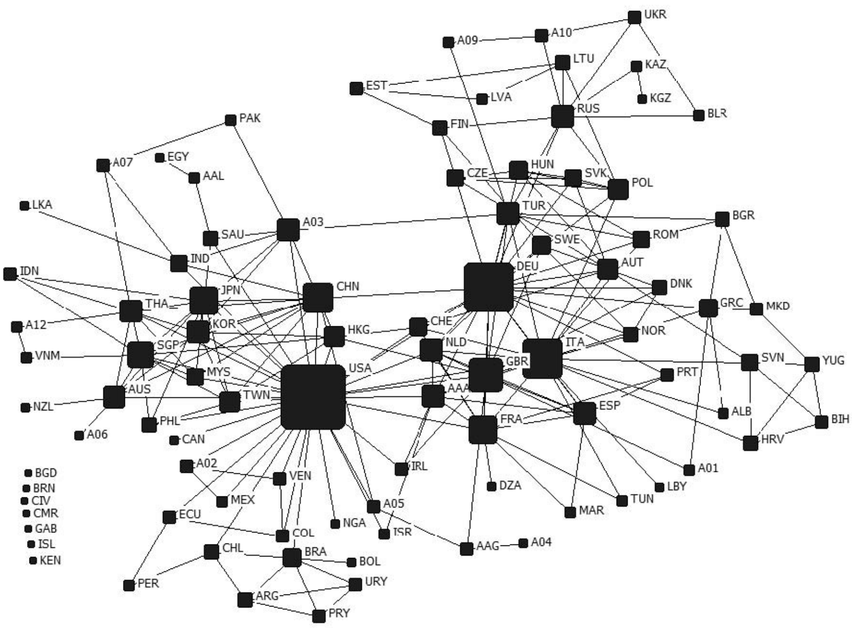
method, despite its inconvenient, provides a clear division based on world trade (Figure 34). Regionalization seems interesting, not in terms of hierarchical positions, but in terms of relational properties. For example, Canada and Mexico are in the same class because their dominant flow (from far) is directed to the USA. Composition of class between 1995 and 2005 illustrates the major changes; the South African region separates itself from Sub Saharan Africa; China and India joins the USA.

A more recent method presents less inconvenient and is thematically more interesting because, instead of looking for structural equivalence (same relations with same actors), the REGE method looks for regular equivalence (same type of relations with same type of actors). We tested it first with the original valued matrix (XX excluded), then with the binary matrix already used (1 if export greater or equal to 5% of total flows). Some aggregations due to the CHELEM database clearly caused a problem, especially sub Saharan countries. Using valued matrices when the level of precision is not always perfectly known is quite a challenge. Testing it with another database should be useful to test its relevance.

Looking for the core(s)

On contrary with equivalence, what is needed here is undirected graph, so we used matrices of dominant trade flows based on the geometric mean. The

Figure 33: Dominant trade flows in 2004-06



Using the dominant trade flows gives a simpler image of this relational system. Its hierarchy is more apparent, the strong core remains and so the medium size countries. Regarding the periphery of the graph, two sub sets emerge; countries related only with one or two other countries (Egypt, Algeria) and isolated ones.

networks are quite dense so searching for k-plexes was useless. Cliques give a good idea of sub cores in the system while k-cores show the hierarchy of the world system. The evolution between 1995 and 2005 is marked by a greater concentration; strong sub cores still exist in 1995, it's less the case in 2005.

These two maps (Figure 35) shows changes in the hierarchy of world trade. The main opposition is the one between poorly connected countries (Canada, Mongolia) and highly connected ones. The emergence of the Mercosur appears, as the decline of Russia. The MAUP problem still makes difficult to comment the evolution of Sub Saharan countries.

5.4 Interest and limits for EuroBroadMap project

Methods from social network analysis perfectly fit to describe a system of relations, the density of its core, or the intensity of relations in a system. They are also really convenient to detect cores and sub cores in a network.

Table 9: Basic indicators for binary exports matrices (export = 1 if greater or equal to 5% of total)

Countries ranked per in-degree					
1995	Indegree	Outdegree	2005	Indegree	Outdegree
Germany	67	8	USA	67	4
USA	53	3	Germany	47	7
Italy	35	4	France	33	6
UK	33	7	UK	30	6
France	31	6	Italy	28	5
Japan	29	4	China	28	3
Russia	15	4	Japan	19	4
Netherlands	12	5	AAA	11	6
China	12	4	Spain	11	5
1995	Indegree	Outdegree	2005	Indegree	Outdegree
Min	0	1	Min	0	1
Med	1	4	Med	2	5
Max	67	8	Max	67	8
Sum	417	417	Sum	438	438
Mean	4.48	4.48	Mean	4.71	4.71
CV	2.40	0.35	CV	2.16	0.32
St dev	10.69	1.54	St dev	10.12	1.52

Taking dominant export flows makes the evolutions less visible. But the ranking remains interesting. Countries on the top of the system have a common properties; the balance between in and out degree is high. In other words, many countries are dependant from them to buy their exports, but the inverse is not true.

What seems harder is to detect peripheries. So, as far as we can say at this point of our research, its utility to provide a complete world division is not perfectly assumed, especially with the Map Areal Unit Problem created by the CHELEM database.

Figure 34: CONCOR method used on dominant exports flows (1995 left, 2005 right)

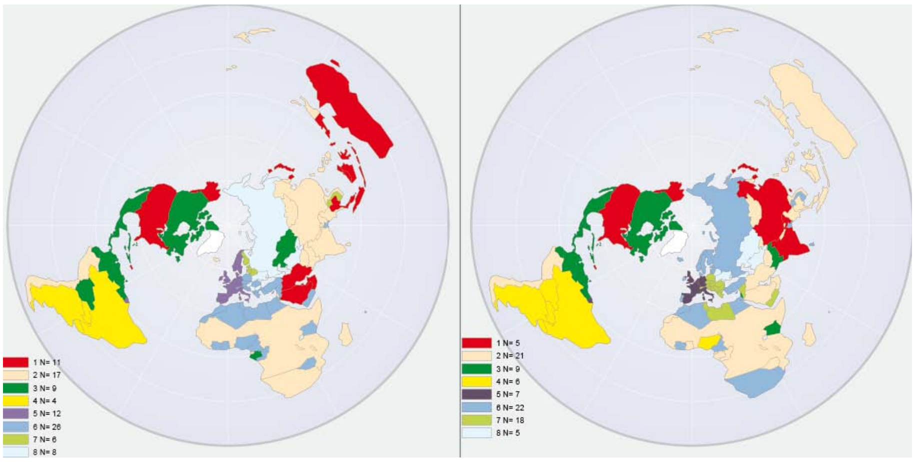
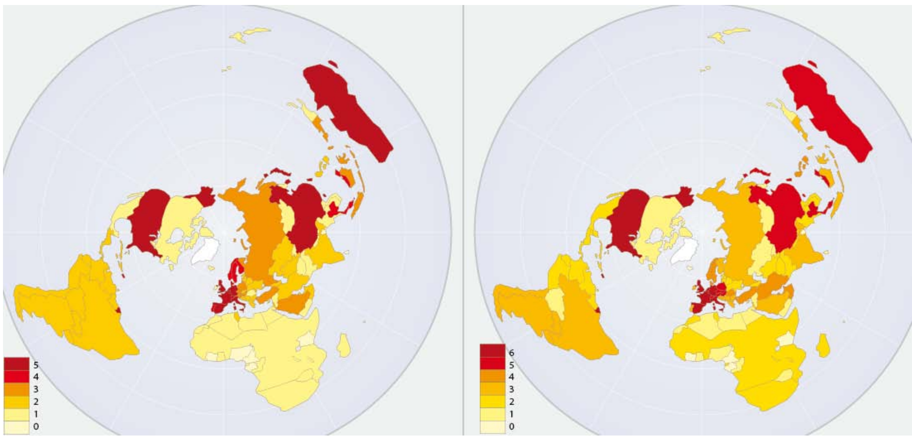


Figure 35: K-cores in 1995 (left) and 2005 (right)



Conclusion

To conclude, it appears that all methods examined here can be divided in three families:

- The international bilateral approach is based on the theoretical assumption that state does still matter and are always major players in the world economy. It ignores therefore the existence of Regional Trade Agreement (RTA) and focus of relations of power that can be derived from state-to-state trade exchanges. The methodological tool related to this point of view is the method of dominant flows.
- The regional structural approach is based on the reverse assumption that world economy is more and more organised in wide integrated cluster of countries that can be related to RTA (like EU or NAFTA) but also inherited from former historical period (former USSR) or related to various forms of distance (cost of transportation, common language, etc.). The methodological tools are related to gravity model assumption, but with a specific focus on residuals of models that controls size effects but not distance effects because they are part of the building of structural regions.
- The centre-periphery approach focus on unequal relations of trade between countries, as part as a more general model of political and social inequalities. From a methodological point of view, a two step approach is developed where countries of the centre are firstly identified and countries of the periphery are derived in a second step according to their dependency to the various cores of the World Economy.

Despite very different theoretical and methodological backgrounds, it appears that many convergences can be observed in the empirical results obtained by each approach. We notice in particular the following empirical discoveries:

- The world is not necessary divided in three dominant areas (the Triad) in 1995 but rather in two major regions: Europe and its southern and eastern neighbourhood versus the rest of the world.
- The organisation in three major areas is more obvious in 2005 with the growing power of China and the building of an Asia-Pacific regions more independent from USA.
- Sub-Saharan Africa but also Persian Gulf and Indian subcontinent are areas that are difficult to assign to the major world regions and are subject to strong fluctuations according to the method and the theory that is applied. But in every case, there is a clear decline of European influence in this parts of the world between 1995 and 2005.

- The direction of trade flows has a strong influence on the results when it is taken into account. The regionalisation obtained through the analysis of export flows (dominant countries are the greater importer) is generally not equivalent to the regionalisation obtained through the analysis of import flows (dominant countries are the greater exporter).

- The disaggregation of trade flows by products reveals very different patterns of regionalisation.

- The aggregation of countries that is used by the CHELEM database (e.g. rest of Gulf, rest of subsaharian Africa) can heavily modify the results of some methods as compared to a pure analysis at state level. Some methods are more robust than other to the MAUP (Modifiable Area Unit Problem).

References

Alvanides S., Openshaw S. and Duke-Williams O., 2000, Designing zoning systems for flows data in Atkinson P., Martin D. (eds.), 2000, *Geocomputation, Part II: Zonation and Generalization*, Taylor & Francis Group, Ed. CRC Press, pp. 115-34.

Amin S., 1976, *L'accumulation à l'échelle mondiale*, Paris, Anthropos.

Anderson J.E., 1979, A theoretical foundation for the gravity equation, *American Economic Review*, 69: 106-16.

Aubigny (d') G., Calzada C., Grasland C., Robert D., Viho G. and Vincent J.M., 2000, Approche poissonnienne des modèles d'interaction spatiale (A statistical approach of spatial interaction models), *Cybergeo*, 126.

Baldwin R.E., 1994, *Towards an integrated Europe*, London, Centre for Economic Policy Research.

Beguin H., 1979, *Méthodes d'analyse quantitative géographique*, Paris, Librairies Techniques.

Bergstrand J.H., 1985, The gravity equation in international trade: Some micro-economic foundations and empirical evidence. *Review of Economics and Statistics*, 67: 474-81.

Bergstrand J.H., 1989, The generalized gravity equation, monopolistic competition, and the factor-proportions theory in international trade, *Review of Economics and Statistics*, 71: 1143-53.

Bergstrand J.H., 1990, The Heckscher-Ohlin-Samuelson model, the Linder hypothesis, and the determinants of bilateral intra-industry trade, *Economic Journal*, 100: 1216-29.

Bonacich P., 2007, Some unique properties of eigenvector centrality, *Social networks*, 29: 555-564.

Borgatti S.P., Everett M.G. and Freeman L.C., 2002, *Ucinet for Windows: Software for Social Network Analysis*, Harvard, Analytic Technologies.

Braudel F., 1979, *Civilisation matérielle, économie et capitalisme : XV^{e}-XVIII^{e} siècle*, Paris, Armand Colin, 3 vol.

Broadbent T., 1969, *Zone size and singly constrained interaction models*, CES-WN-132, Centre for Environmental Studies, London.

Bröcker H. and Rohweder J., 1990, Barriers to international trade - Methods of measurement and empirical evidence, *Annals of Regional Science*, 298-305.

Burt R., 1980, Models of Network Structure, *Annual Review of Sociology*, 6: 235-244.

Degenne A. and Forsé M., 1994, *Les réseaux sociaux*, Paris, Armand Colin.

Doreian P., 2008, A multiple indicator approach to blockmodeling signed networks, *Social networks*, 30: 247-258.

Dorigo G. and Tobler W., 1983, Push-Pull migration laws, *Annals of the Association of American Geographers*, 73(1): 1-17.

Eichengreen B. and Irwin D.A., 1998, The Role of History in Bilateral Trade

Flows, in Frankel J.A. (ed.), 1998, *The Regionalization of the World Economy*, Chicago, University of Chicago Press, pp. 33-62.
Emmanuel A., 1969, *L'échange inégal*, Paris, Maspero.
ESPON Project 3.4.1., 2007, *Europe in the World: Territorial evidence and divisions.*
Faust K. and Skvoretz J., 2002, Comparing networks across space and time, size and species, *Sociological methodology*, 32: 267-299.
Fotheringham A.S. and O'Kelly M.E., 1989, *Spatial Interaction Models: Formulations and Applications*, Dordrecht, Kluwer Academic Publishers.
Frankel J.A. (ed.), 1998, *The Regionalization of the World Economy*, Chicago, University of Chicago Press.
Frankel J.A., Stein E. and Wei S.J., 1995, Trading blocs and the Americas: The natural, the unnatural, and the super-natural, *Journal of Development Economics*, 47: 61-95.
Freeman L.C., 1977, A set of measures of centrality based on betweeness, *Sociometry*, 40(1): 35-41.
Grasland C., 1999, *Interaction spatiale et effets de barrière. 1re partie : Des modèles d'interaction spatiale aux modèles d'interaction territoriale*, Miméo, CNRS, UMR Géographie-Cités.
Haggett P. and Chorley R.J., 1969, *Network analysis in Geography*, London, Edward Arnold.
Hanneman R. and Riddle M., 2005, *Introduction to social network methods*, Riverside, University of California
Head K. and Mayer T., 2002, Effet frontière, intégration économique et forteresse Europe, *Économie et Prévision*, 152-153: 71-92.
Helpman E., 1987, Imperfect competition and international trade: Evidence from fourteen industrial countries, *Journal of the Japanese and international Economies*, 1: 62-81.
Helpman E. and Krugman P., 1985, *Market structure and foreign trade*, Cambridge, MIT Press.
Hirst M. A., 1977, Hierarchical aggregation procedures for interaction data: a comment, *Environment and Planning A*, 9(1): 99-103.
Hummels D. and Levinsohn J., 1995, Monopolistic competition and international trade: Reconsidering the evidence, *Quarterly Journal of Economics*, 110: 799-836.
INSEE, 2004, *Les zones d'emplois.*
De Jong T. and Goetgeluk R., 2007, What about the spatial dimension of subsidiary in housing policy ?, International Conference on sustainable urban areas, Rotterdam[6].
Kim S. and Shin E.H., 2002, A longitudinal analysis of globalization and regionalization in international trade: A social network approach, *Social forces*, 81(2): 445-471.

[6]http://www.enhr2007rotterdam.nl/documents/W02_paper_Goetgeluk_DeJong.pdf

Klaassen L.H., Wagenaar S. and Van der Weg A., 1972, Measuring psychological distance between the Flemings and the Walloons, *Papers of the Regional Science Association*, 29: 45-62.
Kravis I.B. and Lipsey R.E., 1988, National price levels and the prices of tradables and nontradables, *American Economic Review Papers and Proceedings*, 78: 474-478.
Krugman P., 1991a, Is bilateralism bad?, in Helpman E. and A. Razin A. (eds.), *International trade and trade policy*, Cambridge, MIT Press.
Krugman P., 1991b, The move toward free trade zones, in *Policy Implications of Trade and Currency Zones*, Jackson Hole, pp. 7-42.
Krugman P., 2004, The New Economic Geography: Where Are We?, International Symposium on Globalization and Regional Integration-from the Viewpoint of Spatial Economics.
Learner E.E. and Stem R.M., 1970, *Quantitative international economics*, Boston, Allyn and Bacon.
Mackay J.-R., 1958, The interactance hypothesis and boundaries in Canada: a preliminary study, *Canadian Geographer*, 11: 1-8.
Masser I. and Brown P.J.B., 1975, Hierarchical aggregation procedure for interaction data, *Environment and planning A*, 7: 509-523.
Masser I., 1976, The design of spatial systems for internal migration analysis, *Regional studies*, 10: 39-52.
Masser F.I. and Scheurwater J., 1977, Een aggregatie methode voor ruimtelijke interactiegegevens. An aggregation method for spatial interaction data. *Planning, methodiek en toepassing*, 2: 3-9.
Nordlun C., 2007, Identifying regular blocks in valued networks: A heuristic applied to the St. Marks carbon flow data, and international trade in cereal products, *Social networks*, 29: 59-69.
Nyusten J. and Dacey M., 1968, A graph theory interpretation of nodal regions, in Berry B. and Marble D. (Eds.), *Spatial analysis: a reader in statistical geography.* Englewood Cliffs, Prentice-Hall, pp. 407-18.
Openshaw S., 1977, Optimal zoning systems for spatial interaction models, *Environment and Planning A*, 9: 169-184.
Pearson K., 1901, On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, 2(6): 559-572.
Reynaud A., 1981, *Société, espace et justice*, Paris, PUF.
Sarkar P., 2001, The North-South terms of trade debate: a re-examination, *Progress in Development Studies*, 1(4): 309-327.
Sen A. and Smith T.E., 1995, *Gravity models of spatial interaction behaviour*, Berlin, Springer.
Senior M.L., 1979, From gravity modelling to entropy maximizing: a pedagogic guide, *Progress in Human Geography*, 3(2): 175-210.
Serrano M.A., Boguñá M. and Vespignani A., 2007, Patterns of dominant flows in the world trade web, *Journal of Economic Interaction and Coordination*, 2: 111-124.

Singer H., 1950, The distribution of gains between investing and borrowing countries, *American Economic Review*, 40: 473-485.
Singer H., 1984, Terms of trade controversy and the evolution of soft financing: early years in the UN: 1947-1951, in Meier M. and Seers D. (eds), *Pioneers in development*, Oxford, Oxford University Press.
Smith D. and White D., 1992, Structure and dynamics of the global economy: a network analysis of international trade 1965-1980, *Social forces*, 70(4): 857-893.
Snyder D. and Kick E., 1979, Structural Position in the World System and Economic Growth, 1955-1970: A multiple-network Analysis of Transnational Interactions, *American Journal of Sociology*, 84(5): 1096-1126.
Summers R. and Heston A., 1991, The Penn-World-Table (Mark 5), *Quarterly Journal of Economics*, 106: 327-68.
Taaffe E.J. and Gauthier H.L., 1973, *Geography of Transportation*, Englewood Cliffs, Prentice Hall.
Techer G., 1994, Méthode informatisée de recherche et d'analyse des bassins par l'étude des liaisons (M.I.R.A.B.E.L.), Paris, INSEE.
Tinbergen J., 1962, *Shaping the world economy: Suggestions for an international economic policy*, New York, Twentieth Century Fund.
Tobler W., 1983, An alternative formulation for spatial-interaction modeling, *Environment and Planning A*, 15(5): 693-703.
Vandermotten C. and Marissal P., 2004, *La production des espaces économiques*, Bruxelles, PUB, 2 vol.
Van der Zwan J., Van der Wel R., De Jong J. and Floor H., 2005, *Flowmap 7.2: Manual*, Faculty of Geographical Science, Utrecht University[7].
Wallerstein I., 1974-1989, *The modern world system*, New York, Academy press, 3 vol.
Wallerstein I., 1985, *Le capitalisme historique*, Paris, La Découverte.
Wang, Z. and Winters L.A., 1991, The trading potential of Eastern Europe, Discussion Paper 610, London, Centre for Economic Policy Research.
Ward J.H., 1963, Hierarchical grouping to optimize a objective function, *Journal of American Statistical Association*, 58(301): 236-244.
Wasserman S. and Faust K., 1994, *Social Network Analysis: Methods and applications*, Cambridge, Cambridge University Press.
Wei S.J., 1996, Intra-National versus International Trade: How Stubborn are Nations in Global Integration?, NBER Working Paper W5531.
White H.C., Boorman S.A. and Breiger R.L., 1976, Social structure from multiple networks. I. Blockmodels and positions, *The American Journal of Sociology*, 81(4): 730-780.
Wilson A.G., 1971, A family of spatial interaction models, and associated developments, *Environment and Planning A*, 3: 1-32.

[7] http://www.flowmap.geog.uu.nl

List of Figures

Contents