POLITECNICO DI TORINO

M.Sc. in Engineering and Management

Master's thesis

Network analysis of interconnections between economic sectors: the Italian economy and comparison with the main European and world systems

Advisor:
Franco Varetto

Candidate:
Francesca Prato

Academic Year 2019/2020

Contents

Introduction

An economy is made by sectors and their interconnections, understand the nature of these connections is important to comprise how the system works, in fact, the structure of a system determines its functionality. The input-output tables, provided by national statistical institutes, characterize the structure of an economy at the sector level, tracking both direct and indirect supply-demand interdependencies among sectors within the economy. However, packing all the information together in a matrix may limit the possibility of investigating structural features of the economy at the sectoral level.

This work analyzes each national input-output table as a complete and integrated system making use of network science tools.

Network analysis emerges as a data-driven approach for characterizing the structure of network-like complex systems to understand the causality between the system's structure and its functionality; systems frequently studied are for instance social networks, biological and metabolic networks and transportation networks.

In the literature little attention has been paid to study the economy structure as a network, the first paper on this subject was proposed in 1978 by P. Slater, where the author focused on the study of the United States economic structure to identify sector clusters. However, the literature went in silence until the new century. Recent studies on the subject could be classified in three branches. First, the position of the sectors in the networks has been ranked according to different centrality measures. Second, the sectors have been partitioned in clusters with a higher amount of economical interactions. The third area of study is linked to the performance of the economies, it deals with the resilience to the repercussion of a shock within the considered network and with the development of economies.

This paper aims at understanding the Italian economy structure and which sectors play a crucial role in the network.

Chapter 1 introduces the basic concept of the network math, in order to provide the reader the basis to follow the development of the thesis. Chapter 2 describes the history of the input output tables, their economical interpretation and the conversion from a matrix system to a graph system. Chapter 3, 4 and 5 deal instead with the proper network analysis.
At first, the features shared by all the national graphs considered are presented, in order to show the structure of the graphs in question and thus address the field of investigation.
Following, the techniques of analysis used to investigate the Italian economy have been introduced in detail. The Italian network, as all the national graph derived, is weighted and directed, so the most complex graph system possible, there are plenty of measures that try to adapt the unweighted analysis to our case, in this paper it has been chosen to apply more traditional techniques and relatively new ones that seem to take into account the economical interpretation of the network. Chapter 4 is developed around the analysis of the Italian national graph isolated from the other economies, it provides sector rankings according to different measures and studies their evolutions in period 2000-2014.
Finally, sector 5 takes into consideration the other national graphs derived and compares them to Italy; moreover, in this section a global analysis is presented to understand the Italian industries impact on the whole world network.

Chapter 1

Network Science

1.1 Definition of Network Science

Network science is an empirical subject which studies network models. Network theory core assumption is that a cause, an effect, or an association between objects could be modelized as a network, where distinct elements are represented by nodes (or vertices) and connected by links (or edges).

One can't ignore that networks are part of everyday life: the supply chain network, the transportation network, the telephone network, Internet and the banking network are just some of all the networks that characterize our reality. The famous sentence "networks are everywhere" refers, not only, to the possibility of considering many things in the world in relational terms, but it also states that scientific explanation to several problems is aided by abstraction to such a connected representation.

Network science's origins date back to 1736 when Leonhard Euler, a swiss mathematician, faced and solved the problem of how to best circumnavigate the bridges of Königsberg. Its history, according to "Network Science: theory and applications" by Ted G. Lewis, could be divided into three periods: pre-network period (1736-1966), when it was purely the mathematic of graphs; the meso-network period (1967-1998) when applications of networks were emerging from the research literature; and the modern period (1998 - present) when network science became "the new science of networks" and enjoyed a great diffusion. In the last two decades, complex networks reached the center of research interest thanks to an unprecedented computing power, massive data sets and new computational modelling techniques available. The new popularity is proved by the enormous number of citing papers.

1.2 Graph Theory: definitions and concepts

1.2.1 Fundamentals

Graph

A graph is especially a way to describe a relation between the element of a system. As in any mathematical abstraction, describing a phenomenon using a graph representation puts focus on few peculiarities of interest, while many others aren't considered.

Definition 1.1 *A graph G is an ordered pair* $(V(G), E(G))$ *consisting of a finite set of vertices (also called points or nodes)* $V = \{v1, v2, v3, ...\}$*, a set of edges (also called links or arcs)* $E = \{e1, e2, e3, ...\}$*, together with an incidence function* Ψ_G *that associates with each edge of G an unordered pair of (not necessarily distinct) vertices of G.*

Usually the graph is denoted as $G = (V, E)$ where $E \subseteq V \otimes V$, whose elements are not necessarily all distinct.
E is **reflexive** if $(v, v) \in E$ for all v in V, it is anti-reflexive if $(v, v) \notin E$ for all v in V and it is **symmetric** if $(v1, v2) \in E \rightleftharpoons (v2, v1) \in E$.
Let $\{u, v\}$ be an edge, it may be represented as uv or vu. If $e = uv$ is an edge of graph G, it can be said that u and v are **adjacent** (or neighbors) in G, that u and v are **incident** with edge e and that e **joins** u and v.

Isomorphism

Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are said to be **isomorphic** to each other if there is a one-to-one correspondence between their vertices and edges such that any two vertices are adjacent in G_1 if and only if their images in the correspondence are adjacent in G_2: the incidence relationship is preserved. The isomorphism relation is written as $G_1 = G_2$ or $G_1 \cong G_2$.
The required conditions for two graphs to be isomorphic are: same number of

vertices, same number of edges, equal number of vertices with the same degree and same degree sequences. To show that graphs are not isomorphic is sufficient to find a property in one of them that is missing in the other.
In math terms

Definition 1.2 *Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs. A function $f : V_1 \to V_2$ is called a graph isomorphism if*

- *f is a one-to-one function*
- *for all $a, b \in V_1$, $\{a, b\} \in E_1$ if and only if $\{f(a), f(b)\} \in E_2$.*

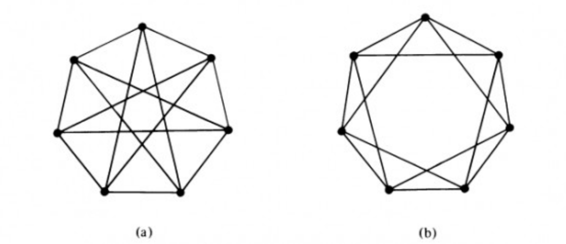


Figure 1.1: Two isomorphic graphs

1.2.2 Directed graphs, undirected graphs and degree

Directed, undirected and weighted graphs

A graph $D = (V, E)$ where every edge $e \in E$ is associated with an ordered pair of vertices, therefore, where each edge of D has a direction, is a **directed graph** (or digraph). In the diagram of directed graphs, the direction is represented by an arrow (or a directed curve) from the initial node to the terminal point. The edge is said to be incident **from** the initial vertex and incident **to** the terminal vertex.
An un-directed graph *G* consists of a set *V* of vertices and a set *E* of edges such that each e $\in$ *E* is associated with an unordered pair of vertices.
Given a graph $W = (V, E)$, whenever a real number can be assigned to an existing arc, the edge has a weight w and the graph W is called **weighted graph**. Both directed and undirected graphs can be also weighted, the weights are positive or negative numbers. Usually, in many practical applications, the weights indicate costs, distances, probabilities, time measures, etc.

In figure 1.2 here is an example of directed graph, it shows the international financial network, where the nodes are major financial institutions belonging to the core and the links give the strongest existing relations among them; node colors indicate different geographical areas: EU (red), US (blue), other countries (green); the thicker edges have a higher weight, they represent a stronger connection.

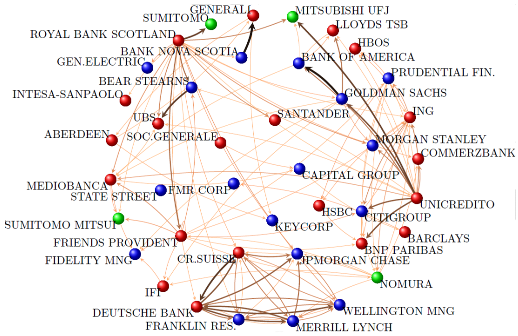


Figure 1.2: Sample of a directed graph: the international financial network

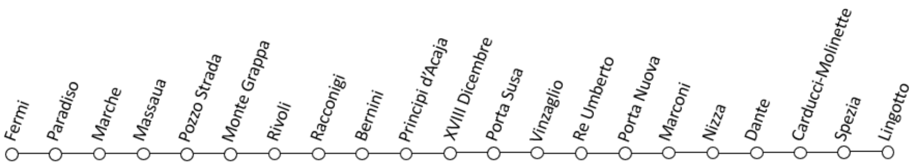


Figure 1.3: Undirected graph representing Turin's subway line

Degree, outdegree and indegree

The degree of a node means the number of incident edges, with edges that join themselves (self-loops) counted twice. The degree of a node is indicated $deg(vertex)$.

In an undirected graph the **total number of links** can be expressed as

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i \tag{1.1}$$

where k_i is the degree of vertex i and the factor $\frac{1}{2}$ is added because every link is counted twice. Another important property is the **average degree**, for an undirected graph it's

$$K = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2L}{N} \tag{1.2}$$

where N is the number of nodes in the vertex set. In directed networks let k^{in} be the incoming degree and k^{out} the outcoming degree. The total degree is the sum of the incoming and the outcoming degrees

$$k_i = k_i^{\text{in}} + k_i^{\text{out}}. \tag{1.3}$$

The total number of links in a directed graph is

$$L = \sum_{i=1}^{N} k_i^{\text{in}} = \sum_{i=1}^{N} k_i^{\text{out}} \tag{1.4}$$

Now the $\frac{1}{2}$ factor is missing because each edge is counted only once since there is a separated addition based on the verse. The average degree value is given dividing the total number of links for the vertices' total.

$$K = \frac{L}{N} \tag{1.5}$$

A vertex with 0 degree is called **isolated**, while a vertex with degree 1 is called **pendent**.

1.2.3 Graph classes

Null graphs

Any graph must have at least one vertex, but a graph may have only isolated nodes. Null graphs have an empty set of edges and are represented as N_n with the n indicating the vertices' number.

Multigraph, pseudo graph and simple graph

In a multigraph loops are not allowed, but more than one edge can join two vertices; these edges are called multiple (or parallel) edges.
A Pseudo graph has a structure in which loops and multiple edges are allowed.
A graph which has no either loops nor parallel edges is called Simple graph.

Complete graph

A graph is said to be complete if every pair of vertices has an edge. It is usually indicated by K_n, where n is the number of vertices. It is complete because is not possible to add any new edge and obtain a simple graph. For every $n > 0$, the degree of each node is $n - 1$ and the number of links is $\frac{n(n+2)}{2}$ edges.

Regular graph

A graph in which all vertices have equal degree r is called regular of degree r or $r-$regular; hence the total number of edges of an r-graph with n vertices is $\frac{nr}{2}$. Note that every null graph is regular of degree 0 and every complete graph is regular of degree $n-1$.

Subgraph

A graph $G' = (V', E')$ is a **subgraph** of $G = (V, E)$ if and only if $V' \in X$, and $V' \in E$. A graph G' with all its vertices and edges in G is represented as $G' \in G$. All the edges and vertices of G might not be present in G'; but if a vertex is present in G', it has a corresponding vertex in G and any edge that connects two vertices in G' will also connect the corresponding vertices in G.

Bipartite graph

Let $r \geq 2$ be an integer, a graph $G = (V, E)$ is called r-partite if it can be partitioned into r different disjoint sets, such that every edge joins different classes.

Definition 1.3 *A graph $G = (V, E)$ is bipartite if exists a partition $V = A \cup B$ of the vertex set such that A and B are not empty and every $(a, b) \in E \rightarrow a \in A, b \in B$.*

A simple way to notice if a graph is bipartite is coloring the graph in two colors, if each edge joins two nodes of the same color then the network is bipartite. In contrast, in non-bipartite graph such a coloring is impossible, e.g. in a triangle Bipartite graphs very often arise in real applications, they are extensively used in modern coding theory, specially to decode codewords received from the channel.

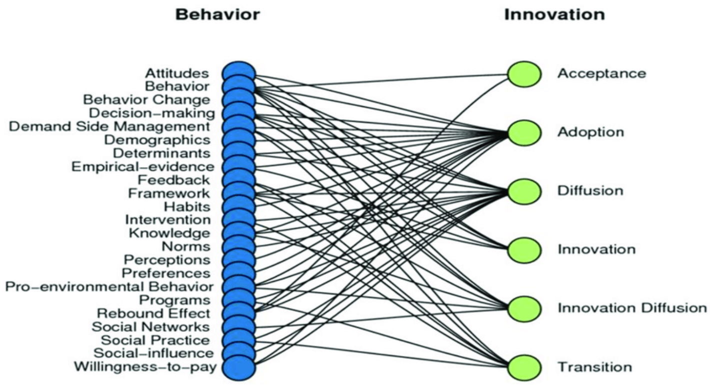


Figure 1.4: Bipartite graph for the leaner representation of the relationship between behavior and innovation stages

1.3 Walks, Paths and Circuits

A graph *G* is connected if every pair of vertices in G are connected, otherwise it's disconnected; in a connected graph is possible to reach any vertex from any other vertex. It follows that a complete graph is always connected and an isolated graph never. Every graph *G* consists of 1 or more connected graphs, each of them is a subgraph. A **walk** is defined as a finite alternative sequence of vertices and edges, of the form

$$\{v_i e_j, v_{i+1} e_{j+1}, ..., v_k e_m\}.$$

In a walk (or trial) each edge in the sequence is incident from the vertex preceding and to the vertex following and no edges appear more than once. A *u*-*v* walk is a walk from vertex u to v, if the first vertex u and the terminal v are coincident the walk is closed, otherwise it is open. An open walk in which no vertex could appear more than once is called **path** or **trail**.

A **circuit** is a path that begins and ends at the same vertex. A circuit that doesn't repeat vertices is called **cycle** or **circuit**.

The number of edges in a walk, and consequently in a path and in a circuit, is called its **length**. A circuit or cycle with *k* length is called *k*-circuit or *k*-cycle.

1.3.1 Eulerian paths

Königsberg's bridges

The Eulerian paths were first discussed by Leonhard Euler while solving the famous Seven Bridges of Königsberg problem, the issue that represented the beginning of network mathematic.

Königsberg, Kalingrad today, was a very important city and trading center during the Middle Age, with its strategical location on the river. The healthy economy allowed the inhabitants to build seven bridges across the river, most of which connected to the island. The people of the city decided to create a game for themselves, their goal being to devise a way in which, while walking in the city, they could cross each of the seven bridges only once.

Even though none of the citizens of Königsberg could invent such a route, still they couldn't prove that it was impossible. Euler found this problem trivial, this simple question interested him, since neither geometry, nor algebra could solve it. He discovered that the problem could be seen as a network, see figure 1.5, with the different parts of the city represented by nodes and the bridges by links connecting them.

Euler explained that was impossible to solve the citizen game, basing his statement on a mathematical explanation. A network with the possibility to cross each link only once, would have had an even number of edges for every central vertex, while the initial and the final one would have had an odd number. This was not possible in the city network, since every zone had an odd number of bridges, from 3 in vertices C, B and D, to 5 in vertex A. Even though this couldn't seem a great result, it's important, since in math proving a problem impossible to solve spare a lot of time spent in trying to find a solution that could not exist.

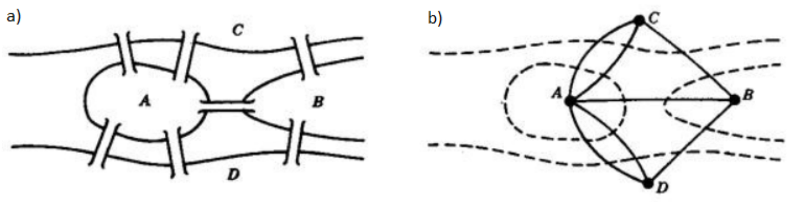


Figure 1.5: The problem of the bridges of Königsberg: the map of the city (a) and its network model (b)

The network's path the inhabitants of Königsberg needed is called **Eulerian path**.

Definition 1.4 *A path in a graph G is called Euler path if it includes every edge exactly once. Similarly, a Eulerian circuit is a Eulerian trail that starts and ends on the same vertex.*

To have a Eulerian trail the graph must be connected and, as specified in the previous discussion, any trail must enter and exit the same number of times at each vertex that means that all degrees must be even. Euler proved that these conditions are necessary and sufficient.

Theorem 1.1 *A connected graph with even degrees has a Eulerian trail*

Considering the case in which the trail departing from a, covering all the edges, must reach point b, it's clear that the initial and the terminal vertex must have degree 1.

Theorem 1.2 *A connected graph as a Trail $T_{a,b}$ covering all edges just once if and only if a and b are the only odd vertices.*

1.3.2 Hamiltonian paths

Hamiltonian paths and cycles are named after that in 1859 William Rowan Hamilton, a famous Irish mathematician, invented a particular puzzle known as the icosian game (or Hamilton's puzzle). Its main part was a regular dodecahedron made of wood; whose corner were marked with names of famous cities. The puzzle consisted in finding a route along the edges of the solid passing through each city just once. The problem could also be analyzed through a graphic representation (it's possible to draw the graph without edge intersections).

Definition 1.5 *A cycle that uses every vertex in a graph exactly once is called a Hamilton cycle, and a path that uses every vertex in a graph exactly once is called a Hamilton path.*

In a graph G the **Hamiltonian cycle** is a closed trail passing only one time through every node. In contrast to eulerian graphs mathematician have so far

failed to find a general criterion.
In 1960 a Norwegian mathematician Øystein Ore discovered a sufficient condition for a graph to be Hamiltonian

Theorem 1.3 *(Ore, 1960). If in a simple graph $G = (X, E)$, with $|X| = n \geq 3$, for every pair x and y of disjoint vertices then G is a Hamiltonian graph.*

$$d(x) + d(y) \geq n, \tag{1.6}$$

then G is a Hamiltonian graph.

1.3.3 Tree

A **tree** is a connected graph and contains no cycles. This implies that there aren't multiple edges and that in a tree there's only a path connecting one vertex with another. Since there aren't cycles the different branches stay apart once they are separated.

Theorem 1.4 *A tree with n vertices has $n - 1$ edges.*

Trees have applications in several fields, for example in game theory are used to represent the extensive form of a game and are really useful in solving dynamic games.
Given a connected graph $G = (V, E)$, a spanning tree S is a tree where V is exactly the same of G and the edge set is a subset of that of G. The problem of finding a spanning tree of minimum (or maximum) weight is called the **minimum (maximum) spanning tree model** and allows the use of an efficient algorithm for any graph.
Given a weighted Graph $G = (V, E)$ the aim is to build a minimum spanning tree graph T (empty at the beginning). First of all, the edge set E must be ordered in increasing order of weights, the algorithm consists in adding every edge one by one, only if it doesn't produce a cycle in T with the previously added edges. The process stops when T is connected.
The algorithm produces a tree because it doesn't include cycles and ends only when T is connected.

1.4 Graph Representation

1.4.1 Matrix

Adjacency and incidence matrix

The diagrammatic representation of a network is convenient for a visual study but it's only possible when the graph is reasonably small. Another possible representation is using a matrix.
There are more types of matrix suitable to represent graphs, among these the adjacency matrix and the incidence matrix.
The **adjacency matrix** of a graph G with n vertices and no parallel edges is represented through a n by n matrix $A = a_{ij}$ where the element

$$a_{ij} = \begin{cases} 1, (i,j) \in E, \\ 0, (i,j) \notin E. \end{cases} \tag{1.7}$$

If there are more edges connecting node i and j the element a_{ij} value is equal to number of arcs from i to j.
From G could be derived n! different adjacency matrices, depending on the order of the nodes; however, every couple of matrices is strictly related, from one matrix it's possible to derive the other just interchanging rows and columns.
The adjacency matrix of an undirected graph is symmetric and the entries along the principal diagonal are all zeros if and only if the graph has no self-loops. A node's degree is a sum over either the rows or the columns of the matrix.

$$k_i = \sum_{j=1}^{N} a_{ij} = \sum_{j=1}^{N} a_{ji} \tag{1.8}$$

The adjacency matrix $A = a_{ij}$ of a directed graph D with n vertices is analogues to the precedent matrix, but the element $a_{ij} = 1$ if the directed arc from node i to j exists in D, 0 otherwise. Note that A is not necessarily symmetric, the sum of entries in row i is equal to the out degree of vertex v_i and the sum of entries in column j

$$k_i^{\text{out}} = \sum_{j=1}^{N} a_{ij}, \; k_i^{\text{in}} = \sum_{j=1}^{N} a_{ji} \tag{1.9}$$

Given a graph $G = (V, E)$ with n nodes and m arcs, is B = b_{ij} its **incidence**

matrix; B is a $n \times m$ matrix where the element

$$b_{ij} = \begin{cases} 1, \text{if } v_i \text{ is vertex of } e_j, \\ 0, \text{if } v_i \text{ is not vertex of } e_j. \end{cases} \tag{1.10}$$

Note that a row with the entries sum null corresponds to an isolated vertex and a row with a single unit corresponds to a pendent vertex; the number of entries in a row is equal to the degree of the correspondent vertex. Each column of B contains exactly two entries, identical columns represent multiple edges and loops are represented using a column with only 1 entry.

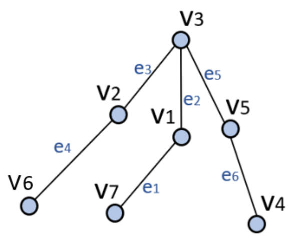


Figure 1.6: Diagrammatic representation of a simple graph

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 1.7: Adjacency Matrix

$$B = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 1.8: Incident Matrix

Eigenvalues, eigenvectors and spectrum of a Graph

The possibility of representing graphs through their adjacency matrices leads to the idea of applying the eigenvectors and eigenvalues theory. The eigenvalues are invariant with respect to the permutation of rows and columns, so they could explain some graph's properties not depending on the naming of the nodes.

Definition 1.6 *Let $A(G)$ be the adjacency matrix of G. An element $\lambda \in F$ is an **eigenvalue** of G if and only if there exists a vector $v \in F^n, v \neq 0$, with $A(G)v = \lambda v$. In this case v is called an **eigenvector** of $A(G)$.*

If λ is an eigenvalue of A every v is an eigenvector of A associated with λ. The multiplicity of the zero λ is denoted by $m(\lambda)$.

Definition 1.7 *Let $A(G)$ be the adjacency matrix of G. The polynomial of degree n in the variable z over the field F is given by:*

$$p_a(z) = det(zI_n - A) = z^n + a_{n-1}z^{n\text{-}1} + ... + a_1 z + a_0 \tag{1.11}$$

The **characteristic polynomial** previously defined is independent of the numbering of the edges in G.
Spectral graph theory is a branch of mathematic based on the study of the properties of a graph in relation to the characteristic polynomial, eigenvalues, and eigenvectors of its matrices, such as the adjacency matrix. The sequence of a graph eigenvalues together with their multiplicities is called its **spectrum**.

Definition 1.8 *Let $\lambda_i, i = 1, ..., n$ be the zeros of $p_a(z)$ in natural order. $\lambda(G) = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_p = \Lambda(G)$. The largest eigenvalue Λ is called the spectral radius of G. The spectrum of a graph G, denoted by Γ, is the set of eigenvalues of $A(G)$ together with their multiplicities:*

$$\Gamma = \begin{pmatrix} \lambda & ... & \lambda_i & ... & \Lambda \\ m(\lambda) & ... & m(\lambda_i) & ... & m(\Lambda) \end{pmatrix} \tag{1.12}$$

1.4.2 Planar graphs

Plane and planar graphs

A graph is called **planar** if it has some geometric representation in the plane $\mathbb{R}^2$ such that none of its edges intersect (except possibly at end points).
In drawing graphs, nodes can be represented by points and edges by Jordan curves. The **Jordan curve** is a continuous curve in the plane which connects two nodes and has no intersection with itself.

Theorem 1.5 *(Jordan Curve Theorem) A closed Jordan curve L partitions the plane into precisely two regions, bounded and unbounded, each having L as boundary.*

A region is **connected** if any pair of its points can be connected by a Jordan curve laying inside the region. A plane embedding divides the plane into connected regions called faces; a **face** is a region delimited by edges that doesn't contain another smaller edge delimited region. Each face is bounded with a cycle and the number of nodes of such cycle is the size of the face.

Theorem 1.6 *(Euler, 1750). Let G be a connected plane graph with n vertices, e edges and f faces, then*

$$n - e + f = 2 \tag{1.13}$$

It's possible to rewrite the relation 1.13 as $f = 2 + e - n$. The reason why the formula is valid for every connected network can be understood looking at a simple graph. Let P be a graph with one edge connecting two nodes: $e = 1, n = 2$ and $f = 1$ meaning the unlimited external region. Adding one edge incident only to one of the nodes, e and n increase by 1 and f stays constant, in fact the there aren't new regions. Adding an edge between the 2 existing nodes, e increases by 1, n stays the same and f increases by 1 because an existing region is divided in two different ones.

The smallest non-planar graphs

Theorem 1.7 *(Kuratowski's Theorem). Let G be a graph. Then G is non-planar if and only if G contains a subgraph that is a subdivision of either $K_{3,3}$ or K_5.*

K_5 and $K_{3,3}$, in figure (1.9), are the most reduced non-planar graphs that every non-planar graph could contain; K_5 in figure is a non-planar graph with the smallest number of vertices, $K_{3,3}$ is a non-planar graph with the smallest number of edges.
One graph is homeomorphic if it's possible to turn it into another one by adding or removing vertices. Kuratowski's theorem states that a finite graph G is not planar, if and only if contains a subgraph homeomorphic to K_5 or $K_{3,3}$.

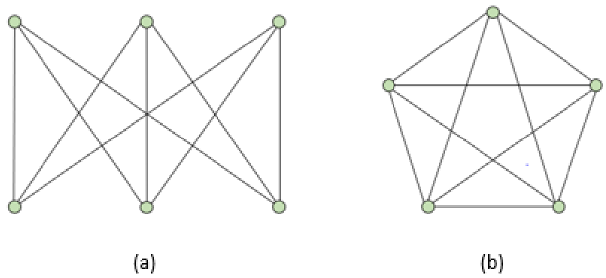


Figure 1.9: Complete bipartite graph on six vertices, $K_{3,3}$ and K_5

It's possible to prove that K_5 is non-planar using Euler's formula. Suppose that K_5 is planar. T is the sum of the sides of every face, since every side could border on a maximum of 2 regions it's possible to state that $T \leq 2e$. Every face is at least delimited by 3 arcs and since there are f faces $3f \leq T$. Joining these relations it could be derived that $3f \leq 2e$. From Euler's theorem is known that $e = n + f - 2$ and substituting the previous results, it's derived $e \leq 3n - 6$. This automatically states that K_5 isn't planar, in fact the edges in the figure are 10, while in order to be planar should be 9 as it's easily derivable from the condition $e \leq 3n - 6$ where n is 5.
$K_{3,3}$ can't be proved non-planar with the previously derived disequation. It is necessary to add another constraint. As it's clear from the graph every cycle in $K_{3,3}$ has the same length and there aren't triangles, assuming that every face is limited by 4 sides and replicating the steps of the previous demonstration is achieved that $4f \leq 2e$. Substituting in Euler's formula lead to the conclusion that a plane drawn of $K_{3,3}$ has $e \leq 2n - 4$. However, replacing the values $n = 6$ and $e = 9$ the condition isn't satisfied and this proves that $K_{3,3}$ is not planar.

1.5 Statistics on Graphs

The possibility for the systematic gathering and usage of data sets on large scale networks prompt to the use of statistical analysis for a mathematical characterization of graphs. The statistical analysis has been initially focused on the small world, the clustering and the degree distribution properties.

1.5.1 Clustering coefficient

The concept of clustering in a graph refers to the tendency in many natural networks to form groups where the density of links is larger than the average (cliques). For an undirected graph the clustering level can be measured with the **clustering coefficient**.

Given an undirected graph G, let i be a vertex with degree k_i and e_i the number of edges between the neighbors of i. The clustering coefficient of node i is

$$c_i = \begin{cases} \frac{2e_i}{k_i(k_i-1)}, ifk_i \geq 0, \\ 0, \text{otherwise.} \end{cases} \tag{1.14}$$

1.5.2 Small-world properties

The small-world effect states that in several large-scale networks it is possible to go from one vertex to any other passing through a small number of intermediaries, so that the average distance between nodes is really short compared to the network's size. The average distance can be measured by the average shortest path l connecting the two nodes. In mathematical terms the small-world property is met if l scales logarithmically with the number of vertices.

The small-world concept is spread in the sociological context by the name of ”six degree of separation”. According to "six degrees of separation", in a world of 6.6 billion people each person is just six introductions away from any other person on the planet. This theory was empirically proved true in more researches. The first study on the small world was conducted by Stanley Milgram in 1967. He decided two targets, two different persons, a stockbroker in Boston and a divinity student in Sharon, then he extracted a number of random people and said to them to send a letter to the most probable person, in their personal network, to know the targets. The experiment leaded to an average number of links sufficient to reach the target of 5.2, a remarkable small number. Later in 1998 Duncan J. Watts and Steven Strogatz proposed the Watts–Strogatz model, a random graph generation model that produces graphs with small-world properties, including short average path lengths and high clustering.

1.5.3 Degree distribution

The simplest statistical characterization of a network is given by the sequence of degrees k_i of its vertices. The **degree distribution** $P(k)$ for an undirected graph is defined as the probability that a randomly selected node has degree k

$$\sum_{k=1}^{\infty} P(k) = 1 \tag{1.15}$$

In the case of directed graphs, one must consider two different distributions, the in-degree $P(k_{in})$ and the out-degree $P(k_{out})$ distributions, defined as the probability that a randomly chosen vertex has in-degree k_{in} and out-degree k_{out}, respectively. For a network with N nodes the degree distribution is the normalized histogram given by

$$P(k) = \frac{N_k}{N} \tag{1.16}$$

Where N_k is the number of degree-k nodes. Therefore, the number of k-degree vertices is easily derived as $N_k = Np_k$

The degree distribution can lead to the individuation of most network properties and its functional form determines many network phenomena, from robustness to the spread of diseases.

1.6 Network Topologies

1.6.1 The Erdős–Rényi model

Random networks were among the earliest studied networks. The Erdős–Rényi model was probably the first model able to describe large networks.

The process to generate an ER random graph starts with N isolated nodes, then for every pair of nodes they are connected with probability $p \in (0, 1)$, if $p = 1$ the nodes are fully connected, for $p = 0$ the nodes stay isolated. The expectation of the number of edges is $\frac{pN(N-1)}{2}$, the larger p, the denser the resultant network. This generation process leads to simple graphs, no loops nor multiple edges, and retrying the process with the same probability builds a different network, due to the randomness.

For stochastic ER networks the probability that a randomly chosen vertex has degree k is the binomial distribution

$$\binom{N-1}{k} p^{\mathrm{k}}(1-p)^{\mathrm{N\text{-}1\text{-}k}} \tag{1.17}$$

Another way to derive the previous distribution is from the degree distribution $P(k)$ of a random network. The degree distribution depends on three terms: the probability that there are k links, the probability that the remaining $(N-1-k)$ are missing and the number of possible k links combinations chosen from the $N-1$ links.

Among the most significant discoveries of Erdős and Rényi, there is the property that for $p \geq p_c$ where the threshold $p_c \sim \frac{ln(N)}{N}$ (p_c is in the same order of $\frac{ln(N)}{N}$), almost all random graphs generated with probability p will suddenly become connected.

The average degree is then easily computed as $K = (N-1)p$, and the clustering coefficient is $c_i = p$ since it is the simple probability that two nearest neighbors of a vertex of degree i have an edge between them. The average of the shortest path lengths between every couple of nodes in the network grows logarithmically with the number of nodes, $l \sim ln(N)$.

Even though the exact form of the degree distribution of a random network is the binomial distribution, for a very large number of nodes $N \to \infty$ the binomial is well approximated by a Poisson distribution

$$p_k = e^{\text{-k}} \frac{k^{\mathrm{k}}}{k!} \tag{1.18}$$

The Poisson distribution peaks at the average degree K and decay exponentially fast at both sides, in particular nodes with a very high degree do not exist. This kind of network is referred to as homogenous network. The procedure seen is effective and has been used to represent many real world networks, but doesn't allow to change the size of the graph during the construction and for this reason it originates a **static** random network.

1.6.2 Small-World network model

Networks with a high clustering coefficient and a short average path length are the so called small network discovered by Watts and Strogatz in 1998.

The process to build a small-world network is:

- start: from a circle network with N nodes, each one is connected to $2M$ other vertices, where M is an integer > 0.

- growth: for every couple of disjoint nodes connect them with probability p where $0 < p < 1$.

This algorithm leads to the construction of a simple graph.
The clustering coefficient is redefined as

$$C_{ws} = \frac{\text{average number of neighboring edges}}{\text{total possible number of neighboring edges}} \tag{1.19}$$

1.6.3 The Barabási-Albert scale-free network model

Static networks are not suitable to describe some kind of real networks, characterized by an evolving number of nodes and disposition of edges. Furthermore, in the growth of a real network, a forthcoming node has the tendency to connect itself to some "big" nodes (with large degrees) that are more important and consequently, more attractive. This mechanism, is named **preferential attachment**, reflecting that the probability of obtaining a higher connectivity in the future growth is proportional to the present connectivity, the so-called "rich gets richer" effect.
Barabási and Albert proposed a new network model, known as the **Barabási-Albert scale-free network model**, that captures two essential features of many complex networks: growth and preferential attachment.
The BA algorithm is as follows:

- Growth: Starting from a connected network of small size $m_0 \geq 1$, introduce one new node to the existing network each time, and this new node is connected to m existing nodes in the network simultaneously, where $1 \leq m \leq m_0$.

- (Linear) Preferential Attachment: The above-referred incoming new node is simultaneously connected to each of the m existing nodes, according to the following probability: for node i of degree k_i,

$$p_i = \frac{k_i}{\sum_{j=1}^{N} k_j} \tag{1.20}$$

The average degree of a BA scale free network is

$$K \sim \frac{ln(N)}{ln(lnN)} \tag{1.21}$$

The node-degree distribution af a BA scale free network is given by

$$P(k) \sim (k^{\text{in}})^{-\lambda} \tag{1.22}$$

where $\lambda = 2 + a/m$. The constant a represents the number of initial nodes, if $a = 0$ the model falls back in the BA network. After t steps, the network will have $N = t + m_0$ nodes and $mt + m^*$ edges, with m^* the initial number of edges.

Theorem 1.8 *Consider a continuously differentiable probability distribution function $f(x)$. If, for any given constant a, there is a constant b such that the following "scale-free" property holds:*

$$f(ax) = bf(x), \tag{1.23}$$

then, with the assumption of $f(1)f'(1) \neq 0$, the function f (x) is uniquely determined by $f(x) = f(1)x - \lambda, \lambda = -f(1)/f'(1)$.

Building a large-scale network with a power-law degree distribution and with $2 \leq \lambda \leq 3$, most nodes result to have very low degrees (i.e., with very few edges) and there are only few nodes having very high degrees (i.e., with many edges). This kind of complex networks is named **heterogeneous networks**, where the high-degree nodes are also referred to as **hubs**.

1.6.4 Comparison

For a given network of any kind, if one node is being removed at a time (consequently all the edges connecting it will also be removed), then at some point the network will be broken to become disconnected. If the network remains connected after some nodes have been removed, then the network is said to be robust against node-removal. Comparing two networks, the ER random network and the BA scale-free network, is found that the BA network is more robust than the ER graph. This outcome is due to the heterogeneous distribution of nodes in the network: most nodes have very small degrees and only a few nodes have large degrees; thus, randomly removing a fraction of nodes will very likely remove some small nodes, which does not affect the network connectivity by too much much. However, for exactly the same reason, any intentional removal of even a very small fraction of high-degree nodes will significantly affect the topology of the network, leading to drastic change of the network connectivity.

1.7 The use of Network Theory in Economics

1.7.1 Economic networks

Networks play a major role in modern life, suffice it to say that our world is more connected than ever. Connected by communication, information and transportation networks.

The economic network theory is an emerging field that applies network mathematics in order to understand economic phenomena, relying on the evolving nature and the representative strength of this instrument. The main reason to solve economy related questions by means of networks is that this perspective doesn't consider only individual components, but allows to study the relations between all the relevant parts in the system.

The economic system reflects the dynamic interaction among a multiplicity of agents, whose increased interconnectivity of behaviors and information leads to a huge complexity, difficult to predict and control. This complexity is increased by the speed in the flow of information across the world, enabled by the technological progress.

Economic network theory is a wide field that contains a set of different networks where each one has its specific meaning (in table 1.1 is possible to see some applications). Nodes can be industries, financial intermediaries, or even countries, and arcs represent their mutual interaction, be it debit-credit relations, trade or ownerships. The choice of nodes and edges identifies the purpose of the network.

Over the past decades applications of networks in economics has significantly increased. The main reason is, as previously exposed, the necessity to analyze the interactions among the agents. Other two important factors are the increased capacity of understanding the features of networks (e.g., how densely connected

a market is, how segregated it is, etc.) and the amount of data available combined with the improvements in computational capabilities that has allowed new tests and applications of models.
Economists' participation in the study of networks has contributed to the develop of the literature. After the 2008 crises, concepts as networks and contagion have entered in the economic and financial lexicon, since the traditional economic theory couldn't explain, neither predict, the collapse of the financial system and its deep effects on the world economy. The 2008 crises highlighted how the propagation of a shock through the financial network, can lead to a domino effect and affect even stable intermediaries.
The research in economic networks advances on two paths. On one hand economics and sociology (micro approach), while on the other hand physic complex system and computer science (macro approach).
The socioeconomic approach centers on understanding how the individual behavior is influenced by the linkages structure, it looks at agent incentives in developing links, while it's not suitable to predict realistic dynamic outcomes. In this model every person can choose between individual incentives and network aggregate welfare, since every individual choice has an impact on the overall structure through contagion, information diffusion, opportunities and other factors. The micro perspective aims at understanding how the strategic behavior of the agents is influenced by network architectures. In recent micro approaches, the network is seen as the result of a network formation game, a link is added or delated according to a specific decision that has been taken by the agent. The agent's choice is influenced by the information at the disposal and guided by the attempt to anticipate what others may do. This approach relies on game theory, whose scope is to identify situation of stability called Nash equilibria.
In contrast, the macro perspective emphasizes the statistical regularities of the network, but can't link them to the economic motivation of individual agents. The complex-systems branch considers how the network-formation rules model the emerging structure.
The study "Network theory and social economics - a promising conjunction?" pointed out some recurring characteristics in socio-economic networks derived from several researches: the high degree of clustering, the small world property and the heavy tail.

Type of economic Network	Meaning of the network
Director network	Common members of the board institutions; corporate structure between institutions
Financial network	Dependency between stocks, stock market interconnectedness
Interbank network	Liability between institutions (banks); interconnectedness of banks
Investment network	Venture capital or investment in firms and corporation
Ownership network	Influence on corporate decisions
Product network	Individuals, co-purchasing products or countries co-producing products
Trade network	Exchange of goods and services; trading relations between countries

Table 1.1: A set of economic networks with the respective explanation

Table 1.1 lists some economic applications of network theory. Studies have shown that social networks have a high degree of clustering if compared to non-social networks. Clusters were identified in many economic networks, e.g. banks in financial markets form clusters of dense trade relationships and countries in the world trading network form clusters of thick relationships. In real networks the degree distribution is highly asymmetric with the number of neighbors, the degree is inversely proportional to the relative frequency of nodes with this number of neighbors. This is similar to a scale free distribution from the fact that the shape doesn't depend on the part of the distribution considered. The probability of tail events in this distribution is high and averaging over a large number of observations could be risky, in case the central limit theorem is not applicable. Corporate ownership and lending networks have been proved heavy tailed, while there are some doubts on other distributions claimed to be scale-free but that could be also log-normal, exponential with cutoff and less regular distribution. In addition, many economic networks present a core-periphery structure. Such a structure indicates a densely connected core and a sparse periphery, so the vertex set is composed by two groups of nodes, one of vertices that are really connected and one of disperse vertices. Actors in the core occupy a dominant position and they are favorite in trading with actor in the periphery. Core-periphery structure can be found in bank networks, in supply chain networks and in international trade networks. Looking for instance at the financial network, the reason why core banks' default tend to be high correlated could be explained by its structure, as the core has a small number of hubs and they are densely interconnected; however, it should be kept in mind that banks are supported by the governments and that the policy of the "too big to fail" has always prevailed.

Chapter 2

From input-output tables to national sector maps

The scope of this chapter is to introduce the national input output tables, explain how they work and the choice of path that has leaded this essay. The first section illustrates the national accounting's categorization of the money flows. Consequently, the input output tables are introduced, from their history to their structure and, finally, their applications. The last paragraph concerns the construction of the national graphs analyzed in this work.

2.1 National accounting

An economy is made of sectors linked each other by commercial exchanges, therefore, it's important to understand the system's structure in order to fully realize its functionality. IO tables are part of the accounting data compiled by national statistical institutes and provide a general overview of the monetary flows between economic sectors.

National accounting, despite not making use of the graph terminology, naturally describes a network with sectors as nodes and money flows as edges. Tracking of money flows involves substantially more complications than measurement of other networks, because it involves different categories of transactions such as output, consumption, income, and investment.

Generally, economies are composed of five sections: households, non-financial

business, financial business, government, and non-profits.
Household consumption (purchases of goods and services) and **value added** are the largest money flows. Value added partly corresponds to purchases of household sector labor by the business sector. They are collectively referred to as the **circular flow** and constitute the backbone of sectoral money flow structure (James McNerneya et Al, 2013). Intermediate consumption, recorded in input/output tables, is the second flow in size and represents purchases made by industries for the production of goods by other industries. Goods purchased for intermediate consumption are required inputs to produce finite goods.
Capital purchases are purchases needed to the production of other goods and that can be used repeatedly for more than one accounting period, these are an important exclusion from transactions classified as intermediate consumption.
The transactions underlying money flows in the accounting system are compiled on an accrual basis (alternative to the cash-flow basis) . Revenues are recognized when they are earned by the transfer of goods or the performance of a service. Expenses are recognized when the associated revenues are earned. Money flows within the full sector-level network are not conserved because they may disappear from accidental loss or destruction and, more importantly, money is regularly created and destroyed by the financial sector. National accounting does enforce a virtual conservation law, though, through the use of balancing items, which are accounting entries that are calculated as the difference between other accounting entries.
The I/O tables do enforce a virtual conservation law where the balancing item is value-added, calculated as the difference between total sales by the business sector and intermediate consumption sales. Value-added measures the value created by production and encompasses all forms of personal income — employee compensation, interest, dividends, and rent, as well as certain kinds of taxes and depreciation.
Exploiting the conservation enforced by the definition of value-added, one can equate money flows in and out of the business sector and deriving the GDP: The figure (2.1) illustrates two approaches to calculating GDP, on the left side the revenue method and on the right one the expenditure one. A third approach is considering the outputs, where value added is calculated as the difference between all business sales and intermediate goods sales. All three approaches are used by

statistical agencies to validate GDP calculations. They also provide equivalent intuitive interpretations of GDP as a measure of total income, a measure of total final expenditures, and as the net output of the business sector.

Value added+ intermediate consumption+ Imports+ business taxes	=	Intermediate consumption +household consumption +government consumption +capital formation +exports +subsidies
GDP = Value added +business taxes -subsidies	=	+household consumption +government consumption +capital formation +exports +imports

Figure 2.1: Gross domestic product

2.2 Input output tables (IOt)

2.2.1 IOt history

The first input-output model is attributable to the 20th-century economist Wassily W. Leontief, that in 1973 earned the Nobel Prize in Economics.

$Input - Output\ Analysis$ (E.Miller & D.Blair, 2009) clarifies that the idea of developing an instrument of this kind is much older and probably dates back to the beginning of the eighteenth century.

In 1758, François Quesnay, a French physician, leader of the Physiocrats[1], wrote the $Tableau\ Economique$, an economic model depicting the income flow between economic sectors. The Tableau is mostly known for its diagrammatic representation of how expenditures can be traced within an economy. However, it must be said that many of Quesnay's and the Physiocrats'views were quite controversial, as the theory that manufacturing and commerce didn't contribute

[1]The Physiocrats were a group of French Enlightenment thinkers of the 1760s that surrounded the French court physician, François Quesnay. The founding document of Physiocratic school was Quesnay's Tableau Économique (1759). To contemporaries, they were often referred to simply as les économistes.

to add value to the economy.
In 1820, Robert Torrens, a British army officer and journalist, suggested that the economic surplus was the key to trace the share of income attributable to sources. In his work on corn trade, he defined the agricultural rate of profit as the ratio between net corn output and corn input. Essentially, he resumed the analytical connection between profits and production factors described in the Tableau Économique, but at the same time, he denied the theory regarding the manufacturing sector not producing added value.
In 1920 Leontief described the economy as a circular flow in his doctoral thesis, drawing on Quesnay's Tableau and other studies. Then, in 1928 he published part of his thesis, where he depicted a two-sector input–output system showing production, distribution, and consumption characteristics of an economy as a single integrated system of linear equations. In 1937, Leontief published the first Input-Output model, revolutionizing both analytics and economics.
Leontief's contributions went far beyond the table of transactions: he derived the analytical basis that transformed the descriptive nature of the Tableau into an empirical analytical tool. Today Leontief's input–output analysis has become one of the most widely applied methods in economics.

2.2.2 Input-Output properties

IOt structure

Every IO table is built around a matrix of monetary transactions where flow values are expressed in currency. The basic principle of IO-tables is that outputs of one industry are inputs of other industries. The rows of the table indicate the amount of outputs sold by each industry, including both intermediate and final goods, while the columns reflect the value of inputs bought by each industry. The sales are inclusive of sales to the government, sales to the consumer, sales to non-profit organization, gross fixed capital information[2], change in inventories and exports. Purchases, otherwise, consist of domestic and imports purchases, value added and a negative flow given by taxes less subsidies on products.

[2]it measures the value of acquisitions of new or existing fixed assets by the business sector, governments and households.

		Intermediate use			Final use			Inventories	Exports	Gross Output
		1	2	RoW	1	2	RoW			
domestic	Industry 1	X_{11}	X_{12}	X_{1R}	F_{11}	F_{12}	F_{1R}	I^A	E^A	GO^A
	Industry 2	X_{21}	X_{22}	X_{2R}	F_{21}	F_{22}	F_{2R}	I^B	E^B	GO^B
	Industry Row	X_{R1}	X_{R2}	X_{RR}	F_{R1}	F_{R2}	F_{RR}	I^R	E^R	GO^C
import	Industry 1	X'_{11}	X'_{12}	X'_{1R}	F'_{11}	F'_{12}	F'_{1R}		0	0
	Industry 2	X'_{21}	X'_{22}	X'_{2R}	F'_{21}	F'_{22}	F'_{2R}		0	0
	Industry Row	X'_{R1}	X'_{R2}	X'_{RR}	F'_{R1}	F'_{R2}	F'_{RR}		0	0
	Taxes less subsidies on products	T^A	T^B	T^R	T'^A	T'^B	T'^R	0	0	0
	Value added	W^A	W^B	W^R	W'^A	W'^B	W'^R	0	0	0
	International transports margins	tm^A	tm^B	tm^R	tm'^A	tm'^B	tm'^R		0	0
	Output	$Output^A$	$Output^B$	$Output^C$	$Output'^A$	$Output'^B$	$Output'^C$	0	0	0

Figure 2.2: Input output table format

The table itself implies a network, in fact its core could be read as a weighted, directed adjacency matrix. Reversing the direction of the links in the network, the network represents the money outflow from one industry to another in order to purchase materials for production inputs. Let T be the adjacency matrix mapping the money flows of the i-o table

$$T = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-1,1} & x_{n-1,2} & \cdots & x_{n-1,n} \\ x_{n,1} & x_{n,2} & \cdots & x_{n,n} \end{pmatrix} \tag{2.1}$$

Where $x_{l,m}$ stands for the monetary flow from industry l to industry m.
Another representation of the sectoral map as an adjacency matrix is through the technical coefficients

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix} \tag{2.2}$$

Each technical coefficient is derived as ratio between the flow from i to j and the sum of all monetary flows from industry i, $a_{ij} = \frac{x_{ij}}{X_i}$ where

$$\sum_{j=1}^{n} a_{ij} = 1 \, \forall \, i. \tag{2.3}$$

An important accounting identity in the table is that gross output (GO) of each industry is equal to the sum of all uses of the output from that industry. In economics, in fact, gross output is the measure of total economic activity in the production of new goods and services in an accounting period. It is a much broader measure of the economy than gross domestic product (GDP), which is limited mainly to final output.

Leontief's quantity model

The output of a single industry can be written as

$$X_i = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n + d_i \tag{2.4}$$

where x_i represents the total output of industry i, d_i is the final output to consumer and a_{ij} are technical coefficients.
The (2.4) equation can be written as

$$x = Ax + d \tag{2.5}$$

Re-arranging:

$$x = (I - A)^{-1}d \tag{2.6}$$

Matrix $L = (I - A)^{-1}$ is known as Leontief inverse matrix.
Note that, x by definition is bigger than d, because some of the goods produced are used as intermediate inputs: not all the output leaves the system.

IOt assumptions and limitations

It's necessary to make a few remarks on IOt limits. As previously mentioned, the tables are filled with currency values. This means that is not possible to distinguish changes in quantities from changes in price. Moreover, some countries only record flow between sector down to a minimum size, with the flows below the threshold recorded as zeros or left blank, giving an incomplete graph topology.
Since several exchanges may pass through small links, the elimination of some minor flows could impact the network analysis. The cutoff flow size may represent as much as 30% of all n^2 links, where n is the number of sectors (J McNerney, 2009).

It should be taken into account, that deriving a limited number of sectors from the industrial scene requires some arbitrary choices, a different definition, in fact, may cause an industry in one country to appear smaller (or larger) than expected in comparison with other countries.

2.3 Input-output table analysis

2.3.1 Classical methods of analysis

Key sector

With the definition of 'key sectors' are meant those sectors supposed to pursue more effective industrial policies and to drive the economy towards increasing sectoral interdependence and higher income levels.
The first one to identify the importance of sectoral linkages for economic development was Hirschman (1958). In his work Hirschman stated that sectoral connections are a measure of production efficiency in an economy. According to his researches, within the industrial system, two mechanism subsist between each pair of industries: the direct backward linkage (or input-provision effects) and the direct forward linkage (or output-utilization effects). The first one reflects the sector potentiality to induce the supply of inputs by other sectors, while the second is a measure of the sector potential effect on the demand. Consequently, key sector could be $key\ pull\ sectors$, with an above average backward strength, or $key\ push\ sectors$, with a higher forward linkage index.
Under the input-output framework, the most known methods to identify key sectors are the Classical Multiplier Method (CMM) based on Rasmussen (1957) and the Hypothetical Extraction Method (HEM) introduced by Paelinck et al. (1965) and Strassert (1968).

Classical Multiplier method (CMM)

The CMM is owed to Rasmussen (Rasmussen, 1957), and it's based on the notions of backward linkages and forward linkages. He identified the column sum of the Leontief inverse as a measure of backward linkages and the row sum as an index of forward linkages. The column sum of L for sector j is given by

$$BL_j = \left[\sum_{i=1}^{n} a_{ij}\right] \tag{2.7}$$

also known as output multiplier or backward linkage. BL_j measures the direct and indirect output from every sector that leads to the production of one unit in sector j. In the same way is possible to derive the input multiplier or forward linkage

$$FL_j = \left[\sum_{j=1}^{n} a_{ij}.\right] \tag{2.8}$$

In order to compare each backward linkage with the average backward linkage in the economy, Rasmussen suggested the use of an index measuring the relative strength of a sector's BL:

$$BL_j = \frac{\sum_{i=1}^{n} a_{ij}}{(\frac{1}{n}) \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}} \tag{2.9}$$

where the numerator is the sector backward linkage and the denominator is the average of all backward linkages.

The output multiplier has been widely used and as a standard measure of backward linkages, while the forward linkage has been much more discussed in the literature. Miller and Blair (2009) evidenced as the method relies on an unrealistic hypothesis: a simultaneous unit increase in final demand in all sectors of the economy. Based on these arguments, the use of the row sums of the Leontief inverse as a measure of forward linkages has been reduced over time.

Hypothetical Extraction method (HEM)

The HEM is a technique developed to measure the value of a sector within an economic system. HEM quantifies the hypothetical output reduction in the overall economy resulting from the elimination of a sector of interest from the economic system.

Let j be the depicted sector, it is extracted from matrix A of technical coefficients. Let A' be $(n-1)(n-1)$ obtained after the elimination of the j-th column and j-th row.

Let X be a vector of outputs and Y a vector of final demands:

X' and Y' are the previous sectors after the extraction of the j-th sector.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{2.10}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{2.11}$$

The IO model can be calculated with the matrix equation (2.6). For matrix A' the Leontief inverse is derived from

$$X' = (I - A')^{\text{-}1} Y' \tag{2.12}$$

The result $i^{\mathrm{t}}X - i^{\mathrm{t}}X'$, given i^{t} an identity row vector of the right dimension, is the difference in the total production with and without the extracted value.
Other methods, instead of deleting the j-th row and columns, replace the interested values with zeros.

2.3.2 Application of network theory in the IO frame

The idea of studying the structure of an economy as a network was first proposed in 1978. Slater applied a maximum flow minimum cut algorithm from network theory to the 1967 United States input-output table to identify sector clusters, but then the literature went in silence until the past decade. Recent studies mainly focused on three areas. First, the positions of particular sectors in the IO network are examined using a variety of centrality measures. Second, researchers apply various methods from network analysis to identify clusters or communities in IO networks. Lastly, the structure of IO networks in relation to the performance of economies they represent: resilience against economic shocks and development level.

2.4 Construction of a country industry network

2.4.1 Graph construction

The national input-output tables processed in this thesis are taken from the World Input Output Database, whose latest version was released in 2016 (WIOD2016R,

for short). The database covers 43 countries for the period 2000-2014 and each national table collects data for 56 sectors, classified according to the International Standard Industrial Classification revision 4 (ISIC)[3].
The code written to derive the national graphs is developed in java and draws on a database derived from the re-elaboration of the IO tables.
The db contains a table of the industries[4], collecting their information: identification code (the primary key) and short description. Moreover, there is a national table for every country that brings together data from the IO historical set of the country: for each couple of sectors the table reports the money flow per year (expressed in Mn of US $)[5].
And thirdly, there's a GO table for each country reporting the gross output of every sector for all the years.
The code has been provided with all the methods developed in order to obtain easily the results needed.

2.4.2 Graphical output

The code includes a graphic interface developed according to the Model-View-Controller design pattern[6]. Launching the program, the graphic interface appears, the user can choose the desired country from a choice box[7], the year and the percentage weight that makes a connection significant to be graphically represented. There are two buttons, one that builds the graph and returns a set of its properties, while the other graphically represents the map.
The process behind the graph building is articulated as follows.
First of all, the program memorizes the country and the year concerned, returned

[3]ISIC is the international reference classification of productive activities, it's the standard for Countries to develop comparable national input-output tables. In the appendix table 2 presents the list of all the sectors with their Id and description.

[4]Each national i-o table has been aggregated considering the same division on 56 sectors.

[5]The period of time is from 2000 to 2014.

[6]Model–view–controller (usually known as MVC) is a software design pattern commonly used for developing user interfaces that divides the related program logic into three interconnected elements. This is done to separate internal representations of information from the ways information is presented to and accepted from the user.

[7]The possible countries are: China, France, Germany, Italy, Japan, Russia, Spain, UK and USA.

by the choicebox, and interacts with the db. The sectors are derived from the industries table and each one is memorized in an object, named $Sector$, with its identification code, name, gross output (GO) and all the relevant information for the analysis; each sector gross output is referred to the country and year under consideration and is obtained through a $join$ with the specific GO table matching the sectors' Id.

Then all the monetary flows are saved in a list as object $Linkage$, a specific class that allows to memorize the seller sector, the purchaser sector and the weight of each linkage.

The kind of graph used to represent the networks is a directed pseudograph[8] a type of graph that allows loops and has directed arcs.

The graph's vertices represent all the sectors, while the edges, rendered as arrows, indicate the money flow and the respective inverse flow of goods and services from one sector to another. Every edge is originated by the purchaser sector and reaches the sector that sells the good, the edge's weight, depending on the graph version considered, is either the monetary flow between the two sectors or its normalized version; the last one allows an analytical comparison between different countries regardless the differences in size[9].

Lastly, all the memorized sectors have been added as vertices and a directed edge for every connection has been added scrolling through the Linkage list. For what concern the graphic representation, nodes are drawn as circles, the higher the node's value, given by the sector GO, the bigger the circle's diameter. The nodes are filled in three different colors, green if the sector GO is less than 1% of the total, red if between 1% and 5%, and blue if larger than 5%.

Every vertex has at its right a label with its identification code and, when positioning the cursor on a vertex, a label with the description appears.

The vertices disposition in the canvas space is the result of a ponderation of the scalar attractive and repulsive forces between the nodes, the function that distributes the nodes considers the distance and the values of each couple of vertices. This permits to obtain a more harmonious and easier to read graph.

The national graphs are mostly fully connected, in fact, their graphical repre-

[8]the proper term should be pseudograph, but the class used in the code is a directed multigraph since java uses a different notation.

[9]Different works applied different normalization using for example the maximum flow in the network as denominator.

sentation doesn't help to identify the differences and similarities between the countries; therefore, it has been chosen to represent only the arcs whose weight is more than a certain percentage of the total amount of commercial exchanges. The default one is 0.01% but the user can choose also 0.05% or 0.1% from the specific radio button. In addition, arcs with a higher weight have been highlighted in dark blue. The nodes location is kept the same refreshing the year or the desired significative percentage.

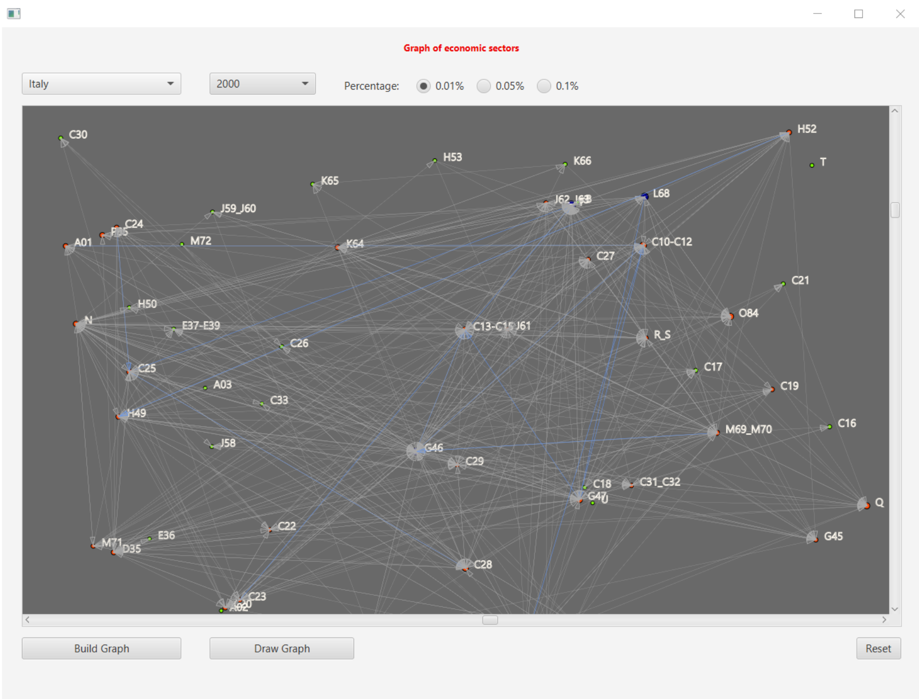


Figure 2.3: Graph representation

Chapter 3

Graph Analysis: illustration of the techniques to analyze the sector maps

This chapter outlines some of the metrics to analyze directed weighted graphs, with the intention of depicting the key sector of the considered economy.
The first section presents the most used measures to valuate a graph, from degree distribution to sector strengths, with a view to gaining a complete picture of the network: the peculiarity is that every national map has homogenous features.
The other sections of the chapter propose and explain measures that will be further applied separately for each national graph.
Sections 2 to 5 introduce each one a new index of analysis: assortativity, modularity, centrality and clustering. While the last paragraph proposes a failure tolerance analysis headed to reproduce the effect of a shock in the country network.

3.1 Properties of the country network

This section illustrates the characteristics of the sectorial maps built, all the observations made are thus related to every country table for the years from 2000 to 2014, unless otherwise specified.

3.1.1 Network topology

Considering all the 135 graphs built (9 countries x 15 years), it appears that they all have a high number of connections, with a density around 85-90%, except for China ($\sim$65%) and Russia ($\sim$35%). The high degree of completeness is a feature proper of highly aggregated I/O tables, as the national tables used in this essay; using a country structure with more industries, for example around 500 industries, the network shows to be 20% complete at that level of aggregation (Carvalho, 2009).

One of the most diffused methods to analyze a weighted directed graph is to approximate it by a binary graph, if the weights are sufficiently similar in size, and/or by an undirected graph, if the adjacency matrix is almost symmetric. However, this simplification is inadequate for the i-o graphs under consideration.

Figure 3.1 illustrates the absence of symmetry between reciprocal edges and the consequent impossibility to conduct this type of analysis[1]. The economic IO network is asymmetric, the size of flow f_{ij} is unrelated to f_{ji}, in the cloud of data there are flow from i to j having their reciprocal several times bigger. It also points out the different sizes of the flows and that an approximation with an undirected graph will compromise the amount of information each table carries. The asymmetry between out-flows and in-flows of industries implies an asymmetry between sectors as providers of goods and users of goods.

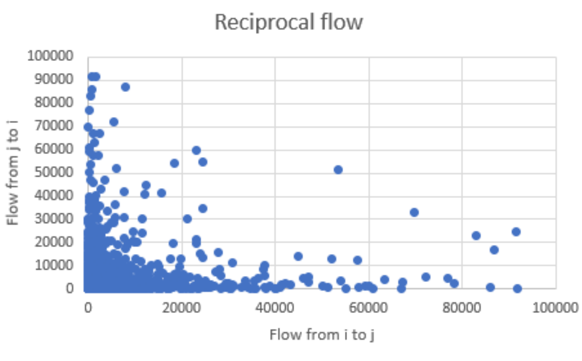


Figure 3.1: The scatter plot of size of a flow versus the size of the reciprocating flow, f_{ij} vs f_{ji}, Italy 2014

[1] The scatter plot represents the Italian economy in 2014, but the same features are evident in all the countries in the whole time horizon.

3.1.2 Nodes' categorization

The country networks are weighted and directed by construction, moreover, each of their nodes could be categorized as transmitter, receiver, ordinary or isolated, depending on its in and out-degree as is synthetically shown in the table.

Type of vertex	Incoming degree	Outgoing degree
Transmitter	= 0	> 0
Receiver	> 0	= 0
Ordinary	> 0	> 0
Isolated	= 0	= 0

Table 3.1: Vertices categorization

Taking into consideration all the 135 graphs, most nodes are ordinary (90.38%), the number of isolated nodes is significative (8.39%), but most of them are in Russia and China. Only France and China have transmitter nodes (just 7 over 7560 vertices). Receiver nodes are 1.14%, most of all in China, Japan and Russia.

3.1.3 Degree distribution

Unlike other network systems such as the internet, where the degree distribution follows the power law, the national tables are characterized by the highly left-skewed degree distributions. Most nodes enjoy high-degree connections because the industries are highly aggregated. Furthermore, every national table is almost complete with an average density of 80-90% (not considering Russia and China). The graphs below represent the (in or out) degree distribution considering all the countries[2] and the entire time period 2000-2014. The frequency spike is 54 for the in-degree and 52 for the out degree. The value 0 has a frequency of 3.5% as in-degree and 4% as out-degree, since the networks have also isolated nodes.

[2]Russia and China weren't considered because of their high number of isolated nodes.

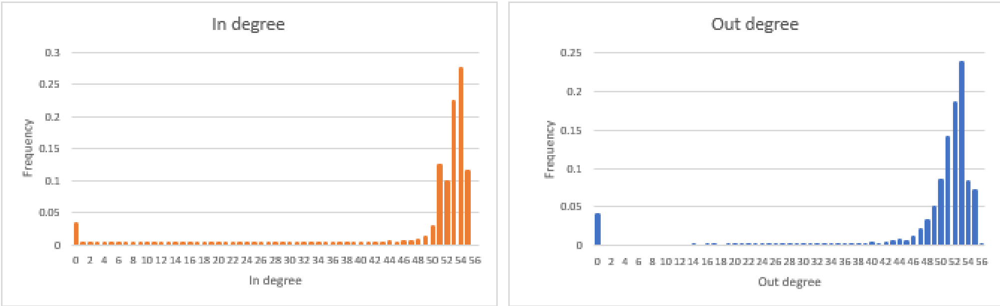


Figure 3.2: In-degree and out-degree distribution

3.1.4 Weights' distribution

To quantify the importance of a sector it has to be valued in the economic context. Measures as gross output and value added do not consider the relationships among the various nodes, so it's important to visualize how the flows are distributed.

Edge weights are the representation of the transfer of money from an industry to another. The magnitudes of money flows in different datasets differs because economies vary in size and inflation or deflation systematically change the nominal value of money[3].

In figure 3.3 the empirical complementary cumulative distribution function is plotted, from which appears that the tails of the national distributions have a power like distribution. The power law behavior showed by the tail characterizes also the global transfer network, as reported by the literature.

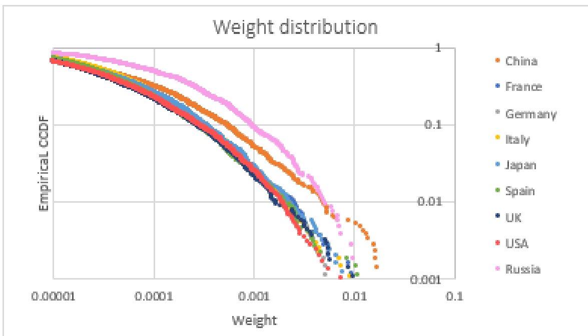


Figure 3.3: Weight distribution for each country, 2014

[3] In the WIOD all the flows are expressed in USA$.

In figure 3.4 three national weights distribution are presented, and, for each one the distribution's lowest moment has been compared with its power trendline.

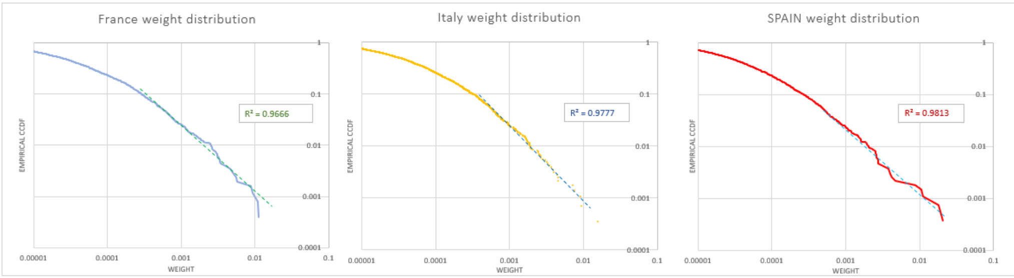


Figure 3.4: Italy, France and UK weight distribution

In *Network structure of inter-industry flows* (James McNerneya et Al, 2013) is illustrated how the weights distribution can be reconducted to a Weibull or to a log-normal distribution, evidencing for each country the best fit; with a sample of 64 countries, 31 appear to be better represented by a Weibull and 34 by a log-normal.

3.1.5 Sector Strengths

$Node\ strength$ generalizes the concept of node degree to weighted networks: it's an important measure to quantify the size of a sector. Node strength is the sum of the weights of the links surrounding a node. A directed graph has both in-strength and out-strength characterizing each node

$$s_i^{\text{in}} = \sum_j f_{ji} = F_{.i} \tag{3.1}$$

$$s_i^{\text{out}} = \sum_j f_{ij} = F_{i.} \tag{3.2}$$

$$s_i = s_i^{\text{in}} + s_i^{\text{out}} \tag{3.3}$$

The normalized in-strength and out-strength distributions are defined as the fractional contribution of each sector to the sum,

$$s_i^{\text{norm}} = \frac{s_i}{\sum_j s_j} = \frac{s_i}{w}, \tag{3.4}$$

with s_i as in(or out)-strength , and w as sum of the in(or out)-strengths in the whole network.

A sector's strength represents how much the sector provides to the other industries in the economy, or how much it uses from them; if a sector uses the outputs of many other sectors for its production, its in-strength tends to be higher (and if its output is used for production by many sectors, its out-strength is higher).
It's to note that the total provided to the economy doesn't include output for final consumption and input from labor: the in-strength and out-strength as defined here are sums over intermediate consumption only (James McNerney, 2009).
The disadvantage is that two vertices with the same strength can be linked to a different number of nodes, but having graphs with a density of 85-90% this issue has less relief.

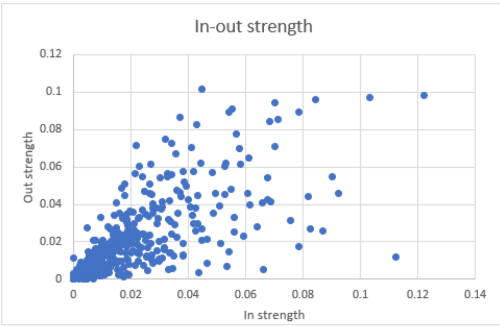


Figure 3.5: Scatter plot in-out strength considering all countries in year 2014

A strong correlation between the two measures is evident from graph 3.5, meaning that a sector receiving a high money value as in flows transmits heavy outflows to the receiving sectors.
The in and out strength distributions, shown in figure 3.6 appear to be exponential, with approximately the same slope for both[4]. The fact that the industries are both highly connected, as derived from the degree distribution, and asymmetrically connected, as captured by the flow distribution, implies that the local shocks are possible to propagate through the national economy and generate a sizable disturbance.

[4]The plots represent data for year 2014, but an analog behavior characterizes the whole period of time 2000-2014.

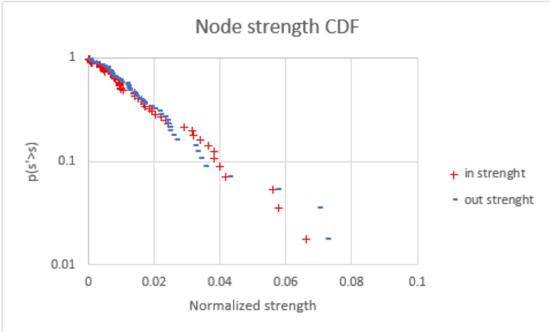


Figure 3.6: Strength Distribution - USA 2014

3.2 Assortativity

Assortativity (or assortative mixing) coefficient (r) measures the tendencies of nodes to connect with other nodes that have similar (or dissimilar) degrees as themselves.
It ranges from –1 to +1, a positive coefficient means that similar nodes are more likely to be connected, a negative one, instead, means that dissimilar vertices are more likely to be connected.
There are two types of network structures: assortative and disassortative. An assortative structure means that high degree vertices tend to connect with their similar, while in disassortative networks, high degree networks tend to connect with low degree vertices. A parameter that indicates whether the network is assortative or not is Knn_i that indicates the average degree of the first neighborhood of a node i of the network.

$$knn_i = \frac{1}{k_i} \sum_{j \in V(i)} k_j \tag{3.5}$$

where $V(i)$ is the set of vertices in the neighborhood of vertex i. From this parameter it is possible to build a graph called Spectrum of correlation that compares each node degree with its knn coefficient. If the network has an assortative structure, the function shows a positive trend, otherwise knn has a negative trend.

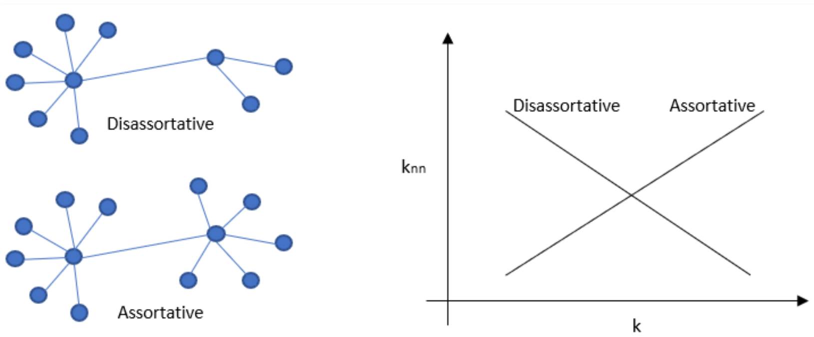


Figure 3.7: Assortative and disassortative structure

3.3 Modularity

3.3.1 Theory

The community detection problem is NP-hard, it's prohibitively difficult to solve exactly for large graphs, but a wide variety of heuristic algorithms exist.
Since the national graphs to analyze are relatively small networks, 56 vertices each, the algorithm doesn't pose restriction.
There are plenty of methods to find clusters, among these the Modularity maximization.
Since the countries' networks are directed, the directed generalization of modularity is seen. Modularity maximization involves searching for partitions of the network into communities that yield high values of the modularity Q over all possible partitions of the network.
The task then is to search over the many possible partitions of the network and find the one with the highest score. In practice, the number of partitions is usually extremely large, so that only a small fraction can be examined directly. Deterministic algorithms are problematic because they fail to show the many alternative partitions. In Network structure of inter-industry flows (McNerney et Al, 2013), a stochastic search algorithm based on simulated annealing[5] addresses this problem and returns a different high-scoring partition in each run. They ran the algorithm 100 times for each country, collecting the alternative partitions, and compare them to test their robustness from run to run.
Then, the frequency with which node i is in the same cluster of node j was mem-

[5] An optimal search algorithm that returns the n most connected clusters.

orized in the *coclassification matrix (CCM)*. The concept is that the presence of a nodes' group repeated frequently will appear as a group of high frequency in the CCM.
The clusters analysis conducted showed, unsurprisingly, that industries have a higher tendency to transact with other industries of the similar type.

3.3.2 Implementation

For the purpose of this dissertation it has been chosen to apply the method illustrated in *Community structure in directed networks* (E. Leicht et Al, 2007).
The method has been shown to be both computationally efficient and highly effective in practical applications. The premise of the modularity optimization method is that a good division of a network into communities will give high values of the benefit function Q,

$$Q = \text{(fraction of edges within communities)} - \text{(expected fraction of such edged)} \quad (3.6)$$

The modularity for directed graphs is derived as

$$Q(c_1, ..., c_n) = \frac{1}{m}\sum_{ij}(a_{ij} - \frac{k_i^{in}k_i^{out}}{m})\delta(c_i, c_j). \quad (3.7)$$

where $A_{ij} = 1$ if there is an edge pointing from node i to node j, c_i is the clustering community of node i, $m = \sum_{ij} a_{ij}$ the total weight of all edges. The delta function, known as Kronecker delta, returns 1 if c_i corresponds to c_j, 0 otherwise.
Note that edges make an higher contribute if k_i^{in} and/or k_i^{out} are small.
The step before the clusters' evaluation is their identification: this procedure starts with the division of the network into just two communities.
Let s_i be 1 if it's assigned to the first community, -1 if it's assigned to the second. Note that this implies that $\sum_{ij} s_i^2 = n$. Equation can be re-written as

$$Q = \frac{1}{2m}s^T Bs \quad (3.8)$$

where s_i is the vector whose elements are the s_i and B the so called modularity matrix with elements

$$B_{ij} = A_{ij} - \frac{k_i^{in}k_i^{out}}{m} \quad (3.9)$$

The goal is now to find s that maximizes Q for any given B.
Since in the undirected case the modularity matrix is symmetric, the lack of symmetry of the direct case is solved by using

$$Q = \frac{1}{4m} s^T (B + B^T) s \tag{3.10}$$

where matrix $B + B^T$ is symmetric per construction. Note that the constant $\frac{1}{4m}$ is conventional, but doesn't influence the positioning of Q maximum.

$$Q = \sum_i v_i^T (B + B^T) \sum_j a_j v_j = \sum \beta (v_i^T s)^2 \tag{3.11}$$

with β representing the eigenvalue of $B + B^T$ corresponding to eigenvector v_i. Assuming the eigenvalues to be labelled in decreasing order $\beta_1 > \beta_2 > .... > \beta_n$, under the constraint $s^T s = n$ the maximum of Q is achieved when s was parallel to the leading eigenvector v_1, condition forbidden by s_i assuming values 1 or -1. So, the optimal solution consists of a s vector whose element $s_i = 1$ if $v_i^{(1)} > 0$ and $s_i = 0$ if $v_i^{(1)} < 0$[6]. where $v_i^{(1)}$ is the i-th element of the eigenvector v_1.
There are a variety of ways of generalizing the approach to more than two communities but the simplest is repeated bisection.
Thus, the algorithm consists of dividing the network in two and then divide the groups and so on, until the modularity coefficient stops to increase. The subdivisions following the first step require a generalization of the method above, considering the change in modularity of the entire network when a community g within in is divided.

$$\Delta Q = \frac{1}{2m} [\sum_{i,j \in g} (B_{ij} + B_{ji}) \frac{s_i s_j + 1}{2} - \sum_{i,j \in g} (B_{ij} + B_{ji})] s \tag{3.12}$$

$$= \frac{1}{4m} \sum_{i,j \in g} [(B_{ij} + B_{ji}) - \sigma_{ij} \sum_{k \in g} (B_{ij} + B_{ji})] \tag{3.13}$$

$$= \frac{1}{4m} s^T (B^{(g)} B^{(g)T}) s \tag{3.14}$$

having

$$B_{ij}^{(g)} = B_{ij} - \sigma \sum_{k \in g} B_{ik} \tag{3.15}$$

[6]if $v_i^{(1)} = 0$ the sign of s_i is equally good.

where $B^{(g)}$ is the submatrix of B having the sum of each row subtracted from the diagonal element. Although $B^{(g)}$ is not symmetric $B^{(g)} + B^{(g)T}$ is symmetric and it's possible to apply the same method as before.
So, after deriving this symmetric matrix, its eigenvalues and eigenvectors, matrix s is valorized and the cluster partitioned. The algorithm stops if doesn't find divisions giving a positive value of ΔQ, when every community reaches this state the algorithm ends.

3.4 Centrality

3.4.1 Centrality for weighted directed graphs

Centrality in network analysis is a measure of a node's importance. There is a number of measures of centrality in use. Three of the most used are betweenness centrality, closeness centrality, and eigenvector centrality.
Most of the centrality indices are only concerned on the presence or absence of a tie between two nodes (binary network), and when applied on weighted networks (in which an attribute is used to weight the tie between nodes) a loss of information occurs. Moreover, betweenness and closeness centrality are based on the shortest path, a concept that has low impact on the national graph studied because of the high number of connections.
In the context economies the most appropriate centrality index is the eigenvector centrality (James McNerney, 2009). In graph theory, eigenvector centrality (also called eigencentrality or prestige score) is a measure of the influence of a node in a network. Relative scores are assigned to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. The con of eigenvector centrality is that it can give strange results for directed graphs, in particular, if a vertex is not in a strongly connected component of size at least 2, then its eigenvector centrality will be 0. If the country network is dense enough, every country i both imports to and exports from every other country j, then this won't be an issue.
In this work it has been chosen to apply the *entropy based centrality*, a new centrality index suitable for weighted directed graphs, introduced and explained in

detail in *A Novel Entropy-Based Centrality Approach for Identifying Vital Nodes in Weighted Networks* (Tong Qiao et Al, 2018).

3.4.2 Entropy based centrality

The total influence of a node is made of two parts: its local power and its indirect power.
First, the complete network has been deconstructed into several serval subnetworks centered on each node. The out-degree centrality of a node interprets the purchasing flow, and the in-degree centrality reflects its sales.
Given a directed weighted graph $G(V, E, W)$, for a vertex v_i, the i-centered subgraph represented by G_i is composed by node i and its neighbors. The subgraph degree centrality of node i and its neighbor j, referred as SDC_i, equals the sum of its in-degree centrality and its out-degree centrality, namely $SDC_i = DC_i^{\text{in}} + DC_i^{\text{out}}$.
If a node has only the in-degree centrality non null, the SDC of that node is given by the $SDC_i = DC_i^{\text{in}}$ and the same goes for the out-degree centrality.
In order to quantify the local power of a given node, the novel definition of entropy centrality takes both structural entropy and frequency entropy into consideration. The structural entropy, based on the topographic properties of the subgraph, evaluates the activity and strength of a given node in specific subnetwork. The frequency entropy, which takes advantage of information contained in the weights of directed links, shows the accessibility of a given node.
The structural information entropy for node i is stated as

$$I_i^{\text{s}} = -\sum_{i=1}^{M+1} \frac{SDC_i}{\sum_{i=1}^{M+1} SDC_i} log \sum_{i=1}^{M+1} \frac{SDC_i}{\sum_{i=1}^{M+1} SDC_i} \tag{3.16}$$

where M denotes the number of neighbors of node i.
The weight of directed links acts could be read as an indicator of the interaction frequency. Of course, it's clear that this is an approximation, since a high edge's weight between two industries doesn't necessary indicate an high frequency flow, but it could be due to the elevate price of the goods.
The previous consideration leads to the definition of the interaction frequency

entropy of node i as follows,

$$I_i^{\text{f}} = -\sum_{j=1}^{M+1} \frac{f_{ij}}{\sum_{j=1}^{M+1} f_{ij}} log \sum_{j=1}^{M+1} \frac{f_{ij}}{\sum_{j=1}^{M+1} f_{ij}} \tag{3.17}$$

where f_{ij} indicates the money flow of the directed edge from i to j.

The local influence of node i on its one-hop neighbors, denoted as LI_i, equals the summation of the structural information entropy and the frequency entropy, multiplied by two coefficients respectively

$$LI_i = w_1 * I_i^{\text{s}} + w_2 I_i^{\text{f}} \tag{3.18}$$

where w_1 e w_2 represent the weight coefficients, respectively, and $w_1 + w_2 = 1$.

The indirect power of a node in the network is calculated through the two-hop influence propagating model, whose core assumption is that meaningful influence can no longer be detectable beyond the boundary of three or four degrees.

The indirect influence of node i on its two-hop neighbor p is

$$II_{ip} = \sum_{p=1}^{N_{ip}} \frac{LI_i * LI_j}{N_{ip}} \tag{3.19}$$

where N_{ip} is the total number of common one-hop neighbor between vertices i and p. The indirect influence on vertex i two-hop neighbors, is denoted as

$$II_i = \frac{\sum_{p=1}^{M_i} II_{ip}}{M_i} \tag{3.20}$$

where M_i corresponds to the number of two-hop neighbors of node i.

Therefore, the total influence of node i is

$$I_i = \theta_i LI_i + \theta_2 II_i \tag{3.21}$$

3.5 Clustering

The study of community structure in networks has a long history and their detection is of great importance in all the disciplines where systems are often represented as graphs.

One of the most used metrics is the clustering coefficient, capable to quantify the tendency of vertices to collect in clusters, with many edges joining vertices of

the same cluster and comparatively few edges joining nodes of different communities. Such clusters can be considered as fairly independent compartments of a graph.
The identification of industrial clusters using input–output tables (IOTs) has received increasing attention from regional economists, economic geographers and policymakers. Clustering allows industries to develop economies of scale, raise their productivity, decrease transaction costs, and encourage innovation (Porter, 1990).

3.5.1 Clustering Coefficient

The clustering coefficient was originally introduced for binary, undirected graphs, and the idea behind is that the extent to which i's neighborhood is clustered can be measured by the percentage of pairs of i's neighbors that are themselves neighbors.
Node's i clustering coefficient (C_i) in undirected networks can be defined as the ratio between all triangles actually formed by i and all the triangles i could have formed. Therefore

$$C_i(A) = \frac{\frac{1}{2}\sum_{j\neq i}\sum_{h\neq(i,j)} a_{ij}a_{ih}a_{jh}}{\frac{1}{2}d_i(d_i-1)} = \frac{(A^3)_{ii}}{d_i(d_i-1)} \tag{3.22}$$

where A is the adjacency matrix, with $a_{ij} = 1$ if there is a connection between i and j, 0 otherwise, $(A^3)_{ii}$ individuates the ith element of the main diagonal of A^3. Each product $a_{ij}a_{ih}a_{jh}$ is meant to count whether a triangle exists or not around i, while d_i is the degree of node i and the denominator quantifies the number of triangles that could possibly exist.
Fagiolo's work (Fagiolo, 2007) presents a weighted, directed version of the clustering coefficient for any type of triangle pattern.
First, the clustering coefficient for a weighted undirected graph is introduced.
Given the NxN matrix W, symmetric, where w_{ij} takes account of the amount exchanged between sectors i and j (normalized as a fraction of the receiving sector money in-flow for the year considered). In this case, without a loss of information, it can be stated that $w_{ii} = 0$ for all i. Let the clustering coefficient be

$$\tilde{C}_i^D(A) = \frac{\frac{1}{2}\sum_{j\neq i}\sum_{h\neq(i,j)} w_{il}^{1/3}w_{ih}^{1/3}w_{jh}^{1/3}}{\frac{1}{2}d_i(d_i-1)} = \frac{(W^{1/3})_{ii}^3}{d_i(d_i-1)} \tag{3.23}$$

where $W^{\frac{1}{3}} = w_{ij}^{\frac{1}{3}}$ is obtained from W by taking the 3rd root of each entry and $\frac{1}{2}d_i(d_i - 1)$ is the maximum number of triangles including i as a vertex.
The clustering coefficient captures the chunkiness of the network at the level of triangles of nodes, so, analyzing a weighted directed network, it's important to consider the edges' directions in each triangle.
Given matrix W, now w_{ij} is a value proportional to the weight of the money flow from i to j in the network. The index considers the number of directed triangles formed by i and valorized through the edges' weights

$$\tilde{C}_i^D(A) = \frac{\tilde{t}_i^D}{T_i^D} = \frac{[(W^{1/3} + (W^t)^{1/3}]_{ii}^3}{2[d_i^{tot}(d_i^{tot} - 1) - 2d_i^{\leftrightarrow}]} \tag{3.24}$$

where d_i is the degree of node given by the sum of its in- and out-degree. Node i can be possibly linked to a maximum of $d_i^{tot}(d_i^{tot} - 1)$ pairs of neighbors and with each pair can form up to two triangles as the edge between them can be oriented in two ways. However, this number also counts "false" triangles formed by i and by a pair of directed edges pointing to the same node $d_i^{\leftrightarrow}$ and for each of them there are 2 false triangles.
The overall clustering coefficient for each type of graph is

$$CC = N^{-1} \sum_i^N C_i \tag{3.25}$$

in particular for the WDG

$$CC = N^{-1} \sum_i^N \tilde{C}_i^D \tag{3.26}$$

The CC for weighted directed networks (3.26) considers all the possible triangles as their direction was irrelevant.
In directed graphs, triangles with edges pointing in different directions have a completely different interpretation. Looking at figure 3.8 there are 4 possible patterns of directed triangles for i's perspective: cycle, middleman, in and out. (i) $Cycle$ triangles with a cyclical path, (ii) $middleman$ having one of i's neighbors reaching both the other not-i nodes directly and passing through i, (iii) in with two inward edges to i and (iv) out where i holds two outward edges.
In order to measure clustering in directed network, it's important to distinguish between the motifs in figure 3.8.

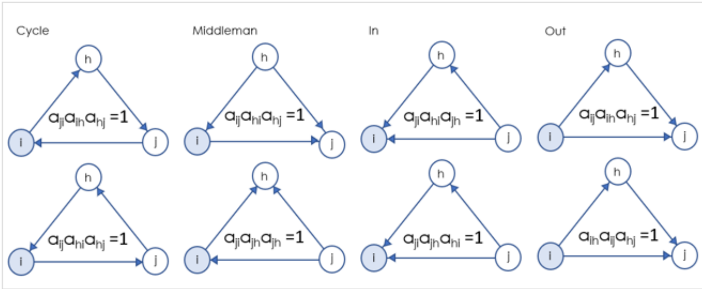


Figure 3.8: All eight different triangles with node i as one vertex

Each CC is derived from the ratio between the number of triangles of that pattern with i as vertex and the total number of triangles of that pattern that i can form. The CC for each pattern is given by

$$\tilde{C}_i^* = \frac{\tilde{t}_i^*}{T_i^*} \tag{3.27}$$

where $*$ individuates the specific typology. Let $T_1^{cyc}, T_2^{mid}, T_3^{in}, T_4^{out}$ the maximum number of triangles per type that i can possibly form, and $t_1^{cyc}, t_2^{mid}, t_3^{in}, t_4^{out}$ the actual number of triangles per type
Notice that,

$$T_i^D = d_i^{tot}(d_i^{tot} - 1) - 2d_i^{\leftrightarrow} = T_1^{cyc} + T_2^{mid} + T_3^{in} + T_4^{out}. \tag{3.28}$$

and the total clustering coefficient is

$$\tilde{C}_i^* = \frac{t_1^{cyc} + t_2^{mid} + t_3^{in} + t_4^{out}}{T_i^D} \tag{3.29}$$

Type	$\mathbf{t_i^*}$	$\mathbf{T_i^*}$
cycle	$\hat{W}_{ii}^3$	$d_i^{in} d_i^{out} - d_i^{\leftrightarrow}$
middleman	$(\hat{W}\hat{W}^t\hat{W})_{ii}$	$d_i^{in} d_i^{out} - d_i^{\leftrightarrow}$
in	$(\hat{W}^t\hat{W}^2)_{ii}$	$d_i^{in}(d_i^{in}-1)$
out	$(\hat{W}\hat{W}^2)_{ii}$	$d_i^{out}(d_i^{out}-1)$

Table 3.2: Number of possible triangles for each pattern

3.6 Cascading failure tolerance Analysis

The cascading failure tolerance analysis allows to classify sectors according to their diffusion properties, a sector with good diffusion properties distributes the effects of a shock to a wider set of industries as oppose to concentrate the effect into itself or just a few others.

The idea to reproduce a cascading failure tolerance analysis has been taken from *Ranking the economic importance of countries and industries* (Wei Li et Al, 2017). Their goal was to introduce a methodology to quantify the importance of a given industry belonging to a given country for the global economic stability with respect to other industries, both in the same countries and in others.

They assumed that the failure of an industry in country A does not reduce the revenue of the other industries in that same country A, because the government is able to make a quick adjustment, such as a central bailout, in order to mitigate the impact to other industries within the country.

Here, the cascading failure analysis has been applied within each country and within the global network, to evidence the connections among the industries. The simple idea behind this analysis is that, if industry i fails, other industries as suppliers won't be able to sell their goods to industry i. The loss of each industry is a reduction of revenues by a fraction p', which is the revenue reduction caused by industry i and divided by the industry j total revenues:

$$p'_{country,year} = \frac{x(i)}{y(j)} \tag{3.30}$$

The tolerance fraction p represents the threshold above which the revenues reduction has a significative impact on the industry, to the point of not manage to operate normally. When the reduced revenues fraction $p'(i)$ is larger than the threshold, industry i fails. But, the failure of industry i, could trigger a series of cascading failures.

$$\begin{cases} fails \quad if \quad p' > p. \\ survives \quad if \quad p' \leq p. \end{cases} \tag{3.31}$$

Here, we assume that the threshold is the same for all industries, and that every industry fails when its $p' > p$.

The methodology can be schematically illustrated as follows. In step 1, industry i fails. This causes other industries to fail if their $p' > p$. Assume that in step

2 industries j, k and l fail. The failure of these industries in step 2 will reduce other industries' revenue and cause more industries to have a reduced fraction $p'(i)$.

Eventually, the system reaches a steady state in which no more industries fail. The surviving industries will all have a reduced revenue fraction that is smaller than the tolerance fraction.

To determine how much the failure of each industry would impact the stability of the economic network, the tolerance fraction p varies from 0 to 1 and measure the fraction of surviving industries left in the network. When the tolerance fraction approaches 0, any revenue reduction caused by the failure of one industry can easily destroy almost all the other industries in the network, and the network will collapse. When the tolerance fraction approaches 1, all the industries can sustain a large reduction of revenue, and the failure of one industry will not affect the others.

Chosen the percentage of the networks that is wanted save after the shock, the algorithm is run for every sector. The result is a threshold for every sector[7] that corresponds to the maximum revenue reduction percentage above which the sector receiving the money flow fails. The higher the threshold the more important the industry is in the spreading of the shock.

[7]The threshold applies to all the money flows reduction in the network.

Chapter 4

Identification of the Italian key sectors

This chapter aims at finding the most important sectors in the Italian economy. The first paragraph presents an outlook at the Italian socio-economic context, while the following sections deal with the analysis of the Italian graph. The cardinal concept in the development of the analysis is that importance of a node in complex networks is based on the view that it is important if it occupies a strategic position in the graph.

4.1 Outlook on the Italian economy and the industries' relations

4.1.1 Evolution of the Gross domestic product

The Italian economy is one of the most influent economies in the world, it occupies the ninth position in the international ranking for size.
In 21st century the Italian economy entered a phase of substantial stagnation, characterized by extremely low growth.
The biennium 2007-2008 were the years of the USA financial crisis, during the early phases of the financial collapse, Italian banks and investors had suffered

little since they were not very heavily involved in highly speculative activities. Italy began to show sign of the crisis in the second quarter of 2008 with the pouring of the financial crisis on the real economy; as it possible to observe in figure 4.1, the GDP growth was positive until 2008 when started its negative path, despite of a brief period of recovery during 2010 and early 2011, with the worsening of the sovereign debt crisis of some European countries and with the growing climate of distrust towards Italy, high-debt country, the economy once again encountered a setback and a new recession began[1].

In the period of the great recession (2008-2013) the Italian economy suffered the loss of a substantial part of its production capacity, but in the main time its structure has evolved. Figure 4.2 shows as GDP starts decreasing in 2008, how there is a slight recovery around 2010, but it also evidences that the Italian GDP has not yet reached the amount it was in year 2000. The last months of 2014 showed weak signs of recovery, in a framework still characterized by recessive trends that affected both the manufacturing industry and the service sectors more related to industrial demand.

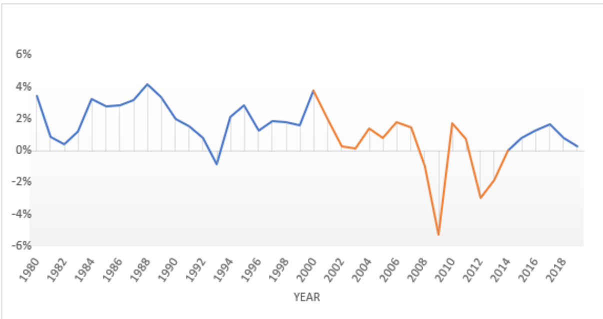


Figure 4.1: Italian GDP growth rate per year

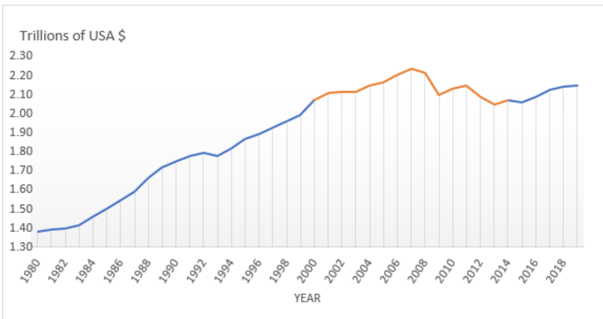


Figure 4.2: Italian GDP distribution in USA$ for the period 1980-2019

[1]The GDP data have been taken from www.worldbank.org

4.1.2 Economy structure and export

Italian industry is dominated by small and medium-sized enterprises, mostly manufacturing, while large enterprises are few. In a developed economic system, services play an increasingly important role; in Italy they explain about 73% of the added value of the economy, compared with the 20% represented by the manufacture.

Much of the service provision is directly related to the production process (about 40% of the overall added value of the economic system). In particular, the use of services as intermediate inputs affect the productivity of the economic system either indirectly (through outsourcing production or abroad displacement of low value added activities towards external suppliers), both directly. In figure 4.3 are presented the six sectors for highest value added complemented by their gross output, the industries don't show the same ranking for the two metrics but there is a positive correlation between the two indices.

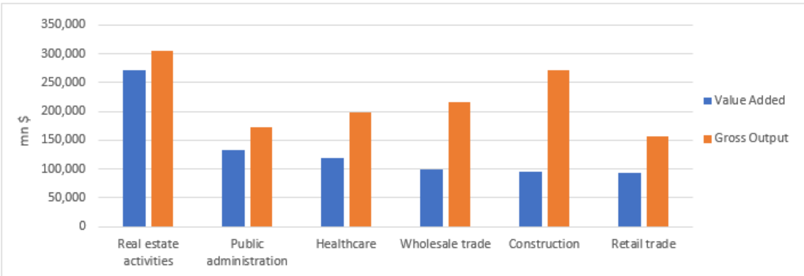


Figure 4.3: The 6 sectors with the highest value added, Italy 2014

Made in Italy is one of the greatest prides of the country, it's a sign of top quality products and exclusive design apart. The impact of exportations on the national GDP varies around 25% and 30%.

Table 4.1 shows the sectors with the highest export amount; these industries have a stable position in the top ranking of exportations, while the others change slightly position in the years. Italy is known worldwide for the products' quality and it isn't surprising that the Italian export is based primarily on manufacturing, with manufacture of machinery, textiles and food leading the way. Table 4.1 ranks the sectors which contribute most to exports.

The export performance in the considered period of time shows a tendency to increase. It's to be noted that the two decreases, in year 2009 and 2012, should be considered in the context of the financial crisis and that, despite of the improvement in 2014 the exports amount is not yet at the pre-crisis level.

Sector	Id	Mn $	Weight %
Manufacture of machinery and equipment	C28	91,521	7.58%
Manufacture of textiles, wearing apparel and leather products	C13-C15	56,391	7.49%
Manufacture of motor vehicles, trailers and semi-trailers	C29	38,180	6.50%
Manufacture of food products, beverages and tobacco	C10-C12	36,430	6.35%
Manufacture of basic metals	C24	30,590	6.16%
Manufacture of chemicals and chemical products	C20	30,412	6.01%
Manufacture of fabricated metal products	C25	28,131	5.59%
Manufacture of electrical equipment	C27	25,696	5.30%
Manufacture of basic pharmaceutical products	C21	24,989	4.88%
Manufacture of furniture; other manufacturing	C31_C32	24,695	4.18%

Table 4.1: Ranking of industries for export amount, an extract of the 10 best exporter Italy 2014

In the biennium 2013-2014, the growth in exportation has involved mainly sectors with low propensity to export, such as manufacture of other transport equipment(no motor vehicles or trailers), health and social activities, postal services and scientific research; all sectors with low export amount. Graph 4.4 shows the export development with consideration on the role the five most important industries play, the height of the bar represents the total export amount and for every year the development of the five main sectors is highlighted; from the graph emerges that in the whole period machinery and textiles industries have played a dominant role with a tendency to increase.

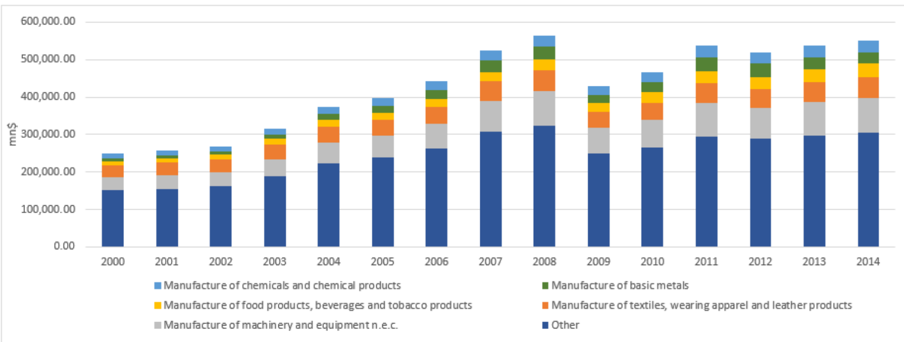


Figure 4.4: Exports in the time period 2000-2014, with focus on the main exporter sectors

4.2 Italian network

4.2.1 Density and graphical representation

A first structural property of the economy is the density of the network, index of the amount of interactions between sectors relative to the total number of connections that could exist if all nodes were linked.
The density of the Italian graph in year 2014 is 89.67% that means the network is really aggregated and each single node is almost connected with all the others. The Italian density has always been around 90%, in the period considered, and it has an increasing trend.

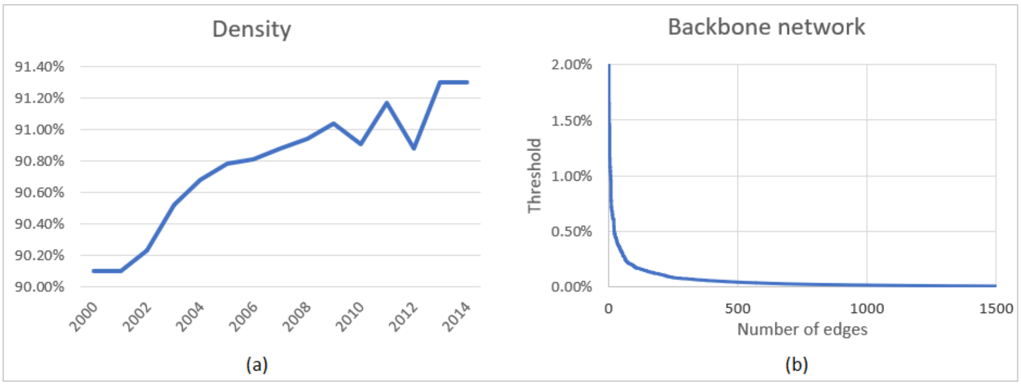


Figure 4.5: (a) Italian graph density evolution in the years (b) number of edges in the national Italian graph with the threshold applied

The graphic representation of the Italian economy, figure 4.6 and 4.8, show only the backbone of the network, to have an easier to read graph, because with a density around 90% the complete graph may be confusing.
Identifying the backbone of a network is particularly interesting for IO networks, by eliminating less important links, one can find the most essential structure without being overwhelmed with massive amount of non-essential links. The backbone of an economy extracted from the corresponding IO network reveals the most essential linkages of all sectors of the economy. The stability of this backbone makes the entire economy resilient even if some non-essential linkages are weakened or broken. Early popular approaches to extract network backbone include the minimum spanning tree method which removes links of a network to minimize the total link weights while keeping all nodes connected or the use of a predefined threshold. Here the second approach was applied, because it allows

to visualize the most important connections and how they change in the time and since the spanning tree would reduce the network flows represented to around 2% of the exiting connections.
The Italian map changes in the years, comparing figure 4.6, representing the situation in 2000, to figure 4.8, dated 2014, it's possible to notice some differences. In the latest graph, some nodes (E37-E39) have a higher gross output, in fact the nodes have a higher radius and a different color, or a lower one(C23)[2].
The number of blue arrows, the heaviest arcs in the network, is almost the same in both periods. In both the years every "blue node" has at least one influent edge, some of these edges also connect red nodes but there are few green nodes connected by such edges, meaning that large money flows are not sporadical but they go with bigger sectors. In addition, in the time period considered there was not an homogeneous growth in the flows, but some connections have grown more in spite of others. However, not all the heaviest connections in 2000 persists also in year 2014, G46, strongly connected to M69-M70, and C13-15, transferring heavy flows to G46 and G47, are no more heavily connected in 2014; In the second graph, G46 develops a new strong connection with sector K64 and keeps the one from C10-C12. K64 receives also a heavy money flow from vertex L68.
In both years the isolated nodes are the same, isolated nodes in this representation could also identify nodes connected but whose connections carry a low amount of money, among these A02 (forestry), A03 (Fishing), T(Extraterritorial bodies) and U (household as employees). The local centrality index shows how the edges in the economy have both in-degree and out-degree around 50, except for these industries. While the forestry and the fishing sector have a discrete number of connections in both years, sector T and sector U stay isolated.

[2]As explained in chapter 2.4 the nodes have a dimension proportional to their gross output and the colors help to figure out the changes; in ascendant order of GO size there are green, red and blue nodes.

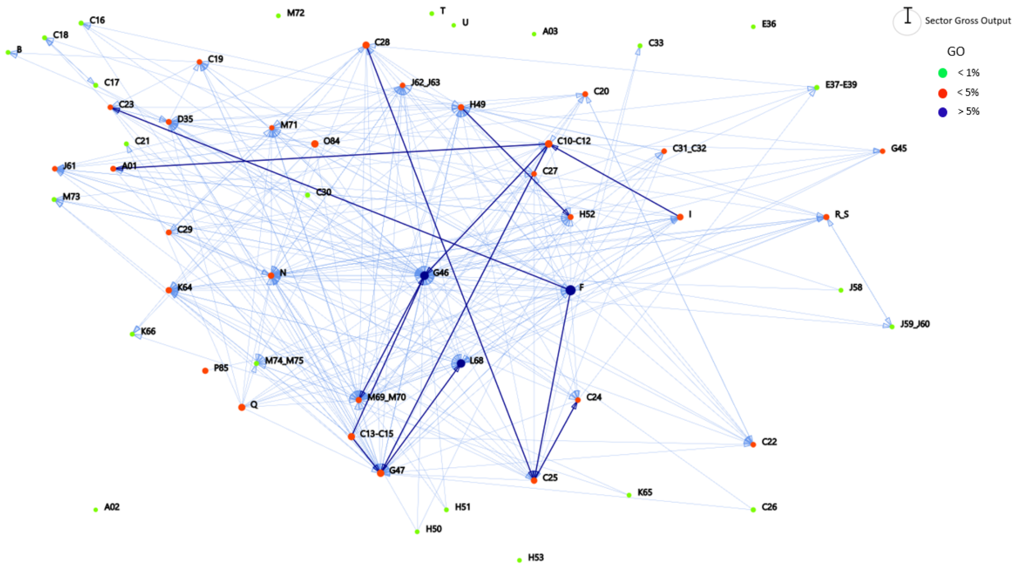


Figure 4.6: Italian graph year 2000, backbone representation with 0.03% threshold

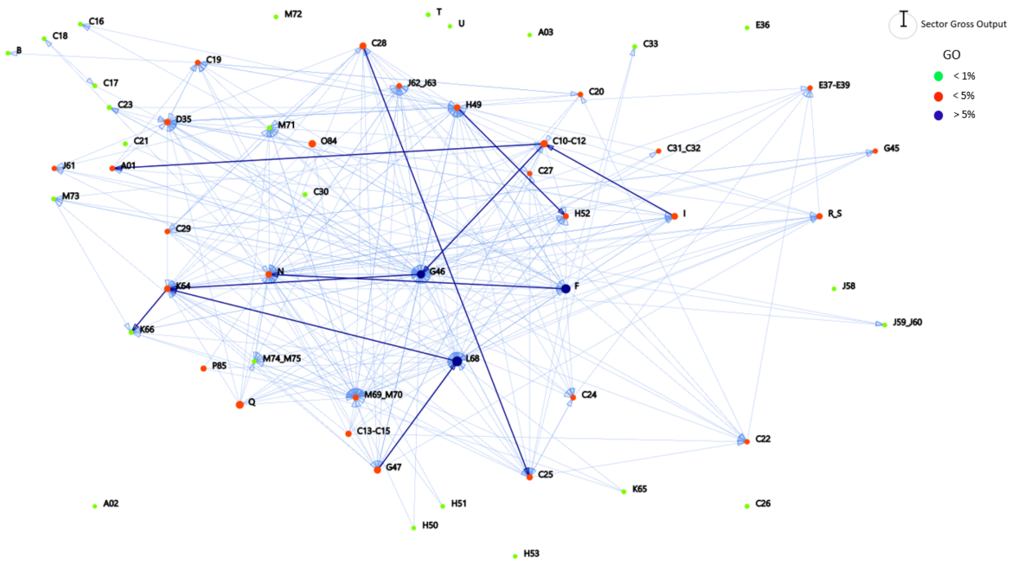


Figure 4.7: Italian graph year 2014, backbone representation with 0.03% threshold

4.2.2 Modularity

This section shows the result application of the partitioning method illustrated in section 3.4.1.
The group extraction is based on the condition that each partitioning increases the modularity of the network, until there is an increase the process run; the procedure has led to the identification of 6 groups listed below, except for the biggest one that includes all the remaining nodes.
Group 1 consists of manufacturing industries of materials (metal and plastic), construction and insurance. Group 2 agglomerates industries in the transport business, water, air and couriers. Group 3 includes legal and administrative services and the forestry and logging sector. Group 4 includes sectors belonging to the social and academical domain, while group 5 encloses several services industries plus manufacture of foods, chemicals and furniture.

- **Group 1:** Manufacture of rubber and plastic products, Manufacture of basic metals, Manufacture of fabricated metal products (except machinery and equipment), Construction and Insurance

- **Group 2:** Water transport, Air transport and Postal and courier activities

- **Group 3:** Legal and accounting activities (head offices and management consultancy activities), Administrative and support service activities and Forestry and logging

- **Group 4:** Education, Human health and social work activities and Other professional, scientific and technical activities; veterinary activities

- **Group 5:** Other service activities, Real estate activities, Computer programming (consultancy and related activities, information service activities), Accommodation and food service activities, Warehousing and support activities for transportation, Retail trade (except of motor vehicles and motorcycles), Wholesale trade(except of motor vehicles and motorcycles), Wholesale and retail trade and repair of motor vehicles and motorcycles, Sewerage; waste collection, treatment and disposal activities; materials recovery; remediation activities and other waste management services, Repair and installation of machinery and equipment, Fishing and aquaculture, Manufacture of food products, beverages and tobacco products, Manufacture of chemicals and chemical products and Manufacture of furniture; other manufacturing

This analysis sees the national network only the presence and direction of the edges, but not their weight. Group 5 is made of 14 sectors where each one is connected with all the others and can't be further divided. Also, the sectors belonging to the 6th group, not listed here for reading reasons, are fully connected; Activities of extraterritorial organizations and Activities of households as employers are instead isolated.

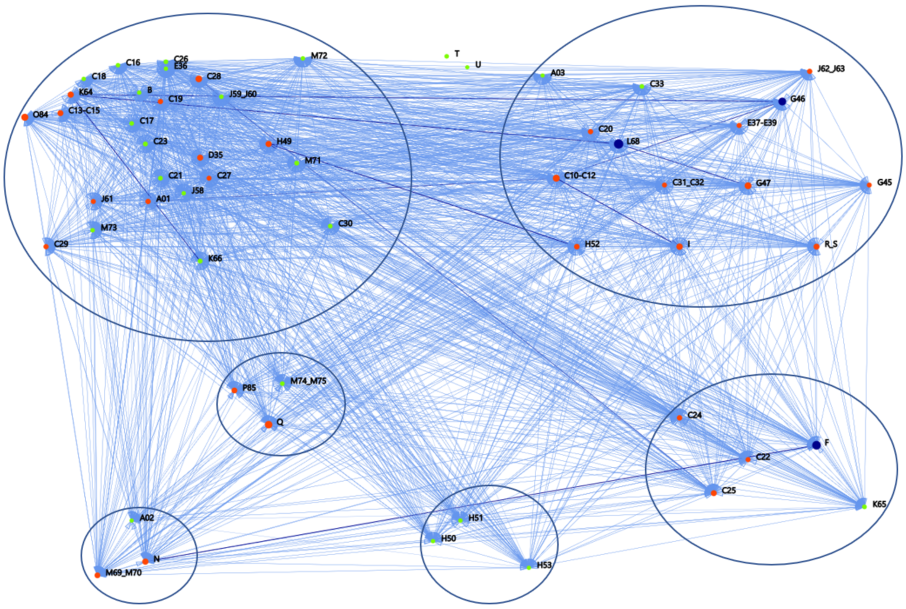


Figure 4.8: Clusters, Italy 2014

4.2.3 Centrality

Centrality measure

The nodes' centrality has been measured through the entropy based centrality index introduced in chapter 3.2.2.

The index is obtained by the combination of the Local influence and of the Indirect influence, however, for the purpose of this study greater importance has been attached to the local index.

The Local Influence coefficient is itself a combination of two analysis, one on the structural connections and the other about the frequency of the interactions. The Indirect Influence coefficient quantifies the ability of a node to reach, through its first level neighbors, nodes not in its subgraph, the reason why it is frequently 0 depends on the nodes' high degree; in fact, the more a node is connected the less number of second level neighbors it has, and if the node is completed connect except for isolated nodes this coefficient happens to be zero, a recurrent condition in the graph. Both indices ranking have a week correlation with the strength distribution.

Rank	Sector	Id	Centrality
1	Manufacture of other transport equipment	C30	0.9537
2	Administrative and support service activities	N	0.9485
3	Manufacture of computer, electronic and optical products	C26	0.9457
4	Repair and installation of machinery and equipment	C33	0.94
5	Wholesale and retail trade and repair of motor vehicles and motorcycles	G45	0.9397
6	Manufacture of basic pharmaceutical products and pharmaceutical preparations	C21	0.9372
7	Wholesale trade, except of motor vehicles and motorcycles	G46	0.9337
8	Publishing activities	J58	0.9222
9	Scientific research and development	M72	0.9214
10	Mining and quarrying	B	0.9185

Table 4.2: The best 10 sector for local centrality index

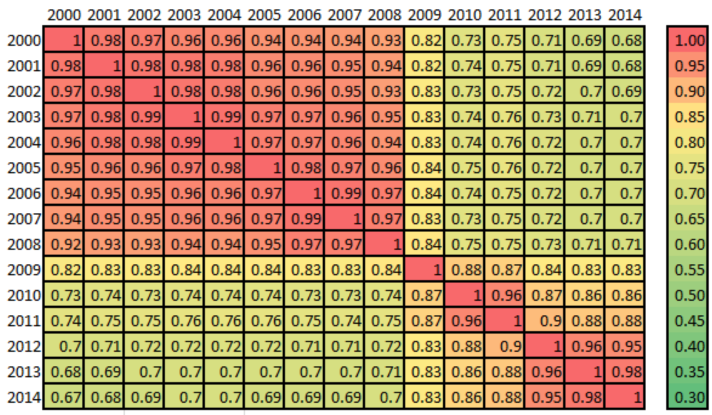
Figure 4.2 lists the ten sectors with the highest centrality index, among these it's possible to individuate similar sectors; there is a set of sectors related to the machinery manufacture and wholesale transport field (C30, C33, G45 and G46); moreover, there are 2 sectors of the research area, M72 and J58; the remaining sectors are two manufacturer industries, pharmaceutical (C21) and electronic (C33), Administrative activities (N) and Mining and quarrying (B). Among these sectors just N, C33, G45 and G46 have 54 connections, the maximum degree possible in the network since over 56 vertices 2 stay isolated, but also the other are strongly connected (both their in and out degree are still more than 50).
The remaining sectors are in fact in the top 6 sectors for indirect centrality in table 4.3. In fact, considering both indices for not fully connected vertices, emerges that the ranking is almost the same, so for the purpose of this analysis it has been chosen to prioritize the local index ranking.

Rank	Sector	Id	Centrality
1	Manufacture of other transport equipment	C30	0.9289
2	Manufacture of computer, electronic and optical products	C26	0.9207
3	Manufacture of basic pharmaceutical products and pharmaceutical preparations	C21	0.912
4	Publishing activities	J58	0.8978
5	Scientific research and development	M72	0.897
6	Mining and quarrying	B	0.8942
7	Other professional, scientific and technical activities; veterinary activities	M74_M75	0.8932
8	Telecommunications	J61	0.8887
9	Manufacture of motor vehicles, trailers and semi-trailers	C29	0.8834
10	Manufacture of electrical equipment	C27	0.8773

Table 4.3: The best 10 sector for indirect centrality index

Centrality development in 2000-2014

The centrality values derived meet the requirements of the statistical correlation, in fact, each sector has a value in every year and a correlation monotone between every pair of sectors persists; moreover, for every year the sectors have been ranked for decreasing centrality values and for each one an ordinal value has been assigned. The application of the Kendall correlation method has led to the heatmap in figure 4.9, where the color code ranges from green (low similarity) to red (high similarity).



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
2000	1	0.98	0.97	0.96	0.96	0.94	0.94	0.94	0.93	0.82	0.73	0.75	0.71	0.69	0.68
2001	0.98	1	0.98	0.98	0.98	0.96	0.96	0.95	0.94	0.82	0.74	0.75	0.71	0.69	0.68
2002	0.97	0.98	1	0.98	0.98	0.96	0.96	0.95	0.93	0.83	0.73	0.75	0.72	0.7	0.69
2003	0.97	0.98	0.99	1	0.99	0.97	0.97	0.96	0.95	0.83	0.74	0.76	0.73	0.71	0.7
2004	0.96	0.98	0.98	0.99	1	0.97	0.97	0.96	0.94	0.83	0.74	0.76	0.72	0.7	0.7
2005	0.95	0.96	0.96	0.97	0.98	1	0.98	0.97	0.96	0.84	0.75	0.76	0.72	0.7	0.7
2006	0.94	0.95	0.95	0.96	0.96	0.97	1	0.99	0.97	0.84	0.74	0.75	0.72	0.7	0.7
2007	0.94	0.95	0.95	0.96	0.96	0.97	0.99	1	0.97	0.83	0.73	0.75	0.72	0.7	0.7
2008	0.92	0.93	0.93	0.94	0.94	0.95	0.97	0.97	1	0.84	0.75	0.75	0.73	0.71	0.71
2009	0.82	0.83	0.83	0.84	0.84	0.84	0.83	0.83	0.84	1	0.88	0.87	0.84	0.83	0.83
2010	0.73	0.74	0.73	0.74	0.74	0.74	0.73	0.73	0.74	0.87	1	0.96	0.87	0.86	0.86
2011	0.74	0.75	0.75	0.76	0.76	0.76	0.75	0.74	0.75	0.87	0.96	1	0.9	0.88	0.88
2012	0.7	0.71	0.72	0.72	0.72	0.72	0.71	0.71	0.72	0.83	0.88	0.9	1	0.96	0.95
2013	0.68	0.69	0.7	0.7	0.7	0.7	0.7	0.7	0.71	0.83	0.86	0.88	0.96	1	0.98
2014	0.67	0.68	0.69	0.7	0.7	0.69	0.69	0.69	0.7	0.83	0.86	0.88	0.95	0.98	1

Figure 4.9: Heatmap of the centrality ranking Kendall correlation for the period 2000-2014

The heatmap in figure 4.9 shows as there is a strong positive correlation among the centrality index ranking for all the couples of years in the period, all having a correlation higher than 68%. However, it highlights also the presence of three different areas of highest correlation, period 2000-2008, the biennium 2010-2011 and the last three year presented. This peculiar conformation is doubtless due to

the changes brought by the financial crisis.
The main difference among the detected areas is that in the period 2000-2008 the sector with the highest centrality is Mining and quarrying(B), whose ranking decreases, reaching the tenth place in the last three years. Another important sector in the first area is Public administration and defense (O84) that loses more than 10 positions. Instead, Publishing activities (J58) and Manufacturer of pharmaceutical (C21), not among the the most central in the first period, gain a high centrality coefficient in 2012-2014. Also the scientific research industries (M72) and Telecommunications (J61) improve their ranking during the last years.

Rank	Sector	Id
1	Mining and quarrying	B
2	Manufacture of computer, electronic and optical products	C26
3	Other service activities	R_S
4	Administrative and support service activities	N
5	Other professional, scientific and technical activities; veterinary activities	M74_M75
6	Wholesale trade, except of motor vehicles and motorcycles	G46
7	Public administration and defence; compulsory social security	O84
8	Repair and installation of machinery and equipment	C33
9	Manufacture of other transport equipment	C30
10	Wholesale and retail trade and repair of motor vehicles and motorcycles	G45

Table 4.4: Sectors with highest centrality ranking in period 2000-2008

4.2.4 Clustering

The Italian clustering coefficient development and its components

The clustering coefficient of every sector per year has been detected using the method introduced by Fagiolo and explained in detail in chapter 3. Among the components of the clustering coefficient, several works showed how out-clustering could provide meaningful insights related to risk spreading, especially in the context of systemic financial risk. In particular, it may reflect higher risk because failure of the purchaser node in an out-triangle can trigger simultaneous non-repayments to the seller nodes, and this, can make them unable to honor their own obligations. The implication of high cycle-clustering is more ambiguous, since nodes in the considered triangles act as both purchaser and sellers in the network, so the consequences of a node failure are unclear.

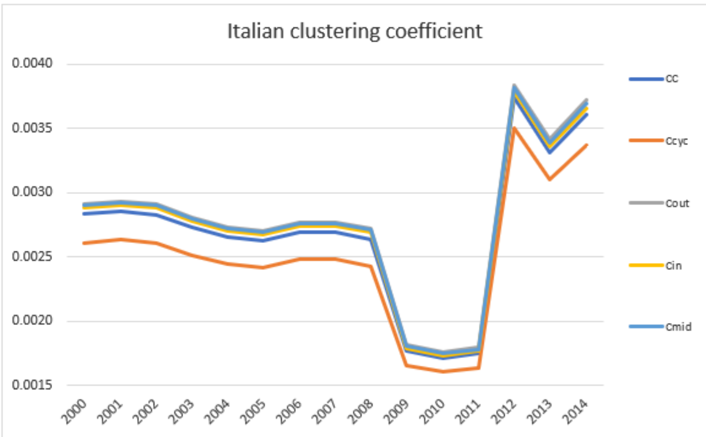


Figure 4.10: Italy clustering coefficient 2000-2014

The Italian economy is characterized by a stable value of different clustering coefficients over time. In figure 4.10 the evolution of the clustering coefficient in period 2000-2014 is shown, emphasizing its 4 components; while the total weight of the clustering coefficient changes over the time, the relative fraction of the different triangles remains approximately the same. The cycle coefficient is consistently lower than all the other, proving that economies tend to be acyclic at the small scale of 3 cycles.

The clustering components steadiness is presumably linked to the stability of structural properties in the economy network.

The clustering coefficient development shows a fall after 2009 whose possible explanation is that the booming economy before 2008 introduced more interactions between industries, hence higher clustering coefficient, and the effects of the financial crisis stifled the excessive relationships.

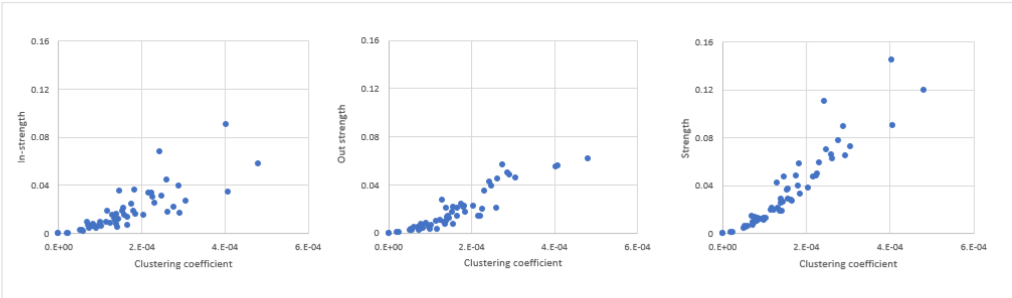


Figure 4.11: Scatter plot strength

Figure 4.11 relates the clustering coefficient of each sector to its strength; for the three typologies of strength considered, in-strength, out-strength and total strength, the clustering coefficient appears to be positively correlated. So, sectors with a higher level of strengths also have a higher level of weighted clustering coefficients, that indicates that sectors with larger total flows tend to participate in more intense trade clusters.

Clustering coefficient ranking through the period 2000-2014

Among the sector with the highest clustering coefficient in year 2014, there are akin industries, sectors related to the legal and administrative domain (N and M69-M70), the Real estate and the Construction sectors (L68 and F), Wholesale trade and land transport (G46 and H49).
It's peculiar how the manufacturing does not stand out for its high clustering, except for manufacture of machinery and metal products(C28 and C25) and manufacture of food, beverages and tobacco (C10-C12); their highest ranking could be justified by the fact that the automotive and food sectors have always been one of the strengths of Italy.
Table 4.5 lists the 10 sectors with the highest clustering coefficient for the year 2014. All these 10 sectors have a clustering coefficient higher than the national level, and the first three sector are two times more clustered than the average in the whole network.

Rank	Sector	Id	Clustering
1	Wholesale trade, except of motor vehicles and motorcycles	G46	0.0110
2	Administrative and support service activities	N	0.0093
3	Construction	F	0.0092
4	Land transport and transport via pipelines	H49	0.0070
5	Legal and accounting activities; activities of head offices; management consultancy activities	M69_M70	0.0067
6	Electricity, gas, steam and air conditioning supply	D35	0.0066
7	Financial service activities, except insurance and pension funding	K64	0.0063
8	Real estate activities	L68	0.0060
9	Manufacture of machinery and equipment n.e.c.	C28	0.0060
10	Manufacture of fabricated metal products, except machinery and equipment	C25	0.0057

Table 4.5: Top 10 sectors for clustering coefficient

Figure 4.12 shows the positive correlation between the clustering coefficient ranking for all the years considered, the minimum correlation is 88% which is a really high value. Moreover, in the heatmap is possible to individuate two areas of higher correlation, the period 2000-2008 and the years 2010-2014.

Even if the clustering coefficient results have been proved positively correlated in the years, there is a gap between the two detected areas that could be connected to the financial crisis and the deep impact it had. Note in figure 4.10 the clustering is distributed over three different values, the time partitioning isn't the same observed in the heatmap, but they are similar, especially considering the stability of the clustering in period 2000-2008.

The high level of correlation shows as for every year the sectors with the highest coefficient are almost the same, with slightly differences in the order. In the whole period the podium is monopolized by Wholesale trade, except of motor vehicles and motorcycles, Construction and Administrative and support service activities. Among the highest ranked sectors there are also Manufacture of machinery, Legal and accounting activities and Land transport and transport via pipelines.

The main change among the ranking is that with the second half of the time period the financial sector (K64) gains a higher coefficient, while the Warehousing sector (H52) loses points. In addition, also energy industries (D35) and Financial service activities (K64) play a major role in the second detected area.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
2000	1	0.99	0.97	0.97	0.96	0.96	0.96	0.95	0.94	0.91	0.91	0.91	0.91	0.90	0.88
2001	0.99	1	0.98	0.98	0.97	0.96	0.96	0.95	0.93	0.91	0.91	0.90	0.90	0.89	0.88
2002	0.97	0.98	1	0.99	0.98	0.97	0.96	0.96	0.94	0.93	0.92	0.91	0.91	0.89	0.88
2003	0.97	0.98	0.99	1	0.98	0.98	0.97	0.96	0.94	0.93	0.92	0.92	0.91	0.90	0.89
2004	0.96	0.97	0.98	0.98	1	0.99	0.97	0.96	0.95	0.94	0.93	0.92	0.92	0.91	0.90
2005	0.96	0.96	0.97	0.98	0.99	1	0.98	0.97	0.96	0.94	0.94	0.93	0.92	0.90	0.90
2006	0.96	0.96	0.97	0.97	0.97	0.98	1	0.99	0.98	0.93	0.94	0.93	0.92	0.91	0.90
2007	0.95	0.95	0.96	0.96	0.96	0.97	0.99	1	0.98	0.94	0.94	0.93	0.92	0.90	0.89
2008	0.94	0.94	0.95	0.95	0.95	0.96	0.98	0.98	1	0.94	0.94	0.94	0.93	0.92	0.91
2009	0.91	0.92	0.93	0.93	0.94	0.94	0.93	0.93	0.93	1	0.95	0.94	0.93	0.93	0.92
2010	0.91	0.91	0.92	0.92	0.93	0.94	0.94	0.93	0.94	0.95	1	0.98	0.96	0.96	0.95
2011	0.91	0.91	0.91	0.92	0.93	0.93	0.93	0.93	0.94	0.94	0.98	1	0.98	0.96	0.96
2012	0.91	0.90	0.91	0.91	0.92	0.92	0.92	0.92	0.93	0.93	0.96	0.98	1	0.97	0.97
2013	0.90	0.90	0.90	0.90	0.91	0.91	0.91	0.91	0.92	0.93	0.96	0.96	0.97	1	0.99
2014	0.89	0.88	0.88	0.89	0.90	0.90	0.90	0.90	0.91	0.92	0.95	0.96	0.97	0.98	1

1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30

Figure 4.12: Heatmap of the clustering coefficient Kendall correlation between each couple of years (2000-2014)

4.2.5 Failure tolerance analysis

Analysis development

The application of the failure tolerance analysis to the Italian graph has required some changes to the procedure explained in the previous chapter.
The implementation is to fail one sector at a time and find the threshold that grants the network survival rate desired. Working on national graphs the failure of an industry results in a loss for all industries that received money flow from the sector under consideration. The adopted interpretation of the failure tolerance analysis looks at the minimum threshold to make more than 30% of the network survive, considering a higher threshold p value as sign of a major impact on the vulnerability of the network. The choice of not focusing on the percentage of nodes survived is justified by the fact that it takes similar values for each node and consequently brings a low significance.
Thus, for the purpose of this analysis, the threshold p of an industry is the measure of its importance in the economic network, in fact, a sector with good diffusion properties is such that it distributes the effects of a shock to a wider set of other sectors as oppose to concentrate the effect into itself or just a few others, so it shows a higher influence on the economy.
Figure 4.13 displays the behavior of the largest thresholds with different levels of aggregation, the largest p increases from 2000 to 2008, and then decreases in 2009 and 2012, the other threshold show a similar, but less mark, behavior. In addition, the average on the 4 largest threshold is still near to the largest thresh-

old, while computing the average on eight values makes the threshold decrease of almost 5%, meaning that there are just few really high thresholds.

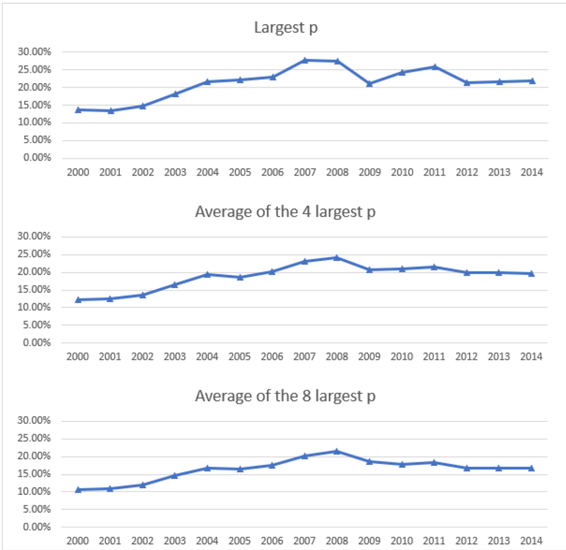


Figure 4.13: Development of (a) the largest threshold in the time period, (b) the average of the 4 largest thresholds and (c) the average of the 8 largest thresholds

The results of the failure tolerance analysis have a positive correlation with gross output, value added and out-strength of the nodes. A sector shows to have a higher threshold if originates heavy weight edges with sectors which in turn are heavily connected. Figure 4.14 evidences the correlation, in particular the comparison out strength-threshold.

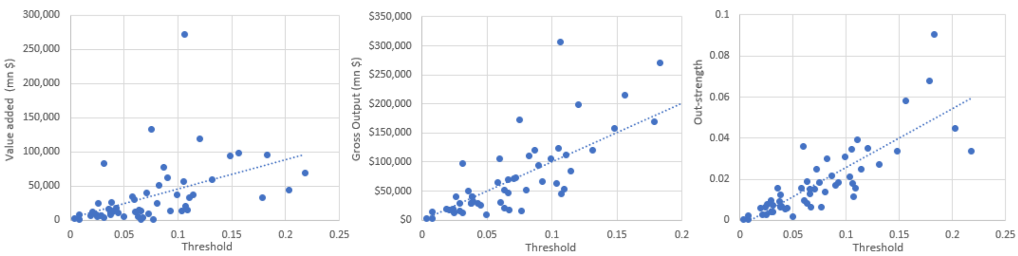


Figure 4.14: Scatter plot of different measures in comparison with the failure tolerance analysis thresholds

Ranking in the different years

The threshold values range from really small values, around 0.1%, to high values as 20-22%. In table 4.6 are listed the ten sector with the largest threshold. For example, the largest threshold is around 22% and refers to sector I representing Accommodation and food service activities industries, this threshold seems really high but in spite of that, in case sector I fell it would result in the failure of the Food and beverage and the Fishing and aquaculture sectors (C10-C12 and A03) just as first step, even with such a large threshold. Another example is the failure of Manufacture of machinery and equipment (C28), where keeping a threshold of 20% would still trigger the failure of Manufacture of fabricated metal product (C25).
It is interesting to note that the percentage of surviving industries is always a very high value, although the value to stop the procedure was just the 30% survival of the network.

Rank	Sector	Id	% Network survived	Threshold
1	Accommodation and food service activities	I	81.48%	21.77%
2	Manufacture of machinery and equipment n.e.c.	C28	90.74%	20.29%
3	Construction	F	94.44%	18.32%
4	Manufacture of food products, beverages and tobacco products	C10-C12	94.44%	17.85%
5	Wholesale trade, except of motor vehicles and motorcycles	G46	94.44%	15.63%
6	Retail trade, except of motor vehicles and motorcycles	G47	92.59%	14.82%
7	Land transport and transport via pipelines	H49	85.19%	13.15%
8	Human health and social work activities	Q	88.89%	12.06%
9	Warehousing and support activities for transportation	H52	94.44%	11.43%
10	Electricity, gas, steam and air conditioning supply	D35	90.74%	11.10%

Table 4.6: Ranking the 10 sectors with the largest threshold derived from the tolerance failure analysis

Comparing industry order rankings for different years allows to catch the similarity in the economic environment across a period of fourteen years.
The heatmap of the Kendall correlation, figure 4.15, shows three areas of high correlation between all the pair of years from 2000 and 2008, biennium 2010-2011, and between year 2012, 2013 and 2014.
All the couple of years in the graph have a discrete correlation with a minimum value around 74%. In the whole time period Construction (F), Manufacturer of machinery (C28) and manufacturer of food product (C10-C12) keep the highest position in the ranking, while Accommodation and food activities stays around the 8th place until 2009. In years 2009-2011 Retail trade (G47) acquires

a large threshold, and Warehousing and support activities (H52) gains a place in the top ten, despite its low ranking in period 2000-2008. Moreover, both the manufacture of textiles and wearing apparel (C13-C15) and manufacture of furniture(C31-C32) are far less important in the second area.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014		
2000	1	0.95	0.93	0.94	0.94	0.92	0.92	0.91	0.9	0.83	0.79	0.79	0.77	0.75	0.74		1.00
2001	0.95	1	0.97	0.96	0.95	0.94	0.92	0.91	0.9	0.85	0.8	0.79	0.78	0.75	0.74		0.95
2002	0.93	0.97	1	0.98	0.96	0.95	0.94	0.92	0.92	0.86	0.81	0.8	0.78	0.75	0.74		0.90
2003	0.94	0.96	0.98	1	0.97	0.95	0.94	0.92	0.93	0.87	0.81	0.8	0.78	0.75	0.74		0.85
2004	0.94	0.95	0.96	0.98	1	0.96	0.95	0.93	0.93	0.86	0.81	0.8	0.78	0.75	0.75		0.80
2005	0.92	0.94	0.95	0.96	0.96	1	0.97	0.95	0.94	0.87	0.83	0.82	0.79	0.76	0.75		0.75
2006	0.93	0.93	0.94	0.95	0.95	0.98	1	0.97	0.96	0.87	0.83	0.82	0.79	0.77	0.76		0.70
2007	0.92	0.91	0.92	0.93	0.94	0.95	0.97	1	0.97	0.85	0.83	0.81	0.79	0.77	0.76		0.65
2008	0.9	0.91	0.93	0.93	0.93	0.95	0.96	0.96	1	0.87	0.84	0.83	0.78	0.77	0.76		0.60
2009	0.83	0.85	0.86	0.87	0.86	0.88	0.87	0.85	0.87	1	0.89	0.88	0.84	0.82	0.81		0.55
2010	0.8	0.8	0.81	0.82	0.81	0.83	0.83	0.82	0.84	0.88	1	0.94	0.89	0.87	0.87		0.50
2011	0.79	0.8	0.8	0.8	0.8	0.82	0.82	0.81	0.83	0.88	0.94	1	0.92	0.9	0.91		0.45
2012	0.78	0.78	0.78	0.78	0.78	0.79	0.79	0.79	0.8	0.84	0.9	0.92	1	0.94	0.94		0.40
2013	0.75	0.75	0.75	0.75	0.76	0.76	0.77	0.77	0.78	0.82	0.88	0.9	0.94	1	0.97		0.35
2014	0.75	0.74	0.74	0.75	0.75	0.75	0.77	0.77	0.78	0.81	0.87	0.91	0.94	0.97	1		0.30

Figure 4.15: Heatmap of the failure tolerance analysis Kendall correlation between each couple of years (2000-2014)

Chapter 5

Comparison with the main European and World systems

This chapter compares the Italian graph structure to the other national graphs considered. The comparison is conducted using the density, the graphical features, the assortativity, in year 2014, and the clustering coefficient development in time period 2000-2014. Moreover, in the paragraph, a global failure tolerance analysis is developed to rank the importance of the Italian sectors in the global environment.

5.1 Outlook

In this study, countries that have strong trade relations with Italy and a high GDP are considered; among these Italy supports increased trades with France and Germany.

Figure 5.1 shows the GDP for each nation, we can see as the United States clearly dominate the economic scene. China, the closest follower of USA, has a GDP score which is 52% of USA's GDP. Japan's gross domestic product is about three quarters of China's value and Spain has the lowest current GDP in the considered sample[1].

When the per capita GDP is considered, figure 5.2, China loses seven position. The first three countries are USA, Germany and United Kingdom, respectively.

[1]Data have been taken from www.worldbank.org

Even the latest per GDP, Russia, has higher per capita GDP than China.

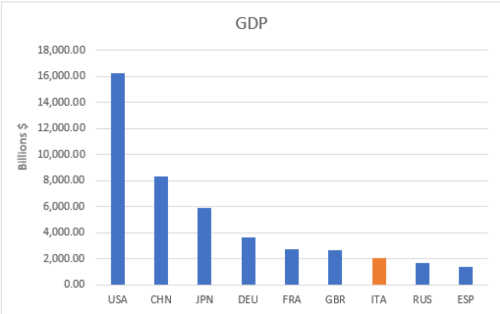


Figure 5.1: 2014 GDP ranking

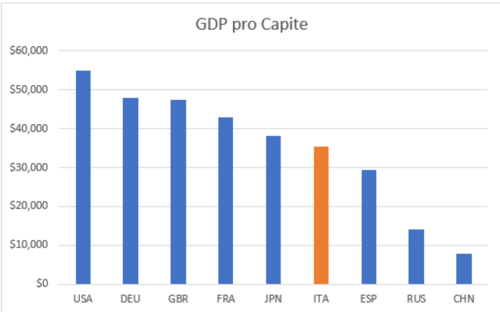


Figure 5.2: 2014 GDP per capita ranking

In 2014 the Global economy keeps showing several signs of weakness and few signs of strength. The country in the best condition seems to be the United States, which may finally be on a sustainable and healthy growth path. During the years of crisis, the world economy has relied on the emerging markets to keep the global economy afloat. Together with the developing countries, they accounted for three-quarters of global growth over the past half decade. However, a growing number of emerging markets are slowing down as the economic cycle turns. Overall, the direction is positive, but global growth is still too low and too fragile.

Figure 5.3 illustrates the GDP growth development in the time period. In the biennium 2008-2009 all the countries exhibit a fall, with a negative growth index, except for China whose growth rate decreases but stays positive. For China after a slight increase in 2011 the growth rate keeps lowering. Russia shows the highest jump from 2008 to 2009, going from a +5% increase to lower than -5%

decrease.

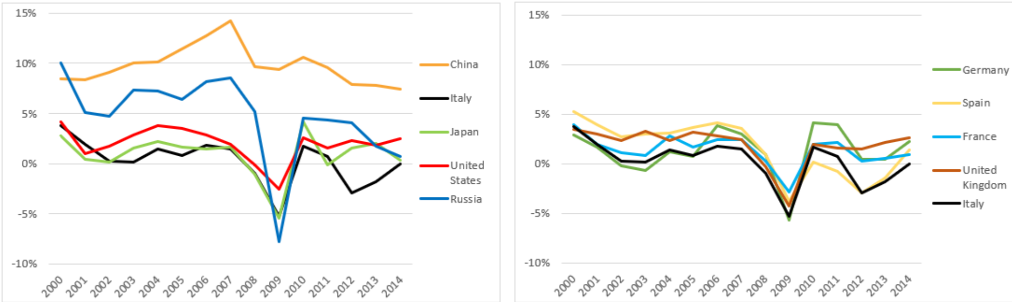


Figure 5.3: GDP growth

5.2 Networks' properties

5.2.1 Graphs comparison

The first measure considered as comparison is the density, that is also indicator of connectivity, table 3 in the appendix. No country has a density coefficient equal to 1, meaning that some domestic sectors in these countries are not connected to one another in terms of intermediate good flow. USA is the most connected country (around 95% density), while the less connected are China and Russia, respectively 67-69% and 33-34%.

The representation of the national input-output tables as graphs has led to the identification of two groups; on one hand the European countries and the USA, while on the other China, Russia and Japan. The main difference among these two groups stays in the number of heavy edges and the fact that Russia and China have a lower density than the other nations. The comparison between figure 5.4 and figure 5.5 well exemplifies the two groups. Russia, compared to Germany, has more high weight edges connecting the sectors with highest gross output and has a higher percentage of disconnected nodes in the backbone network; Russia is the most extreme example of this division, in fact, despite all the other countries it has more nodes with a high gross output heavily connected, which represents an economy centered around a reduced number of sectors.

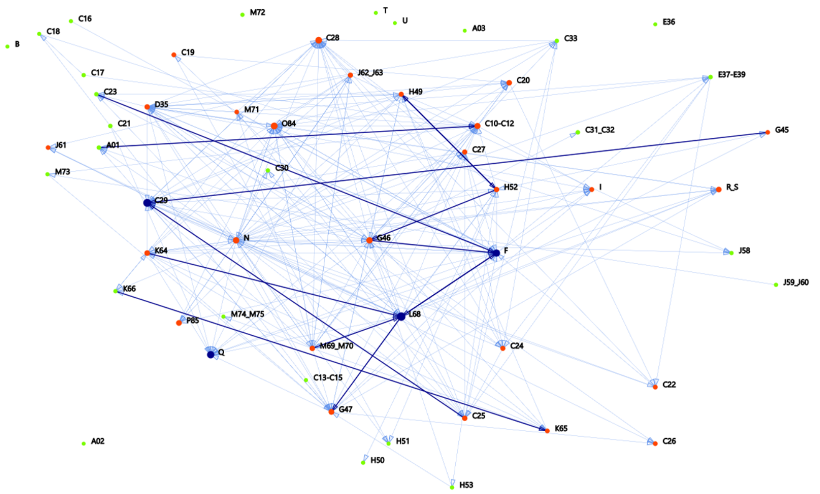


Figure 5.4: Germany, 2014

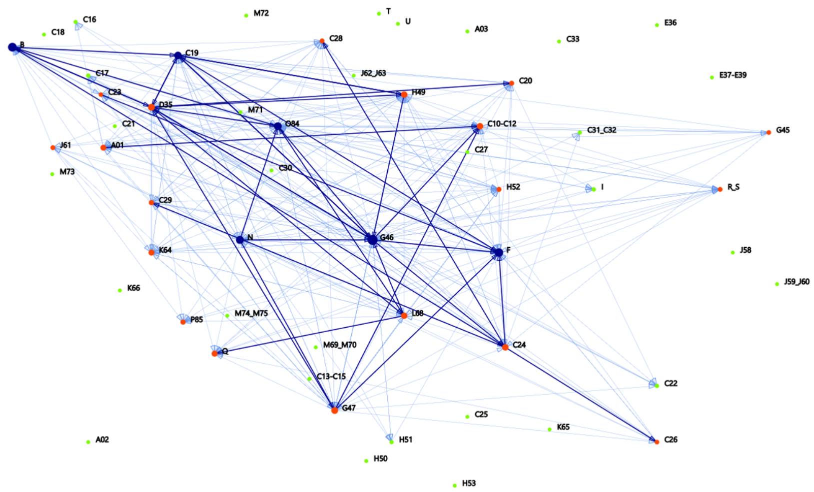


Figure 5.5: Russia, 2014

5.2.2 Assortativity

Assortativity measures the propensity of a node to connect with vertices of similar degree. Notice that when calculating the assortativity for in-degree, out-degree, and total-degree, respectively, we consider the vertices as the neighbors of a given node if they are connected with the given node by only incoming edges, by only outgoing edges, and by either incoming or outgoing edges, respectively.

This measure shows as for all the national graphs considered, high degree nodes tend to be connected with vertices with the same degree. However, this index loses value in the case of high density weighted directed graphs.

An alternative way to calculate the assortativity structure takes into account the in(out)-strength of each node in comparison with the respective average value of its neighbors. The negative assortativity correlation coefficient indicates that there is a disassortative architecture in the network, meaning that there is a core-periphery structure, while if the assortativity correlation coefficient is greater than zero, then there is an assortative structure in the network.

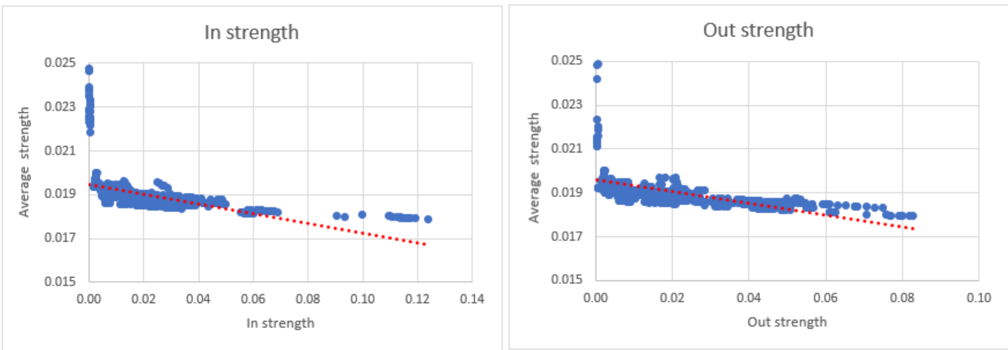


Figure 5.6: In-assortativity and Out-assortativity, Italian graph in period 2000-2014

As it emerges from the graphs in figure 5.6, the Italian economy shows a disassortative behavior; dealing with the same type of analysis on all national maps, it turns out that the disassortative structure is common to all. This finding means that heavy connected sectors are not just connected between their similar, but they have significant connection also with "smaller" sectors. In figure 5.7 is shown the disassortative structure of both the Spanish and French graph as a sample.

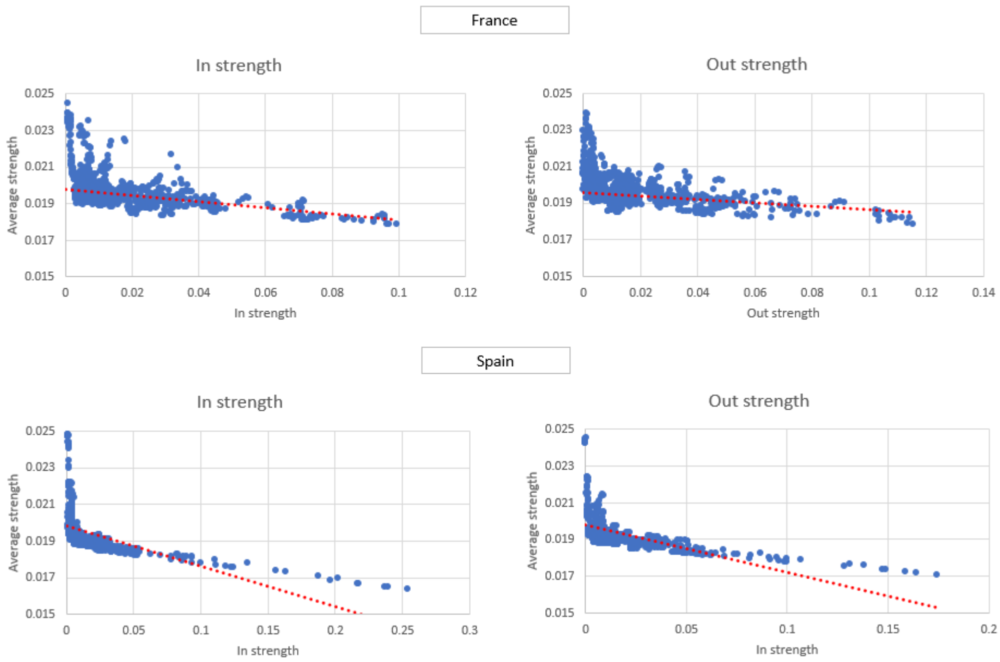


Figure 5.7: In-assortativity and Out-assortativity, France and Spain, period 2000-2014

5.2.3 Clustering

In this paragraph, the static and evolving clustering properties of the Italian graph, derived using the method introduced by Fagiolo, are compared to the other nations' results.

As for the Italian case, for the whole sample of countries the clustering coefficient values has a positive correlation with the sector in(out)-strength.

The graphs in figure 5.8 show the development of the clustering coefficient of the countries in comparison with the Italian case, in black; is evident how Japan and Spain are less clustered than Italy, France, USA, Russia and China have a higher clustering coefficient, the UK has lower values than Italy except for the years 2008-2012, while Germany's coefficient follows almost the same pattern of the Italian one, but it assumes higher values in the financial crisis period

The clustering properties of each country undergo significant temporal change, Japan and China are the most stable, while Russia and USA undergo heavy variations; in addition, there is not a stable trend for each country, especially USA and Russia show to have opposite trends.

The financial crisis has triggered changes in the clustering property of the countries that are not homogenous, either an increase (Russia, Spain, UK and Germany) or a decrease (Italy, France, USA) of the average clustering coefficient.

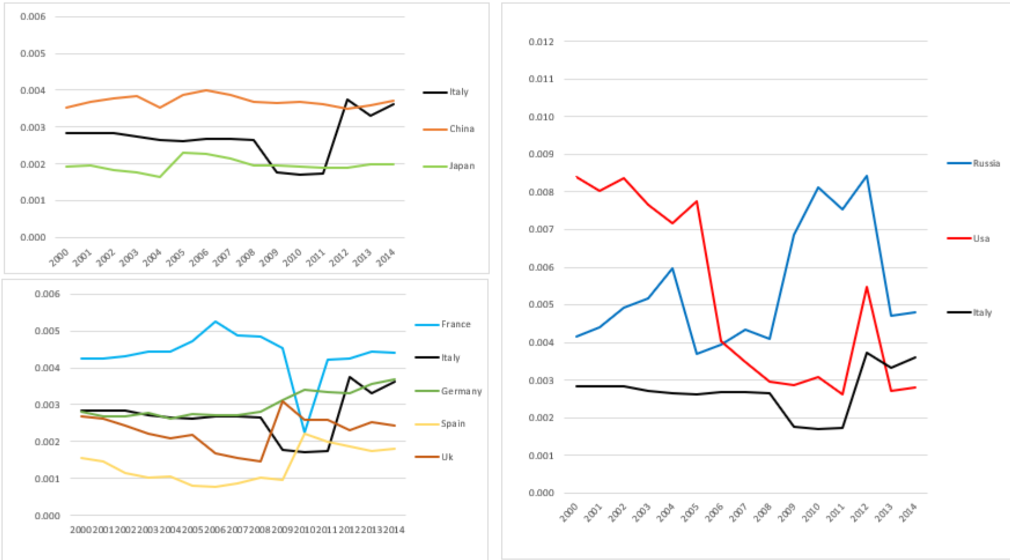


Figure 5.8: Collection of graphs showing the development of the countries clustering coefficient over the time

However, there are properties of the networks which are rather stable over time as the development of the different components of the clustering coefficient. The values of the fractions do slightly change over time, but the structure is stable, as in the Italian graph, over time the relative fractions of the different triangles remain approximately stable, with the cyclic coefficient having a smaller values than the others; the fact that the cycle clustering coefficient is consistently lower than all the others suggests that economies tends to be acyclic at the small scale of 3-cycles. In figure 5.9 are reported UK and France as example, but this behavior effects all the national graphs considered.
In table 4, shown in the appendix, the clustering coefficients and the Herfindahl index[2], for each country in the time horizon, are listed. Small values of H correspond to relatively uniform distribution, while large ones suggest the presence of dominant players. In the group of nations examined, Russia has the highest Herfindahl index, followed by China, while all the other nations have almost the same values. This finding would appear to be supported by the observation of the Russian backbone networks.

[2]The ordinary Herfindahl index is given by $\sum_{i=1}^{N}(\frac{c_i}{\sum_{j=1}^{N} c_j})^2$, N is the number of sector and C_i the clustering coefficient of sector i.

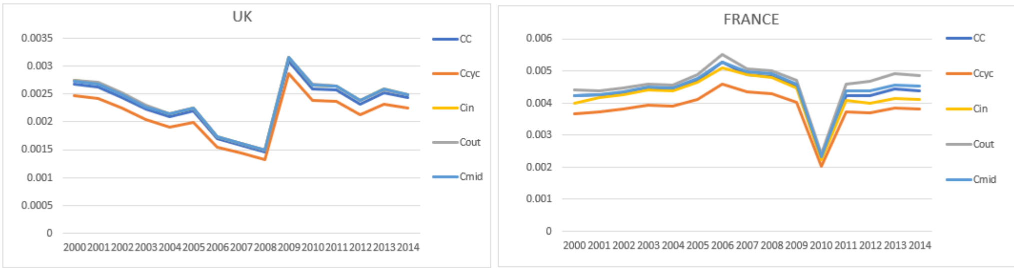


Figure 5.9: The dynamics of the clustering coefficient triangle fractions

5.3 Failure tolerance analysis

Following several decades of rapid globalization, the global economic system has become a complex network, composed by industries in different countries. In such a highly interlinked system, local shocks involving single industrial sectors may lead to a large disruption in the aggregate output. The cascading failure tolerance analysis applied to the global network helps to identify and rank the influence of the different industries in the stability of the world economy[3].

For the analysis, the fourteen countries with the largest gross output dollar values in 2014 have been selected[4] to compose the economical graph into which the shock happens. The graph contains 784 vertices, collection of all the 56 industries of each selected country, and has edges which represent the money flow among all industries regardless of their country, there are so both intra-country and inter-country relations[5].

The basic concept behind this procedure is that, differently from the analysis for the Italian case, the failure of industry i in country C doesn't affect the other industries in that same country C, but only triggers money flow reduction for the other countries; each State is, in fact, supposed to be able to make a quick adjustment to weaken the impact of the shock to other sectors within the nation.

[3]In chapter 4 the failure tolerance analysis has been applied to the Italian graph, not considering the other countries.

[4]Australia, Brazil, Canada, China, France, Germany, India, Italy, Japan, Korea, Russia, Spain, UK and USA.

[5]The data have been taken from the World input output table 2014 released by the WIOD in 2016.

The procedure derives a threshold for each sector that grants at least a 90% survival rate of the global network built; the larger the threshold is, the more influent the sector is in the economy. When the threshold p or a larger value is taken as maximum revenue reduction that doesn't cause failure, once the failure cascade in the system is over, almost all the global network survives (more than 90% of the remaining industries).

Figure 5.10 shows the reaction of the network to the failure of the Chemical manufacturer sector in the USA and the Manufacturer of motor vehicles sector in Germany as their thresholds rise; for both the industries the network survival rate grows gradually with the threshold increase, until at a certain threshold there's a jump to a really high percentage of survival, this behavior affects all the industries under consideration.

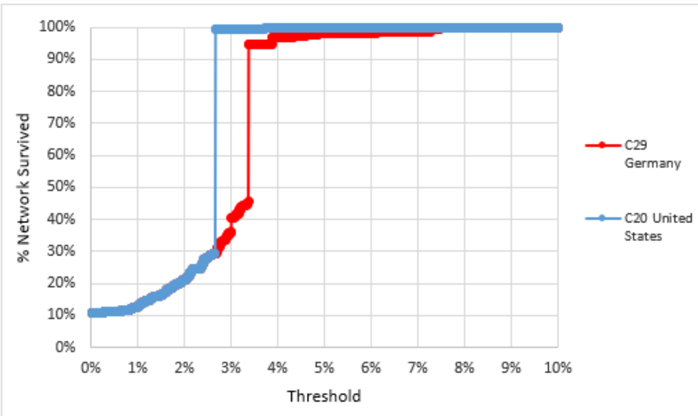


Figure 5.10: Fraction of surviving industries as a function of the tolerance threshold for the failure of a German sector, in red, and of a USA industry, in blue

The analysis has been run for all the countries and has led to a ranking of all the industries. Considering just the Italian case, in table 5.1 are listed the 10 Italian sectors with the highest threshold, with Manufacture of machinery equipment (C28), Manufacturer of electrical equipment (C27) and Manufacturer of chemicals (C20) on the podium.

Comparing the impact that the failure of an Italian industry has at national level with that of the global one we can see that the values of the thresholds differ by an order of magnitude, fully justified by the different sizes of the networks, and, moreover, that the two rankings differ on the typology of sectors with the highest

threshold, while the first ranking collects both services and production sectors, the ranking derived by the global analysis is composed by mainly manufacturer industries.
This may be motivated by the fact that the Italian service sectors are mainly focused on the Italian field, while the manufacturer industries, that have a strongest history and background, have developed strong relationships with the other nations, and are so capable of a major influence on the global economy.
It is interesting to notice how seven on the industries in figure 5.1 are part of the top ten sectors for export amount, except for Wholesale, retail trade and repair of motor vehicles, Construction and Human health activities. In addition, an industry as Manufacture of textiles and wearing apparel that is the second one for exportations is not include in this last ranking.

Rank	Sector	Id	% Network survived	Threshold
1	Manufacture of machinery and equipment n.e.c.	C28	90.05%	3.38%
2	Manufacture of electrical equipment	C27	98.72%	2.28%
3	Manufacture of chemicals and chemical products	C20	97.70%	2.20%
4	Manufacture of motor vehicles, trailers and semi-trailers	C29	99.87%	2.11%
5	Construction	F	99.87%	2.06%
6	Human health and social work activities	Q	99.23%	1.79%
7	Manufacture of basic metals	C24	98.09%	1.77%
8	Wholesale and retail trade and repair of motor vehicles and motorcycles	G45	99.87%	1.75%
9	Manufacture of fabricated metal products, except machinery and equipment	C25	99.87%	1.74%
10	Manufacture of food products, beverages and tobacco products	C10-C12	99.74%	1.70%

Table 5.1: Ranking of the top 10 Italian industries for threshold derived from the global failure tolerance analysis

Considering the global industries rank, it appears that the high threshold of the manufacturer industries is a feature common to all the national economies, especially Manufacture of machinery and equipment (C28), Manufacture of motor vehicles (C29) and Manufacture of other transport equipment (C30). Other two industries with a significant threshold are Manufacture of computer, electronic and optical productsand (C26) and Manufacture of chemicals and chemical products (C21).
Looking at the 30 largest thresholds in 2014, half of the industries belong just to three countries, in order of importance United States, China and Germany[6]. In *Ranking the economic importance of countries and industries* the methodology

[6]To rank the importance of a given country the average of the largest four tolerances of industries is used.

presented here was applied to the time horizon 1995-2011 in order to map the development of the countries from year to year, from their analysis it emerges that the economic importance of China relative to that of the United States shows a consistent increase in the years.
The failure tolerance analysis represents indeed an instrument capable of finding the key sector in the transmission of a shock on the global scale and of illustrating how the economic power architecture in the world's economy is subject to changes over time.

Conclusion

The analysis developed in this thesis has allowed to study the Italian economy as a set of interconnected agents and to overcome the static view provided by the input output tables.
The national tables and the global input-output table have been converted into weighted directed graphs mapping the money flow exchanged between one industry and another.
In the first place, the whole set of graphs has been studied for the purpose of finding the main properties of such networks. All the national maps have been shown to have some common features, such as high density, in-out degree left skewed distributions and power law distribution of the edge weight tail. These findings have helped to conduct more appropriate analysis in order to classify the role each sector plays in the economy. The asymmetric flow distribution has led the choice to model the economies with a weighted directed graph and, together with the symmetry of the in and out-strength distribution have characterized the network as an environment subject to the propagation of shocks that could generate a sizable disturbance.
The most significant among the considered indices are centrality, clustering and the failure tolerance analysis threshold. The first one ranks the importance of a sector considering the in and out degree and its out flows, the second one gives an interpretation of the chunkiness level of the map and the most involved sectors, while the last one classifies the sector regarding the role they play in the spreading of a shock in the considered network. All of them prove to have a positive correlation with the values obtained in the different years, and, especially they show to have two high correlation areas, period 2000-2008 and years 2012-2014; this result shows that the financial crisis of 2008 had an impact also on the relations between sectors in the Italian economy.

The techniques of analysis developed haven't shown a unique ranking of the Italian industries, but it's possible to notice some sectors with a higher position in all the procedures, that is certainly a demonstration of the degree of importance they have in the national economy.
Among these sectors there are the Wholesale trade sector, the Construction industry and the Manufacturer of machinery. The rankings have shown how the strategic place an industry occupies in the economic network doesn't rely on the value added it brings but, on its position, and relation with the whole system.
Consequently, a more targeted comparison with the national graphs of some European and world system has been conducted. It has been found that all the graphs share a common disassortative structure, where industries having high in or out-strength tend to be connected with smaller sectors.
In addition, the clustering coefficient of each nation has a different development, with USA and China having the largest variation over the time, but them all are characterized by a stable proportion of the clustering's components, sign of stability in the economic structure, and by the cycle clustering having a smaller value, index of a network acyclic at the scale of 3. The Herfindahl index takes on similar values for all the graphs, except for the higher values of Russia and China, identifying them as systems with dominant players, property that is also proved by the arrangement of heavy flows in the graphic representation of the backbone of this networks.
To conclude the failure tolerance analysis run at the global level has shown different result regarding the ranking of the Italian industries, since it has attributed a larger threshold mainly to manufacturer sectors. At the global level the analysis has proved how the highest threshold industries belong mainly to three countries: USA, China and Germany.
On the whole, it can be said that the industries belonging to the wider categories of wholesale trade and retail, manufacture of machinery and motor vehicles and construction play a key role in the Italian economy.
A limitation of the work done is that the input output tables covered the period 2000-2014 only and it would be interesting to see the changes that, at the Italian and world level, happened in the following five years.
With regard to the failure tolerance analysis, it could be run on the different years so as to see how the importance of Italy changes over time, in comparison with

the role of the main European and World countries.
Moreover, it could be worthwhile developing a partitioning algorithm that considers the weighted edges of the national graphs to improve the clusters found considering only the edges directions.
A last possible development could be to trace the same analysis making use of different methods for the calculation of the properties seen, for example for centrality and clustering, so as to have a comparison at the level of the individual measures.

Appendix

Additional tables

Id	Description
A01	Crop and animal production, hunting and related service activities
A02	Forestry and logging
A03	Fishing and aquaculture
B	Mining and quarrying
C10-C12	Manufacture of food products, beverages and tobacco products
C13-C15	Manufacture of textiles, wearing apparel and leather products
C16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
C17	Manufacture of paper and paper products
C18	Printing and reproduction of recorded media
C19	Manufacture of coke and refined petroleum products
C20	Manufacture of chemicals and chemical products
C21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
C22	Manufacture of rubber and plastic products
C23	Manufacture of other non-metallic mineral products
C24	Manufacture of basic metals
C25	Manufacture of fabricated metal products, except machinery and equipment
C26	Manufacture of computer, electronic and optical products
C27	Manufacture of electrical equipment
C28	Manufacture of machinery and equipment n.e.c.
C29	Manufacture of motor vehicles, trailers and semi-trailers
C30	Manufacture of other transport equipment
C31_C32	Manufacture of furniture; other manufacturing
C33	Repair and installation of machinery and equipment
D35	Electricity, gas, steam and air conditioning supply
E36	Water collection, treatment and supply
E37-E39	Sewerage; waste collection, treatment and disposal activities; materials recovery; remediation activities and other waste management services
F	Construction
G45	Wholesale and retail trade and repair of motor vehicles and motorcycles
G46	Wholesale trade, except of motor vehicles and motorcycles

Id	Description
G47	Retail trade, except of motor vehicles and motorcycles
H49	Land transport and transport via pipelines
H50	Water transport
H51	Air transport
H52	Warehousing and support activities for transportation
H53	Postal and courier activities
I	Accommodation and food service activities
J58	Publishing activities
J59_J60	Motion picture, video and television programme production, sound recording and music publishing activities; programming and broadcasting activities
J61	Telecommunications
J62_J63	Computer programming, consultancy and related activities; information service activities
K64	Financial service activities, except insurance and pension funding
K65	Insurance, reinsurance and pension funding, except compulsory social security
K66	Activities auxiliary to financial services and insurance activities
L68	Real estate activities
M69_M70	Legal and accounting activities; activities of head offices; management consultancy activities
M71	Architectural and engineering activities; technical testing and analysis
M72	Scientific research and development
M73	Advertising and market research
M74_M75	Other professional, scientific and technical activities; veterinary activities
N	Administrative and support service activities
O84	Public administration and defence; compulsory social security
P85	Education
Q	Human health and social work activities
R_S	Other service activities
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
U	Activities of extraterritorial organizations and bodies

Table 2: Collection of the input-output tables aggregated industries

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
China	65.82%	66.01%	66.10%	66.96%	68.53%	68.75%	68.78%	68.37%	68.72%	69.10%	68.97%	69.01%	68.91%	68.94%	68.94%
France	84.15%	84.95%	85.30%	85.87%	86.07%	85.65%	85.08%	86.93%	86.83%	85.81%	85.11%	83.39%	81.66%	81.22%	81.12%
Germany	86.19%	85.87%	85.84%	86.42%	87.28%	87.21%	87.21%	87.85%	88.46%	88.65%	88.36%	88.87%	88.17%	88.52%	88.55%
Italy	88.49%	88.49%	88.62%	88.90%	89.06%	89.16%	89.19%	89.25%	89.32%	89.41%	89.29%	89.54%	89.25%	89.67%	89.67%
Japan	80.93%	80.36%	80.45%	80.42%	80.29%	80.26%	80.29%	80.23%	80.42%	81.28%	81.03%	81.19%	81.09%	80.90%	80.77%
Russia	32.65%	33.04%	33.51%	33.51%	33.71%	33.99%	34.09%	34.25%	34.41%	34.85%	34.44%	34.44%	34.44%	34.47%	34.60%
Spain	86.93%	86.99%	87.28%	88.11%	88.36%	88.52%	88.68%	88.93%	89.16%	88.87%	87.50%	87.60%	87.34%	87.40%	87.21%
UK	88.49%	88.71%	88.97%	89.13%	89.41%	89.32%	89.57%	89.92%	90.05%	89.51%	89.80%	89.96%	90.08%	90.18%	90.31%
USA	94.32%	94.39%	94.20%	94.20%	94.26%	94.36%	94.39%	94.36%	94.55%	94.45%	94.52%	94.67%	94.58%	94.74%	94.71%

Table 3: Graph density table

	Russia		China		France		Japan		Italy		Germany		Spain		Usa		Uk	
	CC	H	CC	H	CC	H	CC	H	CC	H	CC	H	CC	H	CC	H	CC	H
2000	0.00415	0.03889	0.00352	0.02891	0.00424	0.02458	0.00193	0.02602	0.00283	0.02493	0.00282	0.02459	0.00156	0.02085	0.00841	0.02500	0.00268	0.02546
2001	0.00440	0.03877	0.00367	0.02853	0.00425	0.02415	0.00195	0.02593	0.00285	0.02498	0.00269	0.02465	0.00148	0.02124	0.00804	0.02526	0.00262	0.02576
2002	0.00492	0.03905	0.00377	0.02823	0.00430	0.02439	0.00184	0.02593	0.00283	0.02507	0.00268	0.02445	0.00116	0.02196	0.00837	0.02547	0.00244	0.02584
2003	0.00518	0.03984	0.00384	0.02795	0.00444	0.02460	0.00177	0.02590	0.00273	0.02528	0.00279	0.02471	0.00104	0.02248	0.00766	0.02565	0.00222	0.02634
2004	0.00596	0.04069	0.00351	0.02882	0.00443	0.02488	0.00163	0.02601	0.00265	0.02533	0.00262	0.02481	0.00105	0.02293	0.00718	0.02595	0.00209	0.02638
2005	0.00370	0.04120	0.00387	0.02837	0.00471	0.02499	0.00229	0.02612	0.00262	0.02538	0.00274	0.02485	0.00082	0.02339	0.00777	0.02606	0.00219	0.02652
2006	0.00394	0.04148	0.00400	0.02874	0.00526	0.02510	0.00228	0.02600	0.00269	0.02539	0.00272	0.02487	0.00079	0.02361	0.00403	0.02607	0.00170	0.02665
2007	0.00434	0.04116	0.00387	0.02927	0.00489	0.02545	0.00215	0.02601	0.00269	0.02536	0.00271	0.02494	0.00086	0.02330	0.00348	0.02585	0.00157	0.02699
2008	0.00409	0.04112	0.00368	0.02961	0.00483	0.02571	0.00195	0.02616	0.00264	0.02546	0.00281	0.02514	0.00103	0.02290	0.00297	0.02587	0.00146	0.02752
2009	0.00685	0.04239	0.00365	0.02972	0.00453	0.02610	0.00197	0.02620	0.00177	0.02568	0.00312	0.02532	0.00096	0.02253	0.00288	0.02517	0.00309	0.02643
2010	0.00813	0.04246	0.00369	0.02978	0.00225	0.02675	0.00193	0.02576	0.00171	0.02507	0.00342	0.02562	0.00221	0.02267	0.00308	0.02644	0.00259	0.02624
2011	0.0075	0.04271	0.0036	0.03027	0.0042	0.02684	0.0019	0.02584	0.0017	0.02517	0.0034	0.02579	0.0020	0.02232	0.0026	0.02623	0.0026	0.02644
2012	0.00844	0.04276	0.00350	0.03021	0.00424	0.02690	0.00191	0.02585	0.00374	0.02504	0.00332	0.02560	0.00187	0.02222	0.00547	0.02626	0.00232	0.02623
2013	0.00473	0.04291	0.00360	0.03014	0.00444	0.02700	0.00201	0.02596	0.00332	0.02501	0.00356	0.02563	0.00175	0.02241	0.00272	0.02625	0.00252	0.02589
2014	0.00480	0.04287	0.00372	0.02993	0.00439	0.02709	0.00199	0.02583	0.00361	0.02502	0.00368	0.02566	0.00180	0.02312	0.00282	0.02624	0.00243	0.02616

Table 4: Clustering and Herfindahl coefficient for each national graph, year 2014

Sector Id	Local influence	Indirect Influence
A01	0.8269	0.8024
A02	0.8230	0.7876
A03	0.8519	0.8131
B	0.9185	0.8942
C10-C12	0.7868	0.0000
C13-C15	0.7916	0.7707
C16	0.8368	0.8150
C17	0.8418	0.8199
C18	0.8882	0.8651
C19	0.8400	0.8178
C20	0.8236	0.0000
C21	0.9372	0.9120
C22	0.8992	0.8757
C23	0.8969	0.8736
C24	0.8132	0.7920
C25	0.8561	0.8338
C26	0.9457	0.9207
C27	0.9011	0.8773
C28	0.8557	0.8331
C29	0.9070	0.8834
C30	0.9537	0.9289
C31_C32	0.9020	0.0000
C33	0.9400	0.0000
D35	0.7402	0.7206
E36	0.8529	0.8275
E37-E39	0.8406	0.8180
F	0.8332	0.8115
G45	0.9397	0.0000

Sector Id	Local influence	Indirect Influence
G46	0.9337	0.0000
G47	0.8730	0.0000
H49	0.8398	0.8176
H50	0.8488	0.8219
H51	0.8337	0.8089
H52	0.8280	0.0000
H53	0.8836	0.8573
I	0.8400	0.0000
J58	0.9222	0.8978
J59_J60	0.8498	0.8270
J61	0.9129	0.8887
J62_J63	0.8716	0.8482
K64	0.7713	0.7509
K65	0.6698	0.6503
K66	0.8792	0.8559
L68	0.8418	0.0000
M69_M70	0.8844	0.8614
M71	0.8937	0.8704
M72	0.9214	0.8970
M73	0.8422	0.8203
M74_M75	0.9175	0.8932
N	0.9485	0.0000
O84	0.8861	0.8623
P85	0.8816	0.8579
Q	0.8665	0.8432
R_S	0.8919	0.0000
T	0.0000	0.0000
U	0.0000	0.0000

Table 5: Centrality, Italy 2014

	China	France	Germany	Italy	Japan	Russia	Spain	UK	USA
A01	0.00547	0.00392	0.00231	0.00298	0.00117	0.00660	0.00033	0.00131	0.00194
A02	0.00122	0.00109	0.00060	0.00045	0.00034	0.00000	0.00064	0.00029	0.00040
A03	0.00137	0.00118	0.00029	0.00055	0.00049	0.00000	0.00057	0.00042	0.00032
B	0.00782	0.00121	0.00095	0.00117	0.00073	0.01147	0.00479	0.00329	0.00245
C10-C12	0.00828	0.00783	0.00515	0.00557	0.00306	0.00689	0.00126	0.00506	0.00363
C13-C15	0.00615	0.00211	0.00104	0.00419	0.00132	0.00123	0.00087	0.00359	0.00115
C16	0.00300	0.00185	0.00189	0.00178	0.00057	0.00249	0.00128	0.00344	0.00130
C17	0.00294	0.00254	0.00236	0.00263	0.00202	0.00430	0.00127	0.00298	0.00221
C18	0.00202	0.00220	0.00212	0.00209	0.00159	0.00000	0.00158	0.00348	0.00139
C19	0.00697	0.00395	0.00294	0.00360	0.00187	0.01731	0.00325	0.00153	0.00346
C20	0.01033	0.00461	0.00478	0.00352	0.00358	0.00737	0.00110	0.00425	0.00356
C21	0.00233	0.00207	0.00201	0.00166	0.00123	0.00000	0.00158	0.00350	0.00197
C22	0.00519	0.00350	0.00383	0.00423	0.00271	0.00449	0.00142	0.00136	0.00253
C23	0.00616	0.00339	0.00298	0.00328	0.00151	0.00482	0.00232	0.00081	0.00149
C24	0.00987	0.00352	0.00394	0.00402	0.00329	0.01262	0.00223	0.00127	0.00236
C25	0.00613	0.00484	0.00534	0.00567	0.00302	0.00000	0.00080	0.00129	0.00345
C26	0.00809	0.00189	0.00256	0.00223	0.00329	0.00565	0.00141	0.00177	0.00230
C27	0.00794	0.00196	0.00407	0.00306	0.00200	0.00000	0.00185	0.00170	0.00127
C28	0.00836	0.00250	0.00630	0.00596	0.00242	0.00665	0.00176	0.00106	0.00270
C29	0.00626	0.00256	0.00699	0.00359	0.00241	0.00675	0.00116	0.00215	0.00287
C30	0.00270	0.00343	0.00172	0.00228	0.00114	0.00000	0.00127	0.00174	0.00192
C31_C32	0.00197	0.00160	0.00202	0.00321	0.00168	0.00283	0.00128	0.00203	0.00194
C33	0.00000	0.00381	0.00299	0.00237	0.00000	0.00000	0.00497	0.00527	0.00057
D35	0.00803	0.00683	0.00703	0.00659	0.00361	0.02025	0.00101	0.00109	0.00252
E36	0.00087	0.00205	0.00110	0.00123	0.00087	0.00000	0.00128	0.00085	0.00031
E37-E39	0.00111	0.00391	0.00330	0.00410	0.00080	0.00000	0.00435	0.00154	0.00147
F	0.00953	0.00919	0.00921	0.00923	0.00486	0.01141	0.00160	0.00125	0.00368
G45	0.00000	0.00312	0.00315	0.00312	0.00299	0.00576	0.00503	0.00748	0.00220
G46	0.00883	0.01412	0.00769	0.01099	0.00531	0.02206	0.00261	0.00218	0.00773
G47	0.00293	0.00515	0.00612	0.00509	0.00317	0.01354	0.00316	0.00189	0.00384
H49	0.00562	0.00647	0.00591	0.00700	0.00341	0.01196	0.00051	0.00620	0.00411
H50	0.00188	0.00232	0.00108	0.00174	0.00102	0.00114	0.00089	0.00106	0.00089
H51	0.00140	0.00164	0.00165	0.00187	0.00070	0.00429	0.00273	0.00288	0.00149
H52	0.00283	0.00627	0.00586	0.00530	0.00180	0.00766	0.00077	0.00181	0.00218
H53	0.00075	0.00228	0.00284	0.00139	0.00079	0.00000	0.00272	0.00222	0.00132
I	0.00409	0.00613	0.00267	0.00494	0.00431	0.00353	0.00104	0.00405	0.00437
J58	0.00000	0.00223	0.00285	0.00168	0.00104	0.00000	0.00109	0.00388	0.00197
J59_J60	0.00000	0.00379	0.00132	0.00191	0.00108	0.00000	0.00248	0.00303	0.00207
J61	0.00278	0.00608	0.00394	0.00381	0.00271	0.00632	0.00157	0.00730	0.00404
J62_J63	0.00130	0.00542	0.00532	0.00470	0.00268	0.00000	0.00286	0.00431	0.00413
K64	0.00609	0.00948	0.00692	0.00631	0.00432	0.00855	0.00110	0.00321	0.00401
K65	0.00155	0.00492	0.00356	0.00160	0.00178	0.00000	0.00065	0.00023	0.00367
K66	0.00000	0.00451	0.00160	0.00321	0.00000	0.00000	0.00290	0.00000	0.00305
L68	0.00225	0.00927	0.00859	0.00601	0.00204	0.00907	0.00244	0.00403	0.00675
M69_M70	0.00672	0.01328	0.00695	0.00669	0.00000	0.00000	0.00173	0.00135	0.00786
M71	0.00000	0.00573	0.00384	0.00377	0.00000	0.00000	0.00050	0.00158	0.00392
M72	0.00164	0.00278	0.00146	0.00130	0.00092	0.00000	0.00135	0.00115	0.00262
M73	0.00000	0.00343	0.00229	0.00282	0.00212	0.00000	0.00101	0.00314	0.00263
M74_M75	0.00310	0.00270	0.00235	0.00323	0.00480	0.00000	0.00395	0.00151	0.00124
N	0.00131	0.01499	0.01076	0.00932	0.00180	0.01458	0.00247	0.00385	0.00820
O84	0.00311	0.00601	0.00672	0.00269	0.00330	0.01501	0.00145	0.00247	0.00824
P85	0.00271	0.00488	0.00273	0.00192	0.00113	0.00261	0.00208	0.00217	0.00173
Q	0.00222	0.00433	0.00305	0.00336	0.00286	0.00395	0.00250	0.00064	0.00388
R_S	0.00525	0.00522	0.00476	0.00518	0.00246	0.00585	0.00000	0.00106	0.00418
T	0.00000	0.00000	0.00000	0.00000	0.00161	0.00000	0.00000	0.00174	0.00040
U	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00131	0.00000

Table 6: Industry clustering coefficient per country, year 2014

List of Figures

List of Tables

Bibliography

[1] Barabási A.L. (2014), "Network science"

[2] Barrenas F., Chavali S., Holme P., Mobini M., Benson M. (2009), "Network properties of complex human disease genes identified through genomewide association studies"

[3] Benita F., Sarica S. and Bansal G. (2020), "Testing the static and dynamic performance of statistical methods for the detection of national industrial clusters"

[4] Brandes U., Robins G., McCranie A. and Wasserman S. (2013), “What is network science?"

[5] Caldarelli G. and Vespignani A.(2007), "Large Scale Structure and Dynamics of Complex Networks"

[6] Carvalho V. (2009) "Aggregate fluctuations and the network structure of intersectoral trade"

[7] Cerina F, Zhen Zhu, Chessa A, Riccaboni M. (2015), "World Input-Output Network"

[8] Duc Thi Luu, Napoletano M., Fagiolo G., Roventini A. e Sgrignoli P. (2017) "Uncovering the Network Complexity in Input-Output Linkages among Sectors in European Countries"

[9] Emmert-Streib F., Tripathi S., Yli-Harja O. and Dehmer M.(2018), "Understanding the World Economy in Terms of Networks: A Survey of Data-Based Network Science Approaches on Economic Networks"

[10] Estrada E. and Knight P. (2015) "A First Course in Network Theory"

[11] Fagiolo G. (2007) "Clustering in complex directed networks"

[12] Gräbner, Claudius and Heinrich, Torsten and Kudic, Muhamed (2017) "Network theory and social economics - a promising conjunction?"

[13] Guanrong Chen, Xiaofan Wang, and Xiang Li (2015), "Fundamentals of Complex Networks : Models, Structures and Dynamics"

[14] Guerra Hernandez, Ana Isabel (2012), "Merging the Hypothetical Extraction Method and the Classical Multiplier Approach: A Hybrid Possibility for Identifying Key Distributive Sectors"

[15] E. P. Harvey and D. R.J. O'Neale (2019) "Using network science to quantify economic disruptions in regional input-output networks"

[16] Higgins P.(2011) "La matematica dei social network. Una introduzione alla teoria dei grafi"

[17] ISTAT (2015) "Rapporto sulla competitività dei settori produttivi"

[18] Jackson M. (2014), "The Past and Future of Network Analysis in Economics"

[19] Leonidov A. and Serebryannikova E. (2018) "Dynamical topology of highly aggregated input–output networks"

[20] Leicht E. A. and Newman M. E. J. (2007) "Community structure in directed networks"

[21] Leontief W., "Econlib. Library of Economics and Liberty"

[22] Lewis T. (2008), "Network Science: Theory and Applications"

[23] Lind D. (2014) Value Creation and Structural Change during the Third industrial Revolution. The "Swedish Economy from a Vertical Perspective."

[24] Miller E., Peter D. Blair (2009) "Input-Output Analysis"

[25] McNerney J. (2009) "Network Properties of Economic Input-Output Networks"

[26] J. McNerneya, B. D. Fath and G. Silverberg (2013), "Network structure of inter-industry flows"

[27] Ming Xu and Sai Liang (2019) "Input–output networks offer new insights of economic structure"

[28] Molontay R. and Nagy M. (2020), "Two Decades of Network Science"

[29] Newman M. E. J. and Girvan M. (2003) " Finding and evaluating community structure in networks"

[30] Ore O. and Wilson R.(1996), "Graphs and Their Uses"

[31] Porter M. E. (1990) "The Competitive Advantage of Nations"

[32] Schweitzer F., Fagiolo G., Sornette D., Vega-Redondo F. e White D. (2009) "Economic networks: what do we know and what do we need to know?"

[33] Semanur Soyyigit and Yasemin Asu Çırpıci (2017) "An Input-Output Network Structure Analysis Of Selected Countries"

[34] Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R. and de Vries, G. J. (2015), "An Illustrated User Guide to the World Input–Output Database: the Case of Global Automotive Production"

[35] Tong Qiao, Wei Shan, Ganjun Yu and Chen Liu (2018) "A Novel Entropy-Based Centrality Approach for Identifying Vital Nodes in Weighted Networks"

[36] Vasudev C. (2006) "Graph Theory with Applications"

[37] Vespignani A. (2018) "Twenty years of network science"

[38] Wei Li DrorY. Kenett, KazukoYamasaki, H. Eugene Stanley and Shlomo Havlin (2017) "Ranking the economic importance of countries and industries"

[39] Sink, Todd (2010) "Input-Output Models"

[40] https://www.maa.org/press/periodicals/convergence/leonard-eulers-solution-to-the-konigsberg-bridge-problem

[41] https://www.theguardian.com/technology/2008/aug/03/internet.email

[42] https://en.wikipedia.org/wiki/

[43] https://databank.worldbank.org/source/world-development-indicators

[44] http://www.wiod.org/release16