Juan Carlos Arganaraz (2022301),
Hilario Capettini (2013031),
Gerardo Carmona (2005005),
Luca Rinaldi (2026363).

Characterisation and analysis of the international agro-trade network

---

March 14, 2022

Abstract

Diverse systems that may seem too elaborate at first, can be modelled in an elegant and efficient fashion by means of graphs. Once the theoretical and empirical perspectives on such models are combined, the powerful concepts and techniques of network science emerge. The study of international trade profits extensively from such an approach. Global food safety and sustainability is a growing concern; understanding the underlying structures at different scales is thus essential to promote wellbeing. In this work, we model international agro-food trade as a weighted, directed graph, aiming to extract insightful results from the application of network science methods. At first we perform a full characterisation of the network by the means of the general proprieties and different metrics. Then we deeply investigate the eventual emerging communities and their evolution over time, first with a single-layer approach and then with a multi-layer ones. Finally we analyse the Robustness of the network exploiting different metrics.

1 Introduction

1.1 World Trade Network (WTN) Representation

We model the global trade network as a directed graph in which nodes represent countries and the edges correspond to an import/export relationship for a given good. Therefore, an edge with weight w from A to B represents country A exporting to country B (or B importing from A). The weight of this edge is proportional to the magnitude of the trade. Depending on the scope of the analysis and the related methodology to apply, the topology of the trade network can be adapted (e.g. from directed to undirected graph) to extend the range of potential analyses.
Figure 1 shows the world trade network as of 2015. For clarity, it is plotted as an undirected network here. Note that all edges have the same width, regardless of the size of the bilateral trade flows.

1.2 Relevance of the discussion

The directed, weighted network model for this data proves to be of outstanding usefulness. It is subjected to analysis by economists, statisticians and even policymakers[15].
While the flow of diverse commodities may be studied, such as oil [8] and steel [10], we focus on the agro-food trade network. This network allows for analyses such as identification of vulnerabilities in the international food supply chain [11], the estimation of water virtual trade [22], detection of the international trade dominant flows [21], or even the time evolution of a market, as studied in [7] for meat trade. Comprising many different types of commodities, an appropriate assessment of this network may shed a light on the oncoming global food safety challenges, and even the environmental impact of current trends.

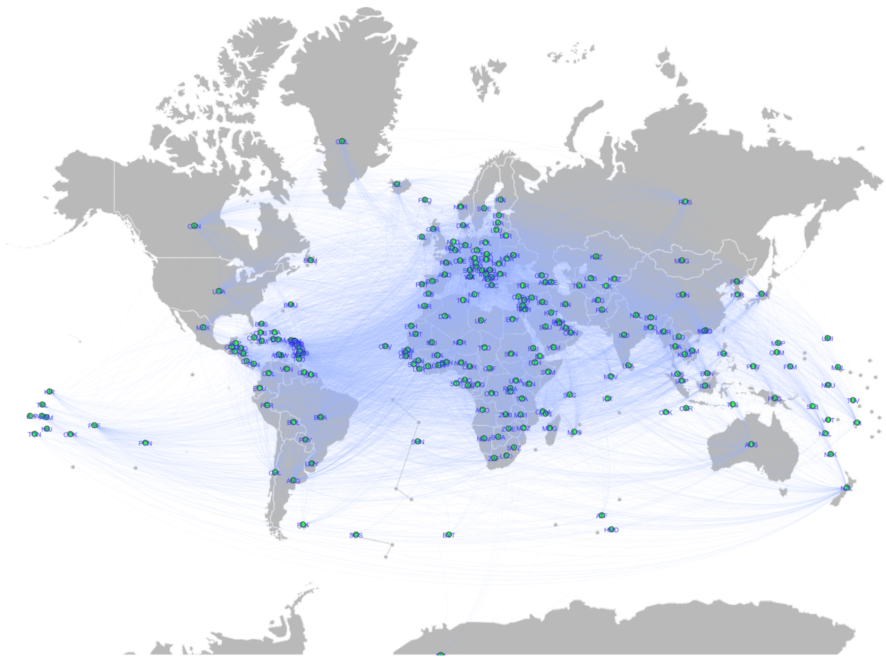


Figure 1: The Agro-Food World Trade Network, 2015.

1.3 Multi-layer networks

Given the variety of commodities involved, a way to endow the network with these as separate degrees is in order. Aiming to keep the different connections meaningful during the analysis of the network, we explore a multi-layer model.
Multi-layer networks are an active research topic[14], which have been proposed to study relations between nodes on different layers (interlayer edges), as well as temporal models for the evolution of networks[3]. We limit our model to a multiplex network [9], that is, a network whose nodes remain fixed throughout the layers. In our model, each layer captures the trade network for one of the commodities. The temporal aspect of such a network is considered in the conventional fashion: snapshots of the (multi-layer) network are considered as separate networks.
We follow techniques presented in [1] for multi-layer analysis. Although conceived for the study of social networks, they lead to interesting results when applied to the international agro-food trade multi-layer network. These results are consistent with characterisations on the single-layer separate networks.

1.4 Structure of the article

We begin by describing the specific implementations and methods used for the analysis and general characterisation of the network. Pseudocode for some the algorithms conceived or adapted for the specific needs of the multi-layer network may be found here.
Then, some representative results are collected and discussed. After concluding remarks, we include further results in an appendix section.

2 Materials

2.1 Dataset

The United Nations Comtrade Database offers global trade data on an annual basis as far back as 1962. Using the Comtrade API *United Nations. (2003). UN comtrade.*, we obtained bilateral trade data for a number of commodities. In accordance with the HS standard we analyse 23 commodities. By means of the `python` package `requests`, we set a script to retrieve data for the period between 2001 and 2020.

The entries of this data include reporter, partner, year, commodity and total import value adjusted for 2010 USD. The full description of the commodities is stated in the Appendix A.

2.2 Network properties

2.2.1 Network metrics

To gain insights from the trade networks, we need to introduce some metrics of network assessment.

- **Adjacency Matrix** it is an $N \times N$ matrix in which each element represent the link between the nodes indexed by the corresponding row and column.
- **Degree** k_i For a binary graph, the degree of a node is its number of links or neighbours.
- **Average Degree distribution** $< k >$ is an important property of the networks and for an undirected network the expression is the following:

$$< k >= \frac{1}{N} \sum_{i}^{N} k_i = \frac{2L}{N}$$

  with N the number of countries. In directed networks, we need to distinguish between the *incoming degree,* k_i^{in}, representing the number of links that point to node i and the *outgoing degree,* k_i^{out}, representing the number of links that point from node i to others. We can define also the *total degree,* k_i, given by:

$$< k >= k_i^{in} + k_i^{out}$$

  Finally, we can defined the *average degree* of a directed network such:

$$< k_i^{in} >= \frac{1}{N} \sum_{i}^{N} k_i^{in} =< k_i^{out} >= \frac{1}{N} \sum_{i}^{N} k_i^{out} = \frac{L}{N}$$

- **Degree distribution** the degree distribution gives the probability that a randomly chosen node in the network has degree k. Since p_k is a probability, its values are in the interval $[0, 1]$.
- **The Scale-Free property** in real (large) networks the degree distribution is heavy tailed. The reason is the presence of certain nodes called *hubs* (few nodes with high degree). From the other side, several nodes have a low degree. This phenomenon is well known as *scale free property.* Plotting the degree distribution in log-log scale, it is verifying that the distribution follows a *power-law* fit, which expression is the following:

$$p_k >= C \times k^{-\gamma}$$

  with the parameter γ called the *exponent* and meaningful in an interval $[k_{min}, k_{max}]$
- **Diameter** the diameter is the largest geodesic (shortest path between two nodes) distance in the network which give us an estimation about how big the network is.
- **Density** is the number of connections a node has, divided by the total possible connections a node could have.

2.2.2 Centrality measures

In graph theory and network analysis, indicators of centrality assign numbers or rankings to nodes within a graph corresponding to their network position [2]. In this work we explore six different centrality measures making usage of the python package `networkx`[12]:

- **Betweenness centrality** looks at all the shortest paths that pass through a particular node. It is fairly good at finding nodes that connect two otherwise disparate parts of a network. If a node is the only thing connecting two clusters, every communication between those clusters has to pass through this particular node. In contrast to a hub, this sort of node is often referred to as a broker.

- **Closeness centrality** of a node measures its average inverse distance ("farness") to all other nodes. Nodes with a high closeness score have the shortest distances to all other nodes. So, a node with high closeness centrality is literally close to other nodes.
- **Degree centrality** is the simplest and the most common way of finding important nodes. A node's degree is the sum of its edges.The nodes with the highest degree in a social network are the people who know the most people. These nodes are often referred to as hubs, and calculating degree is the quickest way of identifying hubs.
- **Eigenvector centrality** give importance to nodes that are hubs, but it also cares how many hubs you are connected to. It's calculated as a value from 0 to 1: the closer to one, the greater the centrality. This centrality is useful for understanding which nodes can get information to many other nodes quickly.
- **Page Rank centrality** is an adjustment of Katz centrality. There are three distinct factors that determine the PageRank of a node: (i) the number of links it receives, (ii) the link propensity of the linkers, and (iii) the centrality of the linkers.

2.3 Community detection

In network science we call a community a group of nodes that have a higher likelihood of connecting to each other than to nodes from other communities [2]. In the case of the network that we are analysing, the international food-trade network, finding such communities may bring insights about the commercial relations that are present in the globalised world with respect to specific commodities such as meat, dairy products or coffee, this approach was recently seen in [23], however the bibliography about this argument is rather scarce; maybe because of the lack of test datasets to check the resulting communities. At first the community detection algorithms are applied on networks for specific years and specific commodities, for example year 2015 and commodity livestock in order to test the different algorithm results and performances. After that, the most performing algorithms were applied to networks in the time range 2001 to 2020 in order to study the evolution of the different networks along the time. Finally we also implemented a multi-layer community detection approach where each layer of the network was built with one commodities.

2.3.1 Single layer approach

To perform the single-layer community detection we inspect different algorithms that are capable of tacking the community detection in a weighted and directed graph. The algorithms that we used can be clustered in three main groups:

1. Based on dynamics
    - *Infomap*: Based on information theory uses the probability flow of random walks on a network as a proxy for information flows in the real system and decompose the network into modules by compressing a description of the probability flow. The result is a map that both simplifies and highlights the regularities in the structure and their relation-ships [20].
    - *Walktrap* is a community detection algorithm based on random walks which was proposed by Pascal Pons [17]. Random walks of a finite length are used to determine the distance between nodes in the graph, after that nodes are assigned to groups where the distance between nodes in the group is small while the distance to nodes in other groups is large. This is an agglomerative algorithm.
    - *Label propagation* is a clustering algorithm proposed by Raghavan [18]. The intuition behind the algorithm is that a single label can quickly become dominant in a densely connected group of nodes, but will have trouble crossing a sparsely connected region. Labels will get trapped inside a densely connected group of nodes, and those nodes that end up with the same label when the algorithms finish can be considered part of the same community [16].
2. Based on statistical mechanics

- *Spinglass* is an algorithm presented by Reichardt and Bornholdt [19] which relies in an analogy between finding the groundstate of an infinite range spin glass and the community detection problem. The community structure of the network is interpreted as the spin configuration that minimizes the energy of the spin glass with the spin states being the community indices.

3. Based optimisation

   - *The Louvain* algorithm is a method to detect communities in large networks. It maximizes a modularity score for each community, where the modularity quantifies the quality of an assignment of nodes to communities. This means evaluating how much more densely connected the nodes within a community are, compared to how connected they would be in a random network. The Louvain algorithm is a hierarchical clustering algorithm, that recursively merges communities into a single node and executes the modularity clustering on the condensed graphs [6].
   - Surprise is another probabilistic method, but rather than the probability of finding dense subgraphs, it focuses on the probability of so many edges within communities. It produces asymptotic approximation, based on the Kullback Leibler divergence [24].

For this project we used two python packages containing implementations of the previously discussed community detection algorithms, `Igraph` package for *Infomap*, *walktrap*, *spinglass* and *label propagation* and `Louvain` for *Louvain modularity* and *Louvain surprise.*
In our approach we only seek for distinct communities, a further step would be to use algorithms capable of detecting overlapping communities which could easily emerge in a trade network such the one in study here. From the algorithms we use, only *spinglass* is capable of such a thing.

2.3.2 Evolution of the communities

In our work we use different community detection algorithms to identify communities in networks built with imports for a specific year and a specific commodity. As our dataset contains information for a period of 20 year a question that arises is how these communities change over time. In order to compare these a first problem to solve is the identification of a community of one year with the following. In other words, how can we say that community A identified in 2001 is the same as community B identified in 2002? Our proposal is to introduce an index that we call *sharedness* s:

$$s(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

which compares the content of both communities and return a values between 0 and 1. Getting the lowest value if the sets are disjoint and the highest if are identical. Once the sharedness $s(A_i, B_j)$ is computed for all the communities A_i in the first year graph combined with all the communities B_j detected in the following year we can associate the communities getting the higher similarities. Though different approaches can be taken, we carry this identification by means of a standard greedy algorithm.

Algorithm 1 Community matching algorithm

Ensure: $labels\_a \cap labels\_b = \emptyset$
  $free\_a \leftarrow communities\_a$
  $free\_b \leftarrow communities\_b$
  while $\max_{A \in free\_a, B \in free\_b}\{s(A, B)\} \geq \sigma$ **do**
    $a, b \leftarrow \operatorname{argmax}_{A \in free\_a, B \in free\_b}\{s(A, B)\}$
    $b.label \leftarrow a.label$
    $free\_a.remove(a)$
    $free\_b.remove(b)$
  end while

2.3.3 Multi-layer approach

For efficient manipulation, we model the time-evolving, multi-layer network using `numpy`[13] arrays of rank 4, two axes for nodes, one for commodities and the last one for time evolution.

We implement an algorithm based on [1], namely, the MNLPA Algorithm, which identifies communities in weighted, directed, multiplex networks such as ours.
The specific implementation pseudocode is reproduced here for completeness.2
In particular, we use a generalization of the Jaccard similarity measure[1]:

Algorithm 2 MNLPA

```
labels ← 1 : N_countries                         ▷ each country is assigned its own label
while condition() ≠ True do
    shuffle(nodes)
    for node in nodes do
        scores[labels] ← 0
        for neighbour in node.filtered_neighbours do
            curr_label ← neighbour.label
            scores[curr_label] ← scores[curr_label] + edges[neighbour, node].weight
        end for
        node.label ← argmax(scores)
    end for
end while
```

$$\text{jacc}(x, y) = \frac{1}{\sum_l \omega^l} \sum_l \frac{1}{2\omega^l} \left[s(\tau_x^l, \pi_y^l) + s(\pi_x^l, \tau_y^l) \right]$$

where x and y are any two nodes, $s(\cdot, \cdot)$ is the sharedness measure introduced earlier, τ_z^l is the set of in-neighbours in layer l for node z, π_z^l is the set of node z out-neighbours on layer l. Although ω^l is set as a weight for layer l, we decided to weight layers equally, i. e. $\omega^l = 1, \ \forall l$.
In our implementation, we used the average value of this measure over each node against all of its in-neighbours to set a reference for the threshold to filter neighbours of nodes throughout the multi-layer network.
The stop condition could be simply a number of iterations, or a maximum number of iterations combined with a bound on the similarity between communities generated in the current iteration and those of the previous one.

2.4 Robustness analysis

The Robustness quantify the ability of the network to withstands failures and perturbations[2]. We distinguish two measures to describe the functioning of the network:

1. The *Largest Strongly Connected Cluster (LSCC)* is the natural extension of LCC for directed graph. LSCC is a widely used measure of the network functioning; it is the highest number of (strongly) connected nodes in the network. This value is normalized with respect to the LSCC of the total network (before the attacks).

2. The IMP ratio, which is an *ad hoc* metrics thought to express how changes the percentage volume of a country's imports on average:

$$IMP = \frac{1}{N} \sum_c^N \frac{\text{tot. import of c after the attack}}{\text{tot. import of c before the attack}}$$

with N the number of countries.

In particular we are interested in the response of the network to the removal of links. The Robustness has been tested for different kinds of attack removal strategies[4]:

- **Rand**: links are randomly removed. This represent the possibility of links failure in the network.
- **Strong**: links are removed in decreasing order of weight, i.e. links with higher weight are removed first and it represents an attack directed to strong links.

- **BC**: links are removed according their betweenness centrality (BC). The betweenness centrality of a link accounts the number of shortest paths from any couple of nodes passing along that link. Links with higher betweenness centrality are deleted first.
- **BCw**: links are removed according to their weighted betweenness centrality (BCw), i.e. links with higher BCw are deleted first.

2.4.1 Robustness over time

Characterising the evolution of the network's Robustness over time requires the definition of a way to compare the response to failures among different networks: we assume that the Robustness of a network is well represented by f_c, the fraction of links to be removed in order to lower the LSCC(IMP) by 99%. The attacks are performed according to the most effective removal strategy for the given Robustness measure; we consider the best link removal strategy as the one able to produce the faster functioning decrease in the network. In other words, the strategy able to select most important links in the networks[4].

2.5 Rich Club

Taking inspiration from the work of Kunal Bhattacharya et al.[5] we tried to investigate the time evolution of the "Rich Club", the lobby of countries controlling half of the world agro-food market. The slice of the market of a given country is assumed to be the total value of its export. For this purpose we first counted the number of countries belonging to the Rich Club year by year. Then we studied how evolves the slice of the market controlled by the top 5 countries.

3 Results and Discussion

3.1 Network properties

In the present section we considered a period from 2012 till to 2015 in order to calculate the metrics referred in the previous section. All networks have been considered as weighted direct network. For the sake of simplicity, the analysed networks are focused consider the livestocks commodity and its trading flows.

3.1.1 Network topology

Table 1 summarise the results obtained based on different algorithms implemented in the `networkx` and `Gephi` libraries.

Metric	2012	2013	2014	2015
$G_{year}(N, L)$	$G_{2012}(192, 3618)$	$G_{2013}(198, 3835)$	$G_{2014}(198, 3850)$	$G_{2015}(194, 3922)$
Average degree	$18,84$	$19,37$	$19,44$	$20,22$
Diameter	5	5	5	5
Density	$0,10$	$0,10$	$0,10$	$0,10$
Strongly Cntd Components	44	47	48	37
Weakly Cntd Components	1	1	1	1
In-Degree Centrality	USA-FRA-NLD	USA-FRA-NLD	NLD-USA-DEU	FRA-NLD-USA
Closeness Centrality	USA-FRA-NLD	USA-FRA-NLD	NLD-USA-DEU	FRA-NLD-USA
Betweenness Centrality	CAN-ARE-LKA	SGP-CAN-ZAF	CAN-THA-SGP	ZAF-THA-IND
Eigenvector Centrality	NLD-USA-FRA	NLD-DEU-FRA	NLD-USA-DEU	FRA-NLD-USA
PageRank Centrality	DEU-USA-NLD	DEU-NLD-USA	NLD-DEU-USA	DEU-NLD-USA
Total Trade (U$S millions)	20.523	21.192	22.845	20.868

Table 1: Agro-Food Trade Networks metrics from 2012 to 2015 (livestock commodity).

We analyzed how the graph evolved on a global scale over a period of four years. We see that between 2012 and 2015 the number of edges increased somewhat steadily but the number of nodes remains constant. This would confirm the hypothesis that world trade economy follows the globalisation process, the fact that nodes (countries) are the same is a particular characteristic of the livestock commodity, the traditional players of this sector are the same. Furthermore, the total amount of trade in a yearly basis confirm the steady grow of the sector.

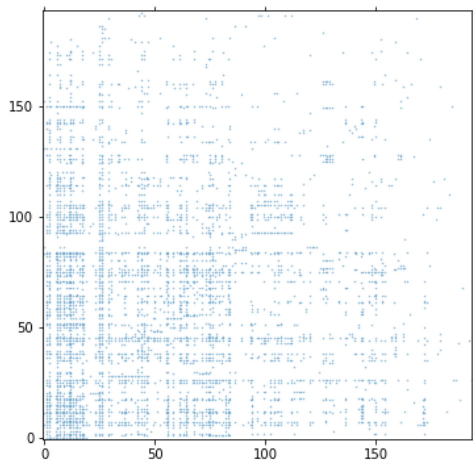


Figure 2: Adjacency Matrix of the WTN 2015, commodity 1 (Meat)

3.1.2 Adjacency Matrix

Figure 2 shows the adjacency matrix used to construct the network we envisaged (2015 and livestocks commodity). We observe that the matrix is non-symmetric and sparse, this layout is typical of weighted directed (real) networks.

3.1.3 Degree Distributions

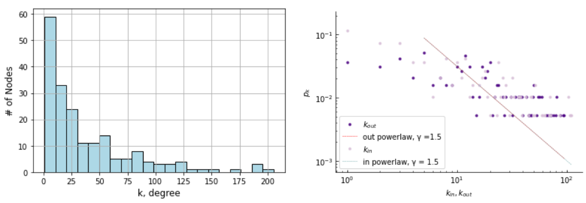
Figure 3 illustrates the degree distributions corresponding to the trade network of livestocks commodity over the period 2012 to 2015. We see that all networks present a heavy-tailed distribution which confirm the presence of the scale-free property.

3.1.4 Power Law

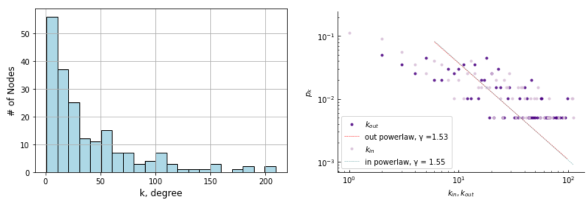
Figure 3 illustrates the in and outgoing degree distributions corresponding to the trade network of livestocks commodity over the period 2012 to 2015. For each case, we determined the γ parameter in order to estimate the power-law function. The fitted lines are showed in the same chart.

3.1.5 Centrality measures

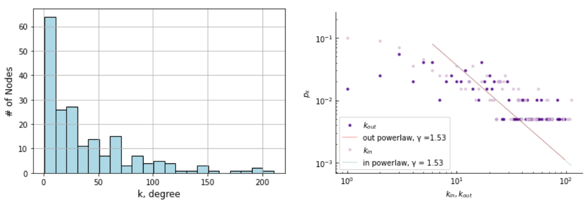
Figure 4 shows the top five countries by its correspondent centrality measurement. In general we can observe that the position of the countries change from one measure to the other as expected given the different definitions of centrality. However there is a group of countries that appear repeatedly in the classifications as Germany, France, the United States of America and the Netherlands among others.



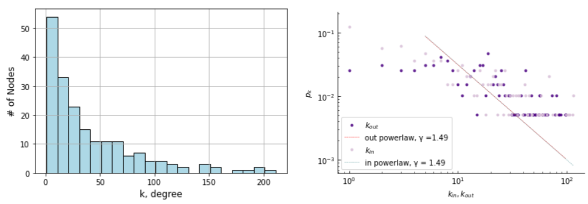
(a) WTN 2012



(b) WTN 2013



(c) WTN 2014



(d) WTN 2015

Figure 3: Degree distributions of the WTN from 2012 to 2015 (commodity 1)

The in degree centrality can be directly interpreted as the top 5 countries that import the most from others countries, here is important to recall that the analysis was done for the year 2015 and commodity livestock. So the countries that have imported the most livestock (in total value) were France and the Netherlands. On the other hand if we look at the out degree centrality the country that dominated the exports of living animals was the USA. We can see it also in the Centrality map in Figure 5b for in degree centrality and in Figure 5c for out degree centrality.

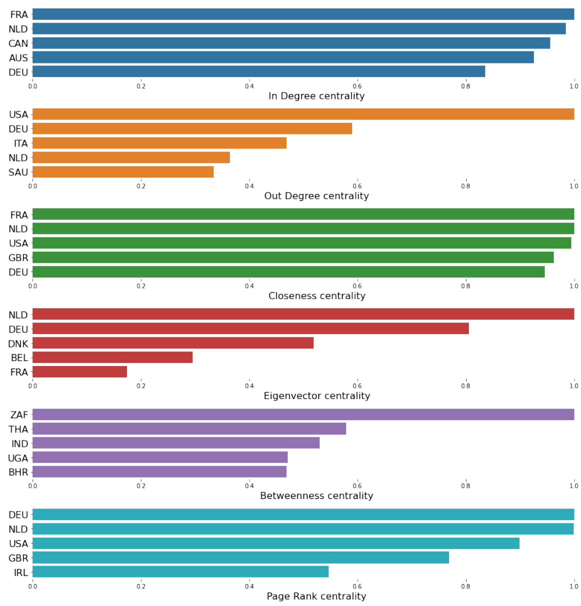
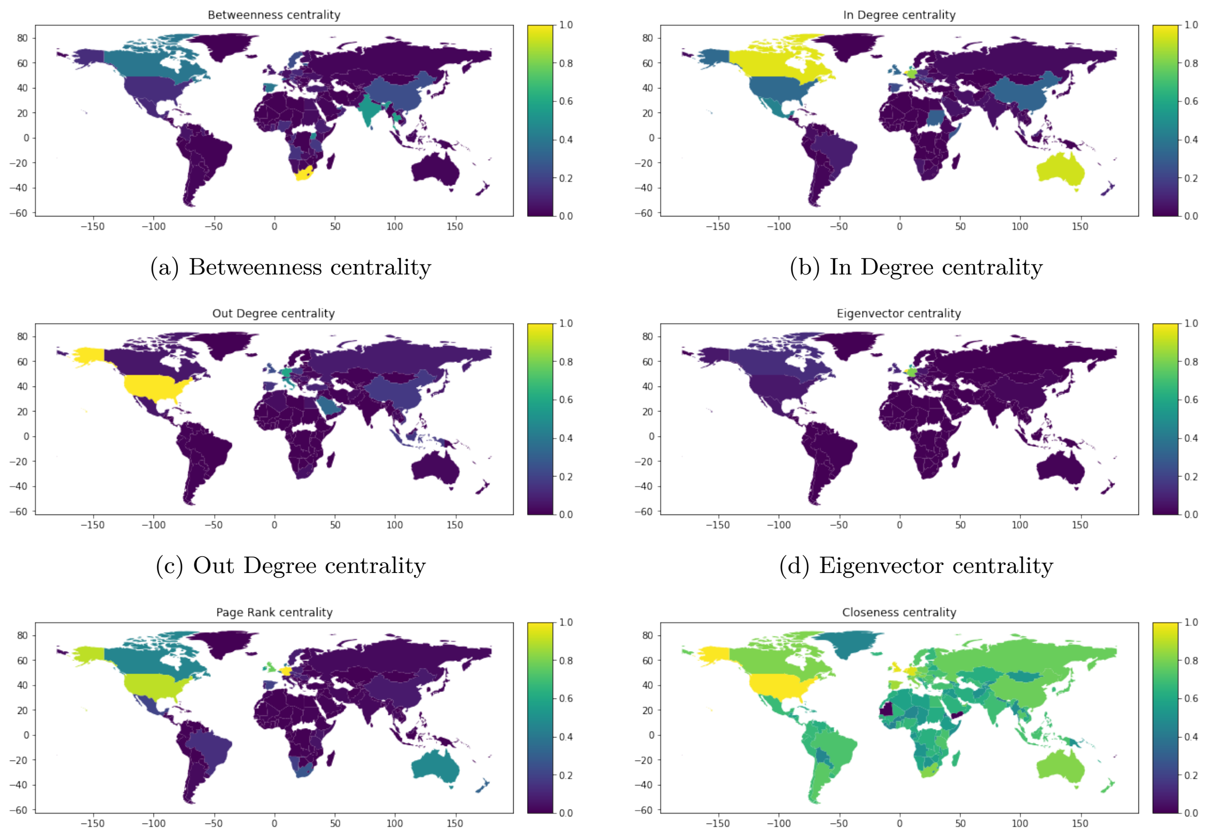


Figure 4: Top 5 countries by centrality.

A similar analysis can be done with the other centrality measures, a country with high closeness centrality can be interpreted in this context as a country who has many close collaborators, given this explanation it results evident why almost all the countries with high closeness centrality are European as seen in Figure 4, however it is also interesting to observe the map in Figure 5f where we can see that almost all the nodes have a high value in this metric, something expectable in such a globalised world.



(a) Betweenness centrality

(b) In Degree centrality

(c) Out Degree centrality

(d) Eigenvector centrality

(e) Page Rank centrality

(f) Closeness centrality

Figure 5: Different centrality measures for the network built for the year 2015 and commodity livestock.

Community detection algorithm	Number of clusters	Modularity
Walktrap	34	0.46
Infomap	30	0.54
Label propagation	14	0.49
Spinglass	5	0.18
Louvain Modularity	9	0.11
Louvain Surprise	156	-0.005

Table 2: Different algorithms results for commodity livestock and year 2015.

3.2 Community detection

The first step in the community detection was to apply the different algorithms implemented in the IGRAPH library capable of dealing with weighted and directed graphs. All the algorithms were implemented in networks built with import information for a specific year and a specific commodity, as we have 20 years of information and 23 commodities this result in 460 implementations. For the purpose of showing an example here we add the results for the year *2015*, and commodities *livestocks*. The results over the years, and commodities added in the appendix B, were used to perform the analysis of the community evolution in time.

Table 2 shows the number of communities and the modularity found for each of the algorithms computed for the whole graph for all the algorithms by exception of *spinglass* algorithm which was applied over the giant component of the network.

The algorithms based on dynamics, *walktrap*, *infomap* and *label propagation* obtained a number of communities of the same order of magnitude, while *spinglass* and *Louvain modularity* are one order of magnitude below. We have observed this trend happening for most of the communities and most of the years, indicating that the algorithms are considering different properties to detect the communities. The last algorithm, *Louvain surprise* was incapable of finding relevant communities assigning to most of the nodes their own community, this can also be seen in the negative modularity.

In general, the trends described for the number of clusters is also observed for the modularity reached by the clusters obtained by each algorithm (Figure 6). Something to remark is the low value obtained by the *Louvain modularity* algorithm, in fact the modularity obtained is among the worsts obtained by the algorithms even though this algorithm is supposed to optimise the modularity value.

The difference in the number of communities obtained by each algorithm has an impact on the size distribution of the communities as can be observed in the KDE Figure 7. In general all the algorithms tend to detect rather small communities with a number of members between 2 to 10. However two of these algorithms, *walktrap* and *infomap* were capable of detecting communities with up to 35 members. Again, this trend was repeated along the years and commodities.

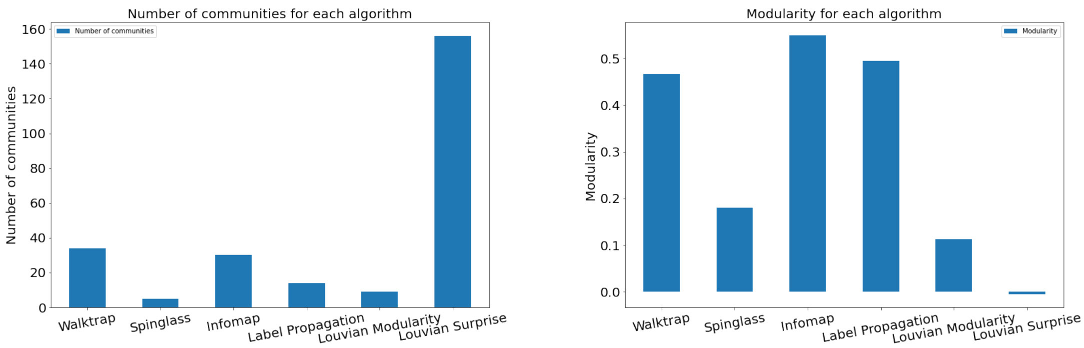


Figure 6: On the left the number of communities found by each algorithm and on the right the modularity reached by the communities found by each algorithm

In Figure 8 we have coloured the countries according to the communities they belong. We did it for the five biggest communities found by each of the algorithms we studied. Maybe the first observation that we make is that most of the clusters belong to the same geographical region; for example, Europe almost always appear in a single comunity or divided between north and south. This is interesting to denote because none of the algorithms was provided with geographical coordinates of the countries, so they are detecting it from the imports they are having. For the particular case of livestock trade it make even more sense because the trade of living animals is usually done regionally.

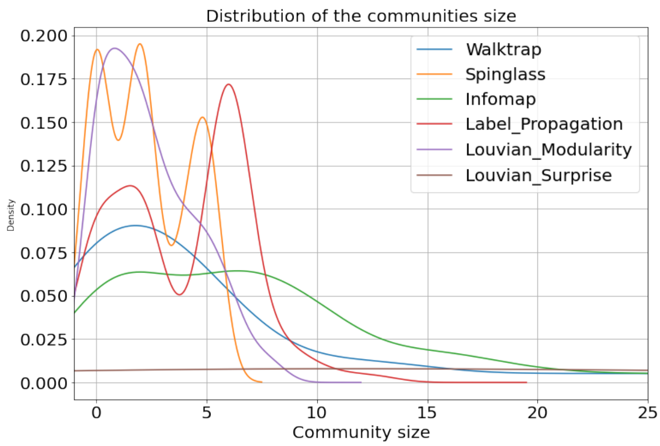


Figure 7: Size distribution of the communities.

In this graphical representation of the clusters we also observed that the communities have different members for different algorithms, which make harder to say which may be real communities. Some times a single community from one of the algorithms is displayed as two communities by other algorithms as in the case of America displayed as green by the *spinglass* and as green and purple by *infomap*. Apart from that we are also aware that none of the algorithms were used to detect overlapping communities and the separated communities are a side effect of that.

In this section we also studied the average betweenness centrality measure for each of the communities detected by the different algorithms, this was only measured in the communities containing more than one country and is displayed in Figure 9.

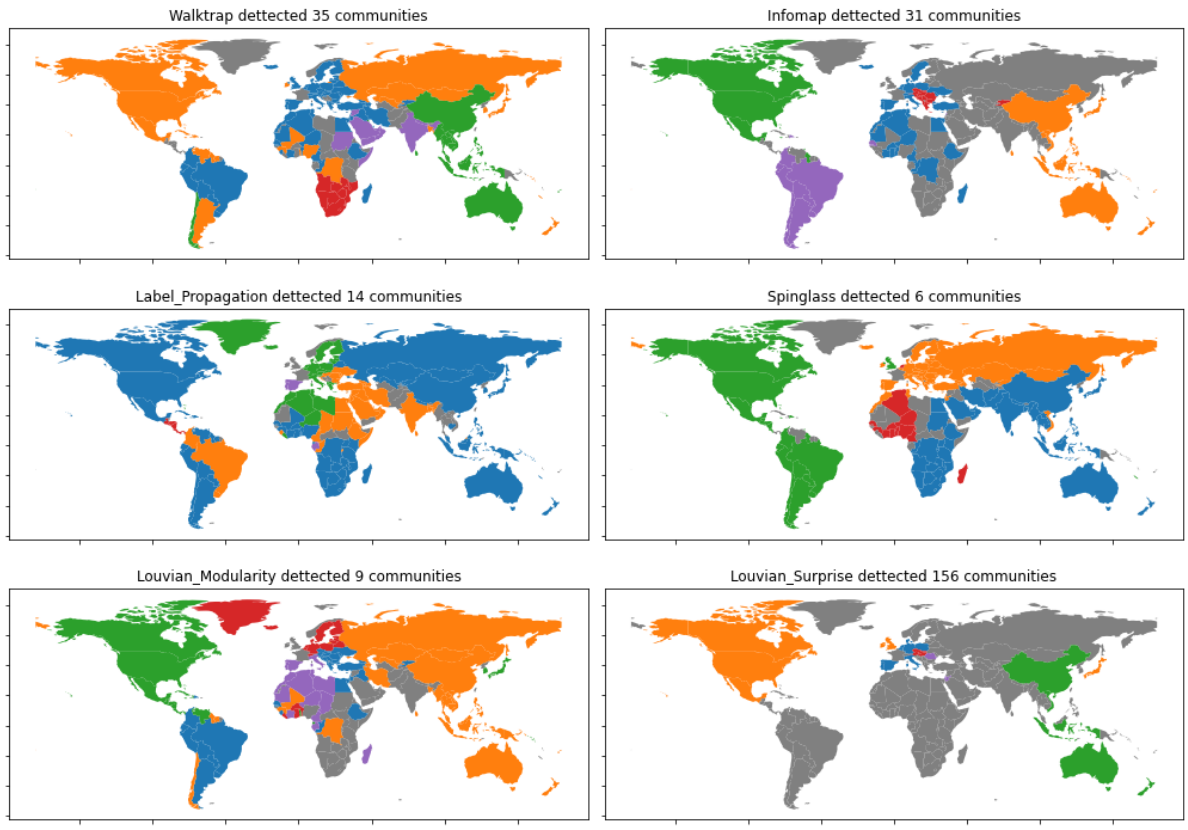


Figure 8: Geographical visualisation of the 10 biggest communities encountered by each algorithm.

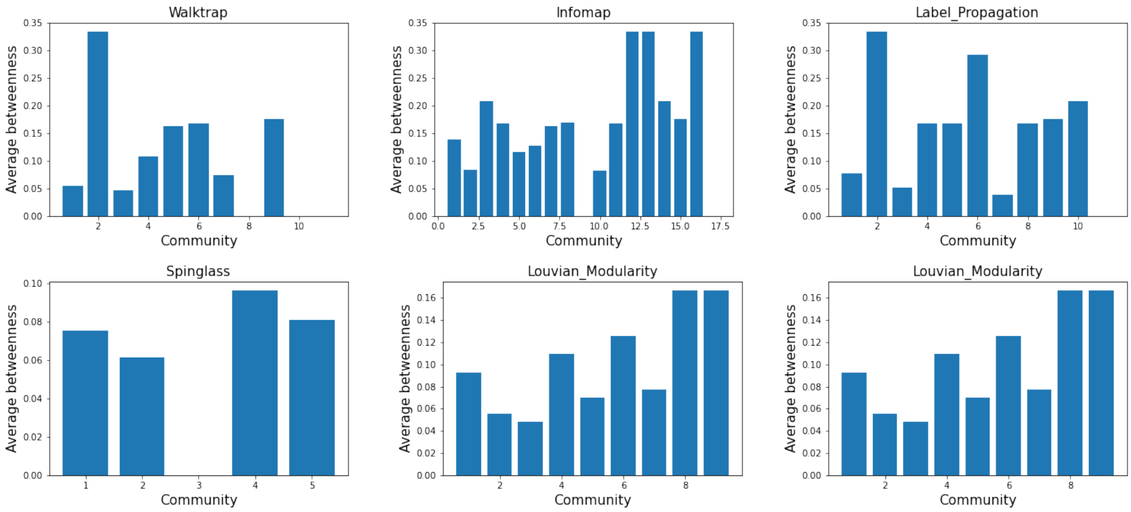


Figure 9: The barplots represent the average betweenness centrality by cluster for different algorithms.

3.2.1 Communities evolution in time

In this work we also implemented a technique to identify the evolution of the communities along the time as described in the methodology. In Figure 10 we are showing the communities that our technique associated as the same community (with more, less or the same countries) sharing the colour along the maps. For example, we can see that the southern part of the African continent was identified as a community by the *spinglass* algorithm for the year 2001 and commodity livestock; we coloured it as red and followed it along the years. In this evolution we see that some core countries remain in time, however there are a few ones that are entering and leaving this community. This kind of observations can be also done for south Asia and Oceania (green), Europe and north Asia (Blue) and America (yellow). This behaviour of having some core countries in a community and some satellites entering and leaving the communities was also observed for other commodities and also for other algorithms as *walktrap*.

3.2.2 Multi-layer communities evolution in time

From applying the 2 algorithm over the multi-layer commodity trade-network, we obtained communities for each year. Then, by means of the (cite identification algorithm), we were able to track an evolution over time for such communities.
Some remarks are worthwhile in 11:

- 2005: The communities from 2004 remain largely unaffected
- 2006: Now, most of South America is found to be in the same community as North America and the South East Asian market
- 2007: Now North America is found to be "a community on its own", while Oceania is predominantly in the European trade network
- 2011: communities are largely reestablished, save for South America, which no longer displays an endogenous trade
- We notice that North America oscillates between an endogenous community and a trans-Pacific one
- In contrast with some of the single-layer community detection algorithms, Europe (Russia included) remains mostly in the same community, although this community displays a growing number of intercontinental members.

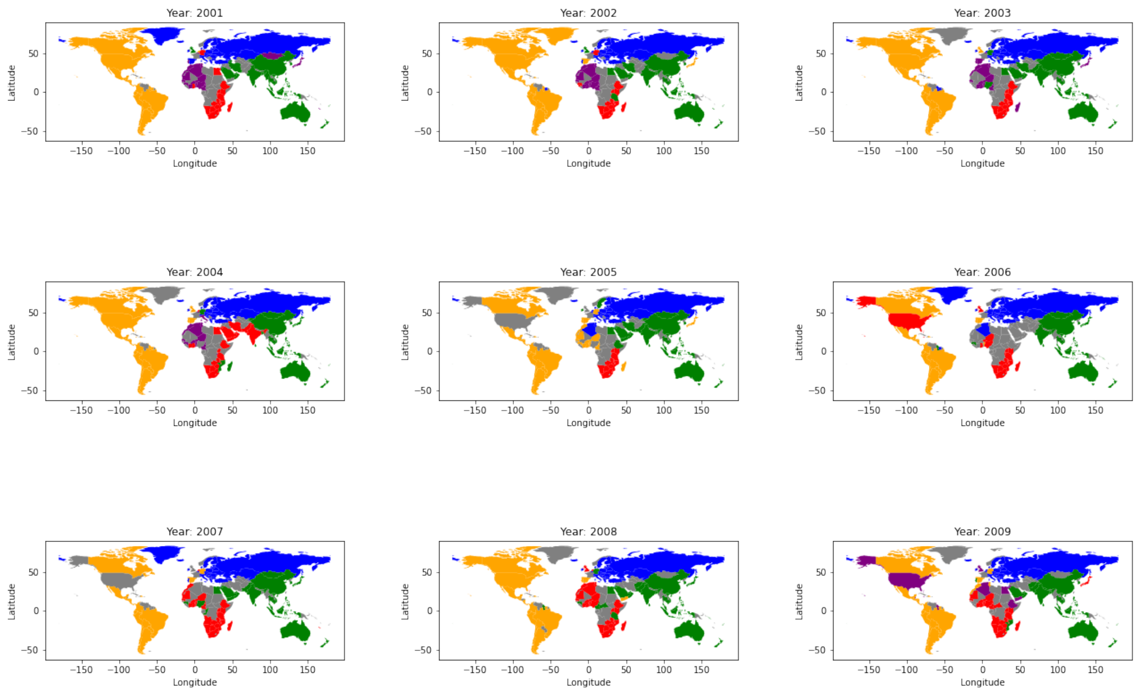


Figure 10: Evolution of the communities identified by the *spinglass* algorithm, the communities sharing colours along the years are the ones identified as the same by our sharedness parameter.

3.3 Robustness analysis

In order to study the Robustness of the Food-Trade network we consider a network built on top of the imports in a given year, for a given commodity. We repeated the analysis on the resulting 460 networks, one for every commodities and every year.

For the sake of brevity we just report the results of such analysis for the year 2015, in Figure 12 for the "livestocks" commodity and in the appendix A for all the remaining. It can be clearly seen that all the networks shares similar responses to attacks.

Measuring the Robustness with the LSCC we observe that Random and Strong attacks are not very effective, leading to a $f_c \simeq 1$, while the attacks based on the betweenness centrality of the links perform better; these results in agreement with [4]. For the LSCC metric the effectiveness of the removal strategy based on betweenness are comparable, in most cases the most effective strategy was **BCw**.

Considering instead the IMP measure it can be seen that the Random, BC and BCw attacks performs almost exactly the same, badly, with a $f_c \simeq 1$, the Robustness is proportional to the residual links in the network. In this case the most effective removal strategy is the **Strong** one.

The IMP measure that we have introduced helps us to get an "economic" interpretation of the Robustness: for the example reported in the Figure 12 it can be seen that removing the 10% of the most heavy links, the average import per country drops to the 20% of its maximum volume. That is to say that if remove the 10% of links in descending order of weight, the total goods that a country imports drops on average by the 80%.

This reflect the heterogeneity in the distributions of the links' weights.

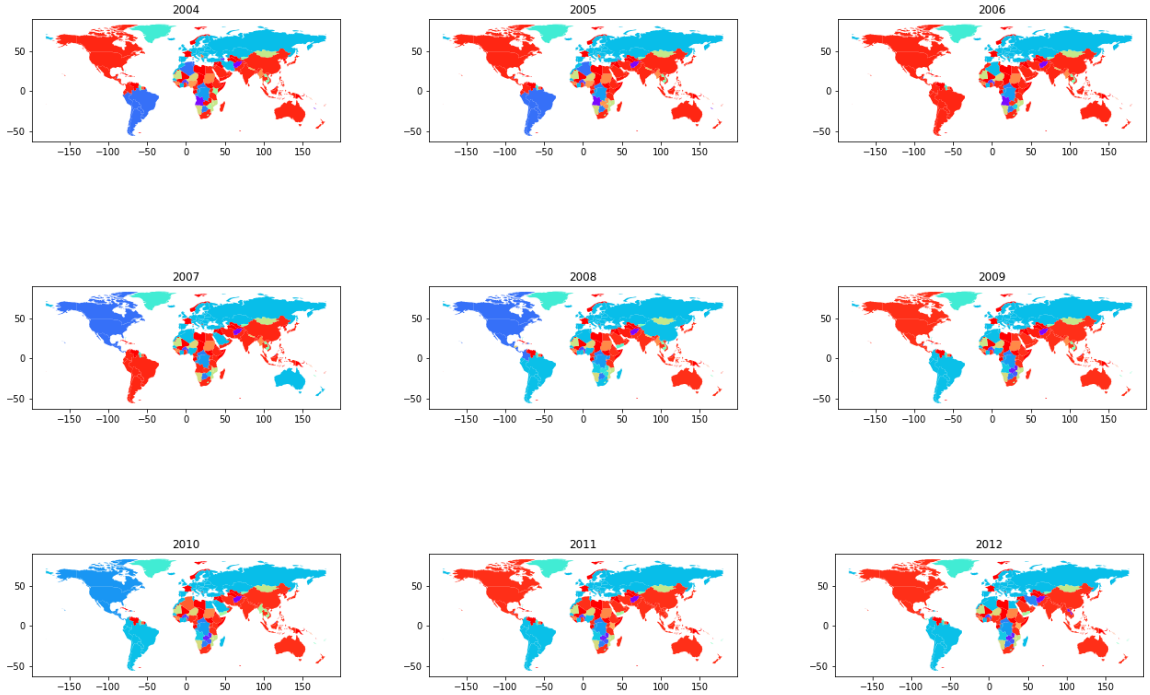


Figure 11: Time evolution of the multi-layer communities found with the MNLPA algorithm

3.3.1 Robustness over time

The study of the evolution of the Robustness over time consist of looking how f_c vary for the network for a given commodity throughout the years. The critical fraction of links to be removed from the network is calculated performing **BCw** attacks to test the LSCC and **Strong** to test the IMP.

The results, for every commodities, are shown in Figure 13. It is clear to notice an overall decreasing trend. Just four commodity networks increased their f_c over time for the LSCC metric: this can be interpreted as the fact the most of the commodity networks are becoming less LSCC Robust over time, In other words nowadays to "break" the network we need to remove less links than 20 years ago.

For the IMP Robustness analysis just one commodity network increased its f_c over the years. In this case is even more clear the trend: 20 out of 21 community networks are becoming less IMP Robust over time. This result can be interpreted as the fact that the stronger link are becoming even more important, so that their removal leads to heavier imports reduction.

3.4 Rich Club

It is notable to remark that for this analysis we consider the aggregated Food-Trade network, consisting of all the trades between countries regardless the commodity. On the left of Figure 14 we can see how changes in time the number of countries belonging to the Rich Club, for various slice of the total market. We can clearly see that there is an increasing trend, for which over the last 20 year the number countries consisting of the Rich Club is increased.

From the plot on the right of the Figure 14 we can appreciate that the total market of the agro-food trade network has growth over the last 20 year. At the same time it can be seen that the relative slice of the market controlled by the top 5 countries is instead shrinking over time.

The two results are in agreement: the members of the Rich Club is increasing over time, in fact the portion of the market controlled by the top 5 country is instead decreasing because new countries entered Rich Club controlling half of the agro-food trades.

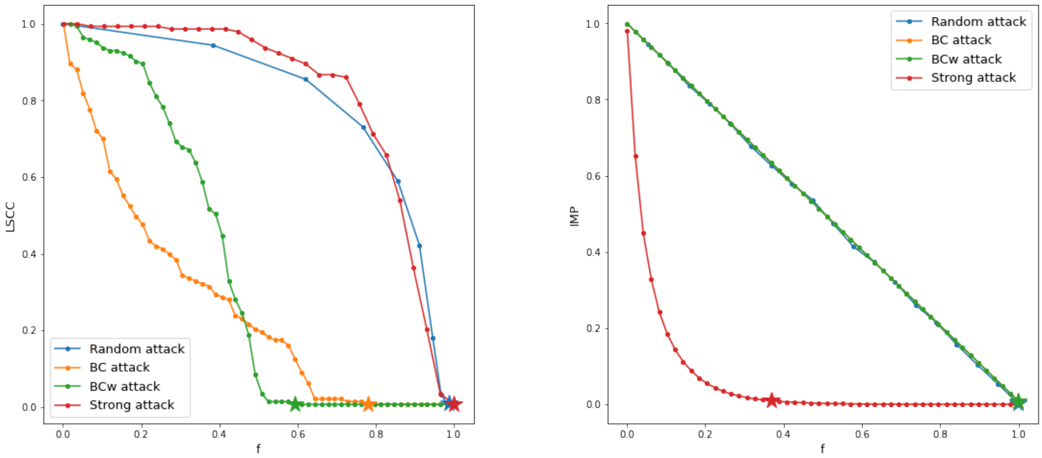


Figure 12: Results of Robustness analysis for the "livestocks" trade-network of 2015. On the left the Robustness has been measured by the means of LSCC; on the right has been empoyed the IMP metric. In both cases it has been performed 4 attack strategies: Random, Strong, BC, BCw. The stars are the marker for the critical fraction f_c, found with a given attack strategy.

4 Conclusion

We successfully calculated the typical measures for properties and structure of the obtained networks. We remark that, although moderately, a scale free behaviour was displayed by the networks. This is however, not conclusive, as the number of nodes in the network is small. Further tests can be performed by setting different networks based on different commodities.
For the centrality measures, the often quoted key-players in global trade are indeed confirmed, whereas for the community detection, we find it to be a bit sensitive to the algorithm, as well as the specific commodity. Given we are dealing with agro-food commodities, some of the commodities, such as livestock and meat, are not as easily imported from certain countries, and thus hinder the formation of communities outside geographic proximity.
When a multi-layer model is considered, the communities that arise coincide with trade treaties, but exhibit the presence of certain trade flows that may not be so endogenous to such treaties, such as France and Sweden not within the same community as the rest of Europe.
Following the sharedness as a measure to map communities between years, we track an evolution of communities in both the multi-layer and single-layer networks. The oscillation of certain groups of countries between a given community and another hints on the need for a shared-communities approach. Furthermore, potential periodic trends may be observed; an analysis over a longer period and a more finely tuned community mapping strategy could lead to insightful characterisations.
The LSCC Robustness analysis supports the hypothesis the agro-food network exhibits the typical behaviour of a scale free network, with the critical fraction f_c near 1 for random attacks. We saw that for most of the commodities the robustness of the network is lowering over time. A possible interpretation is that the heterogeneity in the trades is increasing. Lastly, it has been shown that the the previously defined "Rich Club" is shrinking over time, this could reflect the fact that the market is diversifying.

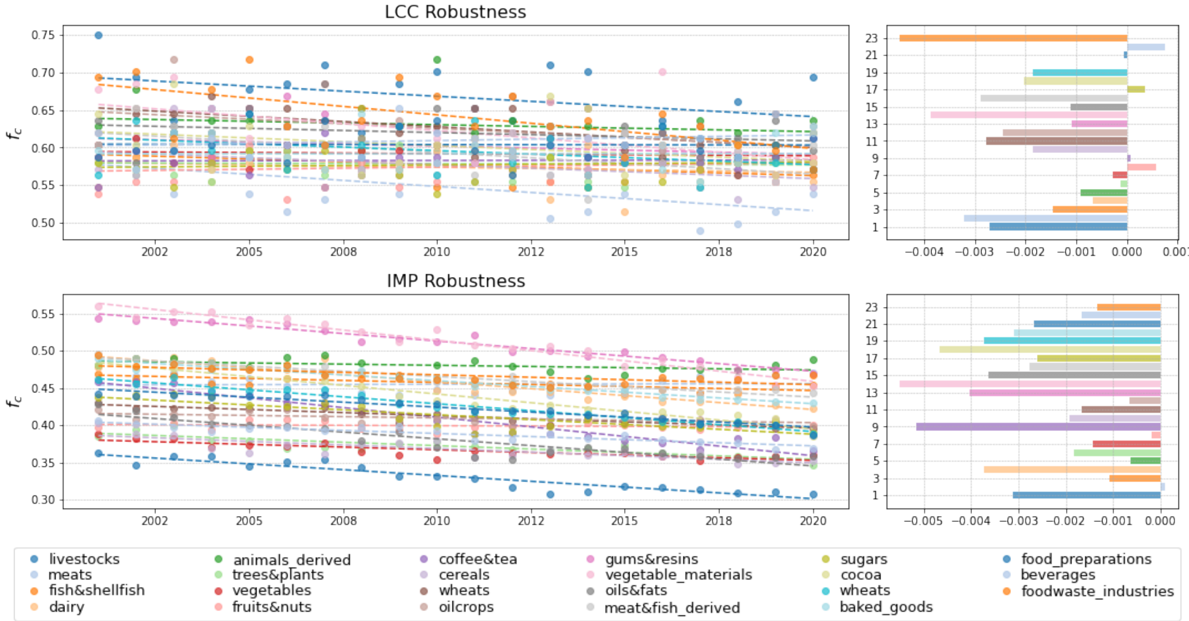


Figure 13: The plots on the left are display the trends for f_c over 20 years for every commodity trade-network by the means of coloured dots. The dashed lines are the linear fit of every time series. On the barplots are reported the slopes of the fitting curves, to quantify the Robustness variations for each commodity. The plots on the top refers to the LSCC Robustness analysis while the ones at the bottom refer to the IMP metric.

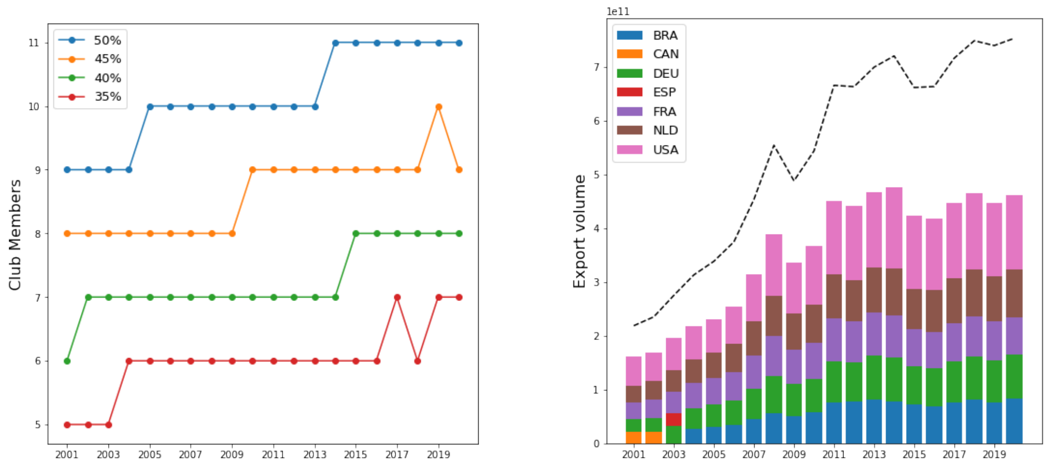


Figure 14: On the left is shown how change the size of the Rich Club over the year, for various threshold of the Rich Club membership: we consider the countries that controls respectively the 50%, 45%, 40% and 35% of the Agro-Food Trades. On the right is displayed the volume of the total export per year, for the top 5 richest countries. The dashed line marks the 50% of the total market. It is easy to see that over the years the top 5 country controls a smaller percentage of the market.

References

[1] Fatemeh Alimadadi, Ehsan Khadangi, and Alireza Bagheri. "Community detection in facebook activity networks and presenting a new multilayer label propagation algorithm for community detection". In: *International Journal of Modern Physics B* 33.10 (2019), p. 1950089. DOI: `10.1142/S0217979219500899`. eprint: `https://doi.org/10.1142/S0217979219500899`. URL: `https://doi.org/10.1142/S0217979219500899`.

[2] Albert-László Barabási and Pósfai Márton. "Network Robustness". In: *Network Science*. Cambridge, United Kingdom: Cambridge University Press, 2016, pp. 00–18.

[3] Marya Bazzi et al. "Community Detection in Temporal Multilayer Networks, with an Application to Correlation Networks". In: *Multiscale Modeling & Simulation* 14.1 (Jan. 2016), pp. 1–41. DOI: `10.1137/15m1009615`. URL: `https://doi.org/10.1137/15m1009615`.

[4] M. Bellingeri et al. "A comparative analysis of link removal strategies in real complex weighted networks". In: *Scientific Reports* 10.1 (Mar. 2020), p. 3911. ISSN: 2045-2322. DOI: `10.1038/s41598-020-60298-7`. URL: `https://doi.org/10.1038/s41598-020-60298-7`.

[5] Kunal Bhattacharya et al. "The International Trade Network: Weighted Network Analysis and Modelling". In: *arXiv.org, Quantitative Finance Papers* (July 2007). DOI: `10.1088/1742-5468/2008/02/P02002`.

[6] Vincent D Blondel et al. "Fast unfolding of communities in large networks". In: *Journal of statistical mechanics: theory and experiment* 2008.10 (2008), P10008.

[7] Min Gon Chung et al. "The spatial and temporal dynamics of global meat trade networks". In: *Scientific reports* 10.1 (2020), pp. 1–10.

[8] Célestin Coquidé et al. "Influence of petroleum and gas trade on EU economies from the reduced Google matrix analysis of UN COMTRADE data". In: *The European Physical Journal B* 92.8 (2019). ISSN: 1434-6036. DOI: `10.1140/epjb/e2019-100132-6`. URL: `http://dx.doi.org/10.1140/epjb/e2019-100132-6`.

[9] Emanuele Cozzo et al. *Multiplex Networks*. Springer International Publishing, 2018. DOI: `10.1007/978-3-319-92255-3`. URL: `https://doi.org/10.1007/978-3-319-92255-3`.

[10] Luca De Benedictis et al. "Network analysis of world trade using the BACI-CEPII dataset". In: *Global Economy Journal* 14.3-4 (2014), pp. 287–343.

[11] Mária Ercsey-Ravasz et al. "Complexity of the international agro-food trade network and its impact on food safety". In: *PloS one* 7.5 (2012), e37810.

[12] Aric Hagberg, Pieter Swart, and Daniel S Chult. "Exploring network structure, dynamics, and function using networkx". In: (). URL: `https://www.osti.gov/biblio/960616`.

[13] Charles R. Harris et al. "Array programming with NumPy". In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: `10.1038/s41586-020-2649-2`. URL: `https://doi.org/10.1038/s41586-020-2649-2`.

[14] Xinyu Huang et al. "A survey of community detection methods in multilayer networks". In: *Data Mining and Knowledge Discovery* 35.1 (2021), pp. 1–45. ISSN: 1573-756X. DOI: `10.1007/s10618-020-00716-6`. URL: `https://doi.org/10.1007/s10618-020-00716-6`.

[15] Matthew O. Jackson. *Social and Economic Networks*. USA: Princeton University Press, 2008. ISBN: 0691134405.

[16] Neo4j. "Label Propagation". In: *https://neo4j.com/docs/graph-data-science/current/algorithms/label-propagation* ().

[17] Pascal Pons and Matthieu Latapy. "Computing communities in large networks using random walks". In: *International symposium on computer and information sciences*. Springer. 2005, pp. 284–293.

[18] Usha Nandini Raghavan, Réka Albert, and Soundar Kumara. "Near linear time algorithm to detect community structures in large-scale networks". In: *Physical review E* 76.3 (2007), p. 036106.

[19] Jörg Reichardt and Stefan Bornholdt. "Statistical mechanics of community detection". In: *Physical review E* 74.1 (2006), p. 016110.

[20] Martin Rosvall and Carl T Bergstrom. “Maps of random walks on complex networks reveal community structure”. In: *Proceedings of the national academy of sciences* 105.4 (2008), pp. 1118–1123.

[21] M ngeles Serrano, Marián Boguñá, and Alessandro Vespignani. “Patterns of dominant flows in the world trade web”. In: *Journal of Economic Interaction and Coordination* 2.2 (2007), pp. 111–124.

[22] Stefania Tamea et al. “Virtual water trade and water footprint of agricultural goods: the 1961–2016 CWASI database”. In: *Earth System Science Data* 13.5 (2021), pp. 2025–2051.

[23] Sofia Torreggiani et al. “Identifying the community structure of the food-trade international multi-network”. In: *Environmental Research Letters* 13.5 (2018), p. 054026.

[24] Vincent A Traag, Rodrigo Aldecoa, and J-C Delvenne. “Detecting communities using asymptotical surprise”. In: *Physical Review E* 92.2 (2015), p. 022816.

A Appendix I : Table of Commodities

Code	Long Description	Short desc.
1	Animal; live	livestocks
2	Meat and edible meat offal	meats
3	Fish and crustaceans, molluscs an other aquatic invertebrates	fish&shellfish
4	Dairy produce; bird's eggs; natural honey; edible products of animal origin	dairy
5	Animal originated products; not elsewhere specified or included	animals_derived
6	Trees and other plants, live; bulbs, roots; flowers and ornamental foliage	trees&plants
7	Vegetables and certain roots and tubers; edible	vegetables
8	Fruit and nuts, edible; peel of citrus fruit or melons	fruits&nuts
9	Coffee, tea, mate and spices	coffee&tea
10	Cereals	cereals
11	Products of the milling industries; malt, starches, inulin, wheat gluten	wheats
12	Oil seeds and oleaginous fruits; miscellaneous grains, seed and fruit;	oilcrops
13	Lac; gums, resins and other vegetable saps and extracts	gums&resins
14	Vegetable plaiting materials; vegetable products not elsewhere specified	vegetables_materials
15	Animal or vegetable fats and oils and their cleavage products; prepared fats;	oils&fats
16	Meat, fish or crustaceans, molluscs or other invertebrates; preparations thereof	meat&fish_derived
17	Sugars and sugar confectionery	sugars
18	Cocoa and cocoa preparations	cocoa
19	Preparations of cereals, flour, starch or milk; pastrycook's products	sweet_wheats
20	Preparations of vegetables, fruit, nuts or other part of plants	baked_goods
21	Miscellaneous edible preparations	food_preparations
22	Beverages, spirits and vinegar	beverages
23	Food industries, residues and waste thereof; prepared animal fodder	foodwaste_industries

Table 3: Agro-Food Commodities

B Appendix II : Community detection results

	Walktrap		Spinglass		Infomap		Label prop		Modularity		Surprise	
	I	II	I	II	I	II	I	II	I	II	I	II
Meats	40	0.4	4	0.2	13	0.4	19	0.4	6	0.1	172	0
Fish&shellfish	36	0.3	5	0.1	2	0	4	0.3	6	0	161	0
Dairy	49	0.3	3	0.1	18	0.3	15	0.2	9	0	161	0
Animals__derived	26	0.3	4	0.14	0	0	9	0	6	0	157	0
Trees&plants	44	0.3	4	0.1	0	0	16	0.2	5	0	180	0
Vegetables	33	0.5	3	0.1	23	0.5	8	0.3	6	0	192	0
Fruits&nuts	43	0.3	3	0.1	11	0.3	3	0	6	0.1	169	0
Coffee&tea	219	0	3	0.1	3	0	9	0.1	6	0	165	0
Cereals	15	0.3	2	0.1	12	0.3	23	0	9	0.1	145	0
Wheats	49	0.4	3	0.1	13	0.4	14	0.1	8	0.1	142	0
Oilcrops	86	0	4	0.1	4	0	7	0.2	6	0.1	199	0
Gums&resins	42	00	3	0	0	0	22	0	5	0	167	0
Vegetables__materials	48	0.2	3	0.2	0	0	26	0.2	6	0.1	163	0
Oils&fats	31	0.3	3	0.1	1	0	7	0	9	0.1	164	0
Meat&fish__derived	42	0.4	3	0.1	11	0.4	15	0.3	6	0.1	165	0
Meat&fish__derived	24	0	3	0.1	12	0.4	14	0.3	8	0	136	0
Sugars	50	0.2	3	0.1	3	0.1	14	0.2	5	0.1	165	0
Cocoa	37	0.3	3	0.1	26	0.4	11	0.4	6	0.1	162	0
Sweet__wheats	33	0.3	3	0.1	10	0.3	10	0.3	6	0.1	170	0
Baked__goods	39	0.4	3	0.1	17	0.3	11	0.3	7	0.1	167	0
Food__preparations	86	0.1	3	0.1	2	0	11	0.1	7	0	181	0
Beverages	37	0.3	4	0.2	2	0	13	0.3	6	0.1	155	0

Table 4: Community detection results for all the products in the year 2015. Column I is the number of communities found and column II is the modularity.

C Appendix III : Robustness analysis results

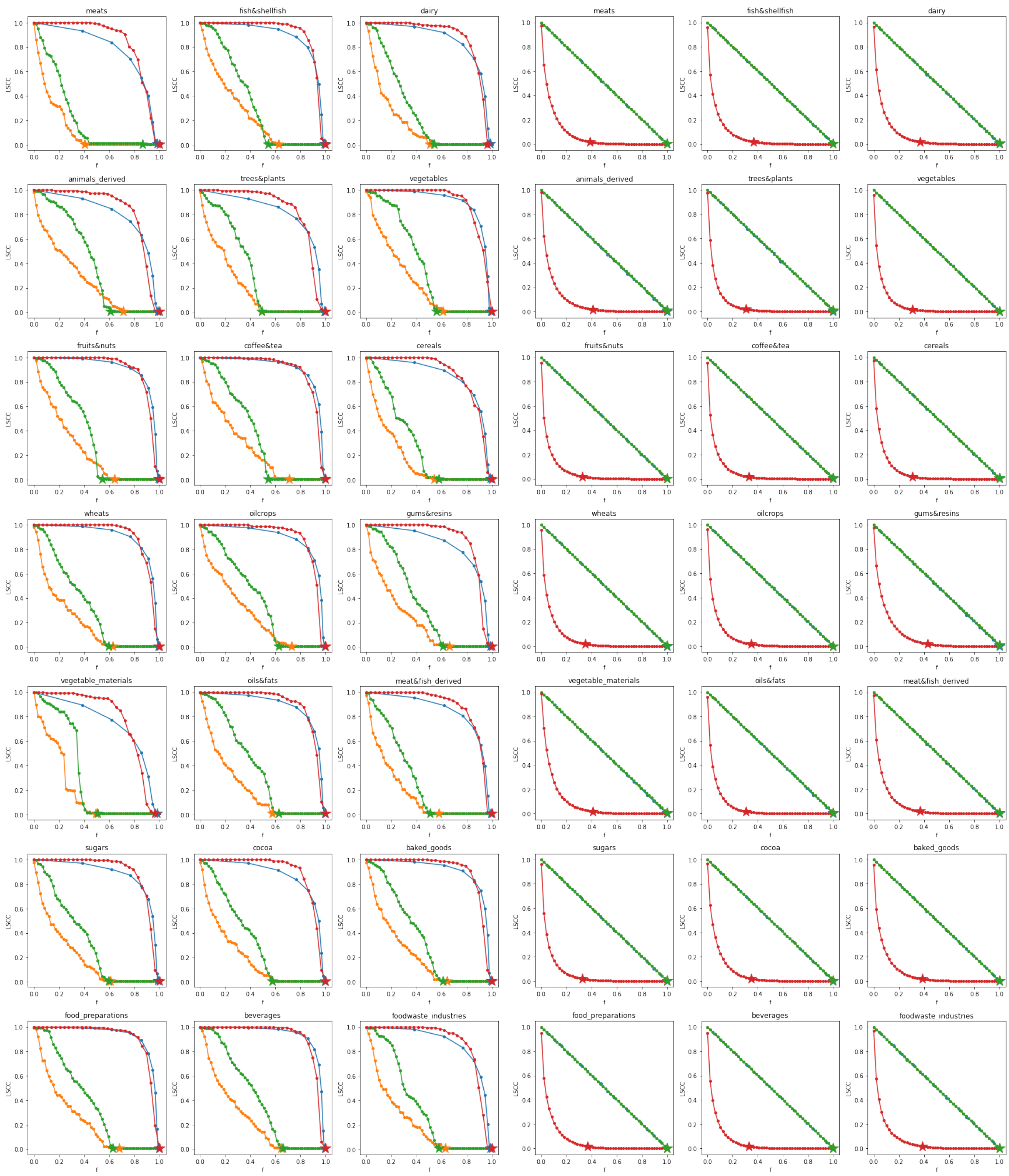


Figure 15: Results of Robustness analysis for the various commodities' trade-network of 2015. On the left the Robustness has been measured by the means of LSCC; on the right has been empoyed the IMP metric. In both cases it has been performed 4 attack strategies: Random, Strong, BC, BCw. The stars are the marker for the critical fraction f_c, found with a given attack strategy.