THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Réseaux, Télécommunications, Systèmes et Architecture

Présentée et soutenue par :

M. SAMER EL ZANT

le vendredi 6 juillet 2018

Titre :

Google matrix analysis of Wikipedia networks

Ecole doctorale :

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (I.R.I.T.)

Directeur(s) de Thèse :

M. DIMA SHEPELYANSKY

MME KATIA JAFFRES-RUNSER

Rapporteurs :

M. MARIO ARIOLI, UNIVERSITA LUM JEAN MONNET

M. PIERRE BORGNAT, CNRS

Membre(s) du jury :

M. PIERRE BORGNAT, CNRS, Président

M. DIMITRI SHEPELYANSKY, CNRS TOULOUSE, Membre

M. JOSE LAGES, UNIVERSITE DE FRANCHE COMTE, Membre

Mme KATIA JAFFRES-RUNSER, INP TOULOUSE, Membre

Mme SANDRINE MOUYSSET, UNIVERSITE TOULOUSE 3, Membre

Acknowledgments

Sincere gratitude to my supervisors Dr. Katia Jaffrès-Runser and Prof. Dima L. Shepelyansky for encouraging my research and supporting my PhD years while also providing the opportunity to enrich my knowledge through participation in international conferences. I thank him for enrolling me in a doctoral courses and schools and giving me the chance for teaching which improved my teaching and research skills.

Furthermore, I would like to thank my PhD committee members, Prof. Pierre Borgnat, Prof. Mario Arioli, Dr. Sandrine Mouysset and Dr. José Lages for reading and reporting my manuscript.

My thanks and appreciations also go to my colleagues in the "IRIT" for their help, support and collaboration to succeed my PhD. It was a pleasure working with all the partners in the project.

Special thanks to Dr. Farouk, Dr. Aziz, Dr. Karim and Eng. Fouad for helping me during my PhD years. Big thanks to all my friends for all their moral and professional support.

Deep appreciation to my family for the great assistance and extended support. I am especially grateful to my mother Samira and my father Amid for their unconditional support during my three years of working hard to complete my thesis and also achieving my career objectives. Thank you both for giving me strength to chase my dreams.

I would like to give special thanks to my brothers and sisters, Nahla, Abir, Manal, Majed, Bassem for encouraging me in all of my pursuits. Sincere words of gratitude and respect are preserved for my eldest sister Dr. Nahla for providing the necessary financial and academic support. Besides, I want to thank my sister Manal and my brother Bassem for being my family in a foreign country working on my thesis and building my future.

A tender thought to my nieces and nephews, Fadi, Ayman, Ryma, Samer, Sirine, Lara, Amid, Lyne, Amir, Sarah and Clara who embellished my life by their innocence.

Résumé

Cette thèse s'intéresse à l'analyse du réseau dirigé extrait de la structure des hyperliens de Wikipédia. Notre objectif est de mesurer les interactions liant un sous-ensemble de pages du réseau Wikipédia. Par conséquent, nous proposons de tirer parti d'une nouvelle représentation matricielle appelée matrice réduite de Google ou "reduced Google Matrix". Cette matrice réduite de Google (GR) est définie pour un sous-ensemble de pages donné (c-à-d un réseau réduit). Comme pour la matrice de Google standard, un composant de GR capture la probabilité que deux nœuds du réseau réduit soient directement connectés dans le réseau complet. Une des particularités de GR est l'existence d'un autre composant qui explique la probabilité d'avoir deux nœuds indirectement connectés à travers tous les chemins possibles du réseau entier.

Dans cette thèse, les résultats de notre étude de cas nous montrent que GR offre une représentation fiable des liens directs et indirects (cachés). Nous montrons que l'analyse de GR est complémentaire à l'analyse de "PageRank" et peut être exploitée pour étudier l'influence d'une variation de lien sur le reste de la structure du réseau. Les études de cas sont basées sur des réseaux Wikipédia provenant de différentes éditions linguistiques. Les interactions entre plusieurs groupes d'intérêt ont été étudiées en détail : peintres, pays et groupes terroristes. Pour chaque étude, un réseau réduit a été construit. Les interactions directes et indirectes ont été analysées et confrontées à des faits historiques, géopolitiques ou scientifiques. Une analyse de sensibilité est réalisée afin de comprendre l'influence des liens dans chaque groupe sur d'autres nœuds (ex : les pays dans notre cas). Notre analyse montre qu'il est possible d'extraire des interactions précieuses entre les peintres, entre les pays et entre les groupes terroristes. On retrouve par exemple, dans le réseau de peintres issu de GR, un regroupement des artistes par grand mouvement de l'histoire de la peinture. Les interactions bien connues entre les grands pays de l'UE ou dans le monde entier sont également soulignées/mentionnées dans nos résultats. De même, le réseau de groupes terroristes présente des liens pertinents en ligne avec leur idéologie ou leurs relations historiques ou géopolitiques.

Nous concluons cette étude en montrant que l'analyse réduite de la matrice de Google est une nouvelle méthode d'analyse puissante pour les grands réseaux dirigés. Nous affirmons que cette approche pourra aussi bien s'appliquer à des données représentées sous la forme de graphes dynamiques. Cette approche offre de nouvelles possibilités permettant une analyse efficace des interactions d'un groupe de nœuds enfoui dans un grand réseau dirigé.

Abstract

This thesis concentrates on the analysis of the large directed network representation of Wikipedia. Wikipedia stores valuable fine-grained dependencies among articles by linking webpages together for diverse types of interactions. Our focus is to capture fine-grained and realistic interactions between a subset of webpages in this Wikipedia network. Therefore, we propose to leverage a novel Google matrix representation of the network called the reduced Google matrix. This reduced Google matrix (GR) is derived for the subset of webpages of interest (i.e. the reduced network). As for the regular Google matrix, one component of GR captures the probability of two nodes of the reduced network to be directly connected in the full network. But unique to GR, another component accounts for the probability of having both nodes indirectly connected through all possible paths in the full network.

In this thesis, we demonstrate with several case studies that GR offers a reliable and meaningful representation of direct and indirect (hidden) links of the reduced network.We show that GR analysis is complementary to the well-known PageRank analysis and can be leveraged to study the influence of a link variation on the rest of the network structure. Case studies are based on Wikipedia networks originating from different language editions. Interactions between several groups of interest are studied in details: painters, countries and terrorist groups. For each study, a reduced network is built, direct and indirect interactions are analyzed and confronted to historical, geopolitical or scientific facts. A sensitivity analysis is conducted to understand the influence of the ties in each group on other nodes (e.g. countries in our case). From our analysis, we show that it is possible to extract valuable interactions between painters, between countries or between terrorist groups. Network of painters with GR capture art historical fact such a painting movement classification. Well-known interactions of countries between major EU countries or worldwide are underlined as well in our results. Similarly, networks of terrorist groups show relevant ties in line with their objective or their historical or geopolitical relationships.

We conclude this study by showing that the reduced Google matrix analysis is a novel powerful analysis method for large directed networks. We argue that this approach can find as well useful application for different types of datasets constituted by the exchange of dynamic content. This approach offers new possibilities to analyze effective interactions in a group of nodes embedded in a large directed network.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Complex networks

A network, as described by Newman in [4], is a collection of nodes (vertices) connected by edges. If transitioning is only allowed from a node i to a node j, a direction can be added to the edge. In this case, edges are directed and together with the vertices they define a directed network. Networks exist everywhere:

Our societies can be described as a network of people connected by friendship, familial or professional relations. Internet is a network of routers connected by physical data connections. World Wide Web (WWW) is a network of pages linked together by hyperlinks. About 10^{11} neurons are connected by biochemical reactions in human's brain. Protein interactions, network of Hollywood actors, power grids, highways and many other fundamentally different systems existing in real life have a network architecture.

According to [5], a complex network is a graph (network) that models real systems with non-trivial topological features. Examples of complex networks are illustrated in Figure 1.1, representing the networks created by US highways and US airline routes.

In this thesis our network of interest is the open access encyclopedia Wikipedia. Wikipedia is a collection of articles, each article referencing encyclopedic information on a given topic. Articles are linked together by WWW hyperlinks. The network of Wikipedia is a graph whose vertices represent the webpages and whose edges the hyperlinks. Understanding and modeling the structure of such networks have spawned the last century, and main results will be underlined briefly next.

1.2 Modeling complex networks

The spread ubiquity of networks makes their study indispensable. Important studies have been led to understand their structure, offer general model to define different types of complex networks and to extract relevant information from them [6–8]. Such studies build on the definition of network properties (or metrics), the core ones being introduced next.

1.2.1 Measuring network properties

There are three main network-wide metrics that play important roles for the study of a network.

1. *The average path length* represents the mean distance between two nodes averaged over all pair of nodes. In societies, the average number of friends in the shortest chain connecting two person is the average path length. In most real networks this number is relatively small [9].

2. *The clustering coefficient* describes how network nodes are assembled together to form flocks of nodes. In another word, for a node i, the clustering coefficient is the average fraction of its pair of neighbors that are also connected to each other. In network analysis, this coefficient lies between 1, for a network having links between each couple of nodes, and $1/N$ in a network having a random linkage between nodes [7].

3. *The node degree distribution* gives the probability for a node j to have k connections with other nodes. The node degree metric for a node i is defined by the number of links it shares with other nodes. For the case of directed networks, out-degree and in-degree can be defined as well. The out-degree represents the number of out-links and the in-degree the number of in-links of a node. For a network structured as a regular lattice, a simple degree distribution is obtained where all nodes have the same degree.

1.2.2 Models of complex networks

Different networks exhibit different degree distributions, average path length or clustering coefficients.

Random networks In the late 1950s, the two mathematicians Erdős and Renyi have defined random networks [6] as a possible way of modeling complex networks. A random network is a collection of nodes with edges, connecting pairs of them at random [10] and whose degree distribution is exponential. Referring to [11, 12], random graphs exhibit a small average path length (proportional to the network's size) along with a small clustering coefficient. An example of a random network is given in Figure 1.1 on the left panel. However, it has been shown since that real life networks are usually not random and that other generic properties of networks are required to model them.

Small world networks In 1929, Frigyes Karinthy wrote the short story "Chains", where he introduced the six degrees of separation [13]. Karinthy supposed that any two person on earth are connected through five intermediaries (or less). The theory means that any two people picked at random from anywhere on the earth will be connected through five intermediary people. In 1960, the six degrees of separation theory had experimental confirmation by Stanley Milgram named small-world experiment [14, 15]. Milgram sent out 300 packages to people both in Boston and in Nebraska. Now what he wanted these people to do, was to try to send their package to a target person in Boston. But they weren't allowed to send it directly to her but had to send it to someone they knew on a first-name basis who they thought had a better chance of knowing the target. Of course, this friend had to forward it again on same basis. Practically, only 64 packages out of 300 made it, but the ones arriving experienced an average path length of around five (avg.*5.2*) [16, 17].

In 1998, Watts and Strogatz [7] identified the concept of small-world networks that lies between completely regular (i.e. lattice) and random networks. Watts and Strogatz show that three different networks belonging to three different fields are small-world networks. Small-world networks are homogeneous and exponential as Erdős and Renyi random networks, but the key difference is the short cut links that exist between some nodes. Small-world networks are highly clustered like regular graphs, yet with small average path length like random graphs. An example of a small world network is the collaboration graph of film actors (two actors are joined by an edge if they have acted in a film together).

Scale-free networks In 1999, Barabási and Albert [8], while mapping the WWW, found that few highly connected web-pages are holding the network, and so the probability that a web-page i has k links follows a power law distribution and not a bell shape exponential distribution as Erdős and Renyi and Watts and Strogatz. Barabási in [18] illustrates the differences between an exponential network and a scale-free network by comparing US roadmap network with US internal flighmap, as shown in Fig.1.1. On the US roadmap [19], the nodes are cities that are connected by highways, in which, each major city has at least one link to the highway system, and there are no cities served by hundreds of highways. U.S. highway system resembles to Random networks. A plot of the distribution of node linkages will follow a bell-shaped curve, with most nodes having approximately the same number of links.

In contrast, scale-free networks, like US airline system [20] consists of airports as nodes connected by direct flights among them. Most nodes have just a few connections and some have a tremendous number of links. In that sense, the system has no scale. In such networks, the distribution of node linkages follows a power law. The independence of nodes degree from the network scale explains the term scale-free.

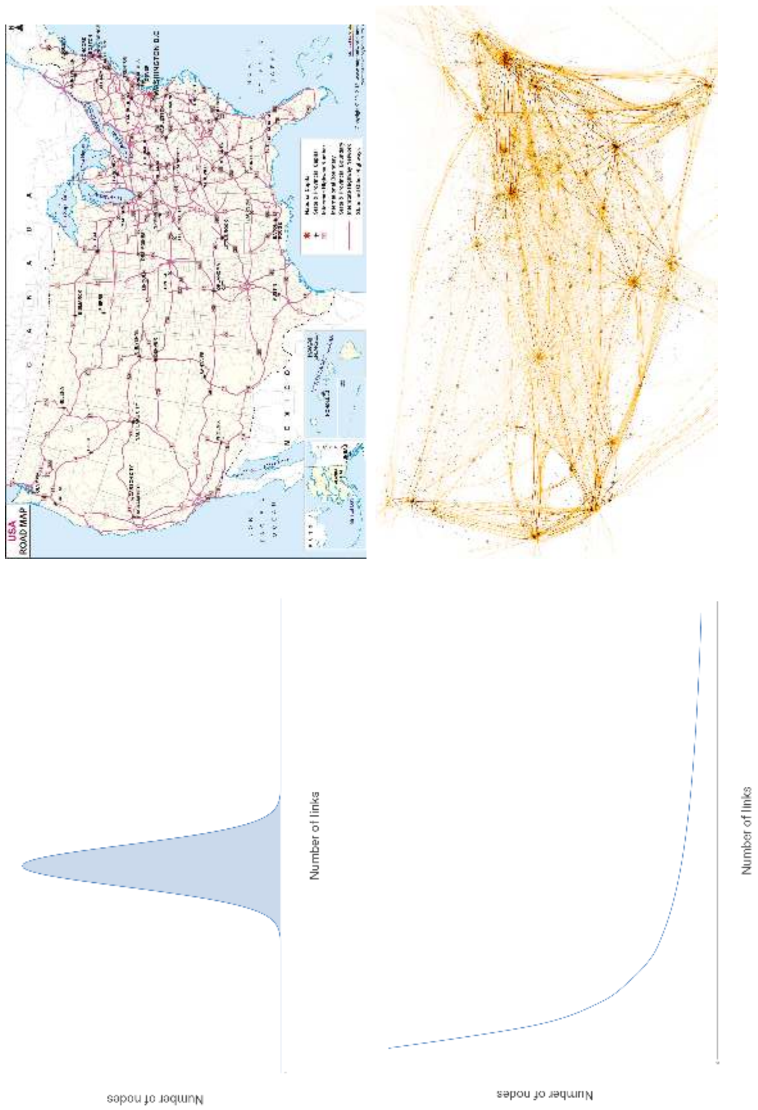

Figure 1.1: **Random vs. Scale-free network.**

Almost all the nodes in Random distribution have the same number of links, while it is not the case for scale-free networks. Left and right side represents roadmap [19] and flightmap [20] networks respectively.

The two basics attribution of Barabási et al. that form the main ingredients of scale-free networks are: *growth and preferential attachments.* Indeed, many real networks are not static, they are dynamically formed by continuous addition of new nodes and new links. Nodes are added to the network progressively. That is what they define as growth. The second ingredient is the preferential attachment. When creating new edges, the probability of a new node connecting to a node i depends on its degree of importance. As example, new research papers are more likely to cite a well-known paper than many other less known ones. For directed networks, scale-free properties apply separately to the in and out-degree of nodes. The score of importance of nodes depends on its number of links. However, in directed networks, each node has two scores of importance, one depends on the number of in-links and the other one depends on the number of out-links [21].

1.2.3 Wikipedia network model

"Wikipedia is a free encyclopedia, written collaboratively by the people who use it. It is a special type of website designed to make collaboration easy, called a wiki." [22].

Each webpage in Wikipedia is related to a clearly defined topic[1]. On each page, there are hyperlinks pointing to other webpages of the same Wikipedia edition that are related to the topic of interest. As such, webpages are interconnected through directed links (i.e. hyperlinks), creating network of webpages. It is common to model this network as a directed graph where vertices represent all webpages and oriented edges represent the hyperlinks. This graph is complex as it can hold up to several millions of vertices and about ten times more edges. In this thesis we have studied the Wikipedia editions listed in Tab.1.1 [23, 24]. They differ in terms of language and of the year they've been collected.

Wikipedia edition	**Number of nodes**	**Number of links**
Arabic 2013	203 326	1 896 621
English 2013	4 212 493	101 611 731
English **2017**	5 416 537	122 232 932
French 2013	1 352 825	34 431 943
German 2013	1 532 977	36 781 077
Italic 2013	1 017 953	25 667 781
Russian 2013	966 284	20 853 206
Spanish 2013	974 021	23 105 758

Table 1.1: **Wikipedia editions and their sizes.**

[1]For instance we have `https://en.Wikipedia.org/wiki/France` for France, `https://en.Wikipedia.org/wiki/United_States` for US, etc.

Similarly to web structure, Wikipedia is an interesting target domain for network analysts due to the hyperlinked structure that provides a direct relationship between web pages and topics. After studying Wikipedia's network structure and based on its properties, it can be defined as a scale-free network. We give herein an example of a French edition[2] to show that its network structure is non-homogeneous and that it follows a power law distribution. Fig.1.2 shows a sample of the French Wikipedia network structure. French Wikipedia network consists of 1 352 825 nodes connected with 34 431 943 hyperlinks. 80% of articles (nodes) have less than 32 out-links, 10% between 33 & 52 out-links, 9% between 53 & 199 and only 1% of nodes have between 200 & 6747 out-links. This distribution (Fig.1.3) exhibits that Wikipedia's network match with scale-free properties.

Wikipedia is, as presented, a huge network of knowledge. In this thesis, we are interested in extracting valuable hidden information from its structure without reading the content of all articles. In order to retrieve such information automatically, we leverage the Google matrix representation of the network.

1.3 Google matrix and PageRank

Previously introduced network models study how network nodes are interconnected. In this thesis, we are more interested in studying the relative importance of nodes and edges in the network. This study looks for the nodes or the edges that are central to the network. Different types of node centrality metrics exist offering different way to measure the importance of nodes.

1.3.1 Node centrality metrics

Typically, one of the most common centrality metric is the node degree. A high node degree node is central as it is known by lots of neighbors. In directed networks, there exist two types of importance scores [21]:

1. Sink importance or in-degree: depends proportionally on the number of in-links. A node absorbing a lot of flow is called *authority*.

2. Source importance or out-degree: depends proportionally on the number of out-links. A node originating a lot of flow is called *hub*.

Another well-known centrality metric is the betweenness centrality. According to Bavelas [26] and Shimbel [27], degree of centrality (importance/reliability) of a node in a network depends on the number of shortest paths passing through it to connect pairs of other nodes. In another word, if a particular

[2] Data collected mid February 2013.

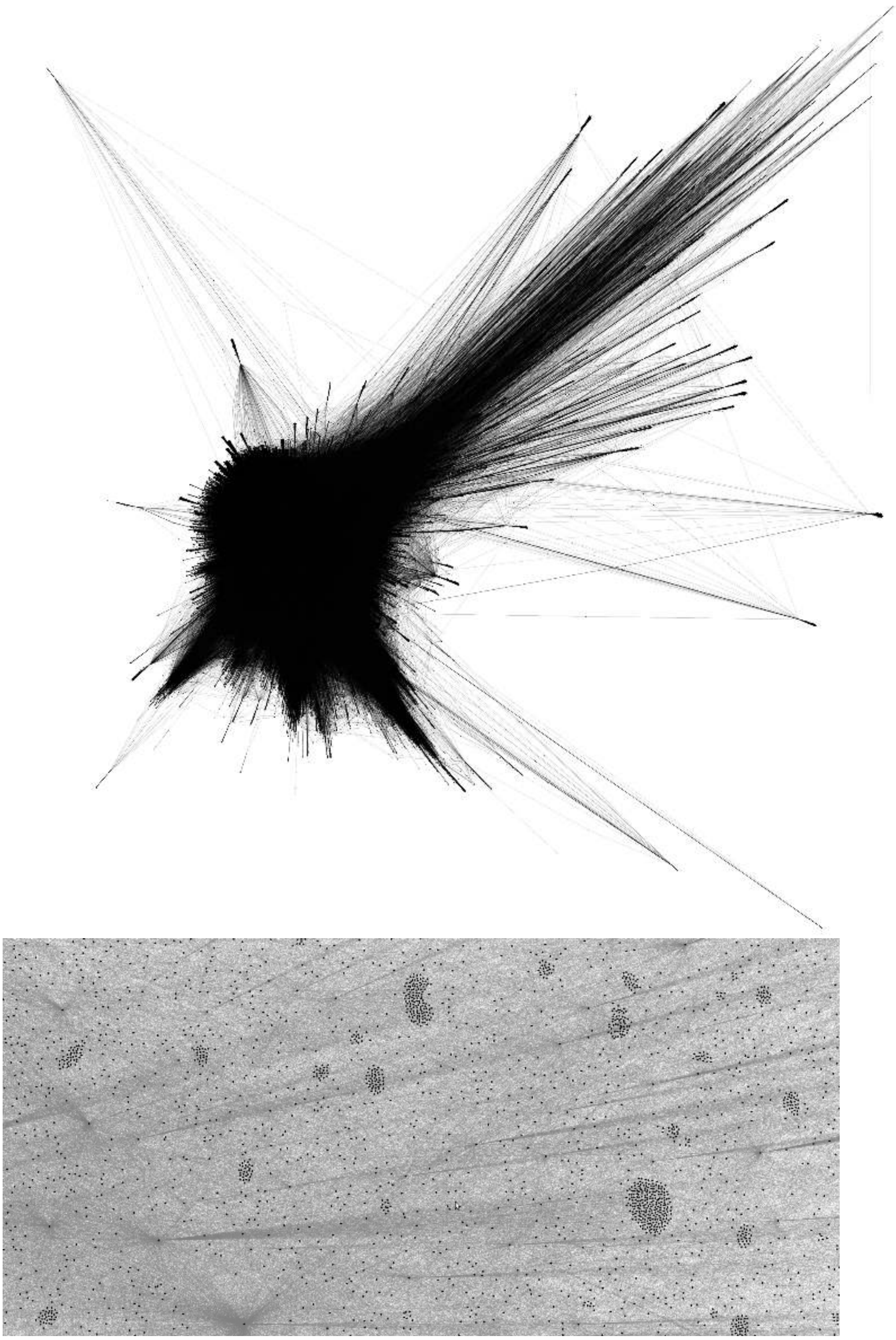

Figure 1.2: **Network structure of French Wikipedia.**

Top panel: only 700 000 hyperlinks are shown representing 2% of the whole network. Bottom panel: A zoomed snapshot from top panel figure. Plotted with Gephi [1] using ForceAtlas2 algorithm [25].

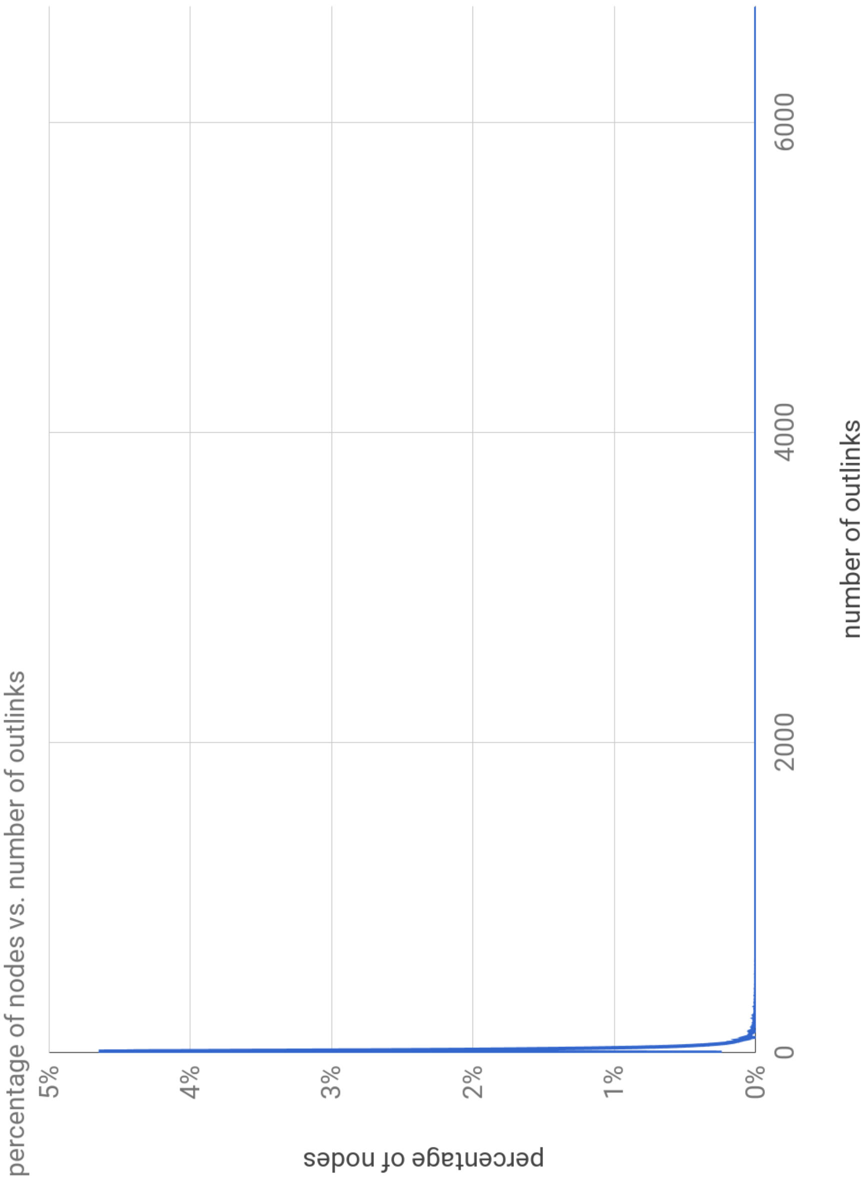

Figure 1.3: **Distribution of out-links for the French Wikipedia edition of 2013.**

person in a group is strategically located on the shortest communication path connecting pairs of others, then this person is in a central position.

Another important centrality metric has been proposed by Brin and Page in their seminal paper [21, 28] introducing the PageRank metric. Each link in a directed network is considered as a recommendation from a node to another one, i.e. a link from node i to node j is i's endorsement of j. Thus, the more in-links (or recommendations) a node has the more important it is. However, the status of the recommender is also important and an endorsement from an important person has more weight (is more significant) than many endorsements from less important ones. However, the weight of each endorsement should be tempered by the total number of recommendations made by the recommender, so that, if an important person has written over $50,000$ recommendations in his life, then his recommendation suddenly drops in weight. This idea is captured in the PageRank centrality metric which is derived from the Google matrix. Google matrix is a particular Markov transition probability matrix that describes the world wide web (WWW) network and is used by Google's PageRank algorithm, a central piece of the Google search engine. Next, we introduce Markov chains, and the specific example of the Google matrix in the context of WWW network modeling.

1.3.2 Markov chains

In 1906, Andrei Andreyevich Markov invented the chains to represent a probabilistic sequence to a process movements between states [29]. Markov chain is a stochastic process. Stochastic process is a set of random variables that describe the state of the process at a time t. The range of available states is referred as state space. Transitions from the current state to the next one are said to be memoryless and thus governed by probabilities usually represented in a *transition probability matrix* $\mathbf{H}$ sized $N \times N$, where N is the size of the state space. The column sum of $\mathbf{H}$ is equal to unity and is composed on non-negative elements, $\mathbf{H}$ being thus a stochastic matrix.

To describe the process of browsing the WWW, each webpage is accounted for as a possible state of the web browsing activity. As such, for an edition of Wikipedia, the state space is composed of all articles. The transition matrix gives the probability of moving from article i to article j using the hyperlinked structure of Wikipedia. An example of transition probability matrix $\mathbf{H}$ construction for WWW browsing is the following. Assuming that ℓ_i is the number of out-links of node i and that all links have same weight, then $\mathbf{H}_{ji} = 1/\ell_i$ if i and j are connected and *zero* otherwise. The resulting transition matrix for the small illustrative example of Figure 1.4 is

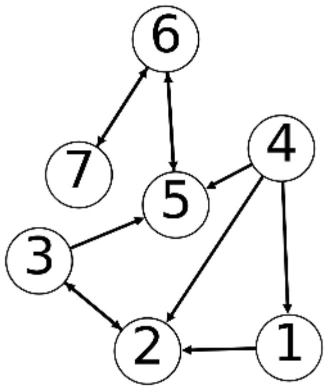

Figure 1.4: **Directed graph representing a WWW network of seven nodes. Arrows represent the direction of the hyperlink existing between two webpages. A double-sided arrow represents the presence of two opposite direction hyperlinks.**

the following:

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 1/3 & 0 & 0 & 0 \\ 1 & 0 & 1/2 & 1/3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 \end{bmatrix}$$

From this matrix, it is possible to derive the steady state probability vector that represents the probability, for the web browsing activity, to end in a specific webpage. This vector Π, composed of N elements, is obtained by solving the following steady state equation:

$$\Pi = \mathbf{H}\,\Pi$$

Since $\mathbf{H}$ is stochastic, its largest eigenvalue equals one and the steady-state vector Π is the corresponding eigenvector. For the example given above, the normalized steady state vector is: $\Pi = [0, 0, 0, 0, 0.25, 0.50, 0.25]$ for eigenvalue 1. In this case, browsing activity is likely to end with probability 0.5 on node 6, and with probability 0.25 on nodes 5 and 7 if the process is repeated at infinity. Intuitively, this is related to the strong interactions between these 3 nodes that will eventually concentrate all moves of the WWW surfer performing this browsing activity.

This Markov chain model has some limitations:

- It can't account for nodes with zero out-links (the so-called *dangling* *nodes* of Brin and Page). A node with no out-links creates a 0-sum column in **H**, and hence a non-stochastic matrix.

- It doesn't capture the fact that browsing may not always follow the hyperlink structure of the network.

Next, we show how these limitations have been solved by Brin and Page in their definition of the Google matrix.

1.3.3 Google Matrix and PageRank

PageRank PageRank is the first algorithm, but not the only one, used by Google to compute the importance of websites in their search engine results. PageRank was named after Larry Page and uses the notion of random surfer. A random surfer is a web surfer who moves randomly between webpages over the hyperlink structure of the network. PageRank vector reflects the probability of the random surfer to end on a given webpage. The decreasing order of PageRank vector gives the PageRank score of a webpage. As such, the lower the PageRank score of a webpage, the larger the probability for the surfer to end on this page. A website's PageRank score depends not only on the number of in-links but as well on the quality of these in-links. It relies on the Markov chain model introduced previously, but Brin and Page have solved the aforementioned issues in the following way.

In order to solve the problem of dangling nodes, Brin and Page introduced the assumption that the random surfer, upon reaching a dangling node, has the same probability to go to any other node (i.e. it can teleport any other webpage with equal probability). Mathematically, the new matrix states as follows:

$$S_{ij} = H_{ij} + (e/N)a^T \tag{1.1}$$

where a is the dangling node vector of size N, with $a_i = 1$ if i is a dangling node and *zero* otherwise. e/N is a uniform distribution vector of size N.

With this modification, they ensure that the transition matrix S is stochastic and it fits a fine description of the network. By applying this modification to the other example of Figure 1.5 where node 2 is a dangling node, we get:

$$S = \begin{bmatrix} 0 & 1/7 & 0 & 1/3 & 0 & 0 & 0 \\ 1 & 1/7 & 1/2 & 1/3 & 0 & 0 & 0 \\ 0 & 1/7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/7 & 1/2 & 1/3 & 0 & 1/2 & 0 \\ 0 & 1/7 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1/7 & 0 & 0 & 0 & 1/2 & 1 \end{bmatrix}$$

This modification solves the problem of dangling node, and then, S describes the case of a random web surfer that follows the hyperlink structure.

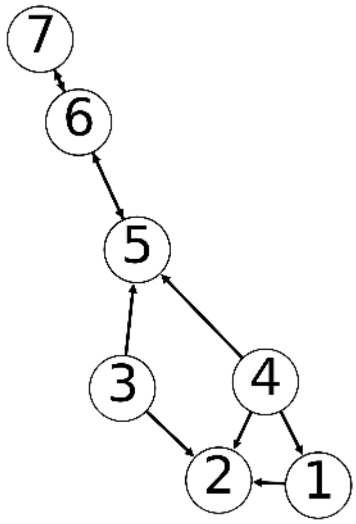

Figure 1.5: **Directed graph representing a network of seven nodes.**

However, in reality, a random surfer may visit a new web-page without following a link (e.g. by directly typing the URL address of a new webpage). In order to take this fact into consideration, Brin and Page use a damping factor $0 < \alpha < 1$, which for a random surfer determines the probability $(1 - \alpha)$ for a random surfer to teleport to a new page without following a hyperlink. The teleportation matrix $E = 1/Nee^T$ where e^T is e transposed, gives an equal weight $1/N$ to all links.

With this final step, the *Google Matrix* G can be defined as follows from S:

$$G_{ij} = \alpha S_{ij} + (1 - \alpha)/N \ \ , \tag{1.2}$$

The Google matrix for the example of Figure 1.5 is thus:

$$G = \begin{bmatrix} 3/140 & 1/7 & 3/140 & 32/105 & 3/140 & 3/140 & 3/140 \\ 61/70 & 1/7 & 25/56 & 32/105 & 3/140 & 3/140 & 3/140 \\ 3/140 & 1/7 & 3/140 & 3/140 & 3/140 & 3/140 & 3/140 \\ 3/140 & 1/7 & 3/140 & 3/140 & 3/140 & 3/140 & 3/140 \\ 3/140 & 1/7 & 25/56 & 32/105 & 3/140 & 25/56 & 3/140 \\ 3/140 & 1/7 & 3/140 & 3/140 & 61/70 & 3/140 & 3/140 \\ 3/140 & 1/7 & 3/140 & 3/140 & 3/140 & 25/56 & 61/70 \end{bmatrix}$$

General properties of eigenvalues and eigenstates The PageRank vector is the right eigenvector of the Google matrix G for the unity eigenvalue[3].

[3]The basic equation is $G\psi_i = \lambda_i \psi_i$. λ_i and ψ_i represents the eigenvalues and right eigenvectors of G respectively [30]

G, a matrix sized $N \times N$ has N eigenvalues (λ_1,λ_2,λ_3,λ_4,...,λ_N) and eigenvectors, where $\lambda_1 = 1$ in stochastic matrices like G. Since G is primitive ($G^k > 0$), $\lambda_2 < 1$ and so are λ_3,λ_4,...,λ_N [21]. Only the PageRank vector is affected by α while other eigenstates are independent of α due to their orthogonality to the unit left eigenvector at $\lambda = 1$.

Perron-Frobenius G is a typical Google matrix of Perron-Frobenius type [21, 31] for a network with N nodes such that $G_{ij} > 0$ and the column sum normalization $\sum_{i=1}^{N} G_{ij} = 1$ are verified. Following Perron-Frobenius theorem, for a positive square matrix like G, the largest eigenvalue is called *perron root*. The *perron root* r is positive and in our case is equal to 1. The *perron vector* is the eigenvector defined by: $Gp = rp$, with $||p||_1 = 1$. The *perron vector* is the PageRank vector.

Power iteration method As said, the PageRank vector is the stationary probability distribution vector of the Markov transition probability matrix G. Brin and Page have chosen the power iteration method to calculate the PageRank vector [21]. In 1960s, the power method became the standard method for calculating eigenvalues and eigenvectors of a matrix using digital computers [32]. The power iteration method algorithm is the following. At each iteration k, the new value of Π, i.e. Π_{k+1} is calculated recursively using $\Pi_{k+1} = G\Pi_k$ relation. At start, Π_0 elements are set to equal values with sum equal to one. Recursive calculation is performed until steady state is reached, i.e. $\Pi_{k+1} \simeq \Pi_k$. Many reasons are in favor of using the power method:

- Programming and implementation are simple,
- G could be expressed in term of H which is sparse,
- H is storage friendly,
- No method can beat the complexity $O(N)$ of each iteration for the power method (one sparse matrix-vector product).
- The number of iterations until convergence for G is limited.

Indeed, the convergence rate is the rate at which $|\lambda_2/\lambda_1|^k \to 0$. In our case $\lambda_1 = 1$, then the convergence rate depends on λ_2. In [33,34], it is shown that:

$$\lambda_k = \alpha\mu_k \quad \textbf{for} \quad k = 2, 3, 4, ..N. \tag{1.3}$$

where the spectrums are: $\sigma(S) = \{1, \mu_2, ..., \mu_N\}$ and $\sigma(G) = \{1, \lambda_2, ..., \lambda_N\}$. Based on [21], for a WWW structure, $|\mu_2| = 1$ and thus the convergence rate depends on α^k. In this thesis, we choose to stop for an accuracy

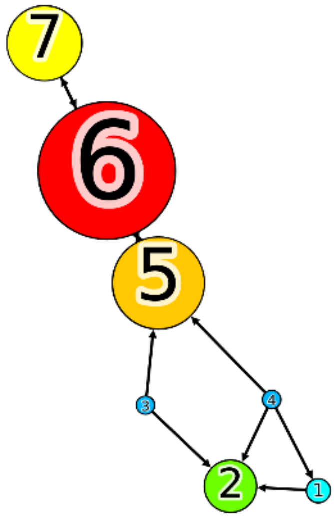

Figure 1.6: **Graphical representation of the resultant PageRank vector for the network of Figure 1.5. The size of the node is proportional to its PageRank vector value.**

obtained after 150 iterations (i.e. *9-10 places of accuracy*): $0.85^{150} = 0.0000000000258725$.

The resultant PageRank vector for our example of Figure 1.5 depicted in Figure 1.6. Its numerical values are:

$$\Pi = \begin{bmatrix} 0.041660 \\ 0.090868 \\ 0.032462 \\ 0.032462 \\ 0.160123 \\ 0.246273 \\ 0.137128 \end{bmatrix}$$

The corresponding PageRank score vector K indicates the rank of each node obtained by the decreasing order of PageRank vector values:

$$K = \begin{bmatrix} 5 \\ 4 \\ 6 \\ 7 \\ 2 \\ 1 \\ 3 \end{bmatrix}$$

K shows that the nodes with identifiers 6, 5, 7, 2, 1, 3 and 4 are in decreasing order of importance in this small network.

CheiRank In addition to the matrix G it is useful to introduce a Google matrix G^* constructed from the adjacency matrix of the same network but with inverted direction of all links [35]. The vector Π^* is called the CheiRank vector [35, 36]. Thus, nodes with many ingoing or outgoing links have small values of $K = 1, 2, 3...$ and $K^* = 1, 2, 3, ...$ respectively [21, 37]. To summarize, PageRank and CheiRank capture the relative importance of nodes in the network. They are extracted from the Google matrix representation of the network of webpages. The Google matrix lists for each link the probability for directly transitioning from one webpage to the other one. The PageRank probability Π represents the probability of ending on a webpage, eventually. Figure 1.7, shows a plotted CheiRank versus PageRank for French (left) and Russian (right) Wikipedia.

An important study, which was introduced by Ermann et al. in [38], is to see what are the other eigenvectors of a Google matrix of Wikipedia, and to what type of information they correspond. Ermann et al. show in their cited paper two examples of node rank corresponding to two eigenvalues $\lambda_2 = 0.97724$ and $\lambda_{80} = -0.8165$. These results show that important information are hidden behind other eigenvectors:

1. For λ_2 linked to the main article *Gaafu Alif Atoll*, the first four nodes are: Gaafu Alif Atoll, Kureddhoo (Gaafu Alif Atoll), Hithaadhoo (Gaafu Alif Atoll), Dhigurah (Gaafu Alif Atoll).

2. For λ_{80} linked to *protein*, the first four nodes are: Photoactivatable fluorescent protein, Kaede (protein), Eos (protein), Fusion protein.

However, analyzing the N eigenvectors is not straightforward if N is large.

In this thesis, we will focus on finding and analyzing these hidden information from Wikipedia networks by using the recent method of K. Frahm and D. Shepelyansky [39] based on Google matrix and called the *Reduced Google matrix* analysis. A complete definition of the Reduced Google matrix is given in Chapter 2.

1.4 Motivation

Networks are part of our life. They are indispensable in the world of culture, business, marketing and many other domains. Social networks are a part of these networks that are changing some of our behaviors.

Academic social networks have become unavoidable for both researchers and research institutions. They are very effective tools and very much used. They transform and improve the visibility of researchers and the work they need to continue and expand their activities. There are different academic social networks used and created by scientists and researchers that make available, in different ways, pedagogical content or research broadly.

Wikipedia as presented before is one of the existing networks that includes, as gathered by Wikimedia projects [40], about 46 million articles in 299 different languages for a publicly accessible user. The different articles (pages) are linked together. Some links are more important than others and can be of importance in different domains.

Research on such networks has derived content-independent effective metrics to rank nodes and edges of the graph based on their relevance to a given criteria such as clustering, importance ranking, etc.. In this thesis we concentrate on one of the most popular network analysis algorithms: the PageRank algorithm [21, 28].

For various language editions of Wikipedia it has been shown that the PageRank vector produces a reliable ranking of historical figures over 35 centuries of human history [23, 36, 37, 41, 42] and a solid Wikipedia ranking of world universities (WRWU) [36, 43]. It has been shown as well that the Wikipedia ranking of historical figures is in a good agreement with the well-known Hart ranking [44], while the WRWU is in a good agreement with the Shanghai Academic ranking of world universities [45].

In order to better understand Wikipedia networks, we study their structures and we search for the importance score of each node. For that purpose, many algorithms are known to analyze networks structures of connected nodes. However, it is really hard to explain interactions between nodes in such large network of several tens of millions of nodes. Just to capture the complexity of analyzing such large networks, we look at a simple metric commonly used to calculate the importance of nodes in a network called the betweenness. This metrics, as said earlier, computes for each node the number of times it is located on the shortest path of any two pair of nodes in the network. A network of N nodes is composed of $\frac{N!}{k!(N-k)!}$ pairs. Dijkstra, a Dutch computer scientist, created an algorithm for finding shortest paths from a source node i to all other nodes in the graph. Dijkstra's shortest path algorithm [46] time complexity is $O(sE\log(E)+N)$ [47] where s is the number of sources, N is the number of nodes and E is the total number of edges. Computing the betweenness centrality [48] for each node has a time complexity of $O(N.E)$ once shortest paths are known [47]. Hence, in addition to the existing research works mentioned in 1.2.3 that prove the reliability of applying PageRank algorithm on Wikipedia networks, it is clear as well that the time complexity of Google's PageRank algorithm of $O(N)$ clearly outperforms a well-known graph centrality like betweenness that accounts for all possible paths in the network.

In this thesis, we will be using the Google matrix to describe and understand the Wikipedia hyperlinked structure. Moreover, we are interested in discovering information hidden in the Google matrix eigenvectors different from the one corresponding to the unity eigenvalue. Therefore, in order to better understand the interaction between a small subset of nodes (compared to the whole network), we propose in this thesis to use the Reduced Google matrix theory [39] presented in Chapter 2.

1.5 Contributions

This section summarizes the scientific contributions of this doctoral research. All contributions listed hereafter have been made with the goal of extracting novel and meaningful knowledge present in different editions of Wikipedia. From these large scale networks collecting many different types of articles belonging to various cultures among the time, we extract fine-grained information to better understand the specificities of networks formed by subsets of pages extracted from Wikipedia.

This thesis presents the following contributions:

- This thesis gives a clear interpretation of the reduced Google matrix theory [39] presented in Chapter 2. It shows that the main derivation steps hold valuable information that be analyzed further and leveraged in the rest of the manuscript.

- From the reduced Google matrix, we can extract a graph of hidden relationships between a chosen subset of nodes. Different types of relationship graphs are studied, originating from the matrix components of the reduced Google matrix or from the Google matrix itself. They are illustrated for a reduced network of painters in Chapter 3, of countries in Chapter 4 and of terrorist groups in Chapter 6.

- A variational analysis of the reduced Google matrix is presented in 5 that underlines how sensitive our selected nodes are to a link variation intensity in the reduced Google matrix. This variational analysis helps us in measuring the impact of a change of relationship intensity between two nodes on the rest of nodes. This study is illustrated for the networks of countries and painters. We show that results obtained by these studies are clearly in line with the common knowledge related to arts history and geopolitics.

- All contributions introduced in Chapters 3, 4 and 5 offer an innovative analysis framework that can be applied to any subset of nodes of a network. We have illustrated how this framework captures novel knowledge within Wikipedia in a study on the world terror network. Reduced networks are here composed of nodes corresponding to the articles describing terrorist groups and articles of countries. Interesting and meaningful ties between terrorist groups and countries are extracted from this study.

These contributions have been published or are under review for possible publication. The list is given below.

1.6 Publications related to this thesis

1.6.1 Journal articles

- K. M. Frahm, S. E. Zant, K. Jaffrès-Runser, and D. L. Shepelyansky, "Multi-cultural Wikipedia mining of geopolitics interactions leveraging reduced Google matrix analysis" Elsevier, PLA, vol. 381, no. 33, pp. 2677 - 2685, September 2017.

- S. E. Zant, K. M. Frahm, K. Jaffrès-Runser, and D. Shepelyansky, "Analysis of world terror networks from the reduced Google matrix of Wikipedia" Springer, EPJB, vol. 91, no.1, pp. 7, January 2018. (Cited among top 5 of 'The Best of the Physics arXiv (week ending October 21, 2017)' by MIT Technology Review.)

1.6.2 Conference proceedings

- S. E. Zant, K. Jaffrès-Runser, and D. Shepelyansky, "Geopolitical interactions from reduced Google matrix analysis of wikipedia" IEEE, MENACOMM, April 2018.

1.6.3 Talks

- S. E. Zant, K. Frahm, K. Jaffrès-Runser, and D. Shepelyansky, "Analyse des interactions géopolitiques par la matrice de Google réduite" AlgoTel 2017, Quiberon, France, May 2017.

1.6.4 Submitted journal articles

- S. E. Zant, K. Jaffrès-Runser, K. M. Frahm, and D. Shepelyansky, "Interactions and influence of world painters from reduced Google matrix of Wikipedia networks" (Submitted to IEEE access).

- S. E. Zant, K. Jaffrès-Runser, and D. Shepelyansky, "Capturing the influence of geopolitical ties from Wikipedia with reduced Google matrix" (Submitted to Plos One). https://arxiv.org/abs/1803.05336

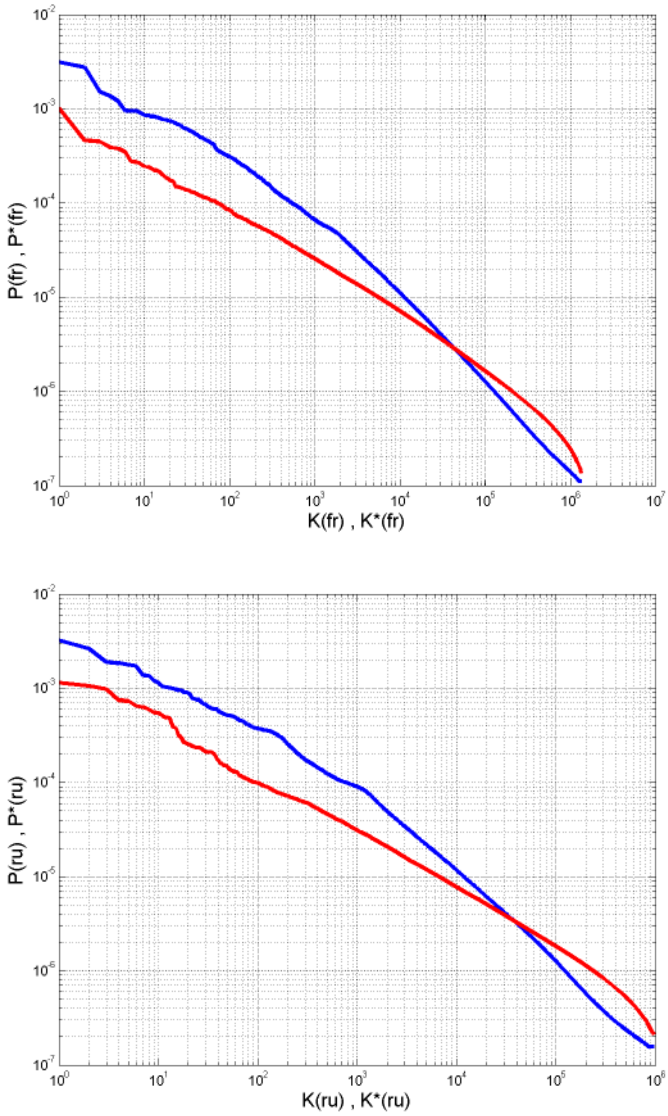

Figure 1.7: **CheiRank versus PageRank for FrWiki and RuWiki.**

Top and bottom panels are FrWiki and RuWiki respectively. Blue and Red lines represent P vs K and P^* vs K^* respectively.

Chapter 2

Reduced Google matrix theory

2.1 Introduction

Google matrix G gives a clear description of the *direct* interactions between nodes of a network. The question answered in this chapter is whether it is possible to snap, for a subset of nodes from a large-scale network, both direct and *indirect* interactions through the rest of nodes of the network in a unique equivalent and reduced network. By indirect interaction, we refer to all possible non-direct (e.g. multi-hop) paths in the full network that interconnect any two nodes of the subset of nodes of interest. In other words, we would like to compute a reduced network matrix representation whose properties for the subset of nodes of interest triggers the same results as the complete matrix G would do.

A solution to this question has been proposed by K. Frahm and D. Shepelyansky in [39]. It is called the reduced Google matrix theory and we show in this thesis that it offers an efficient tool to analyse direct and indirect interactions within a selected subset of nodes. This reduced Google matrix, denoted G_{R}, is of size $N_r \times N_r$ as it is computed for a subset of N_r selected nodes.

As we will see next, G_{R} matrix can be decomposed into three matrix components:

$$G_{\mathrm{R}} = G_{rr} + G_{\mathrm{pr}} + G_{\mathrm{qr}} \ .$$

Each component captures different types of information. The direct interactions between the N_r nodes are given by G_{rr} as it is defined by extracting from G the elements related to these N_r nodes. Matrix G_{pr} is the projector part that mainly captures the PageRank vector contribution and finally, the G_{qr} component captures the indirect interactions between the N_r nodes of interest. This thesis extensively analyses the information held in G_{R} and its components. Their derivation is presented in this chapter.

2.2 Reduced Google matrix

We construct the reduced Google matrix for a certain subset of N_r selected nodes, based on their attachment to a domain and their PageRank, from a

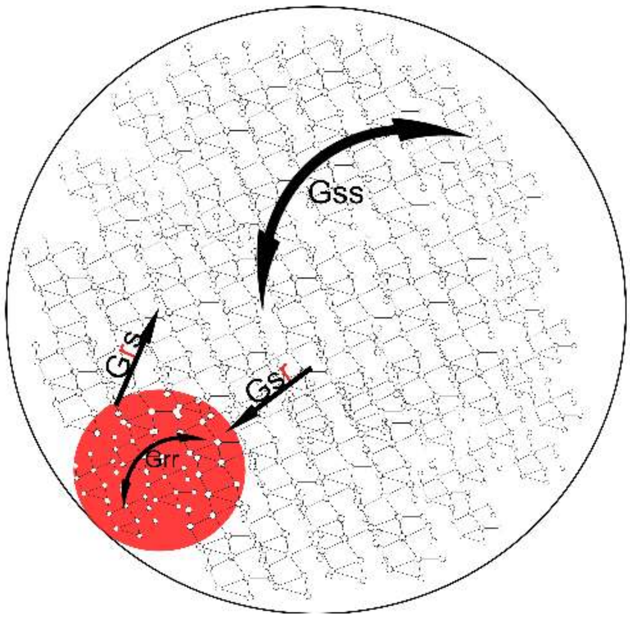

Figure 2.1: Decomposition of G into reduced network and scattering network.

global network of N nodes ($N_r \ll N$). The reduced Google matrix G_R is constructed on the mathematical basis described below. The main element of this construction is to keep the same order of PageRank probabilities of N_r nodes as in the global network and to take into account all indirect links between N_r nodes coupled by transitions via $(N - N_r)$ nodes of the global network.

Let G be a typical Google matrix (1.2) for a network with N nodes such that $G_{ij} > 0$ and the column sum normalization $\sum_{i=1}^{N} G_{ij} = 1$ is verified. We consider a sub-network with $N_r < N$ nodes, called "reduced network". In this case we can write G in a block form:

$$G = \begin{pmatrix} G_{rr} & G_{rs} \\ G_{sr} & G_{ss} \end{pmatrix} \tag{2.1}$$

where the index "r" refers to the selected nodes to be considered in the reduced network and "s" to the other $N_s = N - N_r$ nodes which form a complementary network that we will call the "scattering network". Thus G_{rr} is given by the direct links between the selected nodes N_r, G_{ss} that describes the links between N_s nodes, G_{rs} and G_{sr} represents the links between N_r and N_s. Figure 2.1 represents graphically the decomposition of G into reduced and scattering networks.

PageRank vector P of the full network is the stationary probability distribution vector of matrix G (see Section 1.3.3). Therefore, for a network divided in N_r and N_s nodes, P could be written as:

$$P = \begin{pmatrix} P_r \\ P_s \end{pmatrix} \tag{2.2}$$

The substitution of P and G, from Eq. (2.2) and (2.1), in $G\,P = P$ gives:

$$\begin{aligned} (I - G_{rr})\,P_r - G_{rs}\,P_s &= 0, && (2.3)\\ -G_{sr}\,P_r + (I - G_{ss})\,P_s &= 0. && (2.4) \end{aligned}$$

where I is the unit matrix of corresponding size N_r or N_s. The matrix $I - G_{ss}$ has no two identical columns or rows also it does not contain a row or a column of zeros. Accordingly $I - G_{ss}$:

1. is not singular, i.e. all eigenvalues G_{ss} are, in modulus, strictly smaller than unity.

2. is invertible, then, from (2.4), we obtain that:

$$P_s = (\mathbf{1} - G_{ss})^{-1} G_{sr}\,P_r \tag{2.5}$$

Our aim is to define the reduced Google matrix G_{R} such as the PageRank vector for the N_r selected nodes is preserved. In this case, as we want $G_{\mathrm{R}} P_r = P_r$, and from the substitution of P_s given in Eq. (2.5) into Eq. (2.3), we get the following definition of G_{R}:

$$G_{\mathrm{R}} = G_{rr} + G_{rs}(\mathbf{1} - G_{ss})^{-1} G_{sr} \tag{2.6}$$

Here the contribution of G_{rr} accounts for direct links in the reduced network and the second matrix inverse term corresponds to all contributions of indirect links of arbitrary order. The matrix elements of G_{R} are non-negative since the matrix inverse in (2.6) can be expanded as:

$$(\mathbf{1} - G_{ss})^{-1} = \sum_{l=0}^{\infty} G_{ss}^{\,l} \quad . \tag{2.7}$$

In (2.7), the integer l represents the order of indirect links, i.e. the number of indirect links which are used to connect indirectly two nodes of the reduced network. As we mentioned at the beginning of this chapter, we concentrate on creating a matrix that keeps the order of PageRank of N_r nodes. Hence, intuitively, this matrix must have columns sum normalization being unity. To get the proof that G_{R} also fulfills this condition, we refer the reader to [49].

2.3 Numerical evaluation of G_{R}

This section explains how to evaluate the expression (2.6) of G_{R} in practice as detailed in [49] by Frahm et al. In this thesis, our datasets of networks are composed of millions of nodes, sized between 0.2 million for ArWiki2013 and 5.4 millions for EnWiki2017. This large dataset size has been accounted for in the numerical evaluation method developed herein. The networks studied herein are thus sparse (there are around $\sim 10N$ edges only) and quite large. Moreover, the number of nodes composing the reduced network is of $N_r \sim 10^2$-10^3, which is quite small compared to N. And thus, $N_s \approx N \gg N_r$.

In order to find the matrix inverse $(I-G_{ss})^{-1}$, Gauss algorithm performs elimination steps to the giant matrix $[(I-G_{ss}) \mid I]$ until we reach the identity matrix on the left side, that mean what we have on the right side is the matrix inverse $[I \mid (I-G_{ss})^{-1}]$. If N_s is too large (e.g. $N_s > 10^5$) a direct naive evaluation of the matrix inverse $(I-G_{ss})^{-1}$ in (2.6) by Gauss algorithm is not efficient. In this case we can try the expansion (2.7) provided it converges sufficiently fast with a relatively small number of terms. However, this is most likely not the case for typical applications since G_{ss} is very likely to have at least one eigenvalue very close to unity.

Therefore, the situation where the full Google matrix has a well defined gap between the leading unit eigenvalue and the second largest eigenvalue (in modulus) is considered herein. This is the case if G is defined using a damping factor α, as in (1.2) as the gap is at least $1-\alpha$ which is 0.15 for the standard choice $\alpha = 0.85$ [21]. In order to evaluate the expansion (2.7) efficiently, we need to take out analytically the contribution of the leading eigenvalue of G_{ss} close to unity which is responsible for the slow convergence.

Below we denote by λ_c this leading eigenvalue of G_{ss} and by ψ_R (ψ_L^T) the corresponding right (left) eigenvector such that $G_{ss}\psi_R = \lambda_c\psi_R$ (or $\psi_L^T G_{ss} = \lambda_c \psi_L^T$). Both left and right eigenvectors as well as λ_c can be efficiently computed by the power iteration method in a similar way as the calculation of the PageRank vector. Vectors ψ_R are normalized with $E_s^T\psi_R = 1$ and ψ_L with $\psi_L^T\psi_R = 1$. It is well known (and easy to show) that ψ_L^T is orthogonal to all other right eigenvectors (and ψ_R is orthogonal to all other left eigenvectors) of G_{ss} with eigenvalues different from λ_c. We introduce the operator $\mathcal{P}_c = \psi_R\psi_L^T$ which is the projector onto the eigenspace of λ_c and we denote by $\mathcal{Q}_c = \mathbf{1} - \mathcal{P}_c$ the complementary projector. One verifies directly that both projectors commute with the matrix G_{ss} and in particular $\mathcal{P}_c G_{ss} = G_{ss}\mathcal{P}_c = \lambda_c\mathcal{P}_c$. Therefore we can derive:

$$(\mathbf{1} - G_{ss})^{-1} = \mathcal{P}_c \frac{1}{1-\lambda_c} + \mathcal{Q}_c \sum_{l=0}^{\infty} \bar{G}_{ss}^{\,l} \tag{2.8}$$

with $\bar{G}_{ss} = \mathcal{Q}_c G_{ss} \mathcal{Q}_c$ and using the standard identity $\mathcal{P}_c\mathcal{Q}_c = 0$ for complementary projectors. The expansion in (2.8) converges rapidly since $\bar{G}_{ss}^{\,l} \sim$

$|\lambda_{c,2}|^l$ with $\lambda_{c,2}$ being the second largest eigenvalue which is significantly lower than unity.

The combination of (2.6) and (2.8) provides an explicit algorithm feasible for a numerical implementation for modest values of N_r, large values of N_s and of course if sparse matrices G, G_{ss} are considered. We refer the reader to [49] for more advanced implementation considerations.

2.4 Decomposition of $G_{\rm R}$

On the basis of equations (2.6)-(2.8), the reduced Google matrix can be presented as a sum of three components:

$$G_{\rm R} = G_{rr} + G_{\rm pr} + G_{\rm qr}, \tag{2.9}$$

with the first component G_{rr} given by direct matrix elements of G among the selected N_r nodes. The second projector component $G_{\rm pr}$ is given by:

$$G_{\rm pr} = G_{rs}\mathcal{P}_c G_{sr}/(1-\lambda_c),\ \mathcal{P}_c = \psi_R \psi_L^T\ . \tag{2.10}$$

The third component $G_{\rm qr}$ is of particular interest in this study as it characterizes the impact of indirect or hidden links. It is given by:

$$\begin{aligned} G_{\rm qr} &= G_{rs}[\mathcal{Q}_c \sum_{l=0}^{\infty} \bar{G}_{ss}^{\,l}]G_{sr},\ \mathcal{Q}_c = \mathbf{1} - \mathcal{P}_c, \\ \bar{G}_{ss} &= \mathcal{Q}_c G_{ss} \mathcal{Q}_c. \end{aligned} \tag{2.11}$$

We do characterize the strength of these 3 components by their respective weights W_{rr}, $W_{\rm pr}$, $W_{\rm qr}$ given respectively by the sum of all matrix elements of G_{rr}, $G_{\rm pr}$, $G_{\rm qr}$ divided by N_r. By definition we have $W_{rr}+W_{\rm pr}+W_{\rm qr} = 1$.

Reduced Google matrix has been computed, together with its components G_{rr}, $G_{\rm pr}$ and $G_{\rm qr}$, for the many language edition of Wikipedia and for different fields of study. Density plots for the different components of $G_{\rm R}$ and their weights are given in the chapter of each study. Predominant component is clearly $G_{\rm pr}$ but as we will explain next, it is not the most meaningful.

The meaning of G_{rr} is clear as it is directly extracted from the global Google matrix G. It gives the direct links between the selected nodes and more specifically the probability $G_{rr}(i,j)$ for the surfer to go directly from column j country to line i country. However, since each column is normalized by the number of outgoing links, absolute probabilities cannot be compared to each other across columns.

The sum of $G_{\rm pr}$ and $G_{\rm qr}$ represents the contribution of all indirect links through the scattering matrix G_{ss}. As shown in later results, the projector component $G_{\rm pr}$ is composed of nearly identical columns. Moreover, values of each column are proportional to the PageRank of the nodes if lines and columns are ordered by increasing original K values. As detailed in [49], it is observed numerically that $G_{\rm pr} \approx P_r\, E_r^T/(1-\lambda_c)$, meaning that each column is close to the normalized vector $P_r/(1-\lambda_c)$. As such, $G_{\rm pr}$ transposes essentially in $G_{\rm R}$ the contribution of the first eigenvector of G. We can conclude that even if the overall column sums of $G_{\rm pr}$ account for $\sim$ 95-97% of the total column sum of $G_{\rm R}$, $G_{\rm pr}$ doesn't offer innovative information compared to PageRank analysis.

A way more interesting contribution is the one of $G_{\rm qr}$. This matrix captures higher-order indirect links between the N_r nodes due to their interactions with the global network environment. We will refer to these links as *hidden links*. We note that $G_{\rm qr}$ is composed of two parts $G_{\rm qr} = G_{\rm qrd} + G_{\rm qrnd}$ where the first term gives only the diagonal part of the matrix $G_{\rm qrd}$ and thus represents the probabilities to stay on the same node during multiple iterations of $\bar{G}_{ss}$ in (2.11) while the second matrix captures only non-diagonal terms in $G_{\rm qrnd}$. As such, $G_{\rm qrnd}$ represents indirect (hidden) links between the N_r nodes appearing via the global network. We note that certain matrix elements of $G_{\rm qr}$ can be negative, which is possible due to the negative terms in $\mathcal{Q}_c = \mathbf{1} - \mathcal{P}_c$ appearing in (2.11). The total weight of negative elements is however much smaller than $W_{\rm qr}$. Of course, the full reduced Google matrix G_R has only positive or zero matrix elements. In the following, our study concentrates mainly on the meaning of $G_{\rm qrnd}$, emphasizing the meaning of its largest positive values.

Chapter 3
Hidden relationships between painters

3.1 Introduction

"The art is the expression or application of human creative skill and imagination, typically in a visual form such as painting or sculpture, producing works to be appreciated primarily for their beauty or emotional power" [50]. Artists use different approaches and techniques to create emotions. Since the beginning of mankind, painters have offered masterpieces in the form of paintings and drawings to the world.

Depending on historical periods, cultural context and available techniques, painters have followed different art movements. Art historians group painters into art movements to capture the fact that they have worked in the same school of thought. But a painter could be placed in several movements as his works evolve with time and its individual intellectual path development [51–57].

The major finding of this chapter, is to show that it is possible to automatically extract this common knowledge on art history by analyzing the hyperlinked network structure of the global and free online encyclopedia Wikipedia. The analysis conducted in this work is solely based on a graph representation of the Wikipedia articles.

As a study on the particular interactions between a very small subset of nodes compared to the full network size, in this chapter, we are interested in capturing the interactions of the 30 painters represented in Table 3.2 using the networks extracted from six Wikipedia language editions covering a few millions of articles each. These 30 painters are grouped by categories representing the historical movement they mainly belong to.

In this study, we extract from the reduced Google matrix G_{R} and its decomposition into direct and indirect matrices a high-level *reduced network of 30 painters*. This high-level network can be computed for either direct or hidden (i.e. indirect) interactions. More specifically, we deduce a fine-grained classification of painters that captures what we call the *hidden friends* of a given painter. The structure of these graphs provides relevant information that offers new information compared to the direct networks of relationships.

The aforementioned networks of direct and hidden interactions can be calculated for different Wikipedia language editions. In this study, reduced Google matrix analysis is applied to the same set of 30 painters on networks originating from six different Wikipedia language editions: English, French, German, Spanish, Russian and Italian.

Sections 3.2 describes how we get the names of painters and a calculated PageRank vector of them. Then, Section 3.3 shows our selection of painters. Section 3.4 and 3.5 presents a sample density plot of G_{R} matrices for EnWiki and the reduced Google matrices calculated for 30 painters and from six different language editions respectively. Specific emphasis is given to the very different English, French and German editions. Then, networks of friendship from direct and hidden interaction matrices are created and discussed. We show that the networks of friends completely capture the well-established history of painting by *i*) interconnecting densely painters of the same movement and *ii*) showing reasonable links between painters of different movements. Finally, Section 3.6 concludes this study.

3.2 Top Painters

In order to get the top painters in each Wikipedia edition of our selection, we have created a Matlab code to get the name of all the painters listed in the "List of painters by name" created by Wikipedia that includes painters from all ages and parts of the world [58]. 3334 painter names have been collected after checking their existence in our 6 selected editions. We have manually removed a few names as these persons were not necessarily known for their art painting production (e.g. Hitler). Then, for each Wikipedia edition, the Google matrix is constructed following the standard rules described in Chapter 1. From this Google matrix, PageRank K of all nodes present in all 6 Wikipedia editions is determined. From the 6 PageRank values, we extract the rank of our 3334 identified painters and reorder them by decreasing PageRank value. Table 3.1 shows the list of top 50 painters from the 6 selected Wikipedia editions. The complete list is given in Appendix A.

Not surprisingly, the order of top painters changes with respect to the edition due to cultural bias but main trends are there, e.g.:

- Leonardo da Vinci ranks first place in 5 out of 6 editions,
- Michelangelo and Picasso belong to the top 4 in all editions,
- Russian painters, like Viktor Vasnetsov and Ivan Aivazovsky, are in the top 20 of RuWiki but don't appear before rank 50 in other editions.

Using the PageRank of all 3334 painters computed for the 6 language editions, we have extracted 223 painters by creating the union set of the top

100 painters of each language edition. Figure 3.1 represents on a world map the number of painters born in each country. There is a clear predominance of European painters in this selection with a strong part of Russian artists as well. The full table of painters and other supplementary material is available in Appendix A.

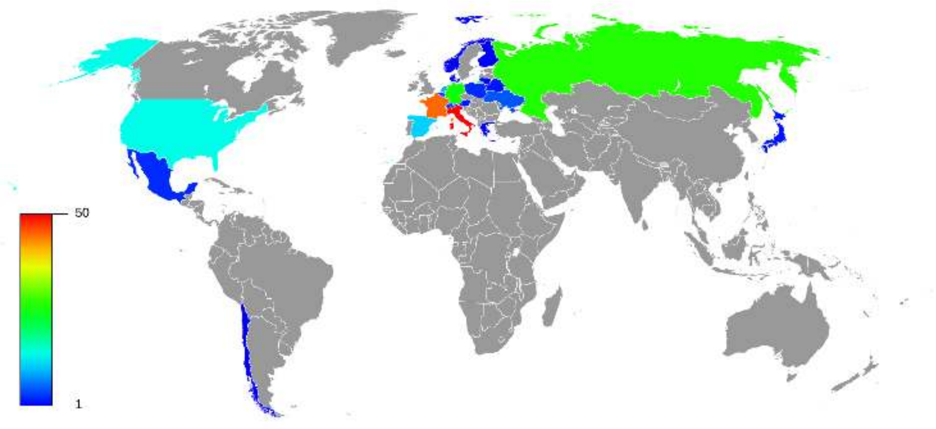

Figure 3.1: **Geographic birthplace distribution of the top 223 painters for the painters appearing at least one time in the top 100 of the 6 language editions analyzed.**

FrWiki	EnWiki	DeWiki	ItWiki	EsWiki	NlWiki	RuWiki
Pablo Picasso	Leonardo da Vinci	Leonardo da Vinci	Leonardo da Vinci	Leonardo da Vinci	Rembrandt Van Rijn	Leonardo da Vinci
Leonardo da Vinci	Pablo Picasso	Pablo Picasso	Michelangelo	Francisco Goya	Leonardo da Vinci	Pablo Picasso
Michelangelo	Michelangelo	Albrecht Durer	Raphael	Pablo Picasso	Peter Paul Rubens	Michelangelo
Claude Monet	Raphael	Michelangelo	Pablo Picasso	Michelangelo	Vincent Van Gogh	Rembrandt Van Rijn
Vincent Van Gogh	Rembrandt Van Rijn	Raphael	Giorgio Vasari	Raphael	Pablo Picasso	Vincent Van Gogh
Jacques-Louis David	Vincent Van Gogh	Rembrandt Van Rijn	Titian	Diego Velázquez	Johannes Vermeer	Raphael
Eugène Delacroix	Francis Bacon	Peter Paul Rubens	Peter Paul Rubens	Salvador Dali	Piet Mondrian	Albrecht Durer
Raphael	Andy Warhol	Vincent Van Gogh	Caravaggio	Peter Paul Rubens	Pieter Bruegel The Elder	Ilya Repin
Henri Matisse	Peter Paul Rubens	Titian	Vincent Van Gogh	Titian	Claude Monet	Peter Paul Rubens
Salvador Dali	Albrecht Durer	Francis Bacon	Giotto Di Bondone	Francis Bacon	Titian	Nicholas Roerich
Paul Cézanne	William Blake	Andy Warhol	Rembrandt Van Rijn	Albrecht Durer	Sandro Botticelli	Titian
Rembrandt Van Rijn	Titian	Paul Klee	Sandro Botticelli	El Greco	Paul Cézanne	Henri Matisse
Peter Paul Rubens	Claude Monet	Paul Cézanne	Albrecht Durer	Rembrandt Van Rijn	Albrecht Durer	Salvador Dali
Andy Warhol	Salvador Dali	Lucas Cranach the Elder	Francisco Goya	Vincent Van Gogh	Frans Hals	Paul Cézanne
Marcel Duchamp	Henri Matisse	Wassily Kandinsky	Giuseppe Arcimboldo	Sandro Botticelli	Giotto Di Bondone	Viktor Vasnetsov
Édouard Manet	Giorgio Vasari	Claude Monet	Piero Della Francesca	Caravaggio	Jan Van Eyck	Ivan Aivazovsky
Giorgio Vasari	Paul Cézanne	Henri Matisse	Edvard Munch	Henri Matisse	Andy Warhol	Diego Velázquez
Paul Gauguin	Francisco Goya	Salvador Dali	Andrea Mantegna	Eugène Delacroix	Anthony van Dyck	Marc Chagall
Albrecht Durer	Joseph Mallord William Turner	Giorgio Vasari	Masaccio	Paul Cézanne	Paolo Veronese	Claude Monet
Pierre Auguste Renoir	Eugène Delacroix	Edvard Munch	Claude Monet	Andy Warhol	Francisco Goya	Valentin Serov
Joan Miró	Caravaggio	Giotto Di Bondone	Jacques-Louis David	Claude Monet	Salvador Dali	Paul Gauguin
Jean-Auguste-Dominique Ingres	Jackson Pollock	Marc Chagall	Samuel Morse	Giorgio Vasari	Édouard Manet	Hieronymus Bosch
Georges Braque	Édouard Manet	Caspar David Friedrich	Wassily Kandinsky	Paul Gauguin	JAMES ENSOR	Henri de Toulouse-Lautrec
Edgar Degas	Anthony van Dyck	Édouard Manet	Diego Velázquez	Diego Rivera	Wassily Kandinsky	Karl Bryullov
Francisco Goya	Pierre Auguste Renoir	Otto Dix	Pieter Bruegel The Elder	Giotto Di Bondone	Paul Gauguin	Eugène Delacroix
Gustave Courbet	Jacques-Louis David	Caravaggio	Fra Angelico	Jacques-Louis David	Henri Matisse	Wassily Kandinsky
Fernand Léger	Diego Velázquez	Francisco Goya	Salvador Dali	Édouard Manet	William Blake	Édouard Manet
Titian	William Hogarth	Pierre Auguste Renoir	Pierre Auguste Renoir	Tintoretto	Rene Magritte	Francisco Goya
Caravaggio	Paul Gauguin	Paul Gauguin	Andy Warhol	Bartolomé Esteban Murillo	Jacob Jordaens	Kazimir Malevich
Jackson Pollock	Hans Holbein The Younger	Max Ernst	Anthony van Dyck	Anthony van Dyck	Gustav Klimt	Andrei Rublev
Wassily Kandinsky	Edgar Degas	Gustav Klimt	Giovanni Battista Tiepolo	Georges Braque	EugÃ¨ne Delacroix	Giorgio Vasari
Nicolas Poussin	Johannes Vermeer	Eugène Delacroix	Paul Cézanne	Edgar Degas	Karel Appel	Jacques-Louis David
Marc Chagall	Marcel Duchamp	Joan Miró	Giovanni Bellini	Joan Miró	Jacques-Louis David	Igor Grabar
Honoré Daumier	Sandro Botticelli	Jan Van Eyck	Domenico Ghirlandaio	Wassily Kandinsky	Giorgio Vasari	Pierre Auguste Renoir
Max Ernst	Giotto Di Bondone	Pieter Bruegel The Elder	Pietro Perugino	Hieronymus Bosch	Henry van de Velde	Samuel Morse
Diego Velázquez	Williem De Kooning	Max Liebermann	Jan Van Eyck	Piero Della Francesca	Henri de Toulouse-Lautrec	Caravaggio
Gustave Doré	Nicolas Poussin	Diego Velázquez	Paolo Veronese	Andrea Mantegna	Paul Klee	Edgar Degas
Sandro Botticelli	Pieter Bruegel The Elder	Sandro Botticelli	Giorgione	Jackson Pollock	Marc Chagall	Mikhail Vrubel
Giotto Di Bondone	John Constable	Marcel Duchamp	Nicolas Poussin	Henri de Toulouse-Lautrec	Joseph Mallord William Turner	Nicolas Poussin
Jean-Baptiste Camille Corot	Wassily Kandinsky	Gerhard Richter	Tintoretto	Johannes Vermeer	Edvard Munch	Anthony van Dyck
Henri de Toulouse-Lautrec	Marc Chagall	Max Beckmann	Paul Gauguin	Francisco De Zurbaran	Roger Van Der Weyden	Joseph Mallord William Turner
William Bouguereau	El Greco	Hans Holbein The Younger	Antonio da Correggio	William Blake	Georges Seurat	Jean-Auguste-Dominique Ingres
Pieter Bruegel The Elder	Lucas Cranach the Elder	El Greco	Edgar Degas	Marcel Duchamp	Nicolas Poussin	Alexandre Benois
Antoine Watteau	Benjamin West	Jacques-Louis David	Édouard Manet	Pierre Auguste Renoir	Joan Miró	Giotto Di Bondone
Georges Seurat	Gustave Doré	Georges Braque	Lucas Cranach the Elder	Hans Holbein The Younger	Gustave Doré	Konstantin Korovin
Rene Magritte	Henri de Toulouse-Lautrec	Johannes Vermeer	Eugène Delacroix	Pieter Bruegel The Elder	Edgar Degas	Isaac Levitan
André Derain	Georgia O'keefe	Henry van de Velde	Gustave Doré	Nicolas Poussin	Georges Braque	Gustave Courbet
Paul Klee	James Abbot Mac Neil Whistler	Edgar Degas	Marc Chagall	Jan Van Eyck	Hans Holbein The Younger	William Blake
François Boucher	Jan Van Eyck	Lovis Corinth	Guido Reni	William Bouguereau	Marcel Duchamp	Tove Jansson
Camille Pissarro	Thomas Gainsborough	Franz Marc	William Blake	Gustave Courbet		Ivan Kramskoi

Table 3.1: **List of 50 top painters from FrWiki, EnWiki, DeWiki, ItWiki, EsWiki, NlWiki and RuWiki by increasing PageRank.**

3.3 Painter set selection

From the set of painters introduced earlier that span the complete history of painting, a reduced set of 30 painters has been selected by choosing six painters for the following five important painting categories: Cubism, Impressionism, Fauvism, Great masters and Modern art (20th century). In order to select the top 5 painters of each category, a global importance score of painters is derived from this global cross-edition global ranking Θ_P defined as:

$$\Theta_P = \sum_E (101 - R_{P,E}). \tag{3.1}$$

Here $R_{P,E}$ is the ranking of top 100 painters P in Wikipedia edition E by PageRank algorithm. The painters with the largest $\Theta-$score are the most important ones for all investigated Wikipedia editions. Based on Θ_P score from English, French and German Wikipedia editions, we have selected the top 5 painters of each category which represents the order of appearance in Table 3.2. Table 3.2 also lists local PageRank index for painters in the English, French, and German Wikipedia editions. Painters that belong to the same movement or having a common piece of history may exhibit stronger interactions in Wikipedia. As such, we have created a color code to each movement (e.g. Fauvism, Cubism, Impressionist) or share a big part of history (e.g. Great Masters, Modern). Color code is as follows: Red, Blue, Green, Orange and Pink represents Cubism, Fauvism, Impressionism, Great masters and Modern (20-21st century) respectively (see Table 3.2).

Name	Category	Colour	FrWiki	EnWiki	DeWiki
Picasso	Cubism	Red	1	2	2
Braque	Cubism	Red	17	20	20
Léger	Cubism	Red	19	24	24
Mondrian	Cubism	Red	25	22	22
Gris	Cubism	Red	29	28	25
Delaunay	Cubism	Red	28	27	26
Matisse	Fauvism	Blue	6	11	12
Gauguin	Fauvism	Blue	13	15	18
Derain	Fauvism	Blue	22	25	27
Dufy	Fauvism	Blue	27	26	29
Rouault	Fauvism	Blue	30	30	28
Vlaminck	Fauvism	Blue	24	29	30
Monet	Impressionism	Green	4	9	11
C'ezanne	Impressionism	Green	8	12	9
Manet	Impressionism	Green	12	13	16
Renoir	Impressionism	Green	15	14	17
Degas	Impressionism	Green	18	16	21
Pissarro	Impressionism	Green	23	19	23
da Vinci	Great masters	Orange	2	1	1
Michelangelo	Great masters	Orange	3	3	4
Raphael	Great masters	Orange	5	4	5
Rembrandt	Great masters	Orange	9	5	6
Rubens	Great masters	Orange	10	7	7
Durer	Great masters	Orange	14	8	3
Dali	Modern 20-21	Pink	7	10	13
Warhol	Modern 20-21	Pink	11	6	8
Kandinsky	Modern 20-21	Pink	20	17	10
Chagall	Modern 20-21	Pink	21	18	15
Miró	Modern 20-21	Pink	16	21	19
Munch	Modern 20-21	Pink	26	23	14

Table 3.2: **List of names of 30 selected painters and their PageRank order for FrWiki, EnWiki and DeWiki, ordered by category.**

3.4 Density plots of G_{R}, G_{rr} and G_{qrnd}

To illustrate the matrices derived by the reduced Google matrix analysis detailed in Chapter 2, we plot G_{rr}, G_{R} and G_{qrnd} in Figure 3.2 for the EnWiki edition. Columns and lines are ordered with the list of painters given in Table 3.2. G_{R} is per-column normalized and dominated by the projector G_{pr} contribution, which is proportional to the global PageRank probabilities. As such, we clearly see that the density of each line of G_{R} is proportional to the importance of the painter in the full network. The matrices are interpreted in the following way: painter of column j is linked with the probability of element (i, j) to the painter of line i.

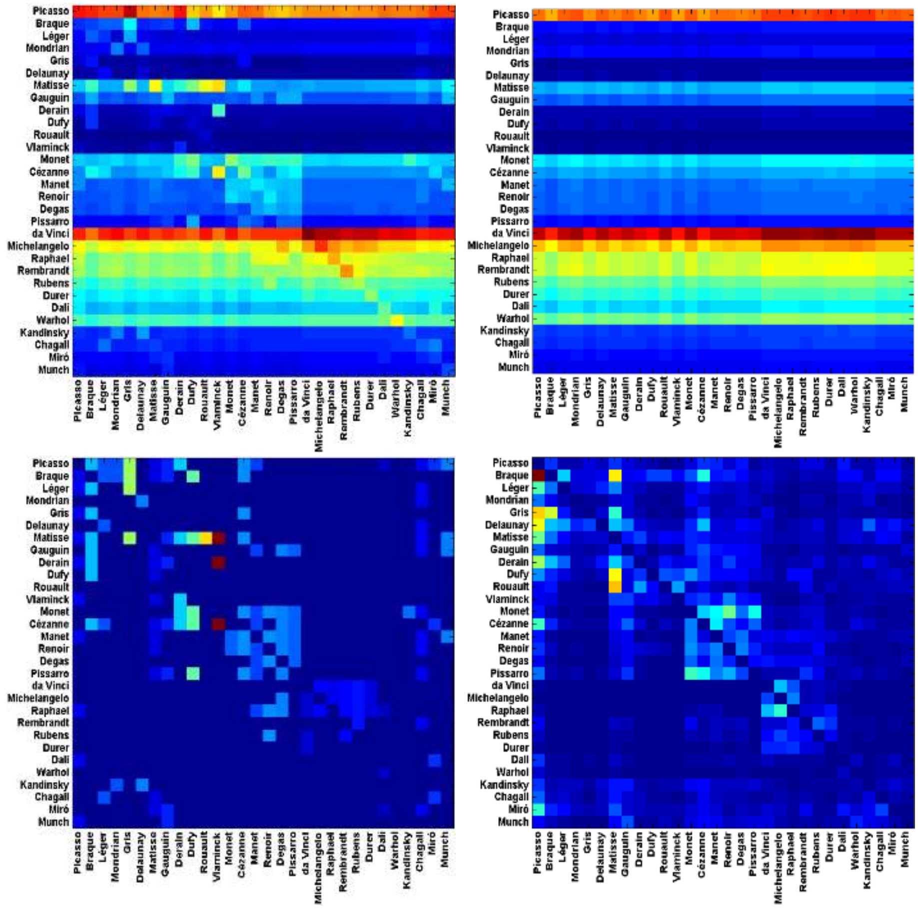

Figure 3.2: **Density plots of matrices for the reduced network of 30 painters in the EnWiki network.** Top left and top right figures are G_{R} and G_{pr}, respectively. Bottom left and bottom right figures are G_{rr} and G_{qrnd} respectively; Color scale represents maximum values in red, intermediate in green and minimum in blue.

G_{rr} provides information only on direct links between painters. In other words, it represents the probability for a random surfer to reach the painter of line i from the article of the painter of column j using a hyperlink linking article j to article i in Wikipedia. On the contrary, G_{qrnd} offers a much more unified view of painters interactions as it captures more general indirect (or hidden) interactions via the $N-30$ other nodes of the full Wikipedia network. In other words, it represents the probability linking the painter of column j to the painter of line i related to all indirect paths linking article j to article i in the full network. An indirect path starts with a hyperlink linking the article of painter j to an article k that doesn't belong to the N_r painter nodes and ends with a hyperlink ending on the article of painter i.

Reading Figure 3.2, we can extract strong and meaningful interactions between painters. New links appearing in G_{qrnd} and being absent from G_{rr} exist. As an example we list the links between Picasso and Braque, Pissaro and Monet, Rouault and Matisse. These relationships are very well known in art history, but looking at the pure structure of the network (i.e. reading G_{rr} matrix), they are absent. They appear clearly in the higher order mathematical analysis of the network using G_{qrnd}. For instance, it is common knowledge that since his visit to Picasso's studio, Braque became impressed by Picasso's paintings. They even became friends [59], which confirms our result. Pissaro and Monet are both Impressionists. Monet succeeded in reaching England after entrusting a number of his works to Pissaro [60]. Rouault and Matisse were both students of Gustave Moreau [61] and were deeply influenced by him throughout their life [62]. Their relationship began in 1906 and lasted all their life. All these interactions can be extracted from the network of Wikipedia webpages using G_{qrnd} matrix.

3.5 Building a friendship network

To deepen our study of the matrices G_{R}, G_{qrnd} and G_{rr} of the painter network extracted from the six Wikipedia editions of interest, we develop a graph representation relying on the following definition of friendship.

3.5.1 Friendship

In order to better capture the interactions provided by G_{qrnd}, one leading painter per group has been selected: Pablo Picasso for Cubism, Henri Matisse for Fauvism [63], Claude Monet for Impressionism [64, 65], Leonardo da Vinci for Great Masters and Dali for Modern. To pick them inside each group, we have chosen the painters whose average ranking score over all 6 selected Wikipedia editions is the highest.

For each leading painter, we extract from matrix G_{qrnd} their top-4 friends. Top friends of a node j are obtained by ordering in descending order the elements of column j of G_{qrnd}. Thus, the top friend of node j is a node for which the probability in G_{qrnd} is the strongest. This means that from node j the probability of reaching top friend i is the largest using indirection interactions. Table 3.3 shows a summary of cross-editions friends from G_{qrnd} for the top painter of each category.

G_{qrnd} seems to emphasize more fine-grained regional interactions and by looking at the interactions (and in addition to relationships explained in Section 3.4), we can see the strong relationship between da Vinci, Michelangelo and Raphael which can be explained by the fact that they were the nucleus of fifteenth-century Florentine art [66]. Another strong relation could be snapped between Mirò and Dali, as both are inspired by Picasso [67].

Top Painter	all 6 editions	5 out of 6 editions	4 out of 6 editions
Picasso	Braque - Gris		
Matisse	Rouault		Braque - Dufy
Monet	Renoir	Pissarro	
da Vinci	Michelangelo - Raphael	Durer	Degas
Dali	Miró		

Table 3.3: **Cross-editions friends from G_{qrnd} for the top-4 painters of each category.** For each top painter, we list the friends present in the top-4 friends list given by all six Wikipedia editions, the ones present in 5 editions out of 6 and the ones present in 4 editions out of 6.

3.5.2 Networks of 30 painters

From G_{qrnd} we extract the top 4 friends of leading painters to plot the graphs of Figure 3.3. Note that Figure 3.3 essentially highlights hidden links as it is extracted from G_{qrnd}. The black thick arrows identify the top 4 friends interactions. Red arrows represent the friends of friends interactions that are computed recursively until no new edge is added to the graph. The graphs are plotted using the Yifan Hu layout of Gephi [1] that groups together nodes more densely interconnected.

Impressionism, Fauvism, Cubism and Great masters create, in all editions, a cluster of nodes densely interconnected. The group of Modern painters plays a role by connecting the other categories: *1)* Dali seems to be the common interconnection node between Fauvism and Cubism categories in EnWiki. *2)* Kandinsky connects Fauvism and Cubism in FrWiki. *3)* Munch connects Impressionism and Fauvism in DeWiki. The networks of G_{qrnd} end up almost spanning the full set of 30 painters. These links show that the interactions between the painters groups are coherent. These graphs picture the essence of painting history by grouping together painters that belong to the same movement and by interconnecting them in a reasonable and close-to historic reality way.

For instance, our graphs are consistent with the history of modern art which starts with the Impressionism movement (1870-1890) that searched for the exact analysis of the effects of color and light in nature. The painters we have selected are among the most important ones of the movement and they create a clear cluster of nodes in Figure 3.3 (see green nodes) as they exhibit a tight relationship (friends) in G_{qrnd}. The Fauvism movement emerged after impressionist (1899-1908) [68–70]. Fauvist painters were concerned with the impression created with colors. This movement was inspired by different artists such as Matisse. The *Fauves* members were a loosely shaped group of artists with shared interests. Henri Matisse became later the leader of the group of artists [63]. He introduced unnatural and intense color into their paintings to describe light and space. The fauvism movement is the precursor of the Cubism movement [71]. Our result shows deep relationships between Fauvism and Cubism, noting that Braque is always the core of this interconnection. Cubism movement (1907- 1922) is pretty distinct from Impressionism, which is underlined as well in our graphs with only a few red links connecting these two clusters of nodes.

3.6 Conclusion

This work offers a new perspective for future art studies. It is possible to extract from multi-cultural Wikipedia networks a global understanding of the interactions between the painters. We have applied the Google matrix (Table 3.1) and the reduced Google matrix analysis (Figure 3.2) to the network of articles of 6 Wikipedia editions to get the top painters in each edition and also to analyze the network structure of 30 painters and the interconnection between painting categories.

This approach takes into account all human knowledge accumulated in Wikipedia, leveraging all indirect interactions existing between the 30 selected articles and the huge information contained by more than 10 millions articles of Wikipedia. The network structure obtained for the painters (Figure 3.3) clearly shows the presence of 5 categories of painters. The main painters in each category are determined from their PageRank. We show that the indirect or hidden links between painters play an important role and are, in many cases, predominant over direct links.

The obtained results, tested on the publicly available data of Wikipedia, are in good agreement with art painting history. The next chapter illustrates the use of reduced Google matrix analysis on a different domain, namely geopolitics. We will concentrate in illustrating then how the network of friends obtained with G_{qrnd} and G_{R} differ. Moreover, we will as well show how ordering row i of one of our matrices of interest (G_{qrnd} and G_{R}) by descending order underlines another type of interaction.

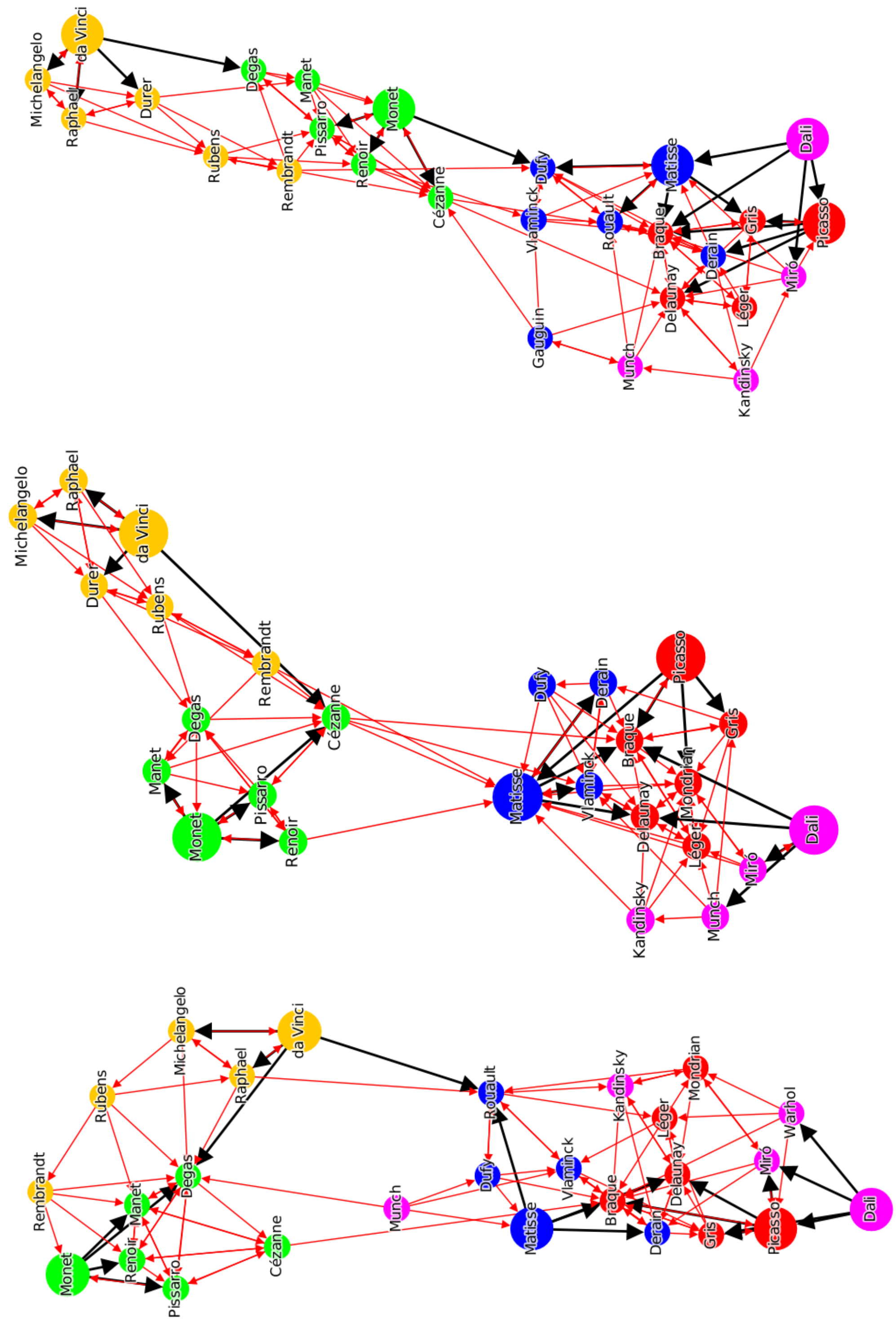

Figure 3.3: **Network structure of friends induced by the top 5 painters of each group in G_{qrnd}.** Results are plotted for EnWiki (Bottom), FrWiki (Middle) and DeWiki (Top). Red, Blue, Green, Orange and Pink nodes represent Cubism, Fauvism, Impressionism, Great masters and Modern (20-21), respectively. The top painter node points with a bold black arrow to its top-4 friends. Red arrows represent the friends of friends interactions computed until no new edges are added to the graph. All graphs are automatically plotted using *Gephi* [1].

CHAPTER 4

Multi-cultural mining of geopolitics interactions

4.1 Introduction

Political and economic interactions between regions of the world have always been of utmost interest to measure and predict their relative influence. Such studies belong to the field of geopolitics that focuses on political power in relation to geographic space. Interactions among world countries have been widely studied at various scales (worldwide, continental or regional) using different types of information. Studies are driven by observing economic exchanges, social changes, history, international politics and diplomacy among others [72, 73].

Complementary to the analysis of chapter 3, this chapter aims to show how the analysis of hidden relationships offers novel and meaningful information for the study of geopolitics. Therefore, we focus on a subset of nodes of Wikipedia representing 40 major countries (that are the top in PageRank) in the world to extract geopolitical ties in between them. All type of information gathered in this collaborative knowledge base can be leveraged to provide a picture of countries relationships, fostering a new framework for thorough geopolitics studies.

Analyzing the interactions within the 40 countries represented in Figure 4.1 is our goal. We will use for that the networks extracted from five Wikipedia language editions covering a few millions of articles each: English (EnWiki), Arabic (ArWiki), Russian (RuWiki), French (FrWiki) and German (DeWiki) editions. Countries were selected as the top 40 countries with respect to the the PageRank of EnWiki.

In this study, we calculate G_{R} and its decomposition into direct and indirect matrices for the selected subset network of $N_r = 40$ countries. In this study, we show how the networks extracted from hidden relationships (e.g. using G_{qrnd}) offer finer grained information on the interaction between the 40 countries compared to the network extracted from G_{R}. Moreover, we introduce another type of interaction analysis we have named the *follower* interaction. Thus, we draw the networks of friends interaction and followers interaction to study the relationship between the 40 countries in different editions of Wikipedia, for both G_{qrnd} and G_{R}.

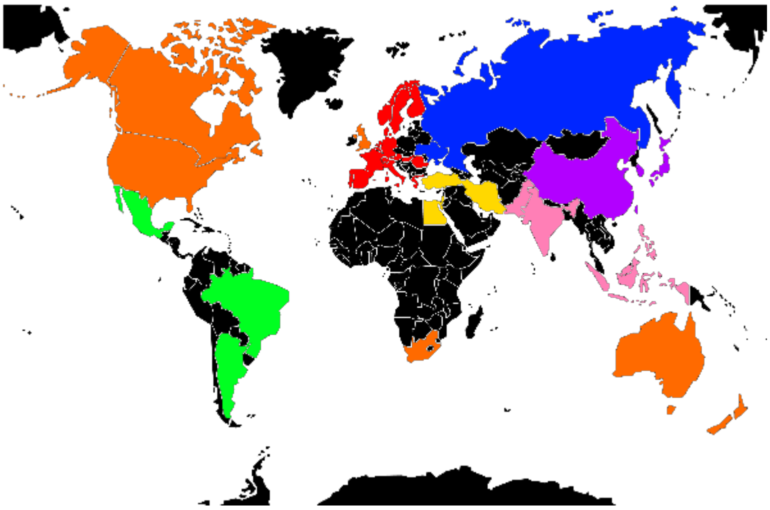

Figure 4.1: **Geographical distribution of the 40 selected countries.** Color code groups countries into 7 sets: orange (OC) for English speaking countries, blue (BC) for former Soviet union ones, red (RC) for European ones, green (GC) for Latin American ones, yellow (YC) for Middle Eastern ones, purple (PUC) for North-East Asian ones and finally pink (PIC) for South-Eastern countries (see colors and country names in Table 4.1 ; other countries are shown in black).

As we will show, the structure of these graphs provides relevant social information: communities of countries with strong ties can be clearly exhibited while countries acting as bridges are present as well. This is mainly the case for the hidden interactions networks of friends (or followers) that offer new information compared to the networks of friends (or followers) extracted from G_{R} whose topology is mainly enforced by top PageRank countries.

Wikipedia language editions are usually modified by authors who mainly belong to the region associated with this language. Thus our study shows the impact of this cultural bias when comparing reduced and hidden networks of friends (or followers) among different language editions. We show that part of the interactions are cross-cultural while others are clearly biased by the culture of the authors.

In Sec. 4.2, we describe G_{R} calculated for 40 countries and for five different Wikipedia editions. Specific emphasis is given to the very different English, Arabic and Russian editions. Networks of friends and followers for direct and hidden interaction matrices are created and discussed in Section 4.3, and conclusions are drawn in Section 4.4.

4.2 Matrices of world countries

Our study focuses on the networks representing 5 different Wikipedia editions[1] from the set of 24 analyzed in [23]: EnWiki, ArWiki, RuWiki, DeWiki and FrWiki that contain 4.212, 0.203 , 0.966, 1.533 and 1.353 millions of articles each.

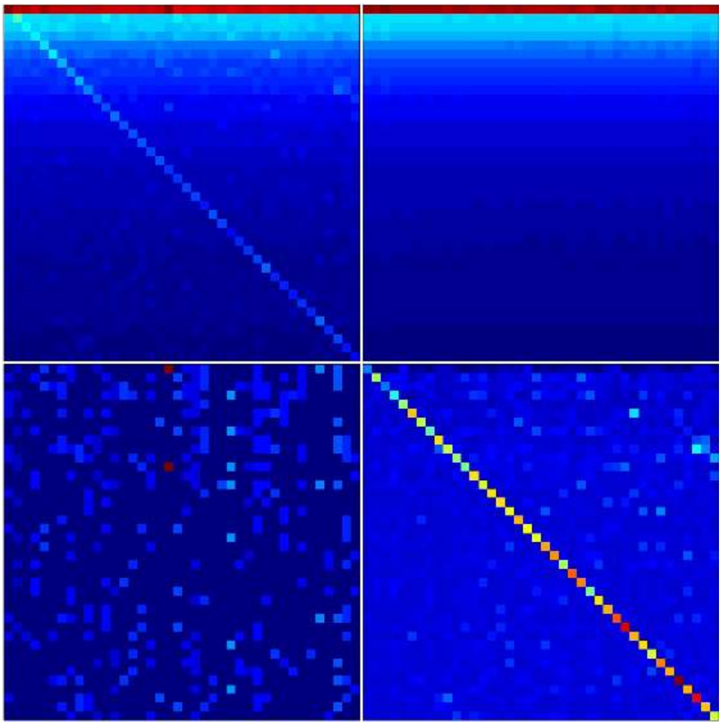

Figure 4.2: **Density plots of matrices for the reduced network of 40 countries in the EnWiki network.** G_{R} (top left), G_{pr} (top right), G_{rr} (bottom left) and G_{qr} (bottom right). The nodes N_r are ordered in lines by increasing PageRank index (left to right) and in columns by increasing PageRank index from top to bottom. Color scale represents maximum values in red (0.15 in top panels; 0.01 in bottom left panel; 0.03 in bottom right panel), intermediate in green and minimum (approximately zero) in blue.

Reduced Google matrix has been computed, together with its components G_{rr}, G_{pr} and G_{qr}, for the English language edition of Wikipedia (EnWiki) and for the $N_R = 40$ countries listed in Table 4.1. These countries are the ones with top PageRank K in the network of EnWiki. Density plots of G_{R}, G_{rr}, G_{pr} and G_{qr}, are given in Figure 4.2 where lines and columns are ordered by increasing K values. Countries are ordered by increasing K

[1] Data collected mid February 2013.

value. The weight of the three matrix components of G_{R} are $W_{\mathrm{pr}} = 0.96120$, $W_{\mathrm{qr}} = 0.029702$ and $W_{rr} = 0.009098$. Predominant component is clearly G_{pr} but as we detailed in Chapter 2, it is not the most meaningful. G_{rr} is directly extracted from the global Google matrix G. G_{qrnd} is the non-diagonal term of G_{qr}, that captures higher-order indirect links between the selected countries due to their interactions with the global network environment.

The meaning of G_{rr} is clear as it is directly extracted from the global Google matrix G. It gives the direct links between the selected nodes and more specifically the probability $G_{rr}(i,j)$ for the surfer to go directly from column j country to line i country. However, since each column is normalized by the number of outgoing links, absolute probabilities cannot be compared to each other across columns.

The sum of G_{pr} and G_{qr} represents the contribution of all indirect links through the scattering matrix G_{ss}. As seen on Figure. 4.2, the projector component G_{pr} is composed of nearly identical columns. Moreover, values of each column are proportional to the PageRank of the countries (lines and columns are ordered by increasing K values). G_{pr} transposes essentially in G_{R} the contribution of the first eigenvector of G. We can conclude that even if the overall column sums of G_{pr} account for $\sim$ 95-97% of the total column sum of G_{R}, G_{pr} doesn't offer innovative information compared to PageRank analysis.

A way more interesting contribution is the one of G_{qr}. This matrix captures higher-order indirect links between the N_r nodes due to their interactions with the global network environment. We will refer to these links as *hidden links*. We note that G_{qr} is composed of two parts $G_{\mathrm{qr}} = G_{\mathrm{qrd}} + G_{\mathrm{qrnd}}$ where the first term gives only the diagonal part of the matrix G_{qrd} and thus represents the probabilities to stay on the same node during multiple iterations of $\bar{G}_{ss}$ in (2.11) while the second matrix captures only non-diagonal terms in G_{qrnd}. As such, G_{qrnd} represents indirect (hidden) links between the N_r nodes appearing via the global network. In the following, our study concentrates mainly on the meaning of G_{qrnd}.

Wikipedia edition			English		Arabic		Russian	
Countries	CC	Color	K	K^*	K	K^*	K	K^*
United States	US	OR	1	9	1	5	2	27
France	FR	RD	2	19	3	31	3	14
United Kingdom	GB	OR	3	25	6	20	7	3
Germany	DE	RD	4	33	8	14	4	24
Canada	CA	OR	5	26	13	19	12	26
India	IN	PK	6	23	9	25	13	8
Australia	AU	OR	7	35	16	22	18	12
Italy	IT	RD	8	15	5	1	6	32
Japan	JP	VT	9	4	11	9	11	7
China	CN	VT	10	8	12	17	9	21
Russia	RU	BL	11	6	7	2	1	2
Spain	ES	RD	12	30	4	8	8	15
Poland	PL	RD	13	12	26	32	10	17
Netherlands	NL	RD	14	37	18	33	15	31
Iran	IR	YL	15	2	14	15	30	22
Brazil	BR	GN	16	3	21	26	20	1
Sweden	SE	RD	17	22	22	7	19	5
New Zealand	NZ	OR	18	28	34	24	36	4
Mexico	MX	GN	19	40	23	38	22	37
Switzerland	CH	RD	20	38	20	34	16	18
Norway	NO	RD	21	32	35	16	27	11
Romania	RO	RD	22	10	19	6	32	36
Turkey	TR	YL	23	7	15	13	21	38
South Africa	ZA	OR	24	24	29	39	35	20
Belgium	BE	RD	25	18	27	37	29	30
Austria	AT	RD	26	39	28	28	14	28
Greece	GR	RD	27	21	10	36	25	25
Argentina	AR	GN	28	1	32	29	33	23
Philippines	PH	PK	29	17	36	21	39	33
Portugal	PT	RD	30	36	24	12	17	9
Pakistan	PK	PK	31	5	25	35	37	29
Denmark	DK	RD	32	16	33	10	31	19
Israel	IL	YL	33	20	17	18	28	6
Finland	FI	RD	34	14	37	4	26	16
Egypt	EG	YL	35	31	2	3	24	39
Indonesia	ID	PK	36	13	31	11	34	10
Hungary	HU	RD	37	11	40	40	23	40
Taiwan	TW	VT	38	27	39	27	40	34
South Korea	KR	VT	39	34	38	30	38	35
Ukraine	UA	BL	40	29	30	23	5	13

Table 4.1: **List of 40 selected countries.** PageRank K and CheiRank K^* for EnWiki, FrWiki and RuWiki. Color code (CC) groups countries into 7 subsets: orange (OR) for English speaking countries, Blue (BL) for former Soviet union ones, Red (RD) for European ones, Green (GN) for South American ones, Yellow (YL) for Middle Eastern ones, Purple (VT) for North-East Asian ones and finally Pink (PK) for South-Eastern countries.

4.2.1 Selected countries

The 40 countries listed in Table 4.1 have been selected from the EnWiki network after computing the PageRank for the complete network. The 40 countries with largest PageRank probability have been chosen, and ordered by a local PageRank index K varying between 1 and 40. The most influential countries are found to be at the top values $K = 1, 2, ...$. In addition we determine the local CheiRank index K^* of the selected countries using the CheiRank vector of the global network [35, 37]. At the top of K^* we have the most communicative countries. Table 4.1 lists K and K^* for EnWiki, ArWiki and RuWiki. Not surprisingly, the order of top countries changes with respect to the edition (for instance, the top country for K is US except for RuWiki whose top country is Russia).

Countries that belong to the same region or having a common piece of history may probably exhibit stronger interactions in Wikipedia. As such, we have created a color code that groups together countries that either belong to the same geographical region (e.g. Europe, Latin America, Middle East, North-East Asia, South-East Asia) or share a big part of history (former USSR; English speaking countries that are the legacy of the former British Empire). Color code can be seen the text color colums of Table 4.1.

It is convenient as well to plot all nodes in the $(K,\ K^*)$ plane to highlight the countries that are the most influential ($K = 1, 2,\ ...$) and the most communicative ($K^* = 1, 2,\ ...$) at the same time. Figure 4.3 plots all 40 countries in the $(K,\ K^*)$ plane for EnWiki, ArWiki and RuWiki editions. This plot is a bi-objective plot where K and K^* are to be minimized concurrently. It is interesting to look at the set of non-dominated countries which are the ones such that there is no other country beating them for both K and K^*. Cultural bias is obvious here as for EnWiki, this set is composed of {US, JP, IR, AR}, for ArWiki of {US, EG, IT} and for RuWiki of only Russia. We note that according to Figure 4.3 some countries with high K value (relative few in-degree and low PageRank probability) act as important diffusers of content (low K^*) even for the language editions being different from the language spoken in those countries (BR, AR, RO). Also countries with low K (having many citations) are relatively poor diffusers (FR, CA). Thus Figure 4.3 demonstrates that different cultures attribute different degree of country popularity (PageRank probability and K index) and different communicative degree (CheiRank probability and K^* index). This indicates the nontrivial features of cultures propagation and interactions.

4.2.2 Density plots of G_{R}, G_{rr} and G_{qrnd}

For the three EnWiki, ArWiki and RuWiki editions, Figure 4.4 plots the density of matrices G_{R}, G_{qrnd} and G_{rr}. We keep for all plots the same order of countries extracted from the EnWiki network. This is meant to highlight

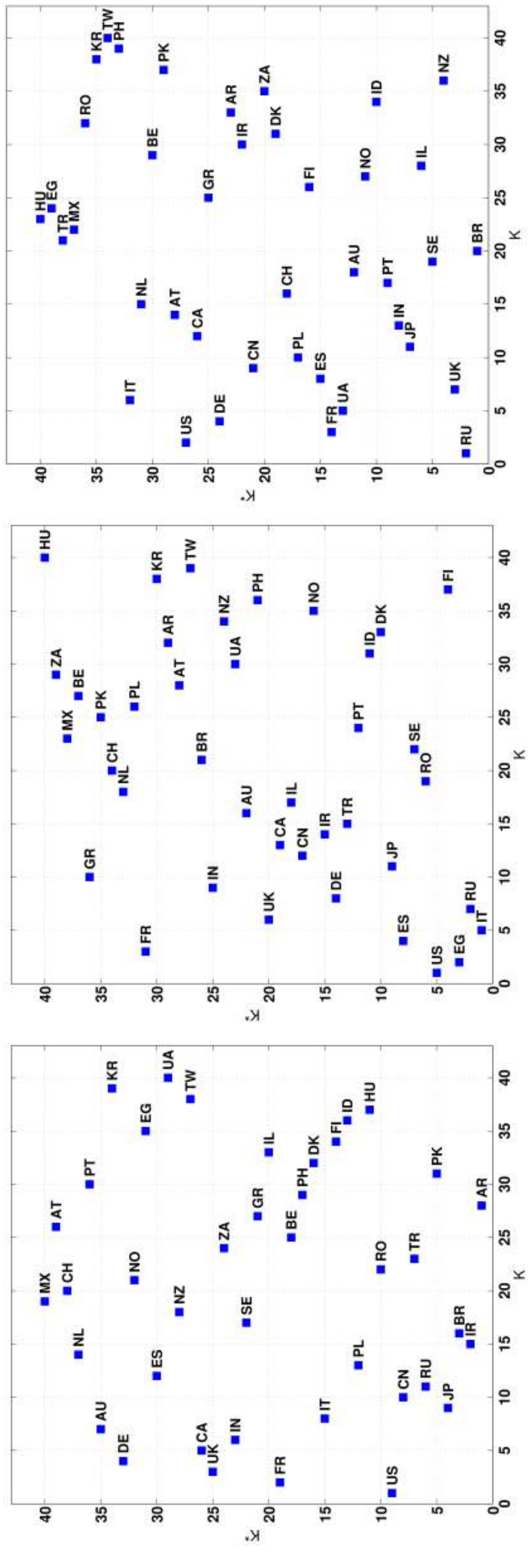

Figure 4.3: **Position of countries in the local (K, K^*) plane of the reduced network of 40 countries.** EnWiki (Bottom), ArWiki (Middle) and RuWiki (Top) networks.

cultural differences among Wikipedia editions. In the G_{R} plots, it is clear that this matrix is dominated by the projector G_{pr} contribution, which is proportional to the global PageRank probabilities. In G_{R}, not surprisingly, the cultural bias is pronounced due to its strong tie to PageRank. For instance, Egypt is much more important in ArWiki than in other editions, and Russia is the top country in RuWiki.

The information from direct links between countries is provided by G_{rr}. As expected, the per-column normalization prevents a meaningful per-line analysis. For EnWiki and ArWiki, the respective columns of Mexico and Hungary are predominant due to their little number of outgoing links. On the contrary, G_{qrnd} offers a much more unified view of countries interactions as it seems to highlight more general interaction that are less biased by cultural views.

For instance, for these three Wikipedia editions, the hidden links connecting Taiwan to China and Pakistan to India are really strong in G_{qrnd}. The link connecting Ukraine to Russia is very strong in EnWiki and ArWiki. It is surprisingly absent from RuWiki (or maybe this is exactly a cultural bias we are observing since during a long time both countries were part of USSR, and there was thus no specific difference between them). Other interesting hidden links are highlighted in EnWiki as New-Zealand is directed to Australia or in RuWiki linking Canada to the USA.

4.2.3 Friends and followers

In order to better capture the interactions provided by G_{qrnd} and G_{R}, we have listed for all 5 Wikipedia editions the top 4 friends and top 4 followers of a set of 7 leading countries. One leading country per group has been selected: US for English speaking countries, Brazil for Latin America, France for Europe, Japan for North-East Asia, India for South-East Asia, Russia for the Soviet block and Turkey for the Middle-East. To pick them inside each group, we have chosen the country whose worst PageRank order over all 5 Wikipedia editions is the highest.

For each leading country, we extract from both matrices G_{qrnd} and G_{R} the top 4 *Friends* (resp. *Followers*) of country j given by the 4 best values of the elements of column j (resp. of line j). In other words, top 4 friends correspond to destinations of the 4 strongest outgoing links of j and the top 4 follower countries are at the origin of strongest 4 ingoing links of j. A summary of relevant results is given in Table 4.2 to show cross-culture interactions.

Looking at G_{R} friends, top friends of leading countries are strongly related to the top PageRank countries (we have a predominance of US, France and Germany). Similarly, cross-edition followers of US are Mexico and Canada, and followers of Japan are China, Korea and Taiwan. On the opposite, higher-order interactions of G_{qrnd} are not as much influenced by

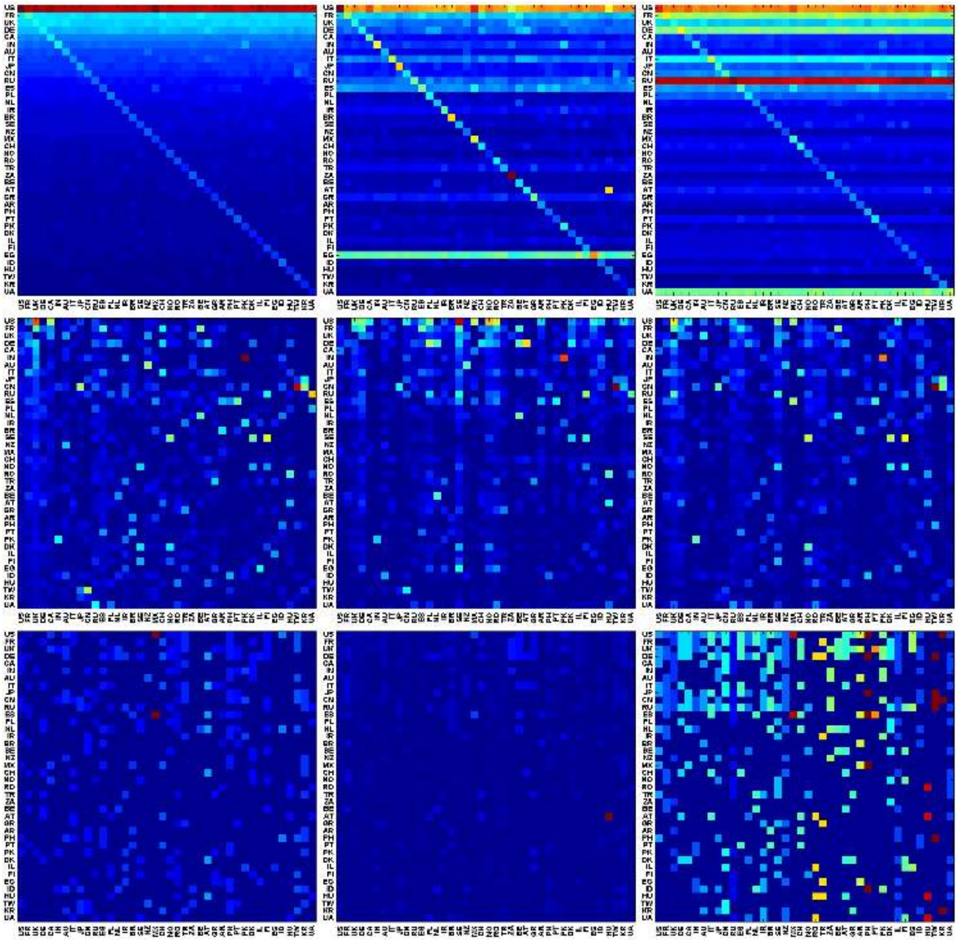

Figure 4.4: **Density plots of the matrix elements for the reduced network of 40 countries of EnWiki, ArWiki and RuWiki.** EnWiki (left column), ArWiki (middle column) and RuWiki (right column). first line: G_{R}, second line: G_{qrnd} and third line: G_{rr} . The country names are given on the axes in the order names in Table 1, thus the nodes N_r are ordered in lines and columns by the reference PageRank of EnWiki. The colors represent maximum (red corresponds to: 0.15, 0.19, 0.13 in top panels from left to right; 0.01, 0.03, 0.012 in middle panels and 0.01, 0.011, 0.006 in bottom panels respectively), intermediate (green) and minimum (blue for zero) values for a give matrix.

Top country	G_{R} Wiki friends present in			G_{R} Wiki friends followers present in		
	all 5 editions	4 out of 5 editions	3 out of 5 editions	all 5 editions	4 out of 5 editions	3 out of 5 editions
US	FR	DE	UK	MX - CA	IL	UK
BR	US - FR	DE		AR - PT		MX
FR	US	DE		BE	ES	IT
RU	US - FR	DE		UA - FI		PL
TR	US - FR	DE		GR - IR		
JP	US - FR	DE		KR - TW - CN	PH	
IN	US - FR		DE	PK - ID		ZA

Top country	G_{qr} Wiki friends present in			G_{qr} Wiki followers present in		
	all 5 editions	4 out of 5 editions	3 out of 5 editions	all 5 editions	4 out of 5 editions	3 out of 5 editions
US	CA	UK - MX	FR		UK	CA - MX
BR	PT - AR	ES - MX		AR	PT	UK
FR	IT - DE		US - BE		UK - BE - ES	CH
RU	UA	DE	CN - PL - US	FI	UA - PL	IR
TR	GR	IR - RU		GR - IR	EG - RO	
JP		CN - TW - US - KR		KR - CN - TW		
IN	PK - CN	ID	IR	PK - ID		CN

Table 4.2: **Cross-edition direct friends and followers extracted from G_{R} and G_{qrnd} matrices for the top countries of each area.** For each top country, we list the direct friends (followers) present in the direct friends list given by all five Wikipedia editions, the ones present in 4 editions out of 5 and the ones present in 3 editions out of five.

PageRank. More subtle but realistic interactions appear: Canada is always identified as a hidden friend of USA while it was never the case in G_{R}. Similarly, Ukraine is always tagged as a hidden friend of Russia; Italy and Spain as friends of France. Thus G_{qrnd} seems to emphasize more fine-grained regional interactions. Next section exploits the concept of friends and followers to create new network representations derived from G_{R} and G_{qrnd}.

4.3 Networks of 40 countries

This study concentrates again on the same 7 leading countries as before. Top 4 friends and top 4 followers of these leading countries are extracted from G_{R} and G_{qrnd} to plot the graphs of Figures 4.5 and 4.6, respectively. Note that Figure 4.6 essentially highlights hidden links. The black thick arrows identify the top 4 friends and top 4 followers interactions. Red arrows represent the friends of friends (respectively the followers of followers) interactions that are computed recursively until no new edge is added to the graph. All graphs are plotted using a force direct layout.

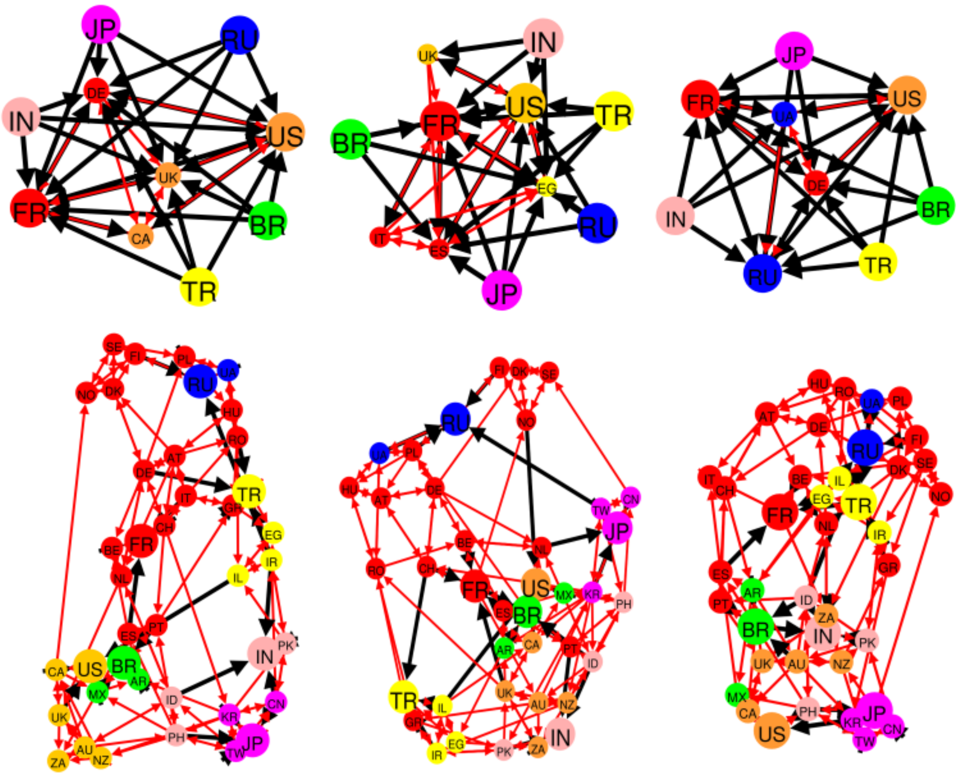

Figure 4.5: **Network structure of friends and followers induced by the 7 top countries of each geographical area (US, FR, IN, JP, BR, TR, RU) in G_{R}.** Friends and followers network are in top and bottom line respectively. Results are plotted for EnWiki (left column), ArWiki (middle column) and RuWiki (right column). Node colors represent geographic appartenance to a group of countries (cf. Table 4.1 for details). Top (bottom) graphs: the top country node points (is pointed by) with a bold black arrow to its top 4 friends (followers). Red arrows show friends of friends (resp. followers of followers) interactions computed until no new edges are added to the graph. Drawn with [1].

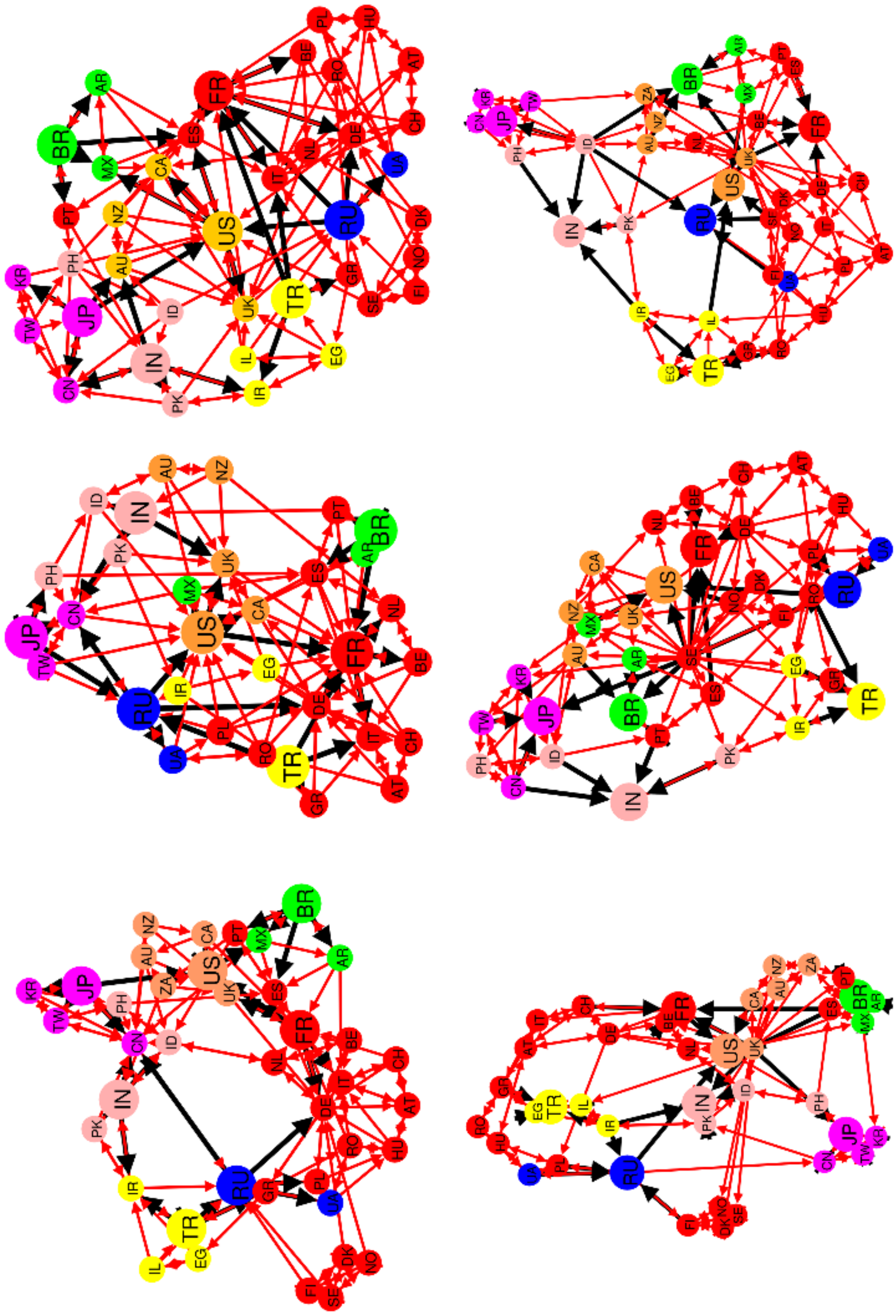

Figure 4.6: **Network structure of friends and followers induced by the 7 top countries of each geographical area (US, FR, IN, JP, BR, TR, RU) in** G_{qrnd}.Friends and followers network are in left and right line respectively. Results are plotted for EnWiki (third row), ArWiki (middle row) and RuWiki (first row). Node colors represent geographic appartenance to a group of countries (cf. Table 4.1 for details). Top (bottom) graphs: the top country node points (is pointed by) with a bold black arrow to its top 4 friends (followers). Red arrows show friends of friends (resp. followers of followers) interactions computed until no new edges are added to the graph. Drawn with [1].

The networks of friends obtained from G_{R} never expand to the full set of 40 nodes. They only concentrate on about 10 countries (including the 7 leading ones). This happens since G_{R} is dominated by the projector component. Looking at the follower graphs, more information can be observed. North-East Asian, Middle-Eastern and Latin American create, in all editions, a cluster of nodes densely interconnected. European countries enclose Russia and Ukraine as these countries are linked to EU countries that were part of the former Soviet Union zone of influence (e.g. Romania, Hungary, Finland, etc.). The networks of followers end up almost spanning the full set of 40 countries.

The networks of friends obtained from G_{qrnd} don't concentrate to a limited set of countries as it is the case for G_{R}. They end up spanning the full set of countries. The hidden friend links show that the interactions between the geographical groups are coherent. North-East Asian countries are linked to South-East Asian countries and to English speaking countries in EnWiki and RuWiki. Interestingly, the set of Baltic countries (SE, NO, DK, FI) create most of the time full meshes, and interconnect Europe and Russia. Cultural bias can be observed as well in these plots. For non-Arabic editions, Middle-Eastern countries create a well-connected cluster of nodes. But for ArWiki, Turkey exhibits a stronger connection with Europe than with the other Middle-Eastern countries. In the view of Arabic countries, Turkey is seen closer to Europe than others for sure.

4.4 Conclusion

This study also contributes to show that Wikipedia is a convenient target for network analysis thanks to its hyperlinked structure. On the basis of the reduced Google matrix analysis of the Wikipedia network, we have determined the relations of friends and followers of countries on purely statistical mathematical grounds obtained from the huge knowledge accumulated by Wikipedia editions in different languages. This approach provides results independent of cultural bias. The reduced Google matrix theory has been shown to capture hidden and indirect interactions among countries, resulting in new knowledge on geopolitics.

In this chapter and the previous one, we illustrated the significance of the reduced Google matrix and mainly its hidden links. The next step will be to study the reduced Google matrix G_{R} link structure, in order to see how nodes are sensitive to a link variation.

Chapter 5

Sensitivity analysis of networks

5.1 Introduction

Results obtained on geopolitical interactions in Chapter 4 lead us to go for a deeper analysis and study the weight of links between countries. In this chapter we show that meaningful results on the influence of country ties can be extracted from the hyperlinked structure of Wikipedia. This study analyses the networks extracted from 5 language editions of Wikipedia to study the influence of countries on each other. We proceed with this analysis for two sets of countries: *i*) the 27 member states of the European Union and *ii*) the top 40 countries according to English Wikipedia PageRank (cf. seen in Chapter 4).

This work goes one step further as it quantifies the influence of a relationship between two countries on the rest of the reduced network using G_{R}. Previous Chapter 4, has identified the strongest ties, but this one focuses on capturing the impact of a change in the strength of a relationship between two countries on the overall network interactions of selected countries via the global network. The impact of this change of tie strength on the overall network structure is measured by calculating the variation of importance of the nodes in the network.

We show that this sensitivity analysis renders a reasonable and meaningful idea of the influence of a given bilateral tie on the whole network. We calculate G_{R} for the group of 27 EU (G_{R} for 40 world countries has already been derived in Chapter 4). Thus, G_{R} reflects in a 40-by-40 or 27-by-27 matrix the complete (direct and indirect) relationships between countries. To identify the relative influence of one relationship between two nations, we propose to compute a logarithmic derivative of the PageRank probabilities calculated from G_{R} and $\tilde{G}_{\mathrm{R}}$, where $\tilde{G}_{\mathrm{R}}$ is the reduced Google matrix obtained after one link has been altered. $\tilde{G}_{\mathrm{R}}$ is almost equal to G_{R}. It only differs by the values of one column. If the relationship going from nation j to nation i is of interest in the study, only the values of column j are changed to relatively inflate the probability $\tilde{G}_{\mathrm{R}}(i, j)$ of nation j ending in nation i compared to the other ones. This is done in practice by increasing $\tilde{G}_{\mathrm{R}}(i, j)$ and then normalizing the column again to unity as it is required by the definition of the Google matrix.

From our sensitivity analysis on both sets of countries, we extract reasonable and really interesting geopolitical influences. Indeed, for instance in the set of 27 EU countries, our data shows clearly that the Nordic group of nations (Sweden, Denmark, Finland) have strong relationships together. If one of them increases its ties to another EU country alone, the remaining ones see their importance drop. The same observation is made for the group created by Austria, Hungary and Slovenia nations. These observations have been made by geopolitical specialists as well in [74] and [75], respectively. Another striking result is the impact of the exit of Great Britain from EU on the other European countries. Our data shows that Ireland will be the most affected country, which is inline with a study delivered recently by the London School of Economics [76]. From our worldwide set of 40 countries, we show that strengthening the relationship between Russia and the United States of America would negatively impact the importance of Ukraine worldwide, which is identical to the interpretation represented by Francis Fukuyama in a recent article [77].

This chapter is structured as follows. In Section 5.2, we explain our selection of nodes and how we group them together. Second, the properties of the calculated reduced Google matrices are shown in Section 5.3 for the set of 27 EU countries. Next, the methodology for our link sensitivity analysis is presented in Section 5.4. A detailed analysis for the two groups of countries is given in the Results section 5.5 that focuses on the sensitivity analysis of important relationships in the group. Results are first given and discussed for the set of 27 EU countries and then for the set of 40 worldwide nations. In order to show that the sensitivity analysis is meaningful in another setting, Section 5.6 leverages the same sensitivity analysis to extract the influence of painters on countries. Finally, conclusions are drawn in the Section 5.7.

5.2 Data Description

The selected countries are the 27 EU countries as of February 2013[1] and the 40 countries selected from the EnWiki network as the top 40 countries of the PageRank probability for the complete network.

Countries that belong to the same region or having a common piece of history may probably exhibit stronger interactions in Wikipedia. For the set of 40 countries, we have created a color code that groups together countries that either belong to the same geographical region (e.g. Europe, South America, Middle East, North-East Asia, South-East Asia) or share a big part of history (former USSR; English speaking countries that are the legacy of the former British Empire) as introduced in Chapter 4. On the other hand, EU countries are grouped upon their accession date to the union (e.g. Founder, 1973, 1981-1986, 1995, 2004-2007). Color code for EU

[1] Croatia joined in July 2013

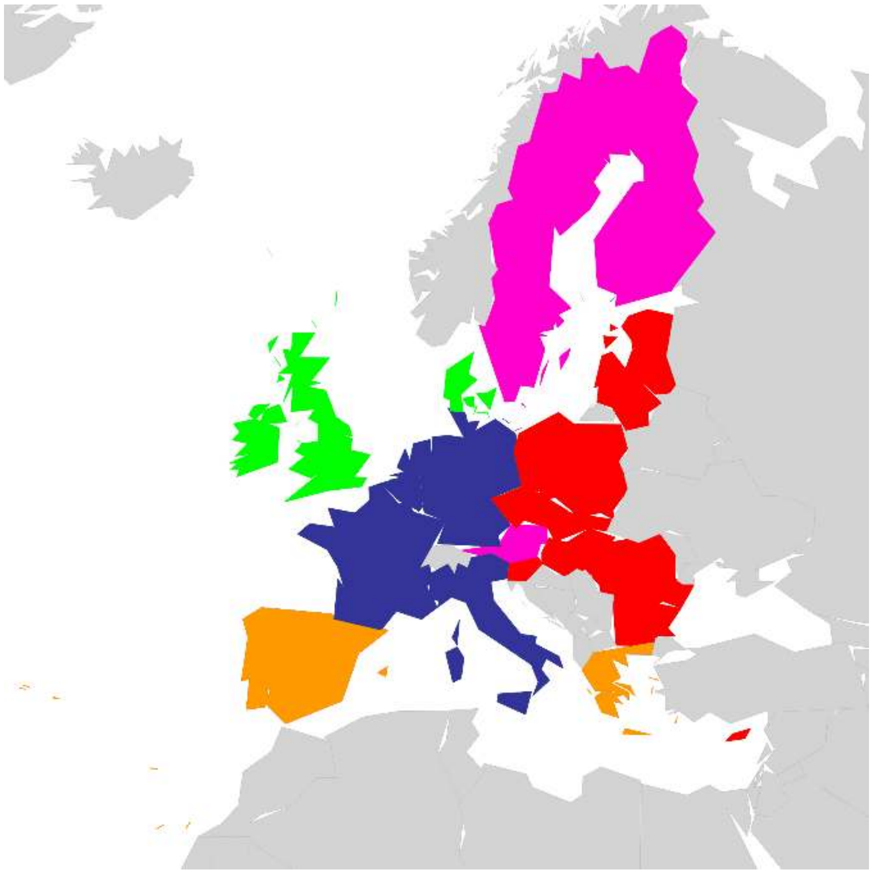

Figure 5.1: **Geographical distribution of the EU countries.** Color code groups countries into 5 subsets: Blue (BL) for Founders, Green (GN) for 1973 new member states, Orange (OR) for 1981 to 1986 new member states, Pink (PK) for 1995 new member states and Red (RD) for 2004 to 2007 new member states.

countries can be seen in Fig 5.1. Color code for the worldwide set of 40 countries is available in Table 4.1.

The two sets of 27 EU and 40 world countries are listed in Tables 5.1 and 4.1 respectively. The set of 40 countries has been chosen by selecting the countries with the largest PageRank probabilities in the full EnWiki network.

Wikipedia edition			English		French		German	
Countries	CC	Color	K	K*	K	K*	K	K*
France	FR	BL	1	10	1	6	2	9
United Kingdom	GB	GN	2	14	4	13	24	27
Germany	DE	BL	3	20	2	7	1	1
Italy	IT	BL	4	6	3	9	4	14
Spain	ES	OR	5	19	5	17	5	15
Poland	PL	RD	6	3	8	5	6	6
Netherlands	NL	BL	7	25	7	12	7	21
Sweden	SE	PK	8	13	11	25	8	18
Romania	RO	RD	9	1	18	4	17	20
Belgium	BE	BL	10	9	6	1	9	4
Austria	AT	PK	11	27	9	23	3	3
Greece	GR	OR	12	11	13	10	14	8
Portugal	PT	OR	13	24	12	2	11	2
Ireland	IE	GN	14	16	19	14	16	26
Denmark	DK	GN	15	7	14	20	10	10
Finland	FI	PK	16	4	17	18	15	7
Hungary	HU	RD	17	2	10	3	13	12
Czech Republic	CZ	RD	18	5	15	24	12	17
Bulgaria	BG	RD	19	22	20	11	20	13
Estonia	EE	RD	20	8	24	15	22	23
Slovenia	SI	RD	21	18	23	21	23	22
Slovakia	SK	RD	22	12	16	8	18	5
Lithuania	LT	RD	23	21	22	27	21	19
Cyprus	CY	RD	24	17	27	26	27	25
Latvia	LV	RD	25	23	25	22	25	24
Luxembourg	LU	BL	26	26	21	19	19	11
Malta	MT	RD	27	15	26	16	26	16

Table 5.1: **List of EU countries.** PageRank K and CheiRank K^* for EnWiki, FrWiki and DeWiki. Color code groups countries into 5 subsets: Blue (BL) for Founders, Green (GN) for 1973 new member states, Orange (OR) for 1981 to 1986 new member states, Pink (PK) for 1995 new member states and Red (RD) for 2004 to 2007 new member states. Standard country codes (CC) are given as well.

In Table 5.1 and Table 4.1, a local PageRank index K is given whose values range between 1 and 27 for EU countries, and between 1 and 40 for the other set. This local ranking keeps the countries in the same sequence as the original ranking over the entire network of webpages. The most influential countries are the top ranked ones with $K = 1, 2, \ldots$. Similarly, the local

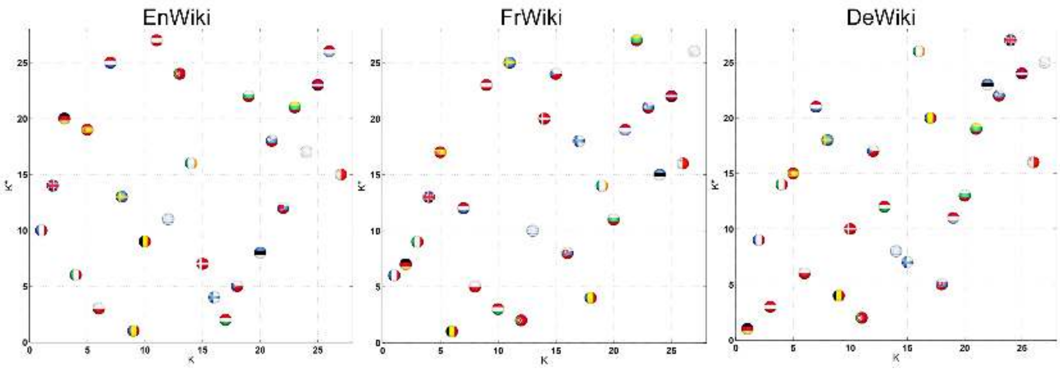

Figure 5.2: **Position of EU countries in the local (K, K^*) plane.** Networks are extracted from EnWiki (left), FrWiki (middle) and DeWiki (right). Countries are marked by their flags.

CheiRank index K^* [35,37] is given in both Tables for the two sets. At the top of K^* we have the most communicative countries. Both local K and K^* are given for EnWiki, ArWiki and RuWiki. Not surprisingly, the order of top countries changes with respect to the edition (for instance, the top country for K is US except for RuWiki whose top country is Russia).

It is convenient as well to plot all nodes in the $(K,\, K^*)$ plane to highlight the countries that are the most influential $(K = 1, 2, \,...)$ and the most communicative $(K^* = 1, 2, \,...)$ at the same time. Fig 5.2 plots EU countries in the $(K,\, K^*)$ plane for EnWiki, FrWiki and DeWiki editions. This plot is a bi-objective plot where K and K^* are to be minimized concurrently. It is interesting to look at the set of non-dominated countries which are the ones such that there is no other country beating them for both K and K^*.

5.3 Results: G_{R} properties

5.3.1 Reduced Google matrix of country networks

	W_{pr}	W_{qr}	W_{rr}	Sum
40	0.96120	0.029702	0.009098	1
EU	0.95332	0.038346	0.008334	1

Table 5.2: **Weights of the three matrices components of G_{R}.**

Reduced Google matrix has been computed, together with its components G_{rr}, G_{pr} and G_{qr}, for the English language edition of Wikipedia (EnWiki) and for the 2 selected sets of 27 and 40 countries listed in Tables 5.1 and 4.1. Countries are ordered by increasing K value in all subsequent matrix representations. The weight of the three matrix components of G_{R} are listed in Table 5.2. The weight of a matrix is given by the sum of all matrix elements divided by N_r. For W_{rr}, W_{qr}, W_{pr} the weights of G_{rr}, G_{pr} and G_{qr}, we have thus $W_{rr} + W_{\mathrm{qr}} + W_{\mathrm{pr}} = 1$. Predominant component is clearly G_{pr} but as already explained, it is not the most meaningful as it holds the projector component and thus is reflects mostly the PageRank vector.

For the three EnWiki, FrWiki and DeWiki editions, Fig 5.3 plots the density of matrices G_{R}, G_{qrnd} and G_{rr}. Several cultural biases can be extracted from G_{R}. For instance, France is the top country in EnWiki and FrWiki, while Germany is the top country in DeWiki.

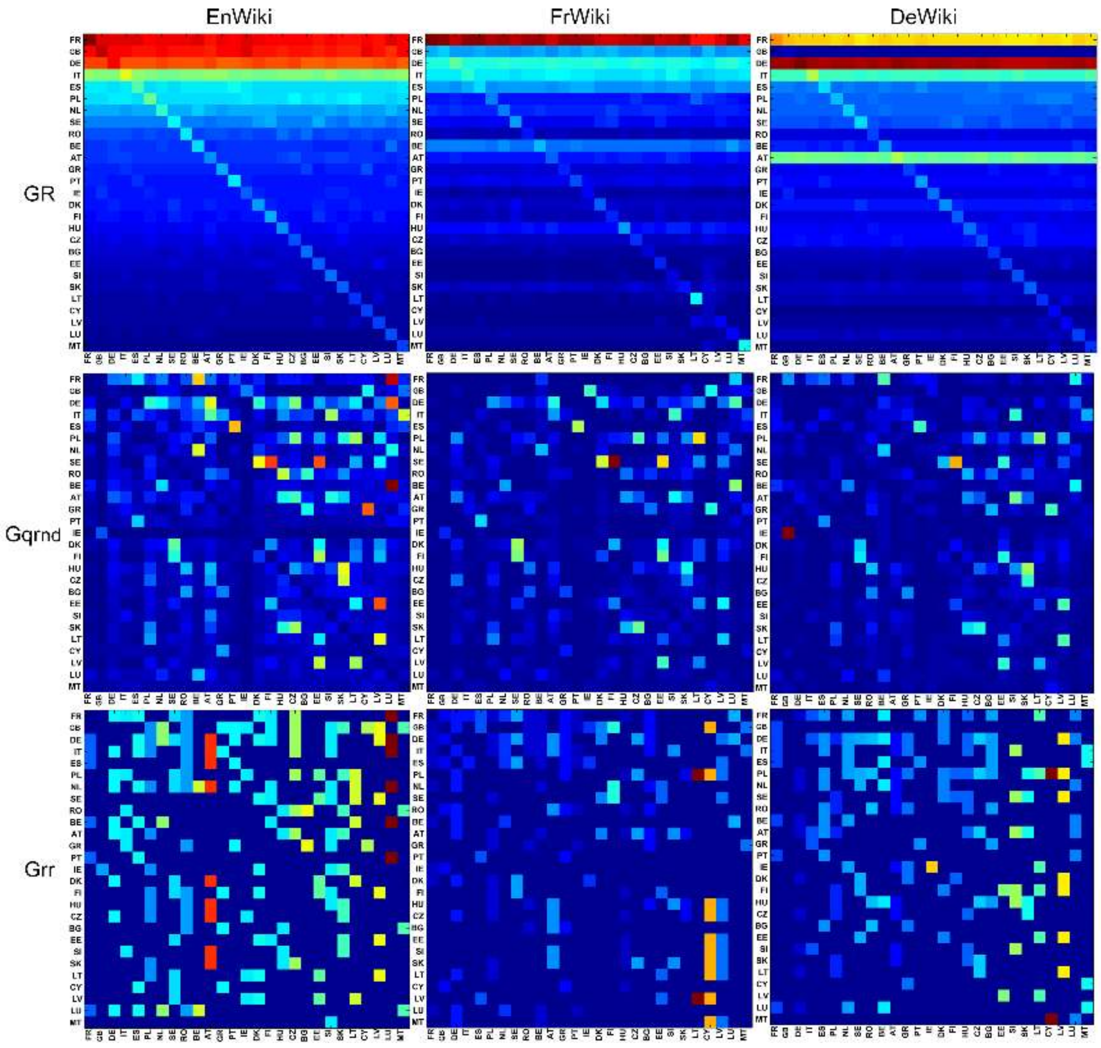

Figure 5.3: **Density plots of G_{R}, G_{qrnd} and G_{rr}.** G_{R} (first line), G_{qrnd} (second line) and G_{rr} (third line) for the reduced network of EU countries of EnWiki (left column), FrWiki (middle column) and DeWiki (right column). The nodes N_r are ordered in lines and columns by the reference PageRank of EnWiki. The colors represent maximum (red), intermediate (green) and minimum (blue) values.

As mentioned before, the information from hidden links between countries is provided by G_{qrnd}. It shows, for the three selected languages editions, the strong hidden links connecting Finland to Sweden. Other interesting hidden links are between Ireland and United Kingdom in DeWiki or in EnWiki linking Luxembourg to France. The reduced Google matrix density plots for the network of 40 worldwide countries are to be found in Chapter 4.

5.3.2 Networks of friends and followers

To create networks of friends and followers, we divide the set of N_r nodes into representative groups as shown in Fig 5.1 for 27 EU country set. EU countries are grouped upon their accession date to the union (e.g. Founder, 1973, 1981-1986, 1995, 2004-2007). One leading country per EU member state group has been selected as follows:

- France for Founders,
- United Kingdom for countries having joined in 1973,
- Spain for countries having joined between 1981 and 1986,
- Sweden for countries having joined in 1995,
- Poland for countries having joined between 2004 and 2007.

For each leading country j, we extract from both matrices G_{qrnd} and G_{R} the top 4 *Friends* (resp. *Followers*) given by the 4 best values of the elements of column j (resp. of line j). In other words, it corresponds to destinations of the 4 strongest outgoing links of j and the countries at the origin of the 4 strongest ingoing links of j. These networks of top 4 friends and followers have been calculated for the five editions of Wikipedia.

Top 4 friends and top 4 followers of EU leading countries are extracted from G_{R} and G_{qrnd} to plot the graphs of Fig 5.4 and 5.5. Results for EnWiki, FrWiki and DeWiki are presented here. Note that Fig 5.5 pictures hidden links. The black thick arrows identify the top 4 friends and top 4 followers interactions. Red arrows represent the friends of friends (respectively the followers of followers) interactions that are computed recursively until no new edge is added to the graph. All graphs are visualized with the Yifan Hu layout algorithm [2] using Gephi [1].

Friends and followers from G_{R} The vertices of the network of friends obtained from G_{R} concentrate, for each Wiki, to about 7 countries, 5 of which being the leading ones. The other vertices are top PageRank countries such as Italy, Germany or Spain. This is due to the predominance of PageRank probabilities in the structure of G_{R}.

A more valuable information could be extracted from the network of followers. In all editions, Benelux and Nordic countries create a cluster densely interconnected. The networks of followers end up spanning the full set of EU countries in this representation. On this representation, it can be noticed that the order of arrival of member states is meaningful. Indeed, nodes of the same color are closely interconnected.

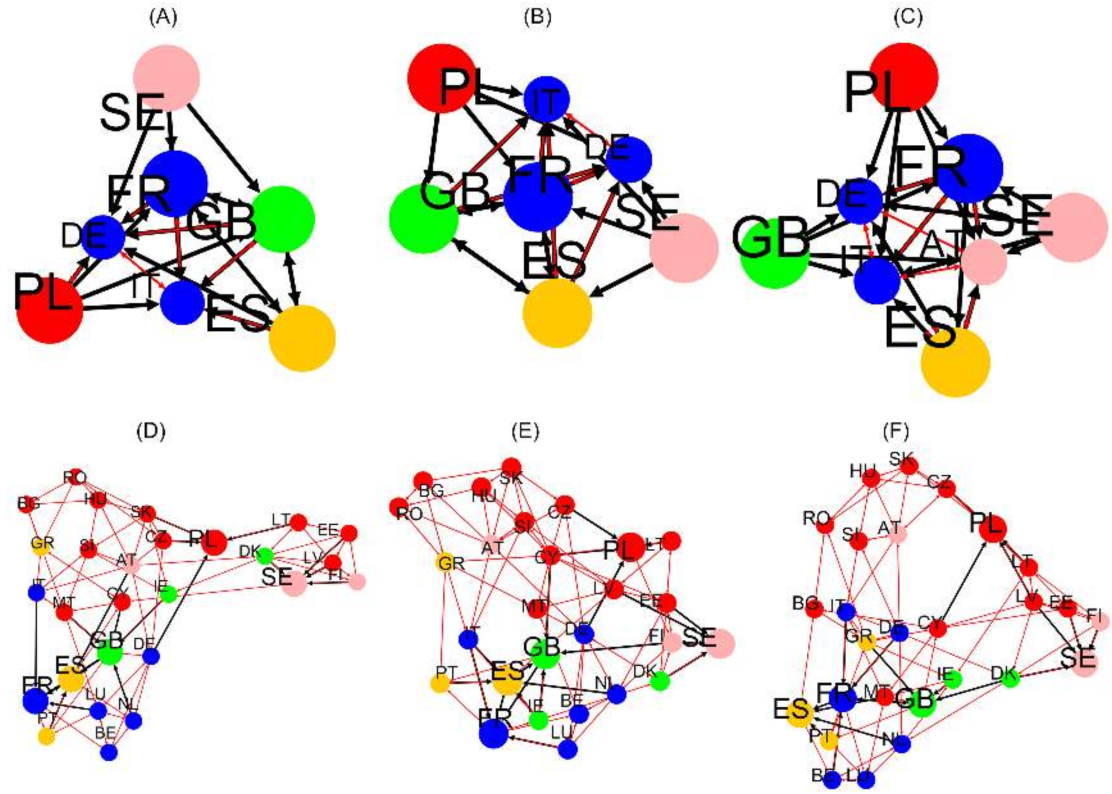

Figure 5.4: **Relationships structure extracted from G_{R} for the network of EU countries.** Friends (top line) and followers (bottom line) induced by the top 4 countries of each group (FR, GB, ES, SE, PL). Results are plotted for EnWiki (A and D), FrWiki (B and E) and DeWiki (C and F). Node colors represent geographic appartenance to a group of countries. Top (bottom) graphs: a country node with higher PageRank probability has a bigger size and points (is pointed by) with a bold black arrow to its top 4 friends (followers). Red arrows show friends of friends (resp. followers of followers) interactions computed until no new edges are added to the graph.

Friends and followers from G_{qrnd} The hidden friends and followers relationships are extracted from G_{qrnd} and illustrated in Fig 5.5. As discussed earlier, G_{qrnd} is not dominated by PageRank, and as such, the resulting network of friends includes more nodes and shows more diversity. It is worth noting that Germany, as one of the Founders, bridges the group of Founders to Sweden (the leader of the countries that have joined EU in 1995) and Poland (the leader of the countries that have joined EU between 2004 and 2007) in FrWiki and EnWiki. From EnWiki and DeWiki, strong ties are seen between Italy and France, while it is not the case from FrWiki authors. This is another example of cultural bias. However, lots of links are seen in

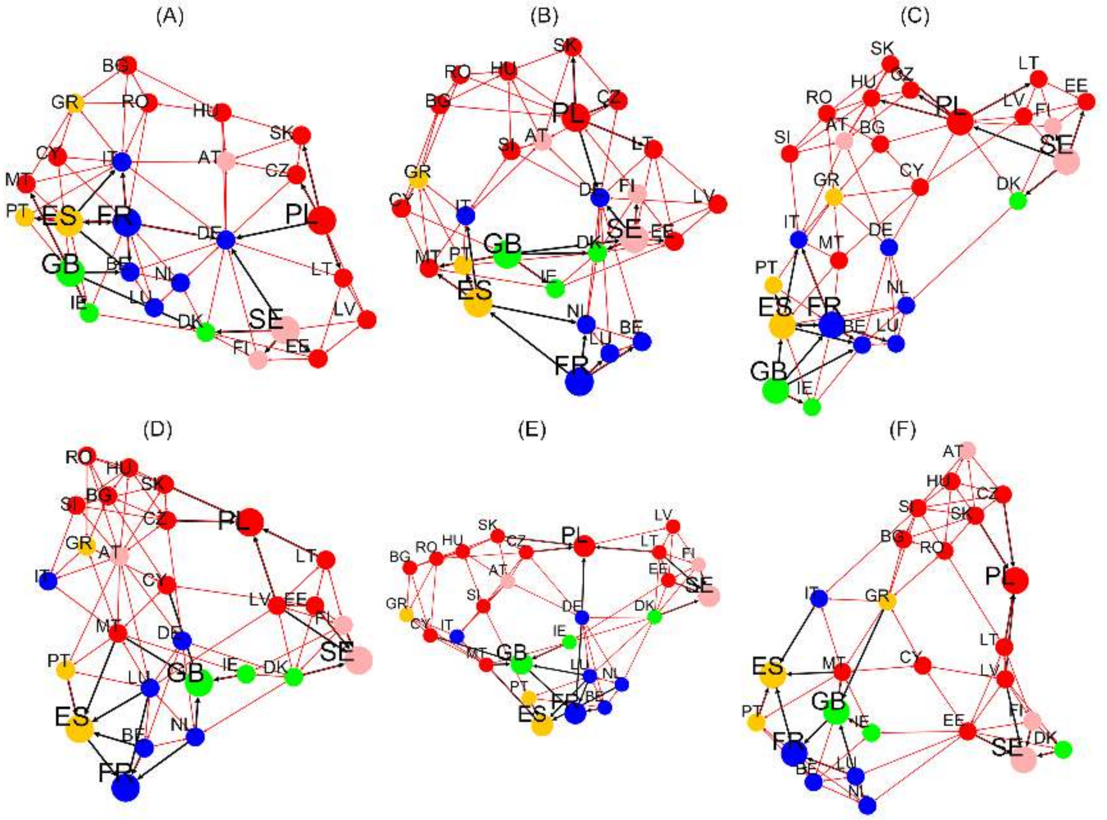

Figure 5.5: **Relationship structure extracted from G_{qrnd} for the network of EU countries.** Friends (top line) and followers (bottom line) induced by the top 4 countries of each group (FR, GB, ES, SE, PL). Results are plotted for EnWiki (A and D), FrWiki (B and E) and DeWiki (C and F). Node colors represent geographic appartenance to a group of countries. Top (bottom) graphs: a country node with higher PageRank probability has a bigger size and points (is pointed by) with a bold black arrow to its top 4 friends (followers). Red arrows show friends of friends (resp. followers of followers) interactions computed until no new edges are added to the graph.

all three editions: GB-IE, SE-FI, ES-PT, PL-LT, IT-GR and many others. To underline this constant presence of links, we give in Table 5.3 the list of friends (resp. followers) that are among the top 4 ones in all 5 editions, in 4 out of 5 and in 3 out of 5 for G_{qrnd} analysis. For each leading country, around 2 to 3 top friends and followers exist across all editions. For the 40 worldwide countries set, networks of top 4 friends and followers are to be found in Chapter 4, calculated for the same 5 editions of Wikipedia as well. Similar observations have been made as for the set of 27 EU countries.

Top	G_{qr} Wiki friends present in			G_{qr} Wiki followers present in		
country	all 5 editions	4 out of 5 editions	3 out of 5 editions	all 5 editions	4 out of 5 editions	3 out of 5 editions
FR	BE -ES	IT		BE	LU - ES	
GB	IE		DK - FR	IE - MT	CY	
ES	IT - PT	FR	BE	MT - PT		LU
SE	DK - FI		EE	DK - EE - FI	LV	
PL	CZ		DE - HU - LT - SK	CZ - LT - SK		LV

Table 5.3: **Cross-edition friends and followers extracted from G_{qrnd} of EU countries per leading country.** For each top country, we list the friends and followers that are identical accros all five Wikipedia editions, in 4 editions out of 5 and in 3 editions out of 5.

5.4 Influence analysis of geopolitical ties using G_{R}

We have now established the global mathematical structure G_{R} and presented how it can be leveraged to extract meaningful geopolitical interactions among countries for the two sets of interest, naming 27 EU and 40 worldwide countries.

These interactions are extracted from Wikipedia and thus stem from all links covering this very rich network of webpages. As such, they encompass not only interactions related to economics or politics, but from any possible domain (arts, history, entertainment, etc.). The strength of this study is to show that just from the structure of the network, relevant and timely information can be extracted. The hyperlinked structure of Wikipedia itself contains an important part of the universal knowledge stored in details on the webpages.

Previous study has shown that G_{R} captures essential interactions between countries. The point is now to see how some ties between countries influence the whole network structure. More specifically, we focus here on capturing the impact of a change in the strength of a relationship between two countries on the importance of the nodes in the network. Therefore we have designed a sensitivity analysis that measures a logarithmic derivative of the PageRank probability when the transition probability of only one link is increased for a specific couple of nodes in G_{R}, relatively to the other ones.

Our sensitivity analysis is performed for a directed link where the relationship going from country i to j is increased. We investigate in the last part of this Section the imbalance between the influence of two opposite direction interactions. In other words, we conduct the aforementioned sensitivity analysis for the link going from country i to j, and for the link going in the opposite direction from j to i. For each pair of countries, we derive from this two-way sensitivity the relationship imbalance to identify the most important player in the relationship.

5.4.1 Sensitivity analysis

We define δ as the relative fraction to be added to the relationship from nation j to nation i in G_{R}. Knowing δ, a new modified matrix $\tilde{G}_{\mathrm{R}}$ is calculated in two steps. First, element $\tilde{G}_{\mathrm{R}}(i,j)$ is set to $(1+\delta)\cdot G_{\mathrm{R}}(i,j)$. Second, all elements of column j of $\tilde{G}_{\mathrm{R}}$ are normalized to 1 (including element i) to preserve the unity column-normalization property of the Google matrix. Now $\tilde{G}_{\mathrm{R}}$ reflects an increased probability for going from nation j to nation i.

It is now possible to calculate the modified PageRank eigenvector $\tilde{P}$ from $\tilde{G}_{\mathrm{R}}$ using the standard $\tilde{G}_{\mathrm{R}}\tilde{P} = \tilde{P}$ relation and compare it to the original PageRank probabilities P calculated with G_{R} using $G_{\mathrm{R}}P = P$. The same process can be applied to the transposed version of $\tilde{G}_{\mathrm{R}}$ to calculate the modified CheiRank probabilities $\tilde{P}^{*}$. Due to the relative change of the transition probability between nodes i and j, steady state PageRank and CheiRank probabilities are modified. This reflects a structural modification of the network and entails a change of importance of nodes in the network.

These changes are measured by a logarithmic derivative of the PageRank probability of node a:

$$D_{(j\to i)}(a) = (\mathrm{d}P_a/\mathrm{d}\delta_{ij})/P_a = (\tilde{P}_a - P_a)/(\delta_{ij}P_a) \tag{5.1}$$

Notation $(j \to i)$ indicates that the link from node j to node i has been modified. Element $D_{(j\to i)}(a)$ gives the logarithmic variation of PageRank probability for country a if the link from j to i has been modified. We will refer to this variation as the *sensitivity* of nation a to the relationship from nation i to nation j. If this sensitivity is negative, country i has lost importance in the network. On the opposite, a positive sensitivity expresses a gain in importance. The computation has been tested for values of $\delta = \pm 0.01, \pm 0.03, \pm 0.05$. Results are almost identical for these three values of δ and thus, we have arbitrarily chosen to use $\delta = 0.03$.

5.4.2 Relationship imbalance analysis

As introduced earlier, sensitivity $D_{(j\to i)}(k)$ of Eq (5.1) measures the change of importance of node a if the link from nation j to i has been changed. The sensitivity of node a to a change in one direction is not necessarily the same as its sensitivity to the change in the opposite direction. We define as such the *2-way sensitivity* of node a which is simply the sum of the sensitivities calculated for both directions:

$$D_{(i\leftrightarrow j)}(a) = D_{(i\to j)}(a) + D_{(j\to i)}(a) \tag{5.2}$$

The two-way sensitivity can be leveraged to find out, for a pair of countries a and b, which one has the most influence on the other one. Therefore, we define the following metric :

$$F(a,b) = D_{(a\leftrightarrow b)}(a) - D_{(a\leftrightarrow b)}(b) \tag{5.3}$$

Here, we measure the 2-way sensitivity for nodes a and b when the link between them is modified both ways in G_{R}. If $F(a,b)$ is positive, it means that the 2-way sensitivity of a is larger than the 2-way sensitivity of b. In this case, a is more influenced by b than b by a. We can say that b is the *strongest* country. If $F(a,b)$ is negative, we can say that a is the strongest country.

5.5 Sensitivity results for country networks

Sensitivity analysis results are shown first for the 27 EU network and then for the 40 worldwide network. For each network, we have identified a set of meaningful links between countries to be modified and observed resulting sensitivity of other nations. We perform as well for each network the relationship imbalance analysis for each pair of nations. More results could be found in our website [78]. Note that if the modified link is clearly identified, we will drop the index $i \to j$ in our sensitivity measure notation for clarity.

5.5.1 27 EU network of countries

In order to better capture the countries' sensitivities from a multicultural perspective, we have calculated the sensitivities for 3 Wikipedia editions: EnWiki, FrWiki and DeWiki. All sensitivity results shown for 27 EU network have been averaged over the three editions as follows:

$$\bar{D} = \frac{1}{3}\sum_{i=1}^{3} D^i \tag{5.4}$$

where index i refers to the Wikipedia edition.

5.5.1.1 Sensitivity analysis

We start this analysis by introducing a first simple example where Italy increases its relationship with France. Then, we analyze the impact on the EU countries of Great Britain's exit (i.e. Brexit) from European Union. Next, we highlight the sensitivity of Luxembourg to the increase of Germany and France's cooperation with other member states. Finally, we present the results that underline the strong ties that exist between groups of countries that function together in Europe.

For each sensitivity analysis, we show two types of figures: *i*) an axial representation of the sensitivity $\bar{D}$ (cf. Fig 5.6, Fig 5.12, Fig 5.14, Fig 5.8, Fig 5.10) and *ii*) a colored map of Europe where countries' color indicate the sensitivity $\bar{D}$ as well (cf. Fig 5.7, Fig 5.13, Fig 5.15, Fig 5.9, Fig 5.11). Color scale for these maps plots lower values of $\bar{D}$ in red, median in green and larger in blue. Each map represents the sensitivity values obtained for a given link variation.

Italy to France relationship Italy is the second top export and import country of Slovenia with \$3.05B and \$3.84B respectively. In 1992, diplomatic relations began between the two countries and in 2012, Foreign Minister of Italy, Giulio Terzi, described the bilateral relationship between Italy and Slovenia as fruitful and dynamic [79]. Politically, Slovenia relies on Italy to become a member of the principal UN, EU and NATO bodies [79]. No doubt Slovenia would suffer if Italy decided to go away from it and increase its relationships with France. The 27 EU network exactly shows the negative impact of Italy increasing its link in G_{R} with France: Slovenia is the nation with lowest sensitivity on Fig 5.6 and 5.7.

Figure 5.6: **Axial representation of $\bar{D}$ for a link modification from {IT} to {FR}.** Here $\bar{D}(IT) = -0.0159$ and $\bar{D}(FR) = 0.0701$ are not represented.

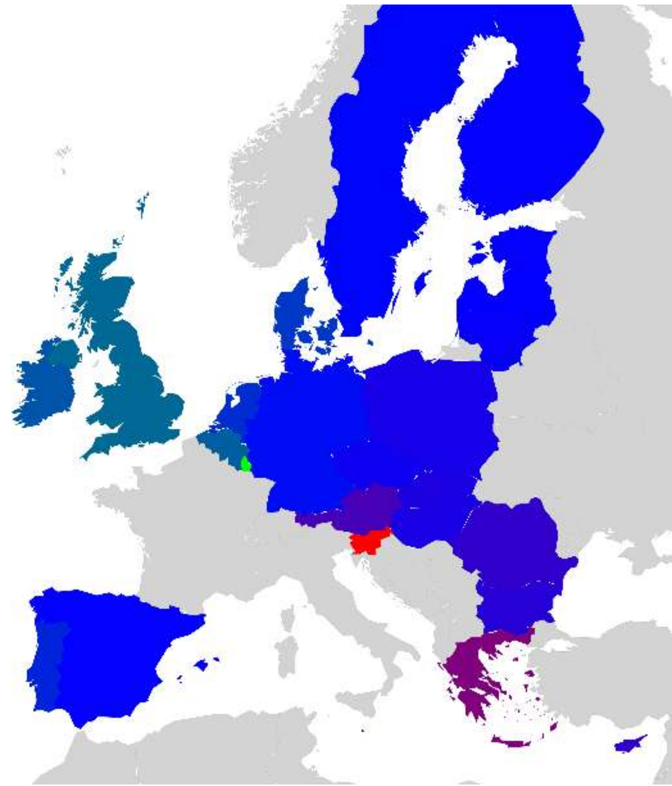

Figure 5.7: **Map representation of $\bar{D}$ for links modifications from {IT} to {FR}.** Lower values of $\bar{D}$ in red, median in green and larger in blue. Values of $\bar{D}$ for IT and FR are not represented.

Impact of Brexit [2] The United Kingdom has triggered article 50 on March 27, 2017 to leave the European Union as a consequence of the referendum of June 23rd, 2016 [80]. To understand its impact on EU countries with our dataset, we have reduced (and not increased as done in other studies) the G_{R} transition probability UK towards France or Germany. We remind that our network is dated by 2013 but it captures the strong UK influence. Results are shown in Fig 5.8 and 5.9 and indicate that Ireland and Cyprus are by far the most negatively affected countries in both cases. Moreover, the sensitivity of UK is negative as it benefits less from France's or Germany's influence. These facts have been recently backed up by specialists. In [76], a study delivered by the London School of Economics discussing the consequences of Brexit forecasts that UK will loose 2.8% of its GDP[3]. Similarly, [76] shows that Ireland will loose as well 2.3% of its GDP, which is the largest proportional loss caused by Brexit. Cyprus-UK Relations are

[2]Brexit is an abbreviation for Britain exit [80].

[3]Gross domestic product (GDP) is the monetary value of all the finished goods and services produced within a country's borders in a specific time period [81].

strong as claimed by the official website of the Ministry of Foreign Affairs of Cyprus [82]. Referring to [83], UK is the 4^{th} top export destination for Cyprus with \$242M and the 2^{nd} import origin with \$508M. As such, this clear bond of UK with Cyprus explains that if GB suffers from Brexit, Cyprus will do as well. Our data strikingly exhibits the same conclusion as shown in Fig 5.8 and 5.9.

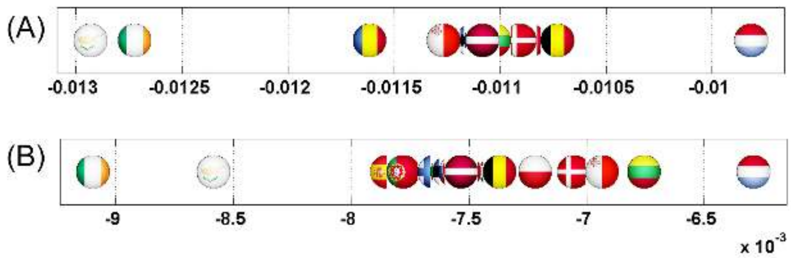

Figure 5.8: **Axial representation of $\bar{D}$ for link modifications from {GB} to {FR or DE}.** (A): GB to FR (Non represented values: $\bar{D}(GB) = -0.0124$ and $\bar{D}(FR) = 0.0577$). (B): GB to DE (Non represented values: $\bar{D}(GB) = -0.0087$ and $\bar{D}(DE) = 0.0606$).

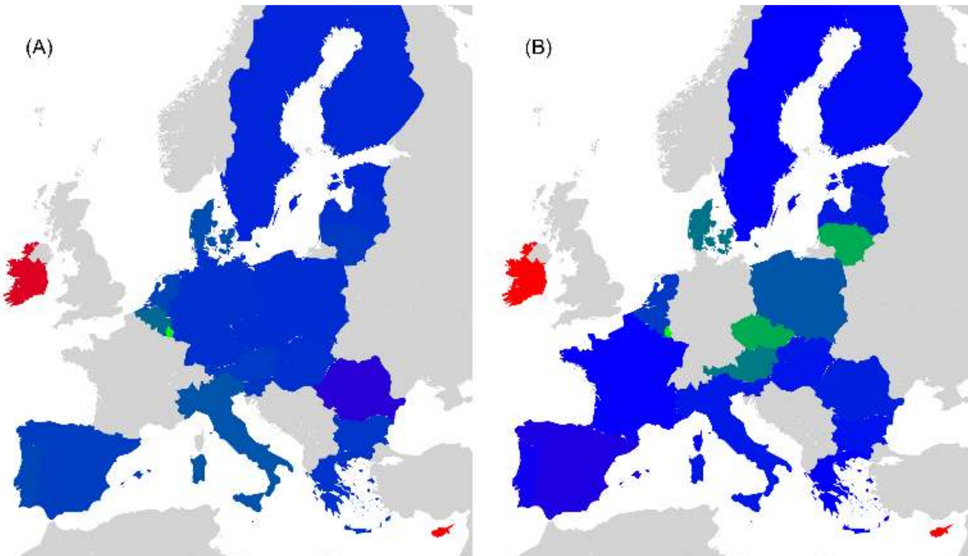

Figure 5.9: **Map representation of $\bar{D}$ for link modifications from {GB} to {FR or DE}.** (A): GB to FR ($\bar{D}$ for GR, FR are not represented); (B): GB to DE($\bar{D}$ for GR, DE are not represented). Lower values of $\bar{D}$ in red, median in green and larger in blue.

Luxembourg's sensitivity to Germany and France Luxembourg shares its borders with Belgium, Germany and France with whom it has strong and diverse relationships. Luxembourg has a very open economy. Together with Belgium, they position themselves as the 12^{th} largest economy in the world. Two of the top three export and import countries of Belgium-Luxembourg are Germany (\$44.6B, \$50.4B) and France (\$43.8B, \$36.8B) [83]. Official languages in Luxembourg are Luxembourgish, French and German. Luxembourg has robust relationships with France [84, 85] and Germany [86] in various areas such as finance, culture, science, security or nuclear power. It is clear that Luxembourg will suffer if one of these European countries reduces its exchanges with it. In Fig 5.10 and 5.11, we clearly show with our sensitivity analysis that Luxembourg is strongly influenced by France and Germany. If France or Germany increases its relationships with Italy or Great Britain, Luxembourg is by far the most negatively impacted country.

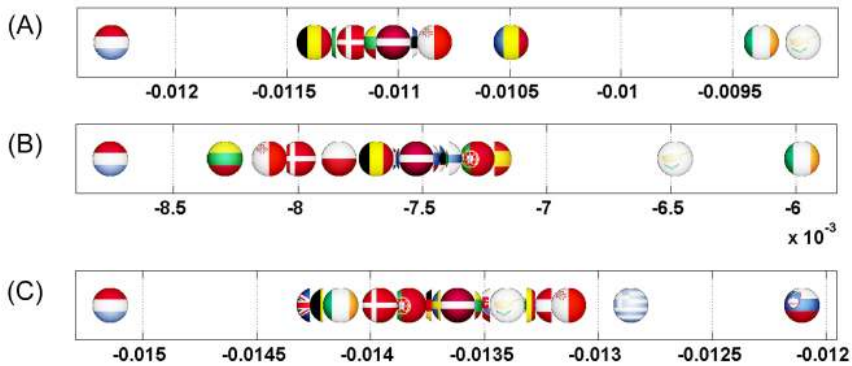

Figure 5.10: **Axial representation of $\bar{D}$ for link modifications from {FR or DE} to {GB or IT}.** (A): FR to GB (Non represented values: $\bar{D}(FR) = -0.0117$ and $\bar{D}(GB) = 0.1572$). (B): DE to GB (Non represented values: $\bar{D}(DE) = -0.0081$ and $\bar{D}(GB) = 0.1248$). (C) FR to IT (Non represented values: $\bar{D}(FR) = -0.0143$ and $\bar{D}(IT) = 0.1508$).

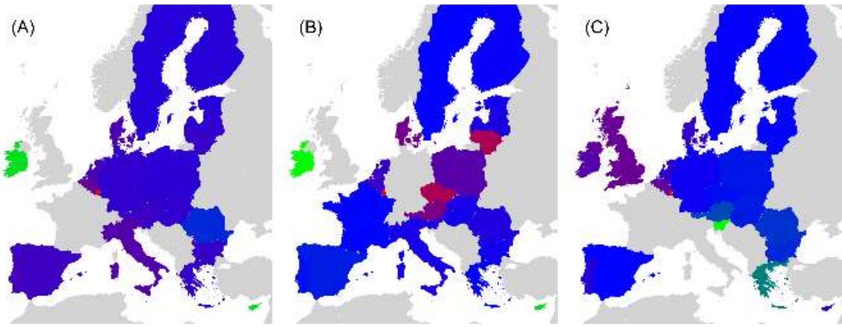

Figure 5.11: **Map representation of $\bar{D}$ for link modifications from {FR or DE} to {IT or GB} : Luxembourg is negatively impacted here.** (A): FR to GB. (B): DE to GB. (C) FR to IT. Lower values of $\bar{D}$ in red, median in green and larger in blue; Values of $\bar{D}$ for FR, DE, IT and GB are not represented.

Clusters of countries By analyzing the sensitivity of countries to various 2-nation relationships, we have noticed that several groups of nations function together. These groups are strongly interconnected, and if anyone of these group members increases its relationship strength with a country outside of the group, all group members loose importance in the network. We highlight two meaningful examples next: the cluster of Nordic countries and the cluster Austro-Hungarian cluster. Other clusters we have identified in our network are for instance the cluster of Benelux countries (e.g. Belgium, the Netherlands and Luxembourg) or the cluster of the Iberian peninsula (e.g. Portugal and Spain).

For both investigated groups, we test the influence of an increase in collaboration from one member of the group to France or to Germany. France and Germany have been chosen as they are central members of European Union.

The Nordic countries Denmark, Finland, and Sweden have much in common: their way of life, history, language and social structure [74]. After World War II, the first concrete step into unity was the introduction of a Nordic Passport Union in 1952. Nordic countries co-operate in the Nordic Council, a geopolitical forum. In the Nordic Statistical Yearbook [74], Klaus Munch illustrates that “The Nordic economies are among the countries in the Western World with the best macroeconomic performance in the recent ten years”. Nordic countries should keep cooperating to stay strong. Thus, if any Nordic country attempts to abandon these relationships in favor of other countries, it will negatively impact the remaining Nordic countries. Our sensitivity analysis illustrates this impact in Fig 5.12 and 5.13. In these figures, we show how the relationship increase between any Nordic country towards France or Germany induces a drop in sensitivity for Nordic countries.

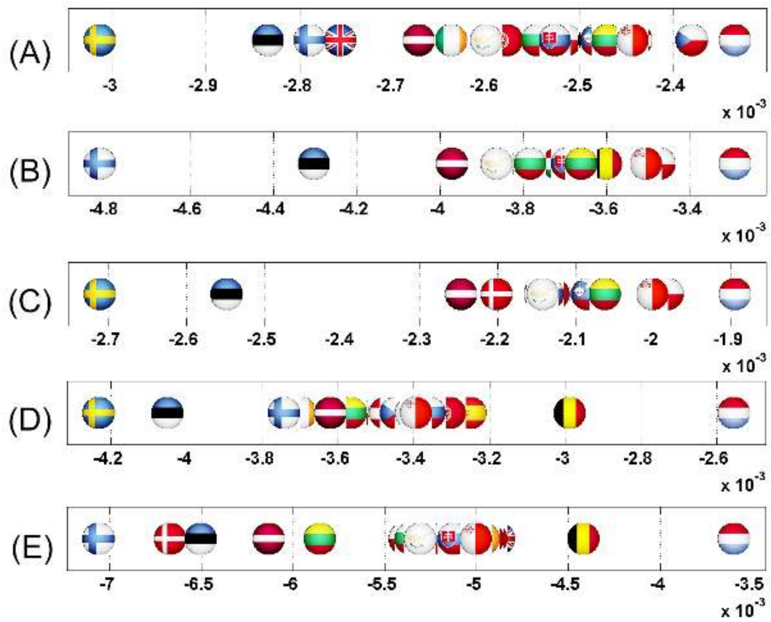

Figure 5.12: **Axial representation of $\bar{D}$ for link modifications from Nordic countries to {FR or DE}.** (A): DK to DE (Non represented values: $\bar{D}(DK) = -0.0050$ and $\bar{D}(DE) = 0.0208$). (B): SE to DE (Non represented values: $\bar{D}(SE) = -0.0064$ and $\bar{D}(DE) = 0.0313$). (C): FI to DE (Non represented values: $\bar{D}(FI) = -0.0046$ and $\bar{D}(DE) = 0.0173$). (D): DK to FR (Non represented values: $\bar{D}(DK) = -0.0077$ and $\bar{D}(FR) = 0.0197$). (E): SE to FR (Non represented values: $\bar{D}(SE) = -0.0100$ and $\bar{D}(FR) = 0.0296$).

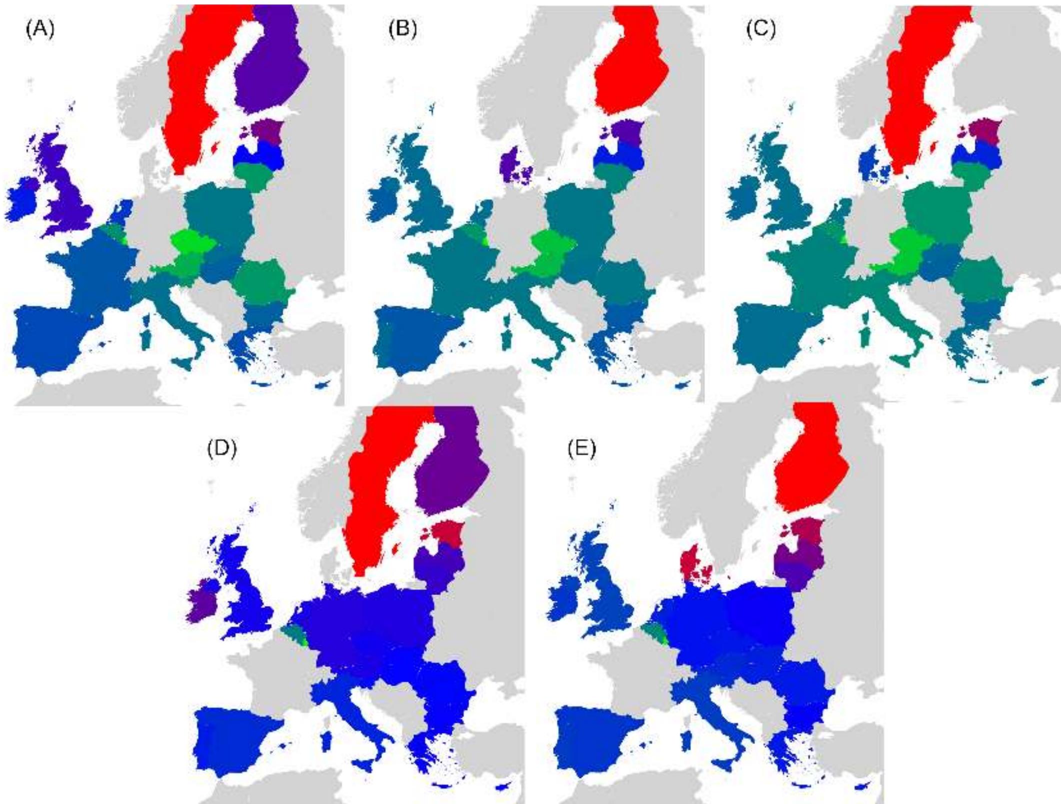

Figure 5.13: **Map representation of $\bar{D}$ for link modifications from Nordic countries to {FR or DE}.** (A): DK to DE. (B): SE to DE. (C): FI to DE. (D): DK to FR. (E): SE to FR. Lower values of $\bar{D}$ in red, median in green and larger in blue. Values of $\bar{D}$ for DK, BE, SE and FI are not represented.

Referring to [75], relations between Slovenia, Hungary and Austria are tight. Hungary has supported Slovenia for its NATO membership applications and Austria has assisted Slovenia in entering European Union. Relationships between Austria and Hungary are important for both countries in the economic, political and cultural fields [87]. Concerning economy [83], Austria is one of the top import origins for Hungary and Slovenia with \$5.54B and \$2.37B respectively. Similarly to the Nordic group of countries, if Austria, Slovenia or Hungary increases its relationships with another European country, the other two will be affected. Sensitivity analysis backs up this statement as seen in Fig 5.14 and 5.15.

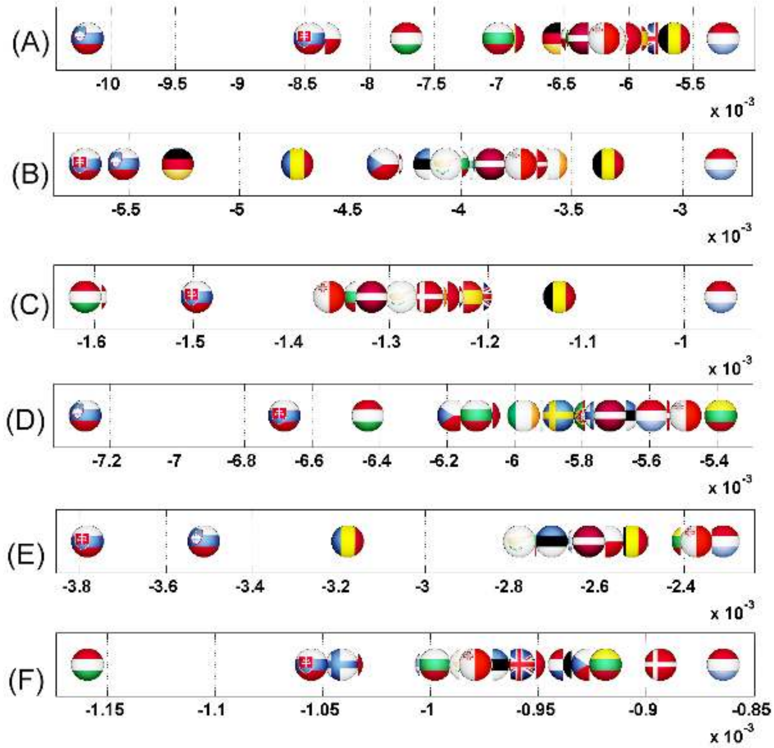

Figure 5.14: **Axial representation of $\bar{D}$ for link modifications from {AT, HU and SI} to {FR or DE}.** (A): AT to FR (Non represented values: $\bar{D}(AT) = -0.0101$ and $\bar{D}(FR) = 0.0373$). (B): HU to FR (Non represented values: $\bar{D}(HU) = -0.0080$ and $\bar{D}(FR) = 0.0205$). (C): SI to FR (Non represented values: $\bar{D}(SI) = -0.0046$ and $\bar{D}(FR) = 0.0075$). (D): AT to DE (Non represented values: $\bar{D}(AT) = -0.0070$ and $\bar{D}(DE) = 0.0393$). (E): HU to DE (Non represented values: $\bar{D}(HU) = -0.0052$ and $\bar{D}(DE) = 0.0311$). (F): SI to DE (Non represented values: $\bar{D}(SI) = -0.0034$ and $\bar{D}(DE) = 0.0081$).

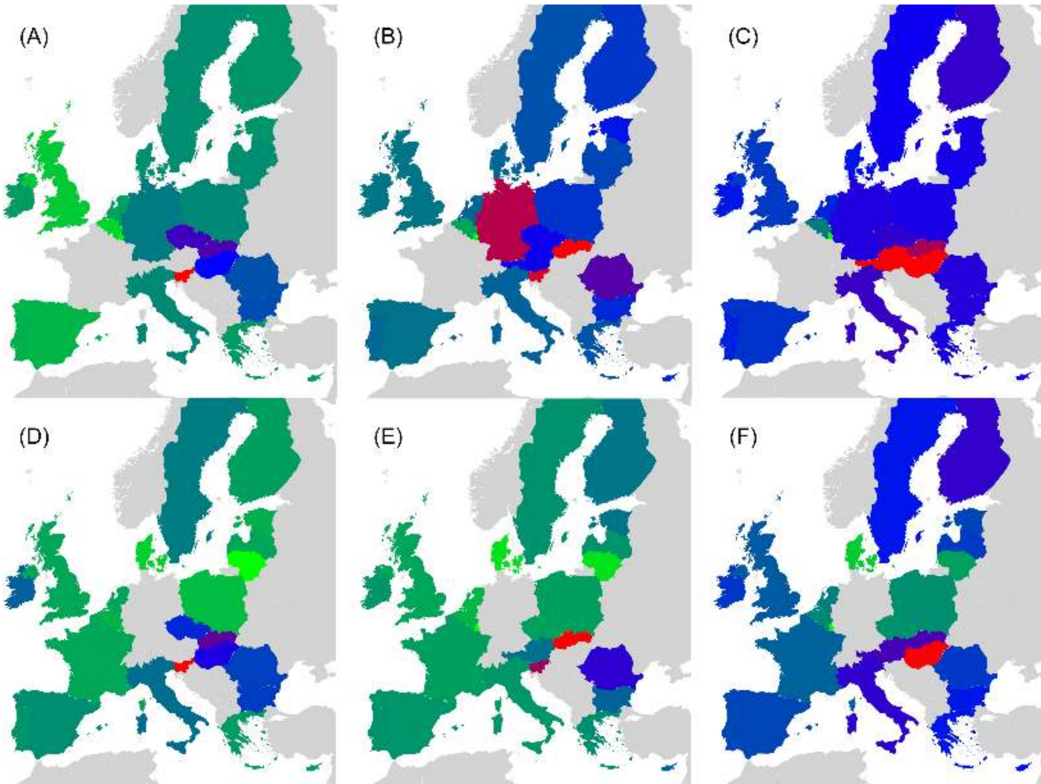

Figure 5.15: **Map representation of $\bar{D}$ for link modifications from {AT, HU and SI} to {FR or DE}.** (A): AT to FR. (B): HU to FR. (C): SI to FR. (D): AT to DE. (E): HU to DE. (F): SI to DE. Lower values of $\bar{D}$ in red, median in green and larger in blue; Values of $\bar{D}$ for AT, FR, HU, SI, DE are notre represented.

5.5.1.2 Relationship imbalance analysis

Relationship imbalance analysis has been derived for all pairs of European countries following Eq (5.3). Fig 5.16 shows a density plot of $F(a,b)$. We recall that if $F(a,b)$ is negative, nation a has more influence on nation b than b on a. If $F(a,b)$ is positive, nation b dominates nation a. According to *The Globe of Economic Complexity* [88] and identical to our results in Fig 5.16, Germany and France are the two largest economies in Europe. From G_{R} we can clearly see the dominance of France and Germany on other EU countries. Another interesting result of Fig 5.16 is the equal influence between all pairs of countries created by one member of {GR, PT, IE, DK, FI, HU} and another of {BG, EE, SI, SK, LT, CY, LV, LU, MT}. These pairs have $F(a,b)$ close to zero and are plotted with orange color in Figure 5.16.

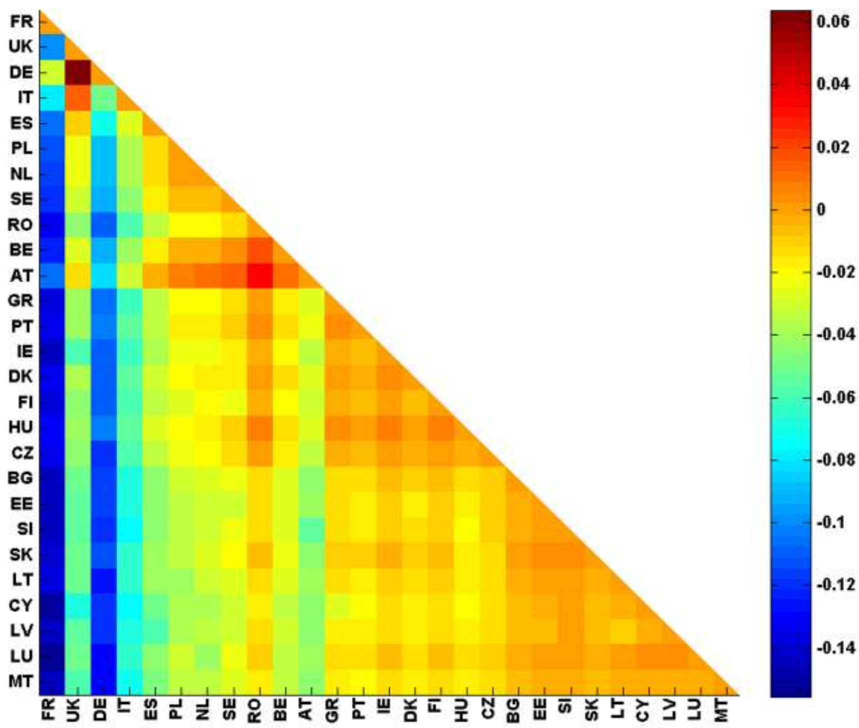

Figure 5.16: **Relationship imbalance analysis: F-representation for 27 EU network.** $F(a,b)$ is given by the colorbar. X-axis and Y-axis represent a and b respectively. If $F(a,b)$ is negative, nation a has more influence on nation b than b on a.

5.5.2 40 worldwide network of countries

Similarly to the 27 EU countries dataset, sensitivity results are averaged over 5 Wikipedia editions: ArWiki, EnWiki, FrWiki, RuWiki and DeWiki. We first show as well the sensitivity analysis for carefully selected links and then conclude this part with the sensitivity imbalance analysis for all pairs of countries.

5.5.2.1 Sensitivity Analysis

In this worldwide set of countries, we have identified relationships whose impact on the network clearly shows how meaningful the sensitivity analysis proposed in this thesis is.

US - Russia. As mentioned previously in the introduction, and according to the results in Fig 5.17 and 5.18, Ukraine would be the most affected country if Russia gets closer to US. This is due to the fact that Ukraine and Russia were both in the USSR and their economies are strongly interconnected. The next influenced country is Finland which also has strong economic relations with Russia being a part of Russian Empire till beginning of 20th century.

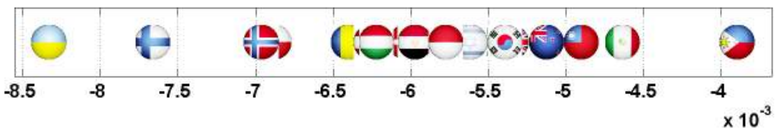

Figure 5.17: **Axial representation of $\bar{D}$ for link modification from RU to US.** Non represented values: $\bar{D}(RU) = -0.0089$ and $\bar{D}(US) = 0.0446$.

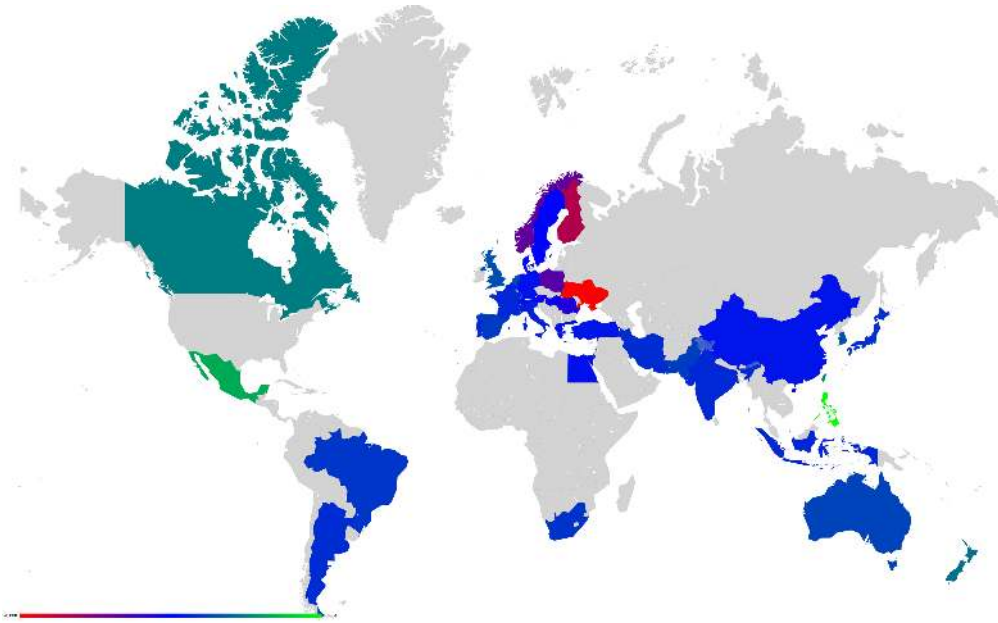

Figure 5.18: **Map representation of $\bar{D}$ for link modification from RU to US.** Values of $\bar{D}$ for RU or US are not represented. Lower values in red, median blue and larger in green

China to US. The effects of an increase in the relationship from China to US are shown in Fig 5.20 and 5.19. Taiwan and Pakistan are the most negatively affected countries. Taiwan is not pictures in Fig 5.20 and 5.19 as it greatly reduces readability of the plots. Indeed, sensitivity of Taiwan is $\bar{D}(TW) = -0.0087$, 4 times the one of Pakistan. BBC's article [89] on the division between China and Taiwan illustrates that US is the most important friend and the only ally of Taiwan. China claims Taiwan as its territory and Taiwan counts on US to establish its full independence to stand up against China. As such, if the ties between China and US get stronger, Taiwan will loose its best ally.

In 1951, Pakistan and China officially established their diplomatic relations and in 2016 they celebrated 65 years of friendship [90]. Regarding security strategy, China has always supported Pakistan in facing terrorism. Politically, Pakistan stands with China on many issues concerning China's core interests (e.g. Taiwan, Tibet, Xinjiang). The trade volume between the two countries reached \$100.11B by 2015 and in 2016 the \$46B China-Pakistan Economic Corridor (CPEC) [91] was constructed. If China strengthens its relationship with US, Pakistan may clearly suffer from it. An article by Ian Price [92] raises a serious question on whether United States aims at sabotaging the CPEC in the near future.

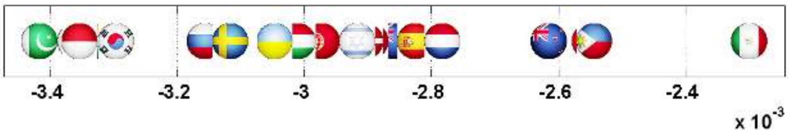

Figure 5.19: **Axial representation of $\bar{D}$ for link modification from CN to US.** Non represented values: $\bar{D}(CN) = -0.0056$, $\bar{D}(US) = 0.0210$ and $\bar{D}(TW) = -0.0087$.

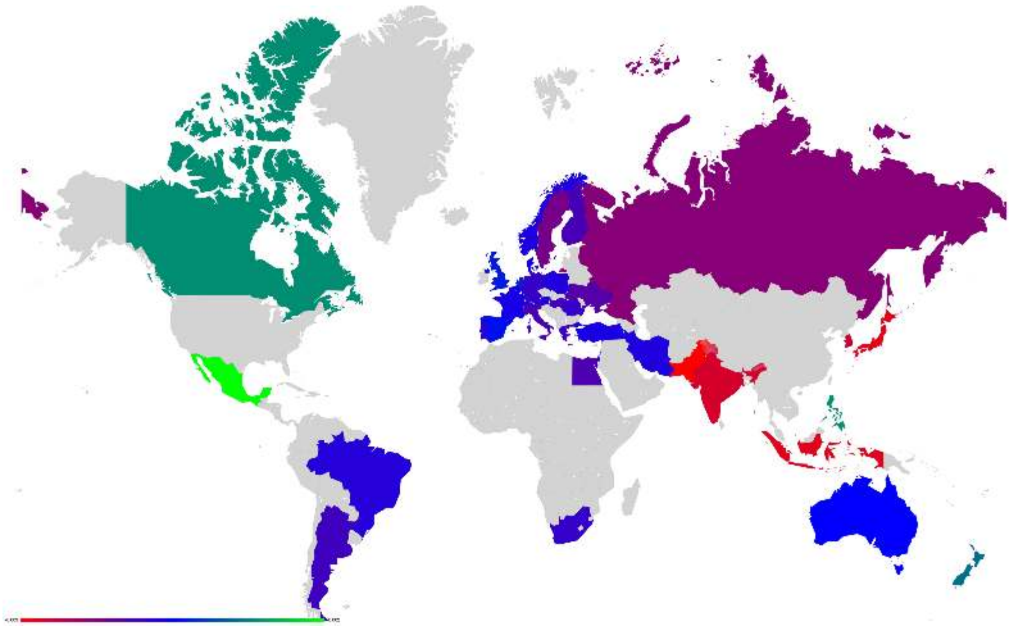

Figure 5.20: **Map representation of $\bar{D}$ for link modification from CN to US.** Non represented values: $\bar{D}(CN) = -0.0056$, $\bar{D}(US) = 0.0210$ and $\bar{D}(TW) = -0.0087$.

United Kingdom to France. The modification of this link gives the most strong effect on New Zealand (see Figs. 5.21, 5.22). Indeed, referring to New Zealand Ministry of Foreign Affairs and Trade [93], UK is the top destination for New Zealand's goods and services exports within the EU, and a base for New Zealand companies doing business in Europe. According to the statistics of March 2015, the total trade in goods between the two countries is $2,807 billion. New Zealand works closely with UK to face terrorism: strategic dialogue talks on security policy issues with UK are held every year. Also, New Zealand shares important cultural and historical links with UK. For New Zealand, UK is the key to Europe. This means intuitively that New Zealand will be strongly affected by the Brexit. These facts are totally in line with our sensitivity analysis conclusions plotted in

Fig 5.22 and 5.21. In order to face the consequences of Brexit together, UK and NZ have started a serious discussion as mentioned in [94, 95].

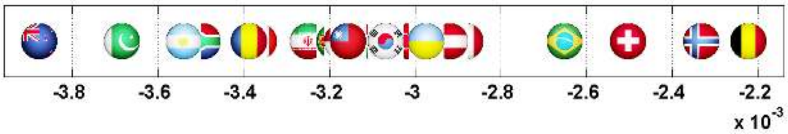

Figure 5.21: **Axial representation of $\bar{D}$ for link modification from GB to FR.** Non represented values: $\bar{D}(GB) = -0.00403$ and $\bar{D}(FR) = 0.0368$.

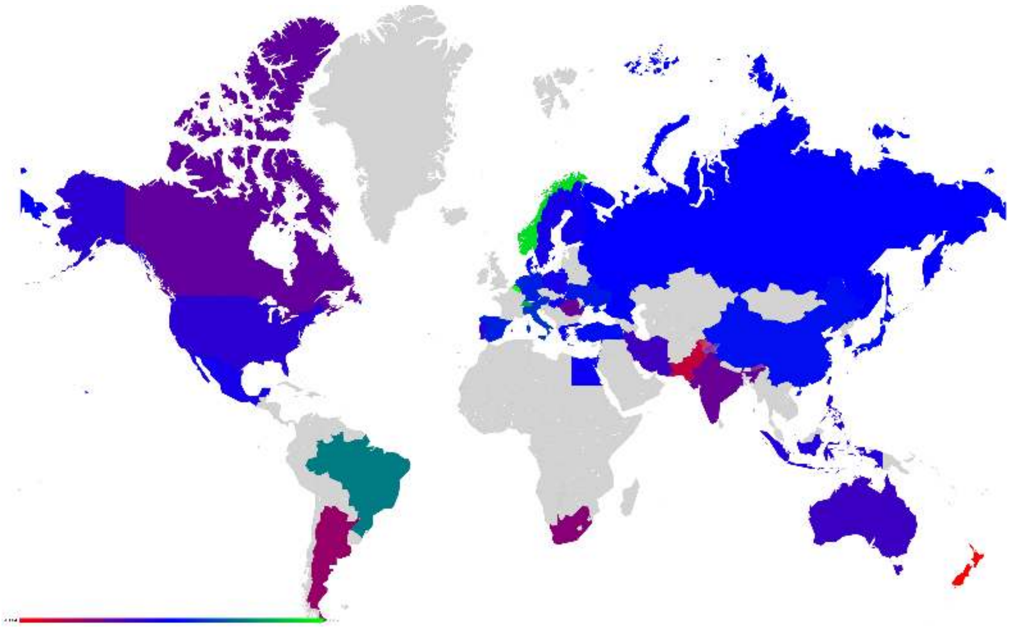

Figure 5.22: **Map representation of $\bar{D}$ for link modification from GB to FR.** Non represented values: $\bar{D}(GB) = -0.00403$ and $\bar{D}(FR) = 0.0368$.

US-Israel-Egypt. The Arab-Israeli relationship has been conflicting ever since the Jewish community has shown interest in establishing a nation-state in Palestine. The 1917 Balfour Declaration favored the establishment of a Jewish national home in Palestine and US supported it [96]. On November 29 1947, the United Nations General Assembly adopted the partition resolution number 181 [97] that would divide Palestinian territory into Jewish and Arab states. Again, US stood aside Israel in supporting the United Nations resolution. Palestinians (and Arabs in general) denounced the partition. Since then, Arab-Israeli did combat in five major wars (1948, 1956, 1967, 1973 and 1982) with Egypt the leader of Arab side in 3 out of 5 wars.

Even though the Camp David Accords [98] between Egypt and Israel were signed on September 17, 1978 followed by a peace treaty on March 26, 1979 [99] (both being signed in US and witnessed by Jimmy Carter), the relationship is still conflicting. It has been called the "cold peace". On the other side, Israeli-US relations are getting stronger according to Jeremy M. Sharp [100]: Israel is the largest cumulative recipient of US foreign aid since World War II. Our results show in Fig 5.23 and 5.24 that Egypt and Israel will be the most affected countries if the other one gets closer to US.

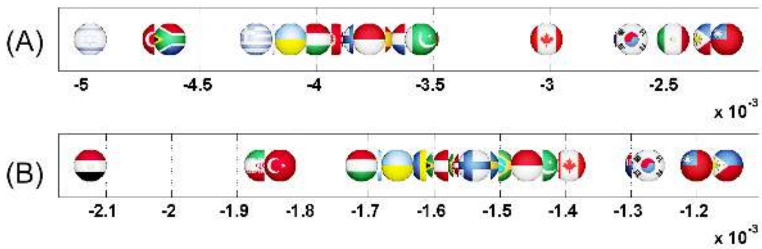

Figure 5.23: **Axial representation of $\bar{D}$ for links modifications from {IL and EG} to US.** (A):EG to US (Non represented values: $\bar{D}(EG) = -0.0080$ and $\bar{D}(US) = 0.0252$). (B):IL to US (Non represented values: $\bar{D}(IL) = -0.0041$ and $\bar{D}(US) = 0.0108$).

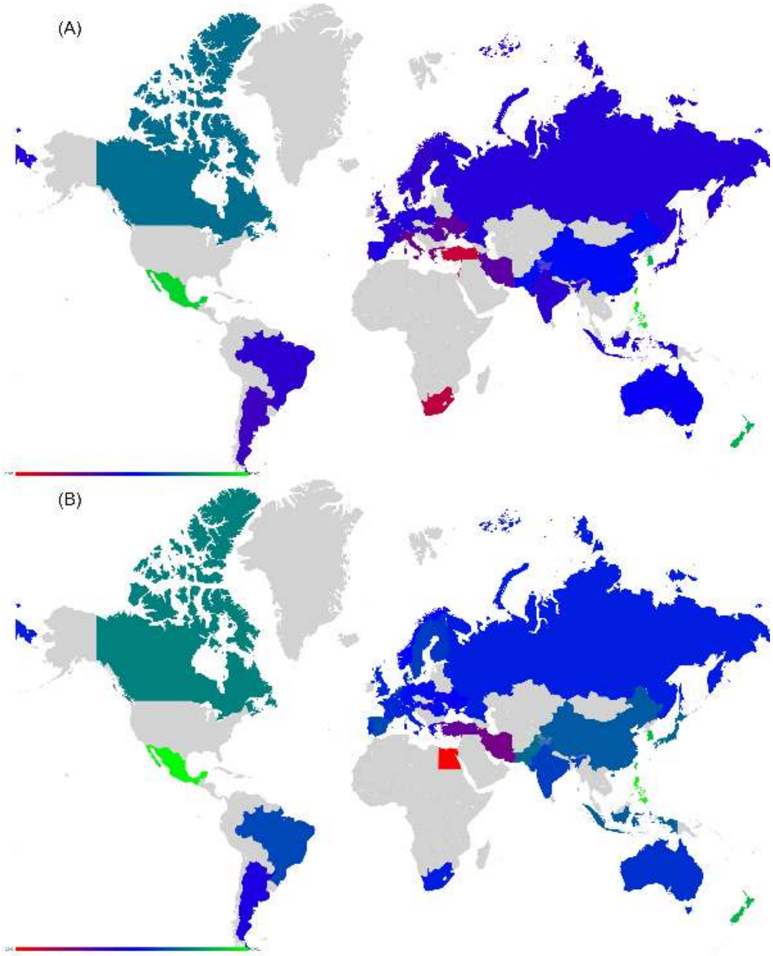

Figure 5.24: **Map representation of $\bar{D}$ for link modifications from {IL and EG} to US.** (A):EG to US. (B):IL to US. Values for EG, US, and IL are not represented.

Argentina and Brazil Their relationship [101] includes all possible fields: economy, history, culture, trade and social structure. As members of the Mercosur sub-regional bloc, Argentina and Brazil relationship offers free trade and fluid movement of goods, people, and currency. Besides that, a Nuclear Cooperation between these two countries was signed on July 18, 1991 and the Brazilian-Argentine Agency for Accounting and Control of Nuclear Materials (ABACC) was created as a binational safeguard organization. Comparing our results (shown in Fig 5.25 and 5.26) with these facts of strong relationship between Argentina and Brazil, we find that any unilateral rapprochement between Argentina or Brazil to US will negatively affect the other country.

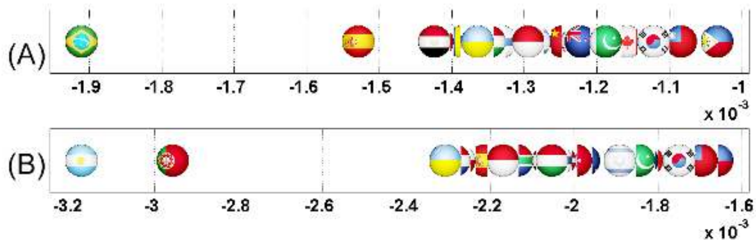

Figure 5.25: **Axial representation of $\bar{D}$ for link modifications from {AR and BR} to US.** (A): AR to US (Non represented values: $\bar{D}(AR) = -0.0050$ and $\bar{D}(US) = 0.0094$). (B): BR to US (Non represented values: $\bar{D}(BR) = -0.0074$ and $\bar{D}(US) = 0.0149$).

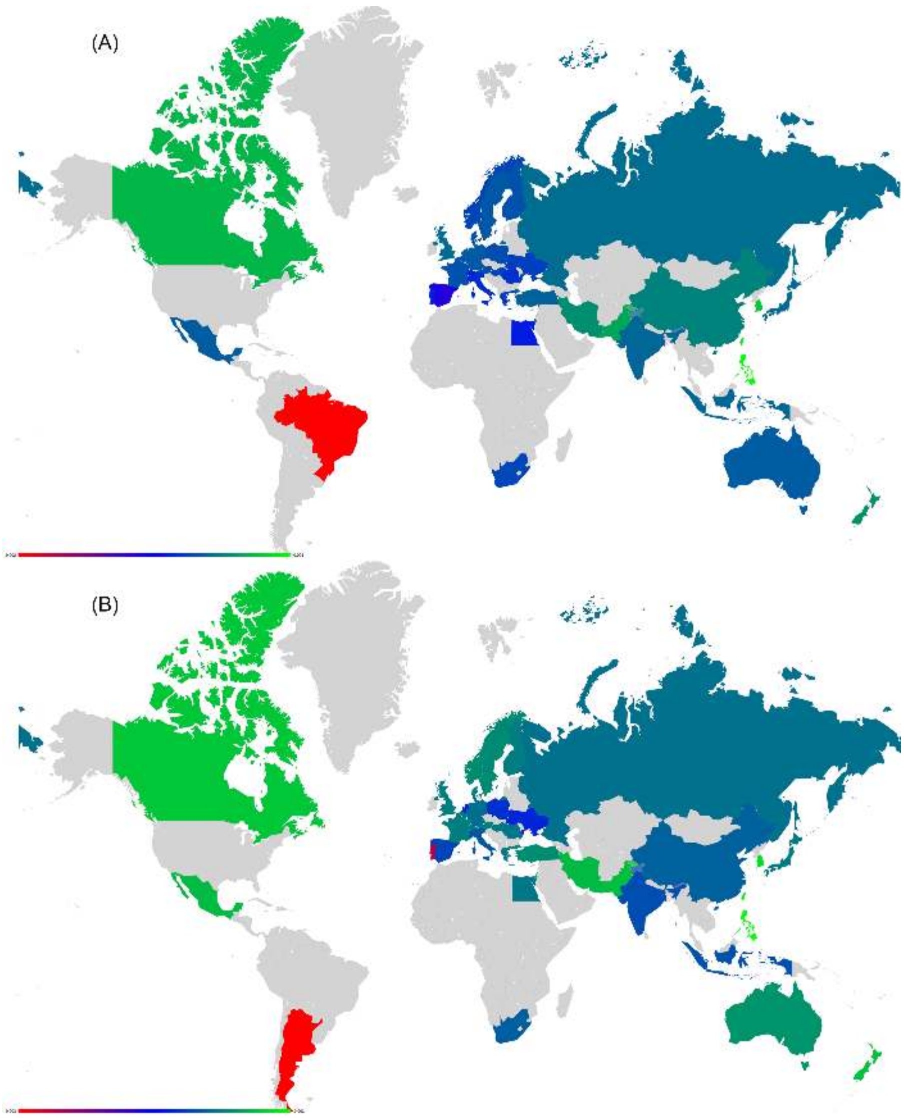

Figure 5.26: **Map representation of $\bar{D}$ for link modifications from {AR and BR} to US.** (A): AR to US. (B): BR to US. $\bar{D}$ values for AR, BR and US are not represented.

5.5.2.2 Relationship imbalance analysis

Relationship imbalance analysis has been derived for all pairs of 40 countries following Eq (5.3) as well. Fig 5.27 shows a density plot of $F(a, b)$. US is clearly the dominant country among all other 39 countries chosen worldwide. Also, Fig 5.27 shows that some countries have a strong influences such as France, Germany, Russia, China and Egypt. Germany and France are the two main players of European Union. Russia has an long history of sovereignty over eastern Europe and Northern Asia, economically, politically and culturally. Egypt plays a central role in the middle east. China, with its large population and strong economy, is dominating several countries. However, its role may be underestimated since no Chinese Wikipedia edition is accounted for in our study.

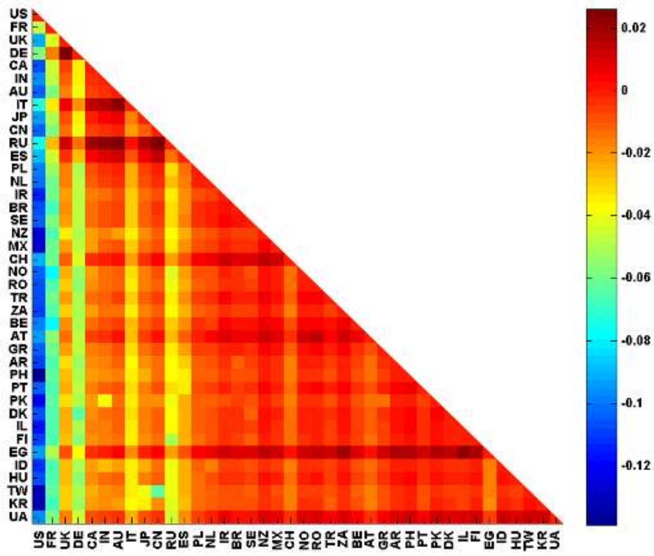

Figure 5.27: **Relationship imbalance analysis: F-representation for 27 EU network.** $F(a, b)$ is given by the colorbar. X-axis and Y-axis represent a and b respectively. If $F(a, b)$ is negative, nation a has more influence on nation b than b on a.

5.6 Influence of painters on countries.

5.6.1 Datasets

Another study is presented here to visualize the influence of painters on countries. To analyze the relation between painters and countries of the world we construct a reduced Google matrix with $N_r = 80$ nodes composed of the top 40 painters shown in Table 5.4 and the group of 40 countries listed in Table 4.1 of Chapter 3. The painters are the ones having top $\Theta-$score for $E = 7$: EnWiki, FrWiki, RuWiki, DeWiki, ItWiki, EsWiki and NlWiki. Hence, in Table 5.4 we give only short names, however, the full painter names are available in Appendix A with the complete list of painters with their $\Theta-$score, birth country and life period are available as well in [102]. The top 40 countries of EnWiki are presented in Table 5.5. The names of countries are given by ISO 3166-1 alpha-2 code (see [103]).

Θ rank	**K_{av} rank**	**Painter**	**Θ rank**	**K_{av} rank**	**Painter**
1	1	Vinci	21	18	Bondone
2	2	Picasso	22	25	Kandinsky
3	6	Gogh	23	19	Botticelli
4	4	Rijn	24	21	Caravaggio
5	5	Rubens	25	23	Velázquez
6	8	Durer	26	30	Degas
7	9	Titian	27	26	Bruegel Eld
8	11	Monet	28	29	Dyck
9	12	Dali	29	28	Renoir
10	14	Cézanne	30	31	Chagall
11	3	Michelangelo	31	33	Lautrec
12	7	Raphael	32	27	Vermeer
13	10	Goya	33	36	Poussin
14	13	Vasari	34	37	Turner
15	16	Matisse	35	38	Braque
16	15	Warhol	36	32	Blake
17	17	Delacroix	37	34	Greco
18	22	Manet	38	39	Miró
19	20	David	39	35	Munch
20	24	Gauguin	40	40	Eyck

Table 5.4: Top 40 painters

Table 5.5: List of PageRank of top 40 countries in EnWiki

Order	**Country**	**Order**	**Country**
1	US	21	NO
2	FR	22	RO
3	GB/UK	23	TK
4	DE	24	ZA
5	CA	25	BE
6	IN	26	AT
7	AU	27	GR
8	IT	28	AR
9	JP	29	PH
10	CN	30	PT
11	RU	31	PK
12	ES	32	DK
13	PL	33	IL
14	NL	34	FI
15	IR	35	EG
16	BR	36	ID
17	SE	37	HU
18	NZ	38	TW
19	MX	39	KR
20	CH	40	UA

5.6.2 Networks of painters and countries

We have for the following top three painters:

1. Leonardo da Vinci with $\Theta = 698$ (Italy),
2. Pablo Picasso with $\Theta = 688$ (Spain),
3. Vincent Van Gogh with $\Theta = 656$ (Netherlands).

The following painters are the most important one for their country of birth:

- Peter Paul Rubens for Germany with $\Theta = 651$ (but worked mainly in Netherlands),
- Claude Monet for France with $\Theta = 605$,
- Wassily Kandinsky for Russia with $\Theta = 515$,
- Joseph Mallord William Turner for United Kingdom (UK or GR) with $\Theta = 386$.

The top 6 countries with the largest number of painters from the global list of 223 painters are Italy (50), France (45), Russia (27), Germany (26), USA (14), Spain (11) (the 223 are given in Section 3.2 of Chapter 3 and in Appendix A).

The geopolitical relations between painters and countries has been analyzed more precisely for EnWiki, FrWiki and DeWiki data. Therefore, we have plotted a network of friendship between our 40 painters and their top 3 most friendly countries. This network has been calculated using G_{rr} and G_{qrnd} calculated for the union of 40 painters and 40 countries. For each painter column, we select the top 3 countries in the sum matrix $G_{rr} + G_{\mathrm{qrnd}}$ to account for direct and indirect interactions and mitigate the effect of the projector component. Resulting networks are shown in Figure 5.28, 5.29 and 5.30 for EnWiki, FrWiki and DeWiki, respectively. In these figures, arrows are colored in red if $G_{\mathrm{qrnd}}(i,j) > G_{rr}(i,j)$ and in black other wise. The network structure is different for each edition due to different cultural views and preferences. However, the central role of France and Italy is well visible in all 3 editions.

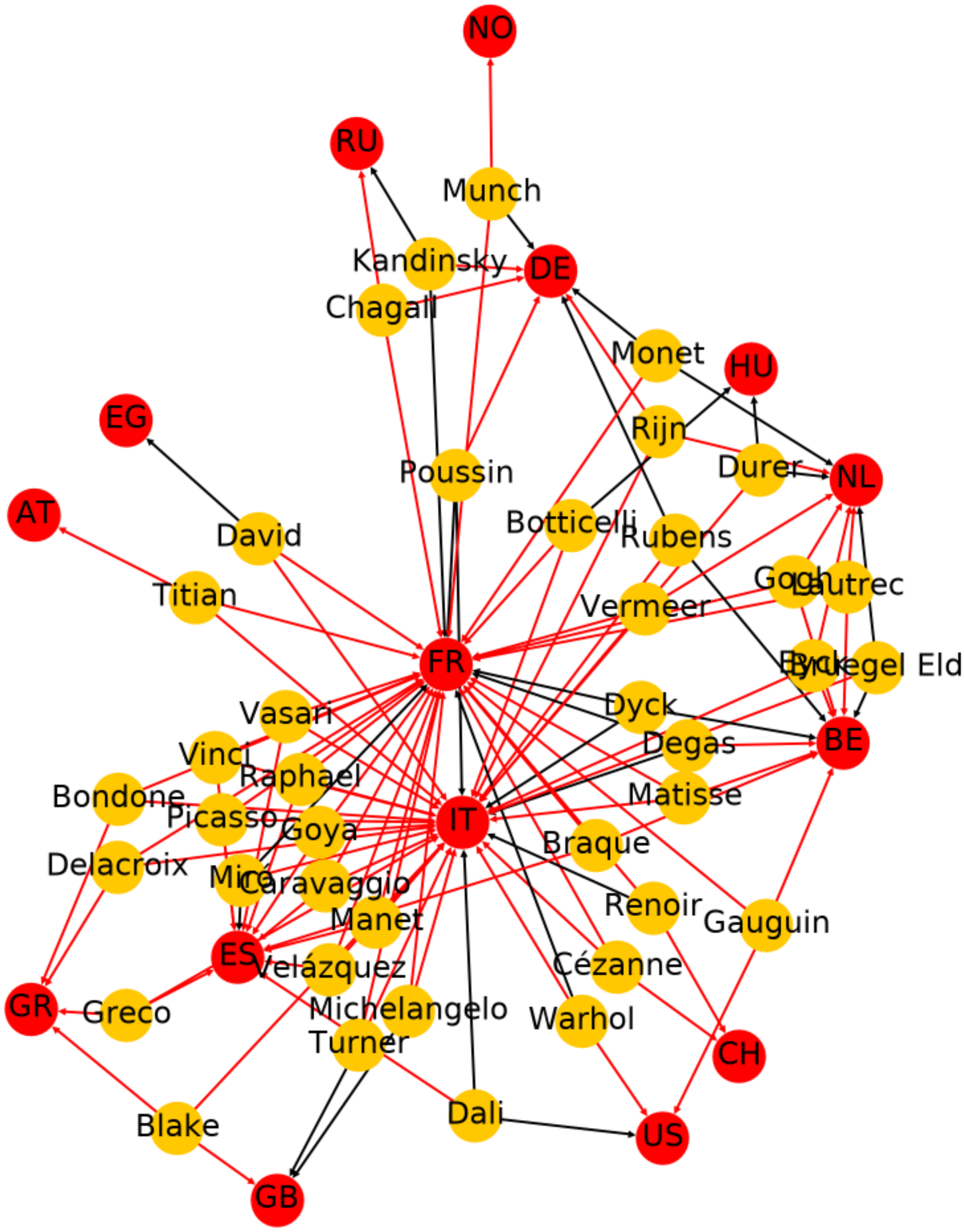

Figure 5.28: **Network structure of top 3 country friends for top 40 painter network for EnWiki** Painters are selected from the global rank list of 7 Wikipedia editions from Table 5.4 for top 40 PageRank countries of EnWiki from Table 5.5. Arrows are showing links only from a painter to top 3 countries, they are given by links of matrix elements $G_{rr} + G_{\mathrm{qrnd}}$, red arrow mark links when an element G_{qrnd} is larger than element G_{rr}, black arrows are drown in opposite case. Countries and shown by red circles and painters are shown by yellow circles.

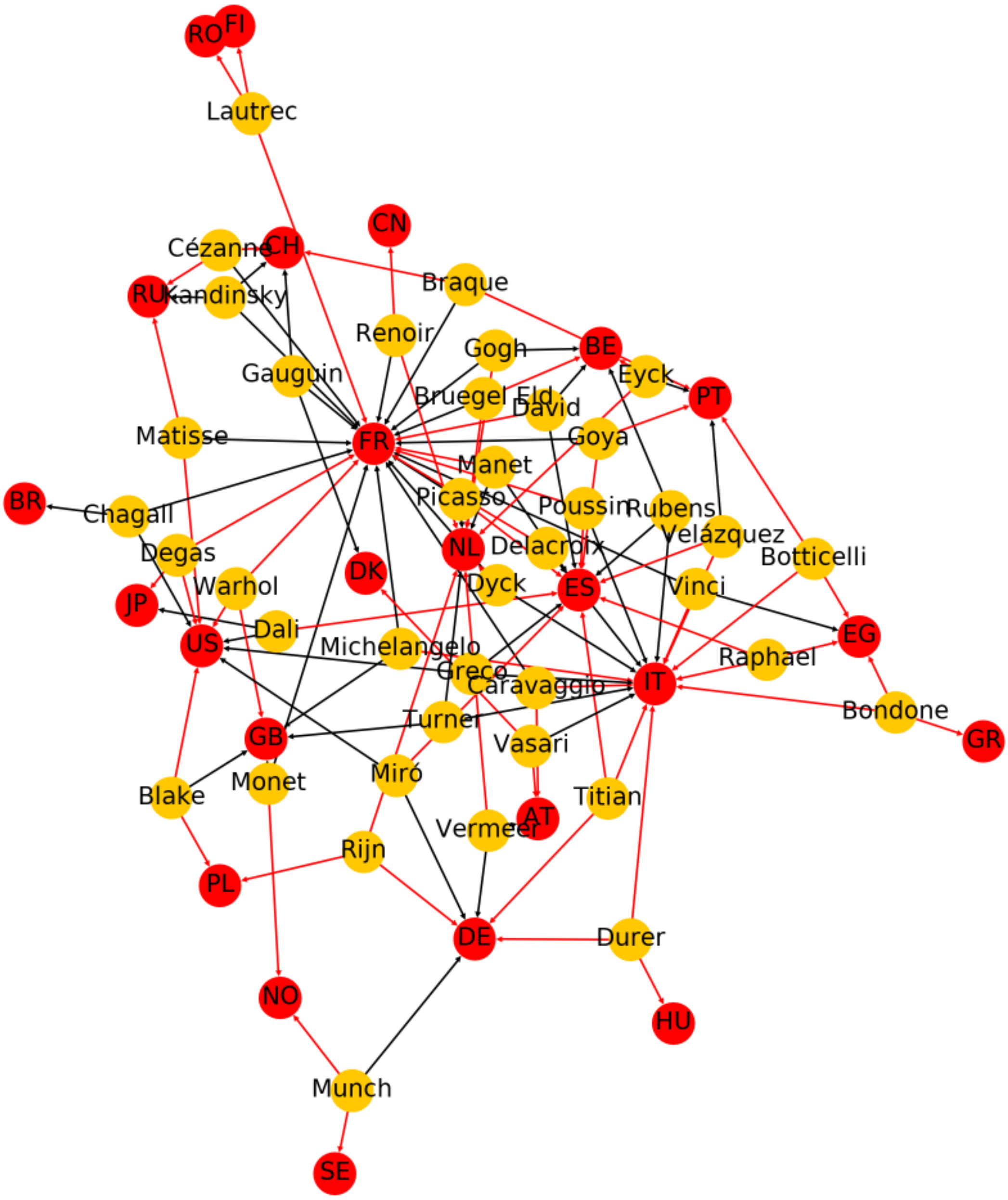

Figure 5.29: **Network structure of top 3 country friends for top 40 painter network for FrWiki** Painters are selected from the global rank list of 7 Wikipedia editions from Table 5.4 for top 40 PageRank countries of EnWiki from Table 5.5.

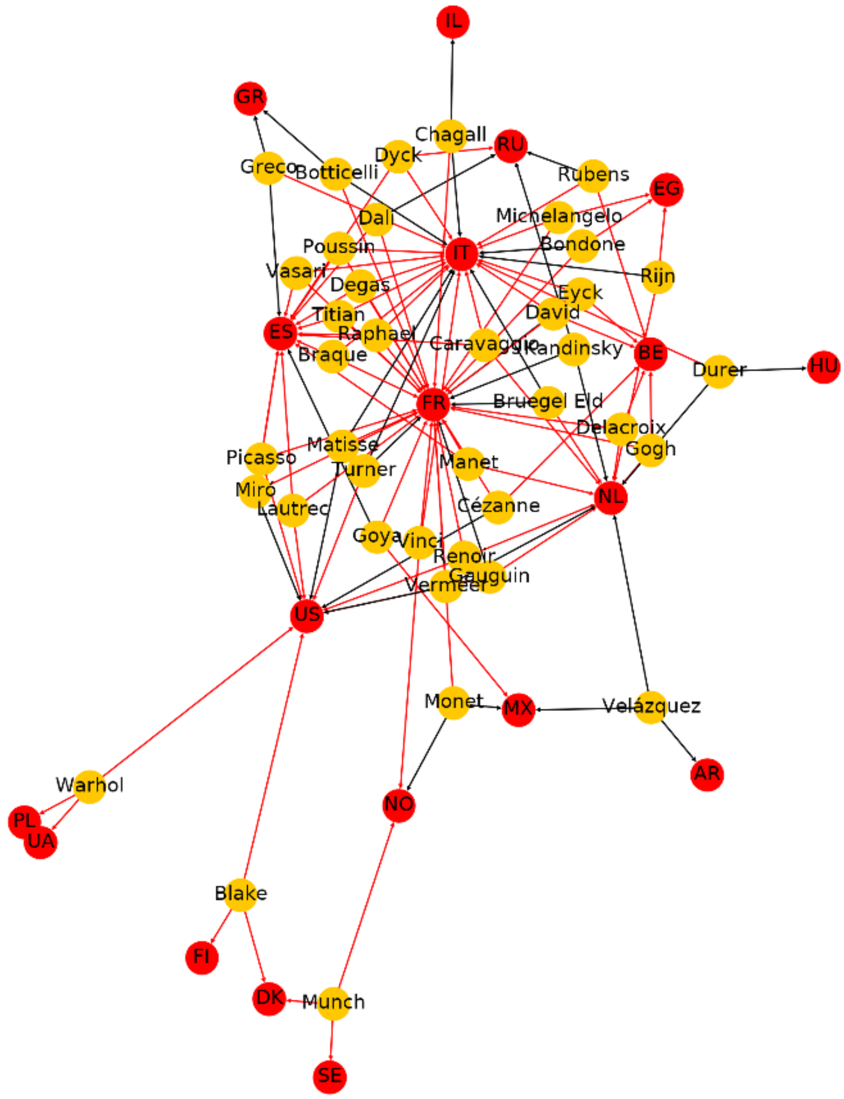

Figure 5.30: **Network structure of top 3 country friends for top 40 painter network for DeWiki** Painters are selected from the global rank list of 7 Wikipedia editions from Table 5.4 for top 40 PageRank countries of EnWiki from Table 5.5.

5.6.3 Influence of painters on countries

To analyze in a more direct way the world influence of painters we average G_{R} matrix and its three components G_{pr}, G_{rr}, G_{qr} over 7 Wikipedia editions that allows us to account for different cultural views on selected 40 painters of Table 5.4 and 40 countries of Table 5.5. The reduced Google matrix $G_{\mathrm{R}av}$ averaged over different editions allows to obtain a balanced view of various cultural opinions of Wikipedia language editions for a selected group of nodes representing Wikipedia articles. We determine the PageRank probability of this averaged G_{R} matrix and compute its logarithmic derivative (sensitivity) in respect a weight variation of a certain link from a given painter to a given country.

Influence of Picasso on Spain and France. Figure 5.31 shows the sensitivity D of 40 world countries with respect to a link variation from Picasso to Spain (top panel) and from Picasso to France (bottom panel). Pablo Picasso, the son of the Spanish painter Don José Ruiz y Blanco, was born in Spain in 1881. Pablo began painting since he was eight, and in 1896, he has joined the art and design school of Barcelona "Escola de la Llotja". In 1904, Picasso married Fernande Olivier a French artist and model. Since that, Picasso spent most of his life in France and died there at 92 years old. This could explain the results we have obtained from our sensitivity analysis, which shows that France and Spain are the most countries affected for a link variation between Picasso-Spain and Picasso-France respectively.

Influence of Van Gogh on the Netherlands and France. Figure 5.32 shows the sensitivity D of 40 world countries in respect to link variation from Van Gogh to Netherlands and from da Vinci to France in top and bottom panels respectively. Even-thought Van Gogh has only spent the last four years of his life in different places in France, these years were important to Van Gogh's painting career. Van Gogh has built strong relationships with leading French painters. Van Gogh has worked with Emile Bernard, Henri de Toulouse-Lautrec, Georges Seurat, Paul Signac and Gauguin. These relationships and the work achieved by Van Gogh in France explain our results in the top panel of Figure 5.32, which shows that France is mostly influenced for a link variation from Van Gogh to Netherlands.

Influence of Leonardo Da Vinci on Italy and France. The Italian painter Leonardo Da Vinci learned painting in the workshop of Verrochio in Florence, and crafted there its first painting between 1472 and 1474. Da Vinci was based in Italy until 1516, when Francois I (King of France) invited him to join the Royal court as: "The King's First Painter, Engineer and Architect". Da Vinci died in France four years after his arrival. Da Vinci's works was highly noted by French statesmen. Louis XII and (later)

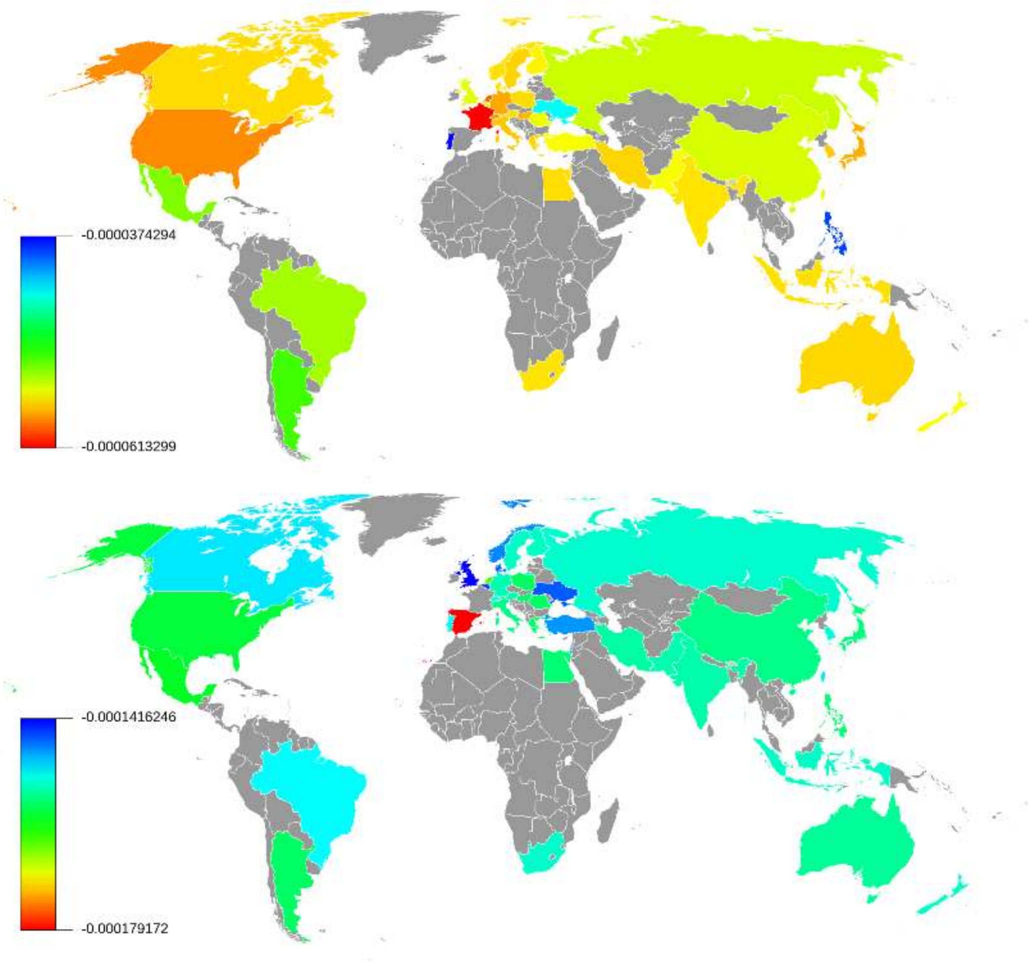

Figure 5.31: **Sensitivity D of 40 world countries to the link Picasso-Spain and Picasso-France** Top panel: Picasso-Spain and bottom panel: Picasso-France. Data is averaged over 7 Wikipedia editions. For a better visibility sensitivity of Spain (top) and France (bottom) are given in Figure 5.33.

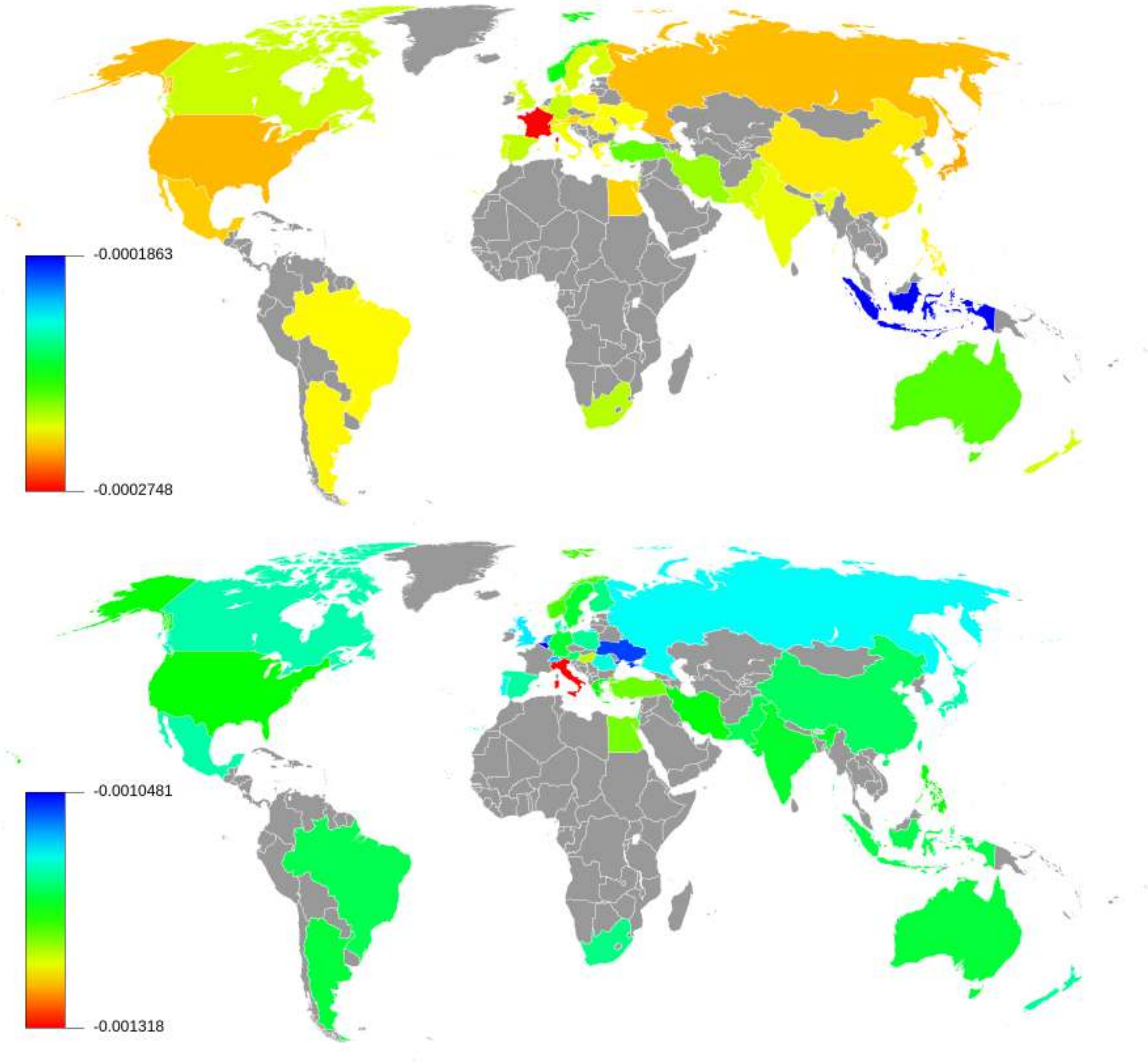

Figure 5.32: **Sensitivity D of 40 world countries to the link Van Gogh-Netherlands and da Vinci-France.** Top panel: Van Gogh-Netherlands and bottom panel: da Vinci-France. Data is averaged over 7 Wikipedia editions. For a better visibility sensitivity of Netherlands (top) and France (bottom) are given in Figure 5.33.

Napoleon though to bring “The Last Supper" to France. “Madonna of the Yarnwinder" is a painting done by Da Vinci to respond the demand of the secretary of state of Louis XII of France. Leonardo brought a version of the "Virgin of the Rocks" to France. One of the most important painting of Da Vinci is "Mona Lisa", currently displayed at Louvre Museum in Paris, was finalized in the Royal court of Francois I. All these elements about the relations between Da Vinci, France and Italy, explain the fact that Italy is mostly influenced for a link variation between Da Vinci and France, as shown in the bottom panel of Figure 5.32.

Diagonal sensitivity of countries Finally in Figure 5.33 we present the diagonal sensitivity of countries to their links with painters. This measure is computed by calculating the 2-way sensitivity of Eq. (5.2) for each painter/country couple (i. e. the sum of the logarithmic sensitivity for the painter to country link and the country to painter link). Thus D is computed as a sum of logarithmic derivatives of PageRank probability variation of a given country when its links with a given painter are varied (in both directions). We see that the strongest influence on countries are produced on average by da Vinci, Picasso and Michelanglo.

5.7 Conclusion

This chapter shows that our sensitivity analysis captures the importance of relationships on network structure. This analysis relies on the reduced Google matrix and leverages its capability of concentrating all Wikipedia knowledge in a small stochastic matrix. We stress that the obtained sensitivity of geopolitical relations between two countries and its influence on other world countries is obtained on a pure mathematical analysis without any direct appeal to political, economical and social sciences. Also, the analysis of the reduced Google matrix of top 40 painters and top world countries allows us to determine the world influence of painters on different countries.

Studies on painters and countries have helped us to understand, manipulate and assess the properties of G_{R} and its components. We've been able to find well-known facts for both sets, showing that reduced Google matrix analysis has good potential for extracting correct knowledge from Wikipedia. In the next chapter, we will apply our sensitivity analysis to another case study that has been chosen for extracting non-obvious knowledge from Wikipedia. We will try to see the influence between terrorist groups and countries.

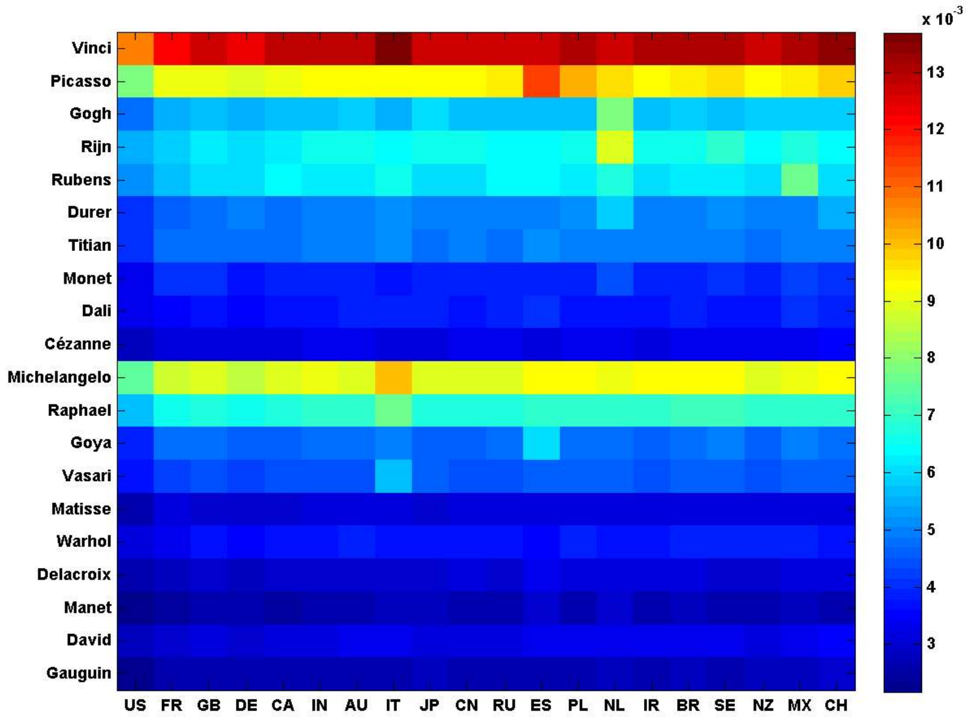

Figure 5.33: **Sensitivity D represents the sum of logarithmic sensitivity value of country for two direction (painter to country) and (country to painter)** Color bar shows the sensitivity values. Data is averaged over 7 Wikipedia editions and are shown for top 20 entries of Table 5.4 and Table 5.5.

Chapter 6

Analysis of world terror networks

6.1 Introduction

"A new type of terrorism threatens the world, driven by networks of fanatics determined to inflict maximum civilian and economic damages on distant targets in pursuit of their extremist goals" [104]. The origins of this world wide phenomenon are under investigation in political, social and religious sciences (see e.g. [104–107] and references therein).

At the same time the number of terrorist groups is growing in the world [108] reaching over 100 officially recognized groups acting in various countries of the world [3]. These numbers become quite large and the mathematical analysis of multiple interactions between these groups and their relationships to world countries is getting of great timeliness. The first steps in this direction are reported in a few publications (see e.g. [109,110]) showing that the network science methods (see e.g. [111]) should be well adapted to such type of investigations. However, it is difficult to obtain a clear network structure with all dependencies which are emerging from the surrounding world with all its complexity.

Here, for the English Wikipedia network, we apply the reduced Google matrix method to analyze interactions between a subset of $N_g = 95$ terrorist groups referenced in Wikipedia articles. Only groups enlisted as terrorist groups by at least two countries (cf. [3]) are selected. Our list of 95 terrorist groups is given in Table 6.1.

In addition, we select the group of $N_c = 64$ related world countries given in Table 6.2. This gives us the size of G_{R} being $N_r = N_g + N_c = 159$ that is much smaller then the global Wikipedia network of $N = 5\ 416\ 537$ nodes (articles) and $N_\ell = 122\ 232\ 932$ links generated by quotation links from one article to another. The obtained results for interactions between terrorist groups and countries are described next .

Using the sensitivity of PageRank to a weight variation of specific links we determine the geopolitical sensitivity and influence of specific terrorist groups on world countries. The world maps of the sensitivity of various countries to influence of specific terrorist groups are obtained this way. Here we present results for English Wikipedia edition only but we think that the extension of this research to other editions will be of significant interest in the future.

Name	KG	Color	Name	KG	Color
Islamic State of Iraq and the Levant	1	BL	Hezb-e Islami Gulbuddin	49	RD
Al-Qaeda	2	BL	Kach and Kahane Chai	50	BK
Taliban	3	RD	Palestine Liberation Front	51	OR
Provisional Irish Republican Army	4	BK	Harkat-ul-Mujahideen	52	RD
Hamas	5	OR	Kurdistan Free Life Party	53	BK
Hezbollah	6	OR	Indian Mujahideen	54	RD
Muslim Brotherhood	7	BL	Abu Nidal Organization	55	OR
Liberation Tigers of Tamil Eelam	8	RD	Hizbul Mujahideen	56	RD
Kurdistan Workers' Party	9	BK	Libyan Islamic Fighting Group	57	GN
Al-Shabaab (militant group)	10	GN	Islamic State of Iraq and the Levant in Libya	58	GN
ETA (separatist group)	11	BK	Revolutionary People's Liberation Party/Front	59	BK
FARC	12	BK	Al-Mourabitoun	60	GN
Houthis	13	PK	Revolutionary Organization 17 November	61	BK
Al-Nusra Front	14	PK	Holy Land Foundation for Relief and Development	62	OR
Boko Haram	15	GN	Ansar al-Sharia (Libya)	63	GN
Ulster Volunteer Force	16	BK	Al-Itihaad al-Islamiya	64	GN
Shining Path	17	BK	Al-Haramain Foundation	65	BL
Popular Front for the Liberation of Palestine	18	OR	Ansar Bait al-Maqdis	66	PK
Lashkar-e-Taiba	19	RD	Ansaru	67	GN
Hizb ut-Tahrir	20	BL	Babbar Khalsa	68	BL
Al-Qaeda in the Arabian Peninsula	21	PK	Jamaat-ul-Mujahideen Bangladesh	69	RD
Tehrik-i-Taliban Pakistan	22	RD	Force 17	70	OR
Islamic Jihad Mov. in Palestine	23	OR	Kata'ib Hezbollah	71	PK
Ulster Defence Association	24	BK	Kurdistan Freedom Hawks	72	BK
Abu Sayyaf	25	RD	Islamic Jihad Union	73	RD
Real Irish Republican Army	26	BK	Abdullah Azzam Brigades	74	PK
Ansar Dine	27	GN	Moroccan Islamic Comb. Group	75	GN
Jemaah Islamiyah	28	RD	Ansar al-Sharia (Tunisia)	76	GN
Al-Qaeda in the Islamic Maghreb	29	GN	Al-Qaeda, Indian Subcontinent	77	RD
Egyptian Islamic Jihad	30	PK	Jund al-Aqsa	78	PK
Al-Jama'a al-Islamiyya	31	PK	Hezbollah Al-Hejaz	79	PK
Jaish-e-Mohammed	32	RD	Jamaat-ul-Ahrar	80	RD
Aum Shinrikyo	33	RD	Jamaah Ansharut Tauhid	81	RD
United Self-Defense Forces of Colombia	34	BK	Islamic State of Iraq and the Levant Algeria Province	82	GN
Armed Islamic Group of Algeria	35	GN	Osbat al-Ansar	83	PK
Continuity Irish Republican Army	36	BK	International Sikh Youth Federation	84	RD
Movement for Oneness and Jihad in West Africa	37	GN	East Turkestan Liberation Organization	85	RD
Quds Force	38	PK	Great Eastern Islamic Raiders' Front	86	BK
Al-Aqsa Martyrs' Brigades	39	OR	Aden-Abyan Islamic Army	87	PK
Com. Party of the Philippines	40	RD	Al-Aqsa Foundation	88	OR
Caucasus Emirate	41	RD	Khalistan Zindabad Force	89	RD
Haqqani network	42	RD	Mujahidin Indonesia Timur	90	RD
Turkistan Islamic Party	43	RD	Al-Badr	91	RD
Ansar al-Islam	44	PK	Soldiers of Egypt	92	PK
Izz ad-Din al-Qassam Brigades	45	OR	National Liberation Army	93	BK
Lashkar-e-Jhangvi	46	RD	Jundallah	94	RD
Harkat-ul-Jihad al-Islami	47	RD	Army of Islam	95	PK
Islamic Movement of Uzbekistan	48	RD			

Table 6.1: **List of selected terrorist groups attributed to 6 categories marked by color.** Source: [3]. *KG* gives the local PageRank index of terrorist groups.

Rank	Name	abr	Rank	Name	abr
1	United States	US	33	Portugal	PT
2	France	FR	34	Ukraine	UA
3	Germany	DE	35	Czech Republic	CZ
4	United Kingdom	GB	36	Malaysia	MY
5	Iran	IR	37	Thailand	TH
6	India	IN	38	Vietnam	VN
7	Canada	CA	39	Nigeria	NG
8	Australia	AU	40	Afghanistan	AF
9	China	CN	41	Iraq	IQ
10	Italy	IT	42	Bangladesh	BD
11	Japan	JP	43	Syria	SY
12	Russia	RU	44	Morocco	MA
13	Spain	ES	45	Algeria	DZ
14	Netherlands	NL	46	Saudi Arabia	SA
15	Poland	PL	47	Lebanon	LB
16	Sweden	SE	48	Kazakhstan	KZ
17	Mexico	MX	49	Albania	AL
18	Turkey	TR	50	United Arab Emirates	AE
19	South Africa	ZA	51	Yemen	YE
20	Switzerland	CH	52	Tunisia	TN
21	Philippines	PH	53	Jordan	JO
22	Austria	AT	54	Libya	LY
23	Belgium	BE	55	Uzbekistan	UZ
24	Pakistan	PK	56	Kuwait	KW
25	Indonesia	ID	57	Qatar	QA
26	Greece	GR	58	Mali	ML
27	Denmark	DK	59	Kyrgyzstan	KG
28	South Korea	KR	60	Tajikistan	TJ
29	Israel	IL	61	Oman	OM
30	Hungary	HU	62	Turkmenistan	TM
31	Finland	FI	63	Chad	TD
32	Egypt	EG	64	South Sudan	SS

Table 6.2: **List of selected countries.**

6.2 Results

In this work we extract from G_{R} a network representation of the 64 countries and 95 groups selected. This network reflects direct and indirect interactions between countries and groups, which motivates us to study the relative influence of group alliances on the other ones and on the countries. The matrix G_{R} and its three components G_{rr}, G_{pr} and G_{qr} are computed for $N_r = 159$ Wikipedia network nodes formed by $N_c = 64$ country nodes and $N_g = 95$ group nodes. The weights of these three G_{R} components are W_{rr}=0.0644, W_{pr}=0.8769 and W_{qr}=0.0587 (we recall that the weight is given by the sum of all matrix elements divided by N_r, thus $W_{rr}+W_{\mathrm{pr}}+W_{\mathrm{qr}} = 1$).

The matrix elements of $G_{\mathrm{R}}, G_{rr}, G_{\mathrm{qr}}$ corresponding to the part of 95 terrorist groups are shown in the color maps of Figure 6.1 (indices are ordered by increasing values of KG as given in Table 6.1, thus element with KG1=KG1 is located at the top left corner). The largest matrix elements of G_{R} are the ones of top PageRank groups of Table 6.1.

According to Figure 6.1 the strong interactions between groups can be found by analyzing G_{qr} looking at new links appearing in G_{qr} and being absent from G_{rr}. As an example we list:

- Tehrik-i-Taliban Pakistan (KG22) and Jundallah (KG94);
- Hamas (KG5) and Izz ad-Din al-Qassam Brigades (KG45);
- Taliban (KG3) and Al-Qaeda in the Arabian Peninsula (KG21);
- Kurdistan Freedom Hawks (KG72) and Kurdistan Workers' Party (KG9).

6.2.1 Network structure of groups

To analyze the network structure of groups we attribute them to 6 different categories marked by 6 colors in Table 6.1:

- C1 for the International category of groups operating worldwide (color BL-blue, top group is KG1 ISIS) ;
- C2 for the groups targeting Asian countries (color RD-red, top group is KG3 Taliban) ;
- C3 for the groups related with the Israel-Arab conflict (color OR-orange, top group is KG5 Hamas) ;
- C4 for the groups targeting African countries (color GN-green, top group is KG10 Al-Shabaab) ;

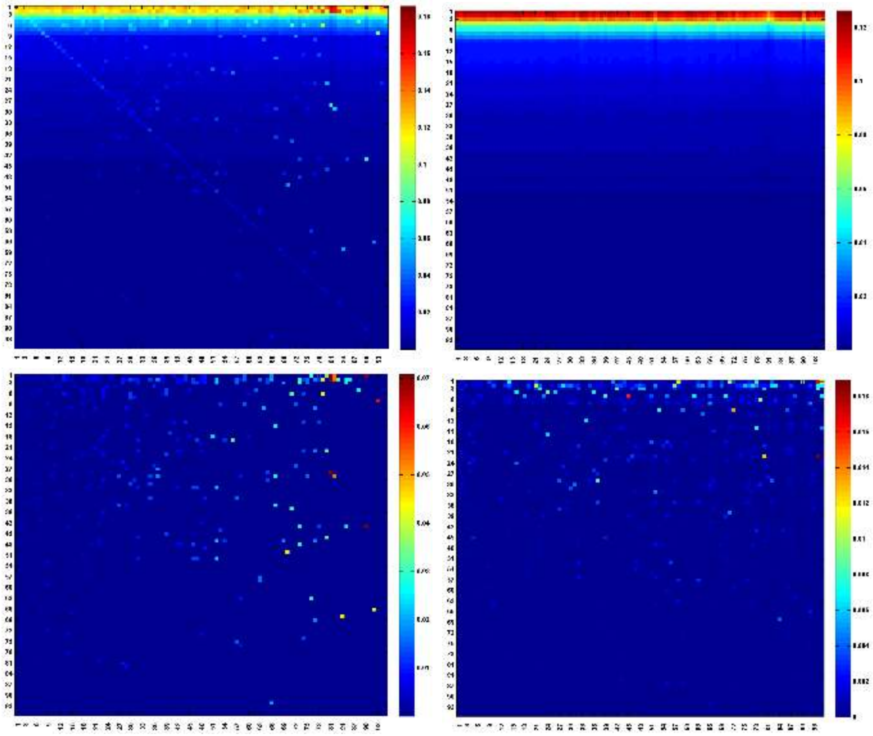

Figure 6.1: **Density plots of matrices G_{R}, G_{pr}, G_{rr} and G_{qrnd}.** Top left and right figures are G_{R} and G_{pr} respectively ; Bottom left and right figures are G_{rr} and G_{qrnd} respectively; color changes from red at maximum to blue at zero; Here we only plot the matrix for the 95 terrorist nodes of Table 6.1.

- C5 for the groups related to Arab countries at Middle East and the Arabian Gulf (color PK-pink, top group is KG13 Houthis) ;
- C6 for all remaining groups (color BK-black, top group is KG4 IRA).

These 6 categories of groups are related to their activity and their geographical location. Only the category C1 has global international activity, other categories have more local geographical activity. We will see that the network analysis captures these categories.

We order the terror groups by their local PageRank index KG in Table 6.1 (highest probability of PageRank vector for groups is at $KG = 1$, G stands for group number). The selected countries are ordered by their local PageRank index K in Table 4.1 (highest probability of PageRank vector for countries is at $K = 1$).

Network of terrorist groups We analyze the network structure of groups by selecting the top group node of each category in Table 6.1 and then, their top 4 friends in $G_{rr} + G_{\rm qrnd}$ (i.e. the nodes with the 4 largest matrix elements of $G_{rr} + G_{\rm qrnd}$ in the column representing the group of interest. It corresponds to the 4 largest outgoing link weights). From the set of top group nodes and their top 4 friends, we continue to extract the top 4 friends of friends until no new node is added to this network of friends. The obtained network structure of groups is shown in Fig 6.2. This network structure clearly highlights the clustering of nodes corresponding to selected categories. It shows the leading role of top PageRank nodes for each category appearing as highly central nodes with large in-degree. We note that we speak about networks of friends and followers using the terminology of social networks. Of course, this has only associative meaning (we do not mean that some country is a friend of terrorist group).

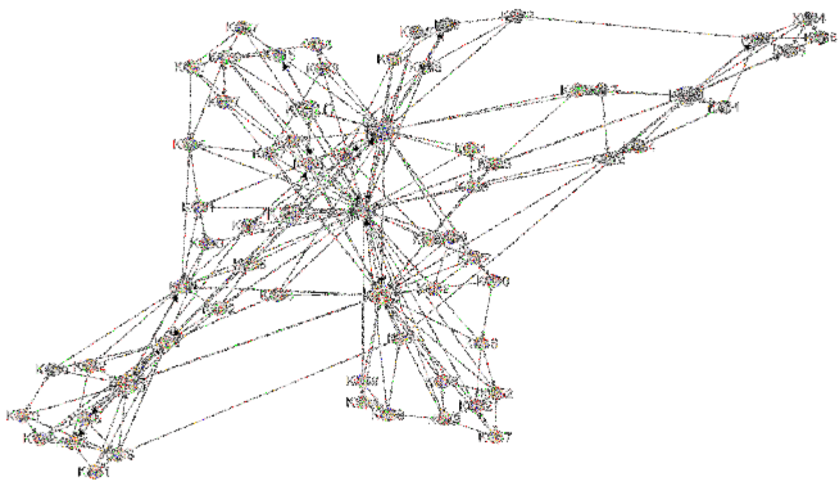

Figure 6.2: **Friendship network structure of terrorist groups obtained from $G_{\rm qrnd}+G_{rr}$.** Colors mark categories of nodes and top nodes are given in text and Table 6.1; circle size is proportional to PageRank probability of nodes; bold black arrows point to top 4 friends, gray tiny arrows show friends of friends interactions computed until no new edges are added to the graph (drawn with [1,2]).

The appearance of links due to indirect relationships between groups is confirmed by well-known facts. For instance, it can be seen that Al-Qaeda in the Arabian Peninsula (KG21) is linking Al-Shabaab (KG10) and Houthis (KG13). Al-Qaeda in the Arabian Peninsula is primarily active in Saudi Arabia. It is well known that Saudi Arabia is an important financial support of Al-Shabaab [112] and that Houthis is confronting Saudi Arabia. As such, it makes sense that Al-Qaeda in the Arabian Peninsula links both groups as it is tied to Saudi Arabia.

Another meaningful example is the one of Hezbollah (KG6) and Houthis that share the same ideology, since they are both Shiite and are strongly linked to Iran. From Figure 6.2, it can be seen that Hezbollah is a direct friend of Houthis. The case of Hamas (KG5) and Hezbollah, that share the same ideology in facing Israel, is highlighted as well in our results. Moreover, Figure 6.2 shows as well that Hezbollah is the linking group between Hamas and Houthis. Finally, the network of Figure 6.2 clearly shows that the groups that are listed as International (blue color) are clearly playing that role by having lots of ingoing links from the other categories.

6.2.2 Relationships between groups and countries

The interactions between groups and countries are characterized by the network structure shown in Figures 6.3 and 6.4. For clarity, we first show in Figure 6.3 the top 4 country friends of the 6 terrorist groups identified as leading each category. In Figure 6.4 we show for the same 6 leading terrorist groups the top 2 country friends and top 2 terrorist group friends. This latter representation shows altogether major ties between groups and countries and in-between groups. Very interesting and realistic relations between groups and countries can be extracted from this network. For instance, Taliban (KG3) is an active group in Afghanistan and Pakistan that represents an Islamist militant organization that was one of the prominent factions in the Afghan Civil War [108,113,114]. As shown in Figures 6.3, 6.4 Afghanistan and Pakistan are the countries that are the most influenced by Taliban.

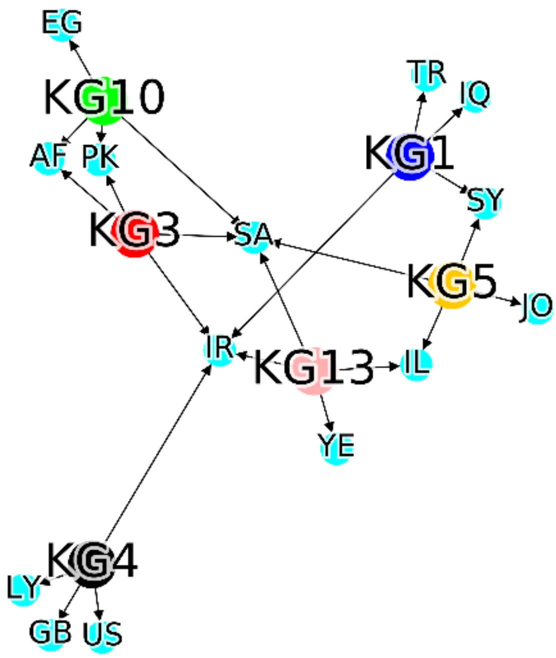

Figure 6.3: **Friendship network structure extracted from $G_{\text{qrnd}} + G_{rr}$ with the top 4 countries per leading terrorist group.** The leading terrorist groups are marked by their respective colors and countries are marked by cyan color. The network structure is construction with the leading terrorist groups of each category and its top 4 friend countries. Networks are drawn with [1, 2].

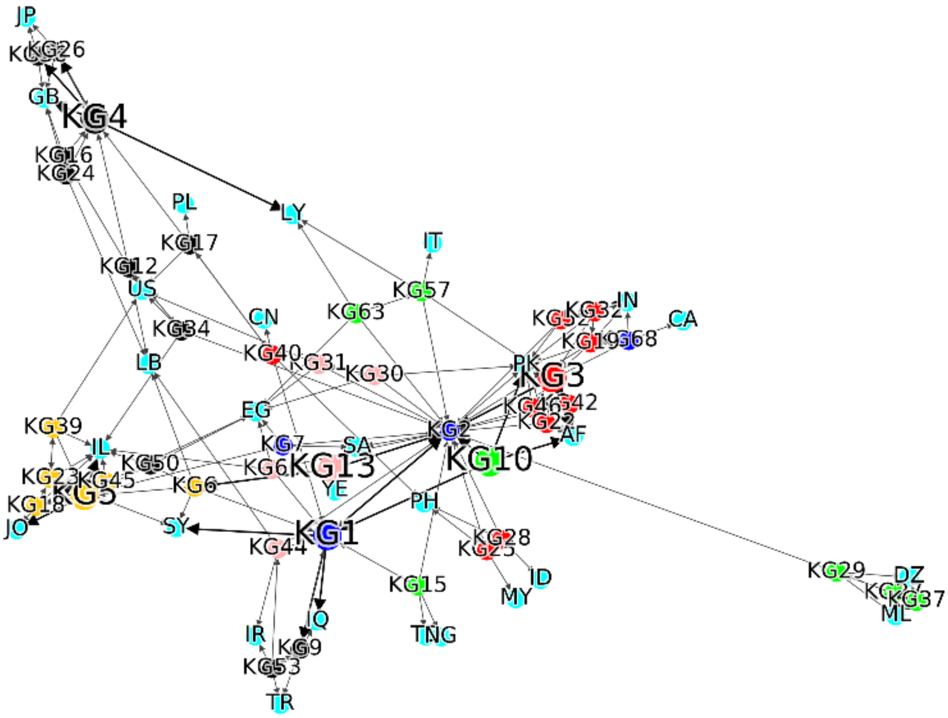

Figure 6.4: **Friendship network structure extracted from $G_{\text{qrnd}} + G_{rr}$ with the top 2 terrorist groups and top 2 countries.** The top terrorist groups are marked by their respective colors and countries are marked by cyan color. The network structure is construction with the leading terrorist groups of each category and its top 2 friend countries and top 2 terrorist groups. Networks are drawn with [1, 2].

The fact that Saudi Arabia links Houthis, Taliban and Al Shabaab can be explained by the fact that Saudi Arabia is in war with Houthis [115,116]. Also, the main funding sources for groups active in Afghanistan and Pakistan originate from Saudi Arabia [117]. Moreover, Al-Shabaab advocates for the Saudi-inspired Wahhabi version of Islam [118]. Referring to [119], ISIS (KG1) was born in 2006 in Iraq as Islamic State of Iraq (ISI). Its main activities are in Syria and Iraq. As shown in Figures 6.3, 6.4 a strong relationship exists among the two countries and ISIS.

Hamas and Hezbollah are the leading groups in MEA facing Israel. As shown in Figs.6.3, 6.4, with the knowledge of the relationship between Hezbollah and Houthis, we can explain why Israel is a linking node between Houthis and Hamas. Finally, we find that Iran links Houthis with ISIS. This could be explained by the fact that both groups are in conflict with Saudi Arabia.

6.2.3 Sensitivity analysis

To analyze more specifically the influence of given terrorist groups on the selected 64 world countries we use the sensitivity analysis described in Chap.5.

Figure 6.5 shows maps of the influence D of the leading groups of each 6 categories on all 64 countries. Here we see that Taliban (KG3) has important influence on Afghanistan, Pakistan, and Saudi Arabia and less influence on other countries. In contrast ISIS (KG1) has a strong worldwide influence with the main effects on Canada, Libya, USA, Saudi Arabia. The world maps show that the groups on the left side of Figure 6.5 (Taliban, Hamas, Houthis) produce mainly local influence in the world. In contrast, the groups on the right panel of Figure 6.5 (ISIS, Al Shabaab, IRA) spread their influence worldwide. Even if IRA mainly affects UK it still spreads its influence on other Anglo-Saxon countries. The presented results determine the geopolitical influence of each terrorist group.

Influence of US and Saudi Arabia on terrorist groups Figure 6.6 shows the influence of a relation between one selected country c and one selected terrorist group i on the other countries j. The results are shown for two countries being US (left panel - column $c = 1$) and Saudi Arabia (right panel - column $c = 46$). Each element (i, j) of given matrices is expressed by $D_{(c \to i)}(j))$. Results show the enormous influence of Saudi Arabia on terrorist groups and other countries (almost all panel is in red). The influence of USA is more selective.

All data for the matrices discussed above, figures and sensitivity are available at [3] and in Appendix C.

We note that above we analyzed the world terror networks. However, at present the statistical data for human crime activity become available

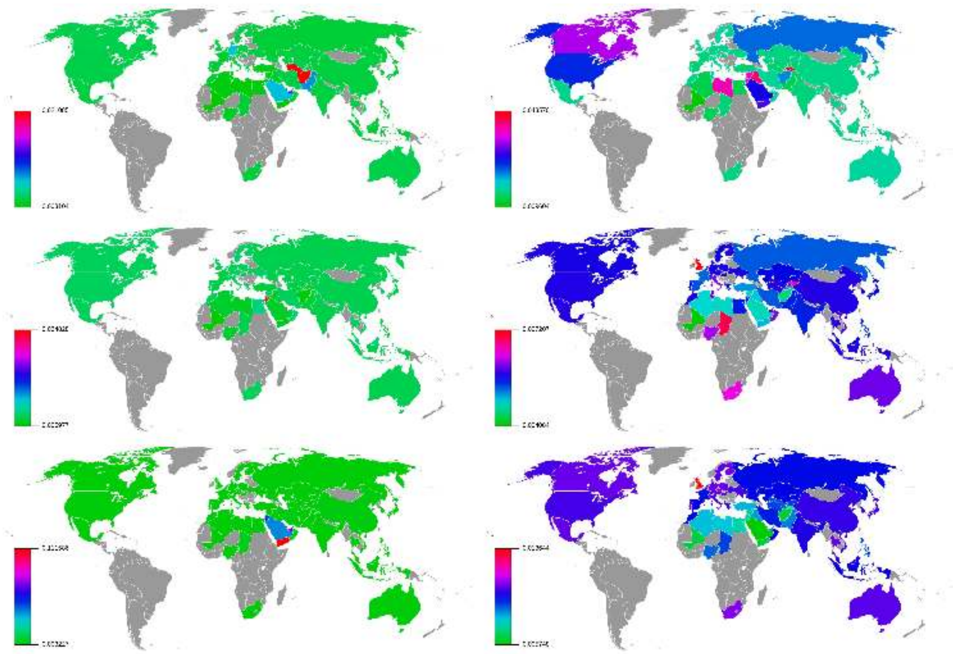

Figure 6.5: **Wold map of the influence of terrorist groups on countries expressed by sensitivity $D_{(j \to i)}(j)$.** Where j is the country index and i the group index. Left column: Taliban KG3, Hamas KG5, Houthis KG13 (top to bottom). Right column: ISIS KG1, Al Shabaab KG10, IRA KG4 (top to bottom). Color bar marks $D_{(j \to i)}(j)$ values with red for maximum and green for minimum influence; grey color marks countries not considered is this work.

[120, 121] and the extension of the described methods to this area would be of particular interest for future works.

6.3 Discussion

We have applied the reduced Google matrix analysis (Figure 6.1) to the network of articles of English Wikipedia to analyze the network structure of 95 terrorist groups and their influence over 64 world countries (159 selected articles). This approach takes into account all human knowledge accumulated in Wikipedia, leveraging all indirect interactions existing between the 159 selected articles and the huge information contained by 5 416 537 articles of Wikipedia and its 122 232 932 links. The network structure obtained for the terrorist groups (Figures 6.2, 6.3) clearly show the presence of 6 types (categories) of groups. The main groups in each category are determined from their PageRank. We show that the indirect or hidden links between terrorist groups and countries play an important role and are, in many cases,

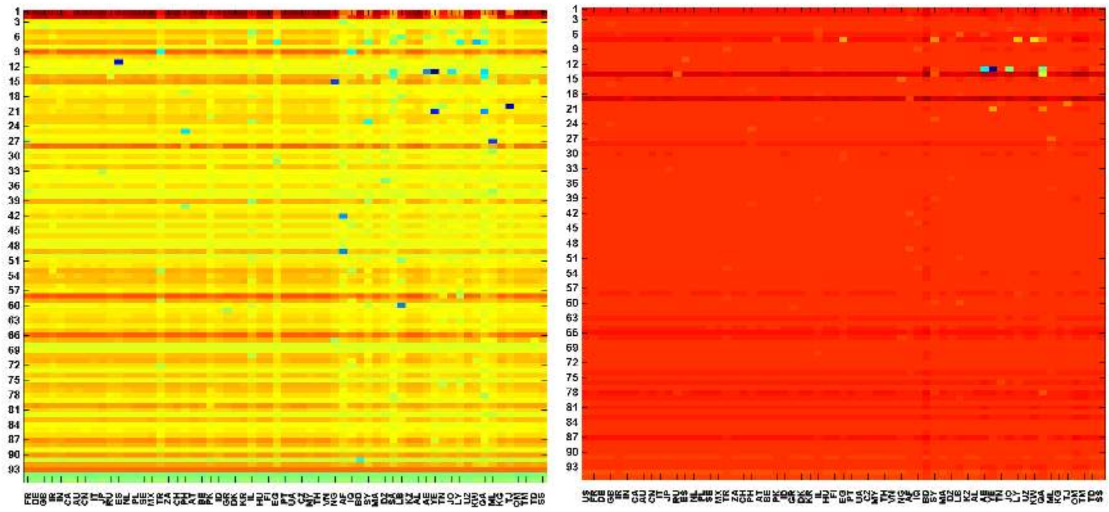

Figure 6.6: **Sensitivity influence $D_{(c\to i)}(j)$ for the relation between a selected country c and a terrorist group i on a world country j.** For two c values: USA (left), Saudi Arabia (right). Terrorist group represented by group index i from Table 6.1 in horizontal axis and countries represented by country index j from Table 6.2 in horizontal axis, $j = c$ is excluded. Color shows $D_{(c\to i)}(j)$ value is changing in the range $(-2.8 \cdot 10^{-4}, 2.1 \cdot 10^{-4}))$ for USA and $(-4.8 \cdot 10^{-3}, 10^{-3}))$ for SA; minimum/maximum values correspond to blue/red.

predominant over direct links.

The geopolitical influence of specific terrorist groups on world countries is determined via the sensitivity of PageRank variation in respect to specific links between groups and countries (Figure 6.4). We see the presence of terrorist groups with localized geographical influence (e.g. Taliban) and others with worldwide influence (ISIS). The influence of selected countries on terrorist groups and other countries is also determined by the developed approach.

The obtained results, tested on the publicly available data of Wikipedia, show the efficiency of the analysis. We argue that the reduced Google matrix approach can find further important applications for terror networks analysis using more advanced and detailed databases.

Chapter 7

General conclusion and perspectives

7.1 Conclusion

In January 2001, the free online collaborative encyclopedia, Wikipedia, began with five different language editions: the French, German, Catalan, Swedish, and Italian. Since that content of Wikipedia has increased to reach more than 46 millions of articles that collect a huge part of human knowledge in 299 different languages. It fully benefits from two key advances: Internet ubiquity and the need of knowledge. The free access to the articles and by enabling edits from different languages on its contents, Wikipedia has drastically reduced obstacles for a web surfer to retrieve reliable information from its contents. One observes that today, Wikipedia has strongly influenced the need of a serious analysis behind its huge structure. It relies on its hyperlinked structure that offers to a surfer to crawl between articles following up existing links.

This thesis dissertation summarizes three years of investigation in this matter, mainly focused on hidden links and sensitivity analysis in Wikipedia networks. The majority of our investigations and the original results presented in this manuscript have been achieved in the context of the GOMOBILE project. The key innovation of this project has consisted in defining and experimenting the concept of Reduced Google Matrix analysis on different domains and different language editions. The major originality of our own contribution concerns the extraction of valuable information that are hidden in the whole Google Matrix of Wikipedia networks and such work could be done using the recent proposed Reduced Google Matrix.

The main outcomes of this thesis are the following:

- New interpretation to classify relationships between nodes based on the different types of connecting links.

- New implementation for a new method of network analysis. This method is based on describing a network structure as hidden and direct relationships. The method is called Reduced Google Matrix and in this thesis we are the first to use it in depth in practical studies.

- New description for hidden links between nodes extracted from the well known Google Matrix.
- New analysis offered on a selected subset of nodes from a much larger network.
- New sensitivity analysis describing the influence of a link variation between two nodes on the rest of a selected subset of nodes.

In a first step, we have provided in Chapter 1 a comprehensive State of the art on Scale free networks, Markov chains, Google Matrix and our network of interest Wikipedia. The detailed description of our selected method Reduced Google Matrix is described in Chapter 2. The remaining of our manuscript has been then divided into three different parts.

The first part, Chapter 3 and Chapter 4, of our manuscript provided a clear methodology to evaluate hidden links through a subset of nodes (articles) carried from different Wikipedia editions. For that purpose, we have selected two subsets from two different domains. Chapter 3 constitutes the content of painters and how hidden relationships links painters and groups them together in an intuitive way. Chapter 4 explains how geopolitical analysis could also be done using the networks of different Wikipedia editions by studying the hidden relationships between a subset of 40 selected countries denoting to different regions in the world.

First of all we have calculated the PageRank vector from the Google Matrix of each Wikipedia editions. Then, we have selected our subsets of painters and countries upon their PageRank score. Second, we have calculated the reduced Google matrix. Third, we regroup our subsets in subgroups according to: *i)* painting categories for painters *ii)* geographical region or shared history for countries. We have assigned a color for each subgroup. Next, we plot the network of friendship to analyse it and to evaluate our choose of subgroups. The aim of this first part is to contribute to a new link analysis model for the selection of nodes according to their domains. This model could be applied with no difficulty to the other domains of Wikipedia networks.

The second part composed of Chapter 5 and Chapter 6, deals with the influence analysis of ties on the importance of nodes using G_{R}. The results presented in these two Chapters represent the most original contribution of this thesis. Chapter 5 introduces Chapter 6. Chapter 5 was dedicated to figure out how sensible our networks of countries and painters are on a particular link variation. Therefore, a link variation is made by modifying slightly a column of G_{R} and calculating a new PageRank vector $\tilde{P}$. Then, we calculate the logarithmic difference between regular P and $\tilde{P}$. As a result, we obtain which countries are the most sensitive to a link variation between two countries. It is exciting to find out that our results are intuitive and harmonious with geopolitical studies. Then, in Chapter 6, we went deeper in our geopolitical study to extract novel knowledge on the ties between terrorist groups and countries. Therefore, we have selected 95 terrorist groups and 64 related countries from English Wikipedia edition of 2017 to create a reduced Google matrix again and use our sensitivity analysis on this network. With this study, we've been able to extract the most influential terrorist groups and on which country they have the most impact. Moreover, we have shown groups are related not only with each other but also to countries. Our results are in line with recent studies derived from reliable sources.

To sum up, for Wikipedia networks, the reduced Google matrix method is suitable and the results derived from different studies on different domains are in line with the proposed contributions. The validity and importance of Hidden relationships are investigated for three different subnetwork: painters, countries and world terror network. The sensitivity analysis applied on two different networks of countries lead to compatible results with many geopolitical studies. The merge between terrorist group nodes and affected world countries gives rise to an original geopolitical analysis.

7.2 Perspectives

The Reduced Google matrix analysis has only been investigated for a few networks in this manuscript. Reduced Google matrix analysis is today in its infancy and it could be seen as a very hot topic. The fact that Reduced Google matrix is based on Google matrix and able to extract hidden links in the network structure appears as a very promising approach for network analysis.

Our study could be extended by analyzing mobile networks in order to have a wider performance view of the Reduced Google matrix method. Also, the resultant graphs describing the direct and hidden interactions between network's nodes could be used in the context of preventing terrorist growth. Thus, if a node has suspicious hidden links, it will be under surveillance. In addition, a deep study on Twitter could be targeted to determine how this huge network of friends and followers is linked, how tweets flows and reach widely, what are the subgroups and how they interact dynamically. Another issue that should be interesting to investigate is the election networks created by electors and how it affects the results and the behavior of the voters as well as how media are biased.

Appendix A

Painters

Table A.1: **Top 223 painters of 7 Wikipedia editions.** It contains the top 100 painters of each edition. POB, YOB, YOD are place of birth, year of birth and year of death. Other columns list the PageRank order for the 7 Wikipedia editions (FrWiki, EnWiki, DeWiki, ItWiki, RuWiki, EsWiki and NlWiki.

Name	**POB**	**YOB**	**YOD**	**Fr**	**En**	**De**	**It**	**Ru**	**Es**	**Nl**
Pablo Picasso	Spain	1881	1973	1	2	2	4	2	3	6
Leonardo da Vinci	Italy	1452	1519	2	1	1	1	1	1	2
Michelangelo	Italy	1475	1564	3	3	4	2	3	4	857
Claude Monet	France	1840	1926	4	13	16	20	19	21	10
Vincent Van Gogh	NL	1853	1890	5	6	8	9	5	14	5
Jacques-Louis David	France	1748	1825	6	26	44	21	32	26	36
Eugène Delacroix	France	1798	1863	7	20	32	46	25	18	34
Raphael	Italy	1483	1520	8	4	5	3	6	5	821
Henri Matisse	France	1869	1954	9	15	17	56	12	17	29
Salvador Dali	Spain	1904	1989	10	14	18	27	13	7	24
Paul Cézanne	France	1839	1906	11	17	13	32	14	19	13
Rembrandt Van Rijn	NL	1606	1669	12	5	6	11	4	13	1

Peter Paul Rubens	Germany	1577	1640	13	9	7	7	9	8	4
Andy Warhol	US	1928	1987	14	8	11	29	52	20	19
Marcel Duchamp	France	1887	1968	15	33	39	98	102	43	54
Édouard Manet	France	1832	1883	16	23	24	44	27	27	25
Giorgio Vasari	Italy	1511	1574	17	16	19	5	31	22	37
Paul Gauguin	France	1848	1903	18	29	29	41	21	23	28
Albrecht Durer	Germany	1471	1528	19	10	3	13	7	11	15
Pierre Auguste Renoir	France	1841	1919	20	25	28	28	34	44	168
Joan Miró	Spain	1893	1983	21	61	33	60	93	33	49
Jean-Auguste-Dominique Ingres	France	1780	1867	22	57	95	94	42	59	73
Georges Braque	France	1882	1963	23	54	45	76	54	31	52
Edgar Degas	France	1834	1917	24	31	48	43	37	32	51
Francisco Goya	Spain	1746	1828	25	18	27	14	28	2	22
Gustave Courbet	CH	1819	1877	26	53	61	75	47	50	65
Fernand Léger	France	1881	1955	27	88	77	104	94	83	55
Titian	Italy	1488	1576	28	12	9	6	11	9	11
Caravaggio	Italy	1571	1610	29	21	26	8	36	16	854
Jackson Pollock	US	1912	1956	30	22	74	114	244	38	66
Wassily Kandin-sky	Russia	1866	1944	31	40	15	23	26	34	27

Nicolas Poussin	France	1594	1665	32	37	85	39	39	47	48
Marc Chagall	Belarus	1887	1985	33	41	22	48	18	97	43
Honoré Daumier	France	1808	1879	34	80	101	147	73	203	80
Max Ernst	Germany	1891	1976	35	68	30	140	106	70	78
Diego Velázquez	Spain	1599	1660	36	27	37	24	17	6	856
Gustave Doré	France	1832	1883	37	45	100	47	95	64	50
Sandro Botticelli	Italy	1445	1510	38	34	38	12	60	15	12
Giotto Di Bondone	Italy	1267	1337	39	35	21	10	44	25	17
Jean-Baptiste Camille Corot	France	1796	1875	40	71	107	108	69	92	76
Henri de Toulouse-Lautrec	France	1864	1901	41	46	62	55	23	39	40
William Bouguer-eau	France	1825	1905	42	91	214	97	174	49	208
Pieter Bruegel The Elder	NL	1528	1569	43	38	35	25	68	46	9
Antoine Watteau	France	1684	1721	44	87	103	179	83	62	84
Georges Seurat	France	1859	1891	45	69	123	83	58	134	47
Rene Magritte	Belgium	1898	1967	46	92	92	96	87	103	31
André Derain	France	1880	1954	47	129	165	155	123	129	195
Paul Klee	CH	1879	1940	48	65	12	66	104	91	42
François Boucher	France	1703	1770	49	117	124	124	97	69	107

Camille Pissarro	US	1830	1903	50	52	69	121	59	96	56
William Hogarth	UK	1697	1764	51	28	88	157	146	116	59
Théodore Géricault	France	1791	1824	52	97	236	154	76	56	184
Maurice de Vlam- inck	France	1876	1958	53	218	304	400	293	345	283
Joseph Mallord William Turner	UK	1775	1851	54	19	53	63	41	55	44
Gustav Klimt	Austria	1862	1918	55	123	31	95	86	67	33
Piet Mondrian	NL	1872	1944	56	67	52	111	179	54	8
Jean- Honoré Frago- nard	France	1732	1806	57	141	195	165	192	63	159
Jean- Léon Gérôme	France	1824	1904	58	110	182	59	112	161	332
Jean Fou- quet	France	1425	1481	59	126	368	134	127	99	303
Anthony van Dyck	Belgium	1599	1641	60	24	65	30	40	30	20
Hieronym- us Bosch	NL	1450	1516	61	70	67	62	22	35	812
Amedeo Modigliani	Italy	1884	1920	62	113	105	65	101	136	121
Antoine- Jean Gros	France	1771	1835	63	219	479	183	186	219	178
Johannes Vermeer	NL	1632	1675	64	32	46	71	57	40	7
Paolo Veronese	Italy	1528	1588	65	60	97	37	61	60	21
Alfred Sisley	France	1839	1899	66	147	119	129	64	193	155

Andrea Mantegna	Italy	1431	1506	67	81	64	18	100	37	83
Claude Lorrain	France	1600	1682	68	62	94	64	70	76	99
Fra Angelico	Italy	1387	1455	69	55	110	26	191	51	839
Jean Dubuffet	France	1901	1985	70	163	178	302	517	162	144
Kazimir Malevich	Ukraine	1879	1935	71	111	56	269	29	88	67
Charles Le Brun	France	1619	1690	72	119	87	171	173	148	119
Jean Siméon Chardin	France	1699	1779	73	247	216	303	114	132	152
William Blake	UK	1757	1827	74	11	86	50	48	42	30
Edvard Munch	Norway	1863	1944	75	75	20	17	74	57	45
Jan Van Eyck	NL	1390	1441	76	49	34	36	915	48	18
Francis Picabia	France	1879	1953	77	221	205	152	322	118	210
Hyacinthe Rigaud	France	1659	1743	78	182	213	109	299	247	256
Man Ray	US	1890	1976	79	100	81	89	241	86	132
Raoul Dufy	France	1877	1953	80	171	194	198	128	239	183
Pietro Perugino	Italy	1445	1523	81	56	98	35	115	80	143
Jean-Baptiste Greuze	France	1725	1805	82	392	378	549	154	410	268
Georges de La Tour	France	1593	1652	83	295	257	345	125	146	392
Williem De Kooning	NL	1904	1997	84	36	163	333	346	94	104

Pierre Puvis de Chavannes	France	1824	1898	85	158	327	209	117	246	240
Jacques Callot	France	1592	1635	86	283	226	321	347	462	348
Hans Holbein The Younger	Germany	1497	1543	87	30	42	73	78	45	53
El Greco	Greece	1541	1614	88	42	43	51	51	12	63
Piero Della Francesca	Italy	1416	1492	89	105	151	16	120	36	82
Egon Schiele	Austria	1890	1918	90	197	72	177	211	113	69
Lucas Cranach the Elder	Germany	1472	1553	91	43	14	45	62	98	88
Francis Bacon	UK	1909	1992	92	7	10	219	370	10	855
Pierre Bonnard	France	1867	1947	93	161	158	173	91	157	169
Jean-François Millet	France	1814	1875	94	106	207	172	96	82	141
Diego Rivera	Mexico	1886	1957	95	74	108	142	159	24	87
Pierre Soulages	France	1919		96	768	616		568	791	249
Canaletto	Italy	1697	1768	97	82	1410	58	161	100	117
Maurice Denis	France	1870	1943	98	300	206	275	116	430	110
Roberto Matta	Chile	1911	2002	99	512	276		462	101	269
Tintoretto	Italy	1518	1594	100	58	57	40	72	28	92
Lucian Freud	Germany	1922	2011	104	89	296	382	218	311	181
Giovanni Battista Tiepolo	Italy	1696	1770	108	84	54	31	131	81	162
Hiroshige	Japan	1797	1858	109	77	201	287	185	388	68

Juan Gris	Spain	1887	1927	113	190	149	250	232	93	123
Ilya Repin	Ukraine	1844	1930	114	76	112	139	8	128	64
Guido Reni	Italy	1575	1642	115	109	122	49	167	72	277
Caspar David Friedrich	Germany	1774	1840	116	86	23	125	105	66	154
Henry van de Velde	Belgium	1863	1957	117	184	47	282	319	186	38
Antonio da Correggio	Italy	1489	1534	118	115	102	42	98	79	242
Frans Hals	Belgium	1580	1666	120	96	78	118	139	71	16
Masaccio	Italy	1401	1428	124	135	82	19	145	53	70
Robert Rauschenberg	US	1925	2008	128	95	84	217	259	175	197
Samuel Morse	US	1791	1872	129	78	1644	22	35	65	75
Giovanni Bellini	Italy	1430	1516	133	108	75	33	109	74	125
Roger Van Der Weyden	NL	1399	1464	135	72	80	82	85	73	46
Thomas Gainsborough	UK	1727	1788	136	50	116	107	108	121	62
James Abbot Mac Neil Whistler	US	1834	1903	137	48	133	182	215	185	85
Hokusai	Japan	1760	1849	141	63	166	99	90	147	57
Francisco De Zurbaran	Spain	1598	1664	143	134	189	132	144	41	234
John Constable	UK	1776	1837	144	39	179	136	66	84	101

Andrei Rublev	Russia	1360	1430	145	152	188	141	30	158	79
Georgia O'keefe	US	1887	1986	156	47	183	387	624	68	245
Kurt Schwitters	Germany	1887	1948	160	231	73	264	355	218	86
Bartolomé Esteban Murillo	Spain	1617	1682	165	125	137	137	118	29	114
Jan Matejko	Poland	1838	1893	167	66	155	195	77	214	361
John Everett Millais	UK	1829	1896	168	51	219	245	197	184	120
Cimabue	Italy	1240	1302	174	143	144	57	229	145	97
Filippo Lippi	Italy	1406	1469	175	150	297	61	182	106	130
Simone Martini	Italy	1284	1344	176	186	217	85		114	279
Domenico Ghirlandaio	Italy	1449	1494	177	93	143	34	181	95	105
Paolo Uccello	Italy	1397	1475	180	196	181	52	243	117	211
John Singer Sargent	Italy	1856	1925	181	73	240	247	258	164	227
Antoni Tàpies	Spain	1923	2012	183	547	171	410	268	90	201
Dante Gabriel Rossetti	UK	1828	1882	185	64	187	153	177	144	58
Andrea del Sarto	Italy	1487	1531	186	189	324	90	176	108	336
Marsden Hartley	US	1877	1943	187	85	484	889	1008	120	742
Annibale Carracci	Italy	1560	1609	188	90	140	80	158	104	171
Bronzino	Italy	1503	1572	190	205	153	72	201	105	176
Georg Baselitz	Germany	1938		194	528	59	462	681	684	263

Antonello da Messina	Italy	1430	1479	199	226	170	68	238	200	343
Benjamin West	US	1738	1820	208	44	224	297	262	222	251
Frida Kahlo	Mexico	1907	1954	211	133	113	167	239	52	112
Norman Rockwell	US	1894	1978	217	94	306	347	257	576	136
Guercino	Italy	1591	1666	218	146	318	53	264	131	427
Pietro da Cortona	Italy	1596	1669	221	103	241	54	295	110	186
Andrea del Verrocchio	Italy	1435	1488	224	201	129	92	219	75	172
Ferdinand Hodler	CH	1853	1918	225	523	66	188	344	320	295
Max Liebermann	Germany	1847	1935	228	258	36	261	290	242	139
Matthias Grünewald	Germany	1470	1528	230	132	70	116	129	87	191
FRANZ MARC	Germany	1880	1916	232	246	50	145	209	212	96
Ernst Ludwig Kirchner	Germany	1880	1938	233	273	63	367	200	383	135
Parmigianino	Italy	1504	1540	235	238	192	77	289	127	194
Filippino Lippi	Italy	1457	1504	238	204	299	87	354	333	170
Giorgione	Italy	1478	1510	239	144	141	38	121	78	164
Giuseppe Arcimboldo	Italy	1527	1593	240	375	295	15	276	77	330
Léon Bakst	Belarus	1866	1924	245	358	588	370	92	440	507
Max Beckmann	Germany	1884	1950	251	228	41	516	591	233	166
Fra Bartolomeo	Italy	1474	1517	257	306	172	91	196	253	

Ivan Aiva-zovsky	Russia	1817	1900	259	279	200	274	16	524	
Pontormo	Italy	1494	1557	263	256	390	79	318	151	346
Emil Nolde	Germany	1867	1956	267	311	51	341	190	211	335
Francesco Hayez	Italy	1791	1882	274	222	450	69		179	702
Otto Dix	Germany	1891	1969	279	330	25	337	221	297	192
David Hockney	UK	1937		284	83	177	295	225	371	254
Albrecht Altdorfer	Germany	1480	1538	291	284	58	256	207	241	421
August Macke	Germany	1887	1914	302	385	79	272	212	202	219
Lorenzo Lotto	Italy	1480	1557	303	280	292	70	245	266	505
Boris Kus-todiev	Russia	1878	1927	304	192	676	185	88	661	712
John Trumbull	US	1756	1843	309	79	366	86	1039	198	848
Arnold Böcklin	CH	1827	1901	310	531	55	131	170	409	311
Viktor Vas-netsov	Russia	1848	1926	315	99	278	240	15	227	292
El Lis-sitzky	Russia	1890	1941	316	237	96	374	198	133	217
Salvatore Rosa	Italy	1615	1673	317	240	495	100	544	317	
William Holman Hunt	UK	1827	1910	319	98	305	419	314	407	165
Alexej von Jawlen-sky	Russia	1864	1941	321	646	93	306	254	569	291
Umberto Boccioni	Italy	1882	1916	323	178	128	122	483	89	329

Masolino da Panicale	Italy	1383	1447	345	371	483	81	594	334	522
Lovis Corinth	Germany	1858	1925	347	373	49	265	390	331	338
Pisanello	Italy	1395	1455	363	203	380	93	330	284	285
Nicholas Roerich	Russia	1874	1947	375	214	294	437	10	424	481
Valentin Serov	Russia	1865	1911	380	360	480	451	20	411	429
George Grosz	Germany	1893	1959	382	170	68	252	381	178	138
Pinturicchio	Italy	1454	1513	393	309	303	67	240	217	293
Karl Bryullov	Russia	1799	1852	409	349	412	373	24		518
Gerhard Richter	Germany	1932		415	217	40	423	406	540	276
Ivan Kramskoi	Russia	1837	1887	422	347	784	568	50	386	551
David Alfaro Siqueiros	Mexico	1896	1974	424	212	220	215	265	58	232
Wolf Vostell	Germany	1932	1998	427	356	60	271		165	94
Igor Grabar	Russia	1871	1960	444	963	1287		33		
Vasily Surikov	Russia	1848	1916	465	265	325	731	53	313	196
Lyonel Feininger	US	1871	1956	466	447	71	228	457	505	327
Aleksandr Deyneka	Russia	1899	1969	480	1082	672		56		
Vasily Vereshchagin	Russia	1842	1904	485	335	574		55		558
Franz Stuck	Germany	1863	1928	498	452	99	176	305	397	131
Mikhail Vrubel	Russia	1856	1910	508	289	419	556	38	686	420
Carlo Crivelli	Italy	1881	1966	520	498	871	88	475	746	516

Renato Guttuso	Italy	1911	1987	531	957	427	74	331	181	
JoaquÃn Sorolla y Bastida	Spain	1863	1923	567	635	315	457	441	61	703
Karl Schmidt-Rottluff	Germany	1884	1976	571	728	91	357	279	520	324
Erich Heckel	Germany	1883	1970	573	648	90	483	686	592	288
Isaac Levitan	Lithuania	1860	1900	575	381	393	868	46	336	156
Sigmar Polke	Poland	1745	1801	578	562	76	467	438	573	213
Thomas Eakins	US	1844	1916	582	59	377	517	310	417	451
Alexander Andreye-vich Ivanov	Russia	1806	1858	591	488	773	673	75	458	253
Franz von Lenbach	Germany	1836	1904	607	460	83	207	384	502	454
Tove Jansson	Finland	1914	2001	624	351	263	335	49	579	460
Mari‘a Fortuny	Spain	1838	1874	636	874	682	440	673	85	446
Ivan Shishkin	Russia	1832	1898	666	440	722	864	81	351	452
Dmitry Levitzky	Ukraine	1735	1822	696	553	1152		89	449	
Mikhail Nesterov	Russia	1862	1942	752	721	731	630	71	265	246
Konstantin Korovin	Russia	1861	1939	775	518	816	625	45	527	557
Vasily Polenov	Russia	1844	1927	784	456	803	789	67	370	585
Alexei Savrasov	Russia	1830	1897	830	1026	1091		80	587	570
Vladimir Borovikov-sky	Ukraine	1757	1825	881		1135	688	79	379	

Francesco del Cossa	Italy	1435	1477	908	952	770	78	600	610	714
Daniel Chodowiecki	Poland	1726	1801	929	604	89		620		
Simon Ushakov	Russia	1590	1649	1022	534		683	99	680	
Dionisius	Russia	1440	1502	1025	326			65	889	
Arkhip Kuindzhi	Ukraine	1847	1910	1069	574	1015	723	84		572
Martiros Saryan	Russia	1880	1972	1099	628	630		82	837	
Carl Bloch	Denmark	1834	1890	1563	321	396	84	1189	1059	811
Alexandre Benois	Russia	1870	1960		376	597	765	43	622	573
Mstislav Dobuzhinsky	Russia	1875	1957		1117	1276		63		

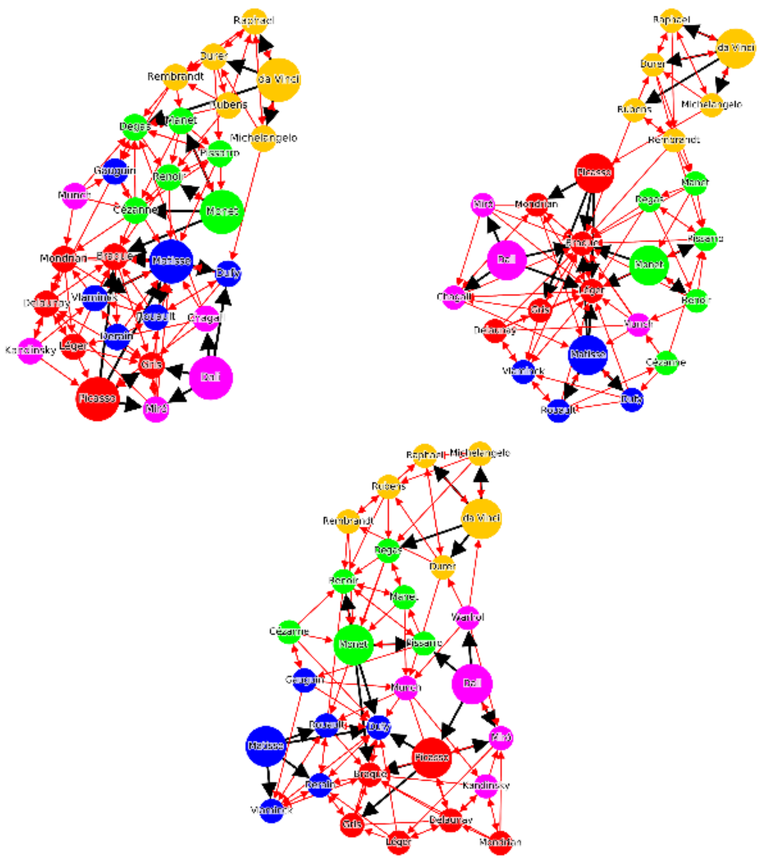

Figure A.1: **Network structure of friends induced by the top 5 painters of each group in G_{qrnd}.** Results are plotted for EsWiki (Top left), ItWiki (Top right) and RuWiki (Bottom). Red, Blue, Green, Orange and Pink nodes represent Cubism, Fauvism, Impressionism, Great masters and Modern (20-21), respectively. The top painter node points with a bold black arrow to its top-4 friends. Red arrows represent the friends of friends interactions computed until no new edges are added to the graph. All graphs are automatically plotted using *Gephi* [1].

Appendix B

Countries

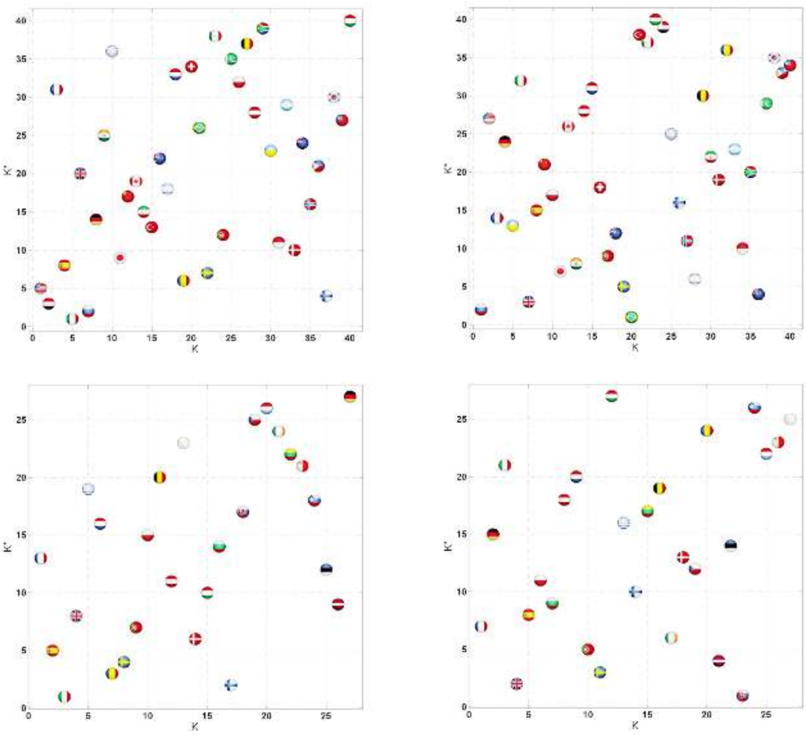

Figure B.1: **Position of 40 and EU countries in the local** (K, K^*) **plane.** First row: 40 countries and second row: EU countries. ArWiki (left) and RuWiki (right) networks. Countries are marked by their flags.

Appendix C

Special groups

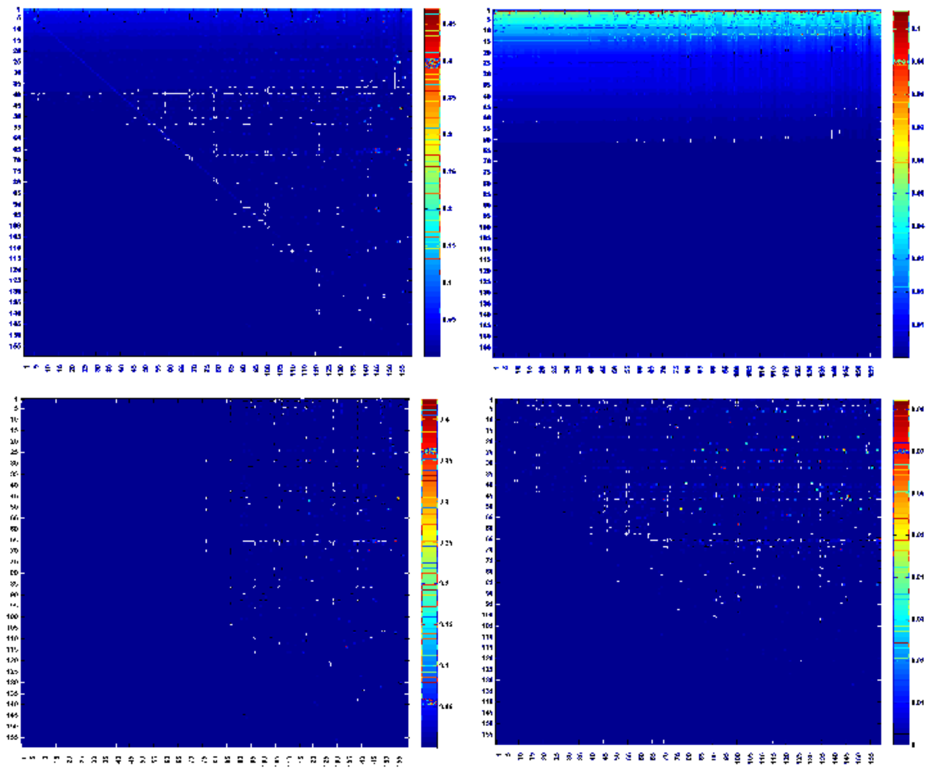

Figure C.1: **Density plots of matrices G_{R}, G_{pr}, G_{rr} and G_{qrnd}.** Top left and right figures are G_{R} and G_{pr} respectively, Top left and right figures are G_{rr} and G_{qrnd} respectively; color changes from red at maximum to blue at zero; 95 terrorist nodes (from Tab. 6.2) and 64 countries nodes (from Tab. 6.1) are shown.

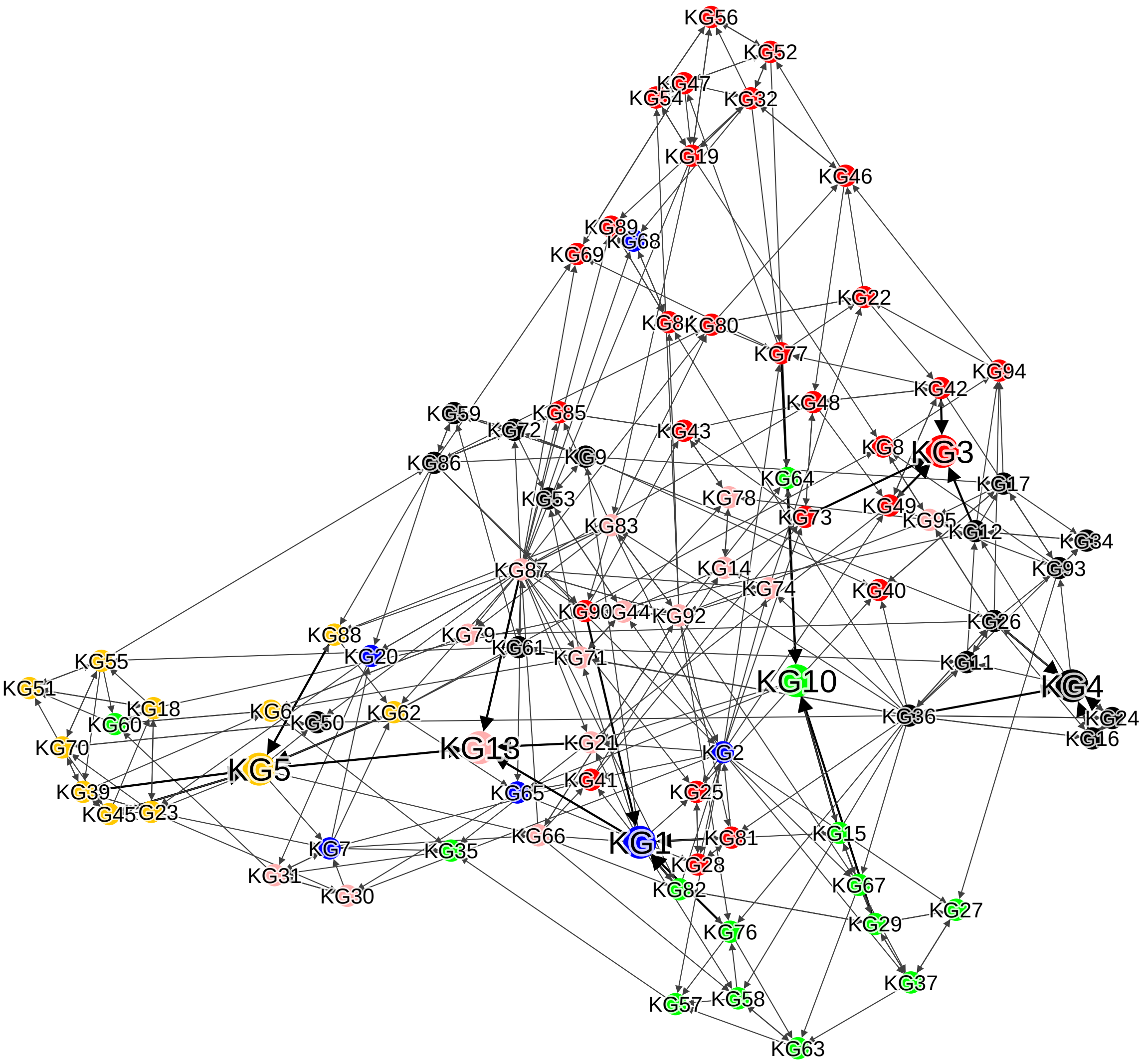

Figure C.2: **Followers network structure between terrorist groups obtained from $G_{\text{qrnd}}+G_{rr}$.** Colors mark categories of nodes and top nodes are given in text and Table 6.1; circle size is proportional to PageRank probability of nodes; bold black arrows point to top 4 friends, gray tiny arrows show friends of friends interactions computed until no new edges are added to the graph (drawn with [1,2].

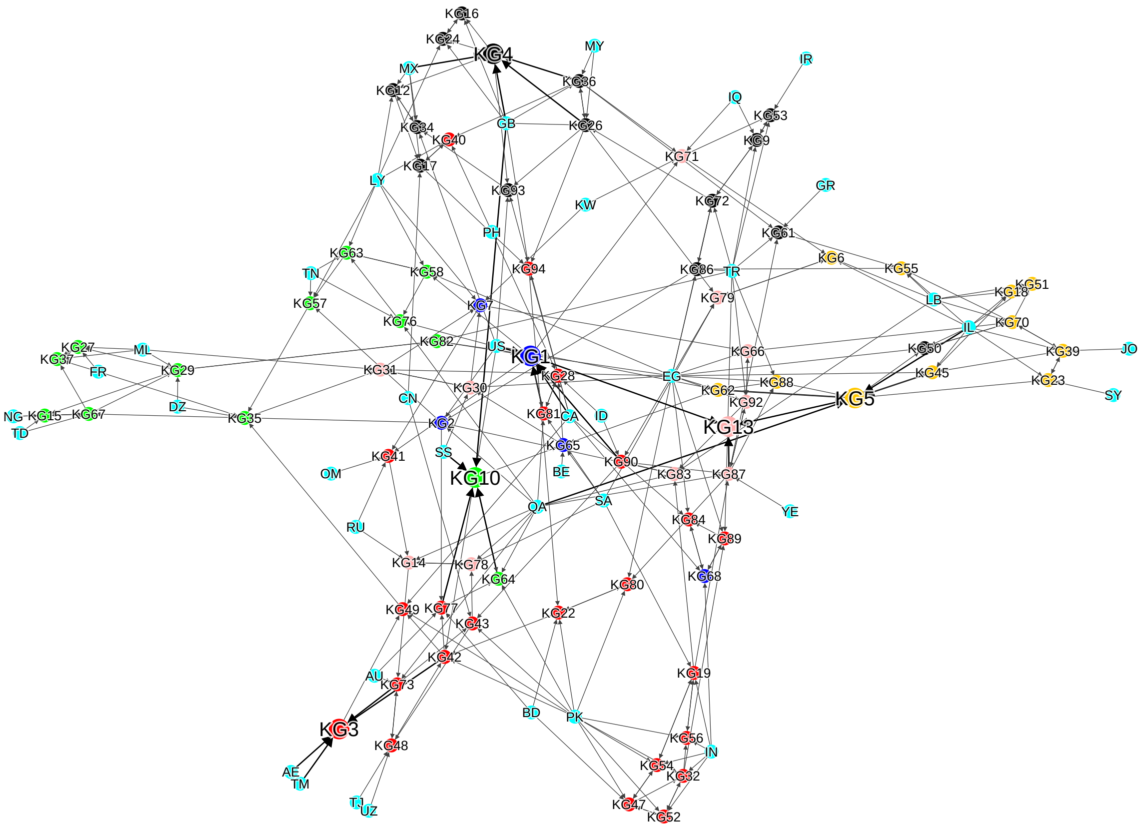

Figure C.3: **Followers network structure extracted from $G_{\text{qrnd}} + G_{rr}$ with the top terrorist groups and countries.** The top terrorist groups are marked by their respective colors and countries are marked by cyan color. The network structure is shown in case of 2 friends for top terrorist groups of each category and top friend 2 countries for each group. Networks are drawn with [1, 2].

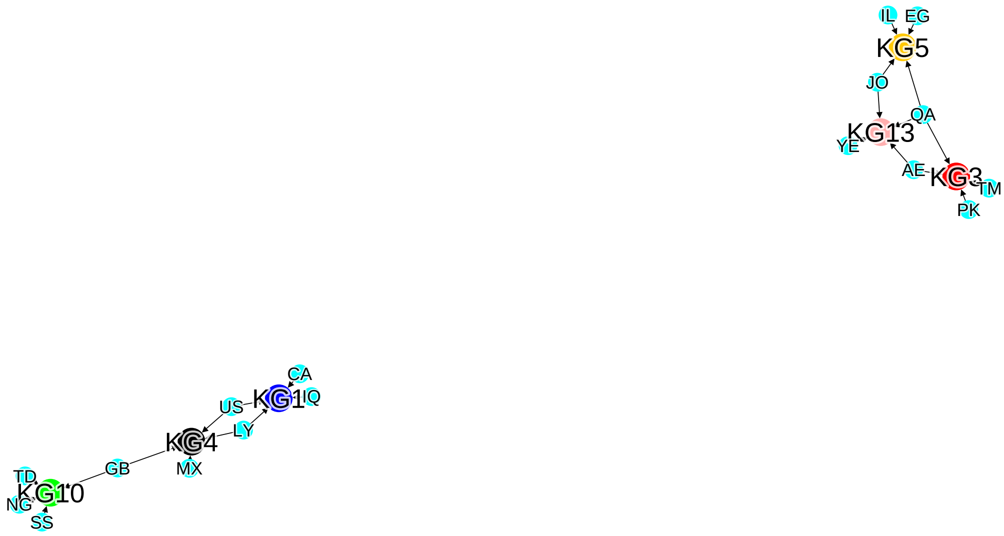

Figure C.4: **Followers network structure extracted from $G_{\text{qrnd}} + G_{rr}$ with the top terrorist groups and countries.** The top terrorist groups are marked by their respective colors and countries are marked by cyan color. The network structure is shown with the top terrorist groups of each category and their top 4 friend countries. Networks are drawn with [1, 2].

Bibliography

[1] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," 2009. [Online]. Available: http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154 (Cited on pages ix, x, xi, xv, xvi, xvii, 7, 35, 37, 49, 50, 60, 103, 105, 106, 128, 132, 133 and 134.)

[2] Y. Hu, "Efficient, high-quality force-directed graph drawing," *Mathematica Journal*, vol. 10, no. 1, pp. 37–71, 2005. (Cited on pages xv, xvi, xvii, 60, 103, 105, 106, 132, 133 and 134.)

[3] W. contributors (2017). List of designated terrorist groups. Accessed 2017-08-23. [Online]. Available: https://en.wikipedia.org/wiki/List_of_designated_terrorist_groups (Cited on pages xx, 97, 99 and 107.)

[4] M. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc., 2010. (Cited on page 1.)

[5] T. C. Silva and L. Zhao, *Machine learning in complex networks*. Springer, 2016, vol. 2016. (Cited on page 1.)

[6] P. Erdős and A. Rényi, "On random graphs i." *Publicationes Mathematicae (Debrecen)*, vol. 6, pp. 290–297, 1959 1959. (Cited on pages 1 and 2.)

[7] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998. (Cited on pages 1, 2 and 3.)

[8] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999. [Online]. Available: http://www.sciencemag.org/cgi/content/abstract/286/5439/509 (Cited on pages 1 and 3.)

[9] X. F. Wang and G. Chen, "Complex networks: small-world, scale-free and beyond," *IEEE Circuits and Systems Magazine*, vol. 3, no. 1, pp. 6–20, 2003. (Cited on page 2.)

[10] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Phys. Rev. E*, vol. 64, p. 026118, Jul 2001. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.64.026118 (Cited on page 2.)

[11] A. Fronczak, P. Fronczak, and J. A. Hołyst, "Average path length in random networks," *Physical Review E*, vol. 70, no. 5, p. 056110, 2004. (Cited on page 2.)

[12] S. Barbarossa. Random graphs. Accessed 2018-04-20. [Online]. Available: http://infocom.uniroma1.it/sergio/Lesson_4_random_graphs.pdf (Cited on page 2.)

[13] T. Blank, *Folk Culture in the Digital Age: The Emergent Dynamics of Human Interaction*. University Press of Colorado, 2012. [Online]. Available: https://books.google.com.lb/books?id=iNe9AwAAQBAJ (Cited on page 2.)

[14] S. Milgram, "The small world problem," *Psychology Today*, vol. 1, no. 1, pp. 61–67, 1967. [Online]. Available: http://snap.stanford.edu/class/cs224w-readings/milgram67smallworld.pdf (Cited on page 3.)

[15] J. Travers and S. Milgram, "An experimental study of the small world problem," *Sociometry*, vol. 32, no. 4, pp. 425–443, 1969. [Online]. Available: http://www.jstor.org/stable/2786545 (Cited on page 3.)

[16] P. Boldi. (2014, June) Graph distance distribution for social network mining. School for advanced sciences of Luchon. [Online]. Available: http://www.quantware.ups-tlse.fr/ecoleluchon2014/slides/boldi3.pdf (Cited on page 3.)

[17] V. Y. channel. (2015, Aug) The science of six degrees of separation. Accessed 2017-08-23. [Online]. Available: https://youtu.be/TcxZSmzPw8k (Cited on page 3.)

[18] A.-L. Barabási and E. Bonabeau, "Scale-free networks," *Sci. Am.*, vol. 288, no. 5, pp. 50–59, 2003. (Cited on page 3.)

[19] M. T. A. Maps. Travel maps and tourist attractions. Accessed 2017-11-20. [Online]. Available: http://www.1830ndaytona.info/us-roadmap/us-roadmap-2-maps-update-1142774-roadmap-of-the-us-road-map-clipart/ (Cited on pages 3 and 4.)

[20] T. Meko. (2016, Dec.) Six maps that show the anatomy of america's vast infrastructure. Accessed 2017-11-20. [Online]. Available: https://www.washingtonpost.com/graphics/national/maps-of-american-infrastrucure/ (Cited on pages 3 and 4.)

[21] A. N. Langville and C. D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton, NJ, USA: Princeton University Press, 2006. (Cited on pages 5, 6, 9, 13, 15, 16 and 24.)

[22] Accessed 2017-11-23. [Online]. Available: http://www.wikipedia.org (Cited on page 5.)

[23] Y.-H. Eom, P. Aragón, D. Laniado, A. Kaltenbrunner, S. Vigna, and D. L. Shepelyansky, "Interactions of cultures and top people

of Wikipedia from ranking of 24 language editions," *PLoS ONE*, vol. 10(3), no. e0114825, p. e0114825, Mar. 2015, 31 pages. 10 figures. Submitted for publication. Supporting information is available on http://www.quantware.ups-tlse.fr/QWLIB/topwikipeople/. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00997898 (Cited on pages 5, 16 and 41.)

[24] D. S. K.M. Frahm. Wikipedia networks of 24 editions of 2017. Accessed 2017-10-09. [Online]. Available: http://www.quantware.ups-tlse.fr/QWLIB/24wiki2017 (Cited on page 5.)

[25] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, "Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software," *PLOS ONE*, vol. 9, no. 6, pp. 1–12, 06 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0098679 (Cited on page 7.)

[26] A. Bavelas, "A mathematical model for group structures." *Applied Anthropology*, 1948. (Cited on page 6.)

[27] A. Shimbel, "Structural parameters of communication networks," *The bulletin of mathematical biophysics*, vol. 15, no. 4, pp. 501–507, 1953. (Cited on page 6.)

[28] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1, pp. 107 – 117, 1998, proceedings of the Seventh International World Wide Web Conference. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016975529800110X (Cited on pages 9 and 16.)

[29] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, ser. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis, 1995. [Online]. Available: http://books.google.com/books?id=TRXrMWY_i2IC (Cited on page 9.)

[30] G. Strang, *Introduction to Linear Algebra.* Wellesley-Cambridge Press, 2016. [Online]. Available: https://books.google.fr/books?id=efbxjwEACAAJ (Cited on page 12.)

[31] S. University. (2009) Lecture 17 Perron-Frobenius theory. Accessed 2018-01-15. [Online]. Available: https://stanford.edu/class/ee363/lectures/pf.pdf (Cited on page 13.)

[32] G. Stewart, *Matrix Algorithms Volume 2: Eigensystems*, ser. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, 2001. [Online]. Available: https://books.google.com.lb/books?id=Y3k5TNTnEmEC (Cited on page 13.)

[33] T. Haveliwala and S. Kamvar, “The second eigenvalue of the google matrix,” Stanford, Tech. Rep., 2003. (Cited on page 13.)

[34] A. N. Langville and C. D. Meyer, “Deeper inside pagerank,” *Internet Mathematics*, vol. 1, no. 3, pp. 335–380, 2004. (Cited on page 13.)

[35] A. D. Chepelianskii, “Towards physical laws for software architecture,” *arXiv preprint arXiv:1003.5455*, 2010. (Cited on pages 15, 44 and 57.)

[36] A. O. Zhirov, O. V. Zhirov, and D. L. Shepelyansky, “Two-dimensional ranking of wikipedia articles,” *The European Physical Journal B*, vol. 77, no. 4, pp. 523–531, Oct 2010. [Online]. Available: https://doi.org/10.1140/epjb/e2010-10500-7 (Cited on pages 15 and 16.)

[37] L. Ermann, K. M. Frahm, and D. L. Shepelyansky, “Google matrix analysis of directed networks,” *Rev. Mod. Phys.*, vol. 87, pp. 1261–1310, Nov 2015. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.87.1261 (Cited on pages 15, 16, 44 and 57.)

[38] L. Ermann, K. M. Frahm, and D. Shepelyansky, “Spectral properties of google matrix of wikipedia and other networks,” *CoRR*, vol. abs/1212.1068, 2012. [Online]. Available: http://arxiv.org/abs/1212.1068 (Cited on page 15.)

[39] K. M. Frahm and D. Shepelyansky, “Reduced Google matrix,” Feb. 2016, working paper or preprint. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01271386 (Cited on pages 15, 17 and 21.)

[40] List of wikipedias @Wikipedia. Accessed 2017-11-29. [Online]. Available: https://meta.wikimedia.org/wiki/List_of_Wikipedias (Cited on page 16.)

[41] Y.-H. Eom, K. M. Frahm, A. Benczúr, and D. L. Shepelyansky, “Time evolution of wikipedia network ranking,” *The European Physical Journal B*, vol. 86, no. 12, p. 492, Dec 2013. [Online]. Available: https://doi.org/10.1140/epjb/e2013-40432-5 (Cited on page 16.)

[42] Y.-H. Eom and D. L. Shepelyansky, “Highlighting entanglement of cultures via ranking of multilingual wikipedia articles,” *PLOS ONE*, vol. 8, no. 10, pp. 1–10, 10 2013. [Online]. Available: https://doi.org/10.1371/journal.pone.0074554 (Cited on page 16.)

[43] J. Lages, A. Patt, and D. L. Shepelyansky, “Wikipedia ranking of world universities,” *The European Physical Journal B*, vol. 89, no. 3, p. 69, Mar 2016. [Online]. Available: https://doi.org/10.1140/epjb/e2016-60922-0 (Cited on page 16.)

[44] M. Hart, *The 100: ranking of the most influential persons in history*. NY, USA: Citadel Press, 1992. (Cited on page 16.)

[45] Academic ranking of world universities @ONLINE. Accessed 2016-08-01. [Online]. Available: http://www.shanghairanking.com/ (Cited on page 16.)

[46] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959. (Cited on page 16.)

[47] igraph. Algorithm space/time complexity. Accessed 2017-12-23. [Online]. Available: http://igraph.wikidot.com/algorithm-space-time-complexity (Cited on page 16.)

[48] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977. (Cited on page 16.)

[49] K. M. Frahm, K. Jaffrès-Runser, and D. L. Shepelyansky, "Wikipedia mining of hidden links between politicalleaders," *The European Physical Journal B*, vol. 89, no. 12, p. 269, Dec 2016. [Online]. Available: https://doi.org/10.1140/epjb/e2016-70526-3 (Cited on pages 23, 24, 25 and 26.)

[50] Oxford dictionaries. Accessed Oct 2017. [Online]. Available: https://en.oxforddictionaries.com/definition/art (Cited on page 27.)

[51] M. Alexenberg, *The Future of Art in a Postdigital Age: From Hellenistic to Hebraic Consciousness - Second Edition*, 2nd ed. Exeter, UK, UK: Intellect Books, 2011. (Cited on page 27.)

[52] J. Fandel, *Pablo Picasso*, ser. Odysseys in Artistry. Creative Education, 2016. [Online]. Available: https://books.google.fr/books?id=EOVZjwEACAAJ (Cited on page 27.)

[53] J. Richardson and M. McCully, *A Life of Picasso: 1881-1906*, ser. A Life of Picasso. Random House, 1991, no. vol. 1. [Online]. Available: https://books.google.fr/books?id=KBlNAAAAYAAJ (Cited on page 27.)

[54] H. Chipp and J. Tusell, *Picasso's Guernica: History, Transformations, Meanings*, ser. California studies in the history of art. University of California Press, 1988. [Online]. Available: https://books.google.fr/books?id=dbhXAQAACAAJ (Cited on page 27.)

[55] J. Richardson, *A Life of Picasso Volume II: 1907 1917: The Painter of Modern Life*. Random House, 2011. [Online]. Available: https://books.google.fr/books?id=4Dw2Uips9B0C (Cited on page 27.)

[56] N. Wolf and U. Grosenick, *Expressionism*, ser. Basic Art Series. Taschen, 2004. [Online]. Available: https://books.google.fr/books?id=MbOdSrOFfpEC (Cited on page 27.)

[57] T. Benson and F. Josenhans, *Expressionism in Germany and France: From Van Gogh to Kandinsky.* Los Angeles County Museum of Art, 2014. [Online]. Available: https://books.google.fr/books?id=PbMhngEACAAJ (Cited on page 27.)

[58] List of painters by name (wikipedia). Accessed Oct 2017. [Online]. Available: https://en.wikipedia.org/wiki/List_of_painters_by_name (Cited on page 28.)

[59] Georges braque and pablo picasso. Accessed Oct 2017. [Online]. Available: https://www.masterworksfineart.com/blog/georges-braque-and-pablo-picasso/ (Cited on page 34.)

[60] Monet and pissarro in london. Accessed Oct 2017. [Online]. Available: http://www.visual-arts-cork.com/history-of-art/claude-monet-camille-pissarro-in-london.htm (Cited on page 34.)

[61] Rouault / matisse, correspondances au musée d'art moderne de la ville de paris. Accessed Oct 2017. [Online]. Available: http://lucileee.blog.lemonde.fr/2007/01/08/rouault-matisse-correspondances-au-musee-dart-moderne-de-la-ville-de-paris/ (Cited on page 34.)

[62] Matisse-rouault : correspondance 1906-1953, "une vive sympathie d'art". Accessed Oct 2017. [Online]. Available: https://critiquedart.revues.org/13275 (Cited on page 34.)

[63] P. Schneider, *Matisse.* Random House Incorporated, 1984. [Online]. Available: https://books.google.fr/books?id=dlMkAQAAMAAJ (Cited on pages 34 and 36.)

[64] D. Wildenstein, *Monet Or the Triumph of Impressionism.* TASCHEN, 2014. [Online]. Available: https://books.google.fr/books?id=mYwAnwEACAAJ (Cited on page 34.)

[65] C. Monet and N. Metropolitan Museum of Art (New York, *Monet's Years at Giverny: Beyond Impressionism.* Metropolitan Museum of Art, 1978. [Online]. Available: https://books.google.fr/books?id=H76Nl_fNgOUC (Cited on page 34.)

[66] Michelangelo and leonardo da vinci. Accessed Oct 2017. [Online]. Available: https://www.michelangelo.org/michelangelo-and-da-vinci.jsp (Cited on page 35.)

[67] Picasso, miró, dalí: The birth of modernity (independant). Accessed Oct 2017. [Online]. Available: https://goo.gl/fuVW4B (Cited on page 35.)

[68] J. Ferrier, *The Fauves: The Reign of Colour : Matisse, Derain, Vlaminck, Marquet, Camoin, Manguin, Van Dongen, Friesz, Braque, Dufy*, ser. Glorious world of art. Terrail, 1995. [Online]. Available: https://books.google.fr/books?id=xmFQAAAAMAAJ (Cited on page 36.)

[69] N. Brodskaia, *The Fauves : The Masters Who Shook the World of Art (Schools & Movements Series).* Parkstone Press, 1996. (Cited on page 36.)

[70] S. Whitfield, *Fauvism*, ser. World of art. Thames and Hudson, 1996. [Online]. Available: https://books.google.fr/books?id=DoxwQgAACAAJ (Cited on page 36.)

[71] M. Antliff and P. Leighten, *Cubism and Culture*, ser. World of art. Thames & Hudson, 2001. [Online]. Available: https://books.google.fr/books?id=KYl2QgAACAAJ (Cited on page 36.)

[72] E. Jones, *The European Miracle: Environments, Economies and Geopolitics in the History of Europe and Asia*, ser. The European Miracle: Environments, Economies, and Geopolitics in the History of Europe and Asia. Cambridge University Press, 2003. [Online]. Available: https://books.google.fr/books?id=iPKIM-JNc_YC (Cited on page 39.)

[73] J. Dittmer and K. Dodds, "Popular geopolitics past and future: Fandom, identities and audiences," *Geopolitics*, vol. 13, no. 3, pp. 437–457, 2008. [Online]. Available: https://doi.org/10.1080/14650040802203687 (Cited on page 39.)

[74] N. C. of Ministers, *Nordic Statistical Yearbook 2014: Nordisk statistisk årsbok 2014.* Nordic Council of Ministers, 2014. [Online]. Available: https://books.google.fr/books?id=fUl0BQAAQBAJ (Cited on pages 54 and 70.)

[75] I. com and U. International Business Publications, *Slovenia Business Law Handbook: Strategic Information and Laws.* International Business Publications USA, 2013. [Online]. Available: https://books.google.fr/books?id=SjdaUe9bDtUC (Cited on pages 54 and 72.)

[76] S. Dhingra, G. Ottaviano, T. Sampson, and J. van Reenen, "The consequences of brexit for uk trade and living standards," London School of Economics and Political Science, LSE Library, LSE Research Online Documents on Economics, 2016. [Online].

Available: https://EconPapers.repec.org/RePEc:ehl:lserod:66144 (Cited on pages 54 and 67.)

[77] F. Fukuyama, "La fin de la démocratie libérale," *Courrier International, N1360*, Nov 2016. [Online]. Available: https://www.courrierinternational.com/article/politique-trump-ou-la-fin-de-la-democratie-liberale (Cited on page 54.)

[78] S. E. Zant. (2017, Dec) Sensitivity analysis: Capturing the influence of geopolitical ties - supplementary materials. Accessed 2017-12-23. [Online]. Available: https://sites.google.com/view/reducedsens (Cited on page 65.)

[79] I. M. of Foreign Affairs. (2012, Mar.) Italy-slovenia: 20 years of fruitful relations, says terzi. Accessed 2017-11-20. [Online]. Available: http://www.esteri.it/mae/en/sala_stampa/archivionotizie/approfondimenti/2012/03/20120327_itslov.html (Cited on page 66.)

[80] B. W. (BBC) Alex Hunt. (2017, Mar.) Brexit: All you need to know about the uk leaving the eu. Accessed 2017-11-20. [Online]. Available: http://www.bbc.com/news/uk-politics-32810887 (Cited on page 67.)

[81] Investopedia website. Gross domestic product - gdp. Accessed 2017-11-20. [Online]. Available: http://www.investopedia.com/terms/g/gdp.asp (Cited on page 67.)

[82] C. M. of Foreign Affairs. Cyprus - UK relations. Accessed 2017-11-20. [Online]. Available: http://www.mfa.gov.cy/mfa/highcom/london.nsf/hcrcl10_en/hcrcl10_en?OpenDocument (Cited on page 68.)

[83] T. O. of Economic Complexity official Website. Accessed 2017-11-20. [Online]. Available: http://atlas.media.mit.edu (Cited on pages 68, 69 and 72.)

[84] F. M. of Economy Treasurer. Les échanges commerciaux entre la France et le Luxembourg en 2014. Accessed 2017-01-15. [Online]. Available: http://www.tresor.economie.gouv.fr/File/411730 (Cited on page 69.)

[85] F. M. of Foreign Affairs and I. Development. La france et le luxembourg. Accessed 2017-01-15. [Online]. Available: http://www.diplomatie.gouv.fr/fr/dossiers-pays/luxembourg/la-france-et-le-luxembourg/ (Cited on page 69.)

[86] F. F. Office. Bilateral relations - luxembourg. Accessed 2017-01-15. [Online]. Available: http://www.auswaertiges-amt.de/EN/Aussenpolitik/Laender/Laenderinfos/01-Nodes/Luxemburg_node.html (Cited on page 69.)

[87] I. Kőrösi, "Austrian and hungarian relations since 1989, the current situation and future perspectives," *Centre for Economic and Regional Studies HAS Institute of World Economics*, Aug 2013. (Cited on page 72.)

[88] T. G. of Economic Complexity. Accessed 2017-11-20. [Online]. Available: http://globe.cid.harvard.edu (Cited on page 75.)

[89] F. F. Office. (2016, Dec) What's behind the China-Taiwan divide? Accessed 2017-11-20. [Online]. Available: http://www.bbc.com/news/world-asia-34729538 (Cited on page 77.)

[90] (2016, May) Pakistan and china relations: 65 years of friendship to strategic partnership @Pakistan Today. Accessed 2017-11-20. [Online]. Available: http://www.pakistantoday.com.pk/2016/05/15/pakistan-and-china-relations-65-years-of-friendship-to-strategic-partnership/ (Cited on page 77.)

[91] A. Stevens. (2015, Apr) Pakistan lands $46 billion investment from china @CNN Money. Accessed 2017-11-20. [Online]. Available: http://www.pakistantoday.com.pk/2016/05/15/pakistan-and-china-relations-65-years-of-friendship-to-strategic-partnership/ (Cited on page 77.)

[92] I. Price. (2016, Sep) Is the us trying to sabotage the china-pakistan economic corridor? @The Diplomat. Accessed 2017-11-20. [Online]. Available: http://thediplomat.com/2016/09/is-the-us-trying-to-sabotage-the-china-pakistan-economic-corridor/ (Cited on page 77.)

[93] N. Z. M. of Foreign Affairs and Trade. United kingdom. Accessed 2017-11-20. [Online]. Available: https://www.mfat.govt.nz/en/countries-and-regions/europe/united-kingdom/ (Cited on page 78.)

[94] S. C. Arj Singh. (2017, Jan) UK will attempt to agree post-brexit trade deal with new zealand, theresa may says @Independent. Accessed 2017-11-20. [Online]. Available: http://www.independent.co.uk/news/uk/politics/theresa-may-new-zealand-trade-deal-bill-english-brexit-downing-street-a7526956.html (Cited on page 79.)

[95] (2017, Jan) UK keen for NZ trade deal after brexit @NewsHub. Accessed 2017-11-20. [Online]. Available: http://www.newshub.co.nz/home/politics/2017/01/uk-keen-for-nz-trade-deal-after-brexit.html (Cited on page 79.)

[96] O. of the Historian. The arab-israeli war of 1948. Accessed 2017-11-20. [Online]. Available: https://history.state.gov/milestones/1945-1952/arab-israeli-war (Cited on page 79.)

[97] U. N. R. 181. Resolution adopted on the report of the ad hoc committee on the Palestinian question. Accessed 2017-11-20. [Online]. Available: http://www.un.org/en/ga/search/view_doc.asp?symbol=A/RES/181(II) (Cited on page 79.)

[98] Egypt and M. H. H. le-heker ha-Mizrah ha Tikhon Veha-Diplomatyah, *The Camp David accords and related documents*, ser. Collection series. Chaim Herzog Center for Middle East Studies & Diplomacy, 1998. [Online]. Available: https://books.google.fr/books?id=oI9tAAAAMAAJ (Cited on page 80.)

[99] Egypt, *The Egyptian-Israeli Peace Treaty, March 26, 1979*, ser. Department of State publication. Department of State, Bureau of Public Affairs, Office of Public Communication, 1979. [Online]. Available: https://books.google.fr/books?id=qGN9NK9go_0C (Cited on page 80.)

[100] J. Sharp, *U.S. Foreign Aid to Israel*, ser. Major studies and issue briefs of the Congressional Research Service: Supplement. DIANE Publishing Company, 2010. [Online]. Available: https://books.google.fr/books?id=kyI7nn6jGPMC (Cited on page 80.)

[101] K. L. P. Mariano, "Two to tango: an analysis Brazilian-Argentine relations," *Brazilian Political Science Review*, vol. 7, pp. 118 – 121, 00 2013. [Online]. Available: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1981-38212013000100006&nrm=iso (Cited on page 81.)

[102] Accessed Feb 2018. [Online]. Available: http://www.quantware.ups-tlse.fr/QWLIB/wikipainternets/ (Cited on page 85.)

[103] Accessed Feb 2018. [Online]. Available: https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2 (Cited on page 85.)

[104] M. Sageman, *Understanding Terror Networks*, ser. Bienestar y proteccion social. University of Pennsylvania Press, Incorporated, 2011. [Online]. Available: https://books.google.fr/books?id=iCoYDUv63L8C (Cited on page 97.)

[105] G. Kepel, *Jihad: The Trail of Political Islam.* I.B. Tauris, 2006. [Online]. Available: https://books.google.fr/books?id=OLvTNk75hUoC (Cited on page 97.)

[106] G. Kepel and A. Jardin, *Terreur dans l'Hexagone. Genèse du djihad français*, ser. Hors série Connaissance. Editions Gallimard, 2015. [Online]. Available: https://books.google.fr/books?id=vesACwAAQBAJ (Cited on page 97.)

[107] M. Sageman, *Turning to Political Violence: The Emergence of Terrorism*. University of Pennsylvania Press, Incorporated, 2017. [Online]. Available: https://books.google.fr/books?id=wNPGDgAAQBAJ (Cited on page 97.)

[108] S. U. project. Mapping militant organizations. Accessed 2017-08-23. [Online]. Available: http://web.stanford.edu/group/mappingmilitants/cgi-bin/ (Cited on pages 97 and 104.)

[109] D. L. Hicks, N. Memon, J. D. Farley, and T. Rosenørn, *Mathematical Methods in Counterterrorism: Tools and Techniques for a New Challenge*. Vienna: Springer Vienna, 2009, pp. 1–5. [Online]. Available: https://doi.org/10.1007/978-3-211-09442-6_1 (Cited on page 97.)

[110] V. Latora and M. Marchiori, "How the science of complex networks can help developing strategies against terrorism," *Chaos, Solitons and Fractals*, vol. 20, no. 1, pp. 69 – 75, 2004, towards a new vision of complexity. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960077903004296 (Cited on page 97.)

[111] S. Dorogovtsev, *Lectures on Complex Networks*, ser. Oxford Master Series in Physics. OUP Oxford, 2010. [Online]. Available: https://books.google.fr/books?id=DrMGQgAACAAJ (Cited on page 97.)

[112] U. L.T.C. Geoffrey Kambere. Financing al shabab: the vital port of kismayo. Accessed 2017-08-23. [Online]. Available: https://globalecco.org/financing-al-shabaab-the-vital-port-of-kismayo (Cited on page 103.)

[113] M. M. @CBS News. (2006, Aug) Mapping militant organizations. Accessed 2017-08-23. [Online]. Available: https://www.cbsnews.com/news/the-taliban-in-afghanistan (Cited on page 104.)

[114] S. Ogata, *The Turbulent Decade: Confronting the Refugee Crisis of the 1990's*. W.W. Norton, 2006. [Online]. Available: https://books.google.fr/books?id=OUVz-xwmhf4C (Cited on page 104.)

[115] (2017, Mar) Yemen crisis: Who is fighting whom? @BBC. Accessed 2017-08-23. [Online]. Available: http://www.bbc.com/news/world-middle-east-29319423 (Cited on page 107.)

[116] (2016, Nov) Middle east conflict: Why is saudi arabia attacking yemen? important questions surround houthis and civil war in sana'a @IBT. Accessed 2017-08-23. [Online]. Available: http://www.ibtimes.com/middle-east-conflict-why-saudi-arabia-attacking-yemen-important-questions-surround-2447121 (Cited on page 107.)

[117] (2010, Dec) Wikileaks cables portray saudi arabia as a cash machine for terrorists @The Guardian. Accessed 2017-08-23. [Online]. Available: https://www.theguardian.com/world/2010/dec/05/wikileaks-cables-saudi-terrorist-funding (Cited on page 107.)

[118] (2016, Dec) Who are Somalia's Al-shabab? @BBC. Accessed 2017-08-23. [Online]. Available: http://www.bbc.com/news/world-africa-15336689 (Cited on page 107.)

[119] C. W. E.M. Saltman, *Islamic State: The Changing Face of Modern Jihadism*. Quillam, 2014. (Cited on page 107.)

[120] D. Helbing, D. Brockmann, T. Chadefaux, K. Donnay, U. Blanke, O. Woolley-Meza, M. Moussaid, A. Johansson, J. Krause, S. Schutte, and M. Perc, "Saving human lives: What complexity science and information systems can contribute," *Journal of Statistical Physics*, vol. 158, no. 3, pp. 735–781, Feb 2015. [Online]. Available: https://doi.org/10.1007/s10955-014-1024-9 (Cited on page 108.)

[121] M. R. D'Orsogna and M. Perc, "Statistical physics of crime: A review," *Physics of life reviews*, vol. 12, pp. 1–21, 2015. (Cited on page 108.)