Network Science Project

The project concerns the study of global migration via network analysis. For each of the year considered a weighted and directed graph is built, reporting countries as nodes and the number of migrants as edges' attributes. A comparison between the metrics of the different resulting networks tries to capture the evolution of the phenomena over a thirty-year period. Specifically, the parsed statistics are the weighted degree, both in and out, and the partitioning in communities. The graphs are first constructed using Python and then visualized and compared with Gephi.

I. Data Sets and Networks Construction

The first dataset employed comes in the form of a csv file, retrieved from a public available database from the World Bank[1]. For each record it reports the country of origin, the country of destination and the number of migrants, in absolute terms, for three different years, 1960, 1980 and 2000. The countries considered are 110, only those most relevant, in order to avoid to obtain over-saturated graphs. For the sake of comprehensiveness the reported migrant stocks are defined as follow: *"Global matrices of bilateral migrant stocks spanning the period 1960-2000, disaggregated by gender and based primarily on the foreign-born concept."*[2] The last statement, the concept of foreign born, makes reference to: "*people who have residence in one country but were born in another country."*[3].

Table 1. Head of the first dataset.

Orig. Name	Orig. Code	Dest. name	Dest. Code	1960	1980	2000
Andorra	AND	Andorra	AND	0	0	0
Andorra	AND	Afghanistan	AFG	0	0	0
Andorra	AND	U.A. Emirates	ARE	0	0	0
Andorra	AND	Albania	ALB	0	0	0
Andorra	AND	Argentina	ARG	0	7	0

The second dataset, downloaded once again as a csv, is retrieved from Kaggle[4] and it collects the geographical coordinates of all worldwide countries, plus those of the USA states. While these former are employed as nodes attributes, these last are not useful for the purpose of the analysis.

Table 2. Head of the second dataset.

latitude	longitude	country
42.546245	1.601554	Andorra
23.424076	53.847818	U.A. Emirates
33.939110	67.709953	Afghanistan
41.153332	20.168331	Albania
-38.416097	-63.616672	Argentina

The first dataset is used to build the networks, after some pre-processing, while the second is exploited for a matter of visualization. Indeed, adding the geographical coordinates as nodes attributes, enables some useful plugins provided by Gephi, as Geo Layout, which displays the countries coherently with their world location.

[1] https://databank.worldbank.org/source/global-bilateral-migration.
[2] World Bank metadata.
[3] World Bank glossary, under: International migrant stock, total.
[4]https://www.kaggle.com/datasets/paultimothymooney/latitude-and-longitude-for-every-country-and-state

Once both the datasets are uploaded on Python they are pre-processed. The missing values, together with some unnecessary columns, are dropped. Moreover, from the second dataset, only those countries of interest, therefore the ones present in the first dataset, are retained. A directed and weighted graph for each period is then finally created by using the package Networkx[5].

II. Network Metrics

The networks are exported from Python as GEXF files and then are uploaded on Gephi for the computation of the metrics of interest. The statistics are the following:

Weighted in-degree: The sum of the weights of all ingoing edges to a specific node.
Weighted out-degree: The sum of the weights of all outgoing edges from a specific node.
Communities: Groups of densely interconnected nodes that are only sparsely connected with the rest of the network. The intuition behind is that a node within a community has a greater probability of connecting to a different node of the same community, with respect to an external one. Their detection is based on the concept of modularity[6], which measures some eventual deviation of a specific subgraph, or partition, from a random configuration.

Here below the graphs, two for each epoch, are reported, along with their metrics. For all the graphs, the sizes of the nodes and of the labels are proportional to their weighted in-degree, while nodes' opacity is representative of their weighted out-degree. Also, the edges' thickness is proportional to their weights. Instead, in those representations visualizing communities, the nodes' colour signals the class membership.

[1960]

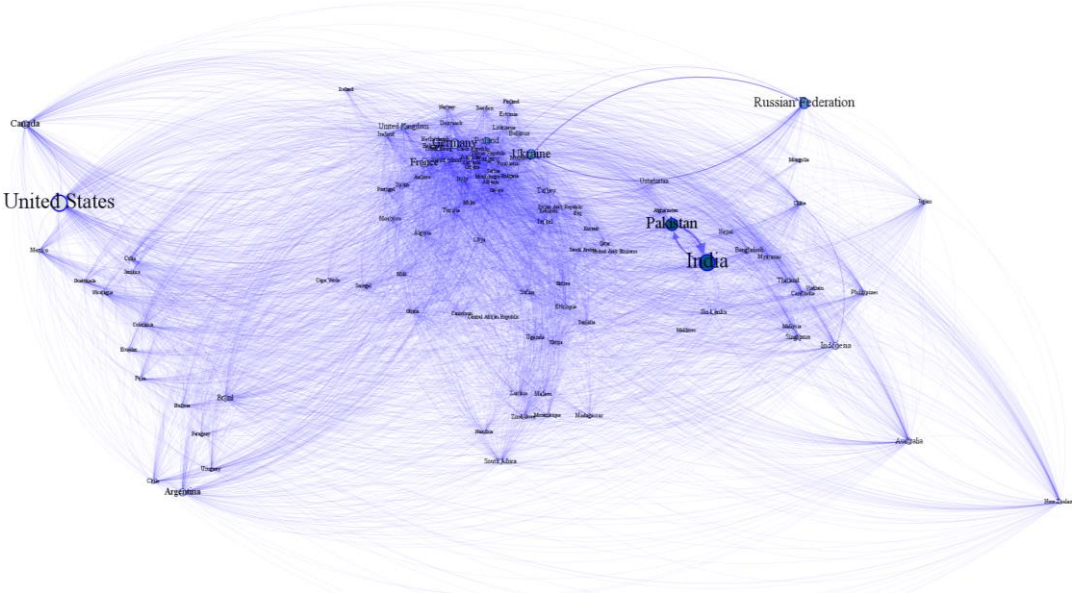

Figure 1. Migration stocks 1960. Order: 106, size: 6120.

[5] https://networkx.org/documentation/stable/index.html.

[6] $Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \rho_{ij}\right)\delta(C_i, C_T)$. The partitioning is given by its maximization, which is however a NP hard problem. The Louvain algorithm, an heuristic technique, can be used instead. References: *Fast unfolding of communities in large networks, Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P10008 (12pp)*.

The countries with the highest weighted in-degree are the United States, India, Pakistan, the Russian Federation and Germany. On the other hand those scoring highest in their weighted out-degree are India, Pakistan, the Russian Federation, Ukraine and Poland. Among all the foreign minorities residing in each of the considered country, the largest, in absolute terms, is the one of Pakistanis living in India, counting 8662538 migrants. The minority formation, according to historians, can be attributable to the partition of India occurred in 1947. Finally, the overall number of international migrants is 73148798.

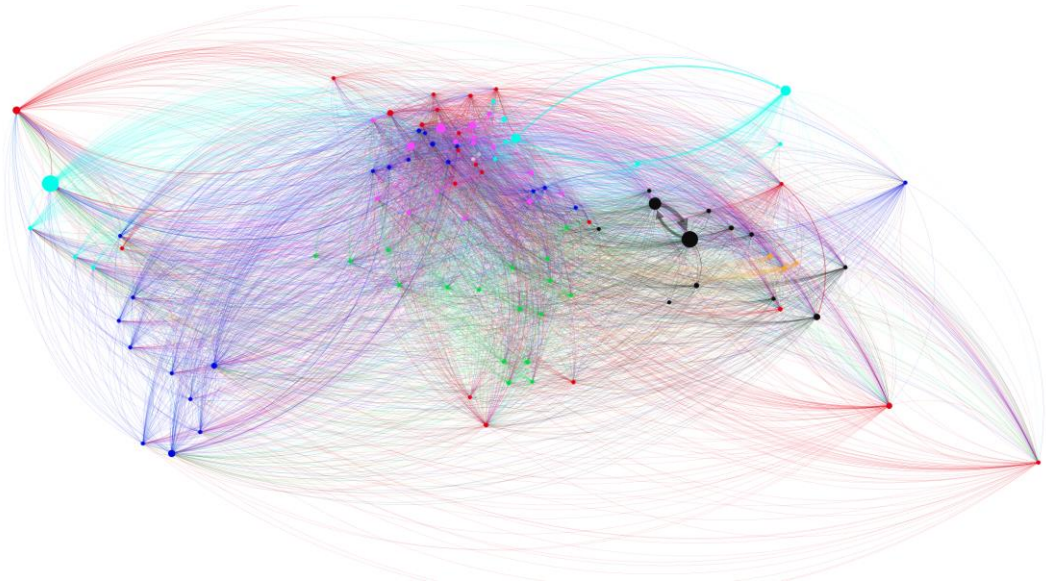

Figure 2. Communities detection, 9 clusters.

[1980]

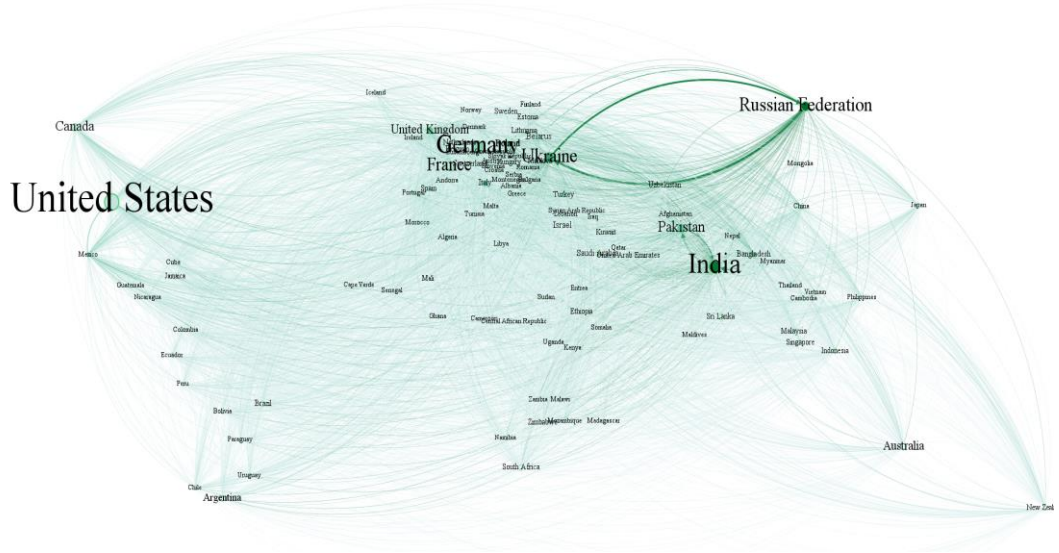

Figure 3. Migration flaws 1980. Order: 106, size: 6939.

The countries with the highest weighted in-degree are, once again, the United States, India, Germany, the Russian Federation and Ukraine. Instead those scoring highest in their weighted out-degree are the Russian Federation, India, Ukraine and Bangladesh. Now the greatest minority is the Russian one residing in Ukraine, counting 4803152 migrants. It should be noticed however that at that time Ukraine was part of the Soviet Union, therefore they were not seen as foreign born. Finally, as expected, the overall number of international migrants has increased in 20 years, reaching an amount of 86783225 persons.

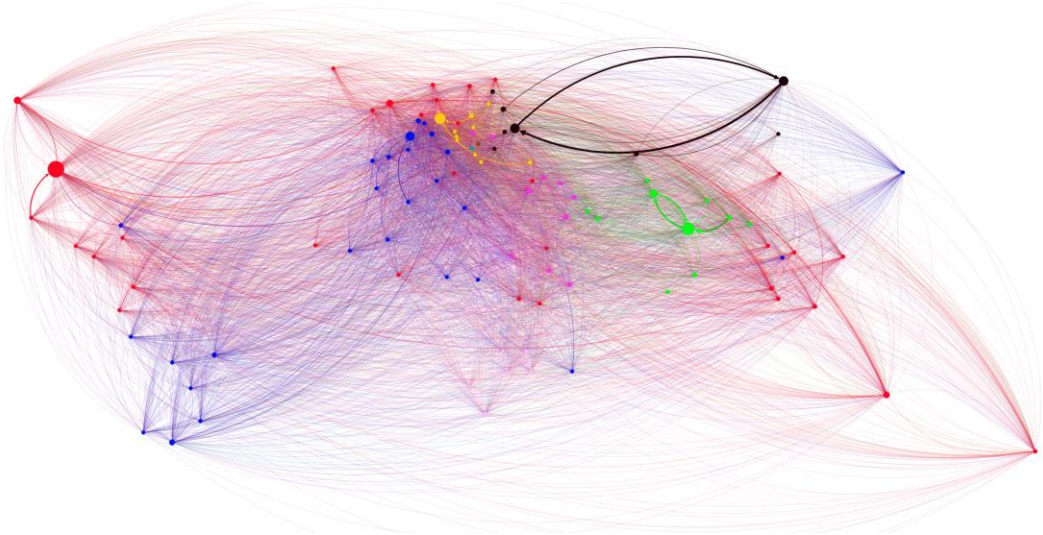

Figure 4. Communities detection, 9 clusters.

Differently from the previous partitioning, the US and the Russian Federation do not belong anymore to the same community. The result is surprising given that the 60s were the period of greatest tension between the two states, so one would expect the opposite. Moreover, it can be noticed that several countries in south Asia are now clustered with those of north America and north Europe, unlike before. The rest seems pretty much unchanged.

[2000]

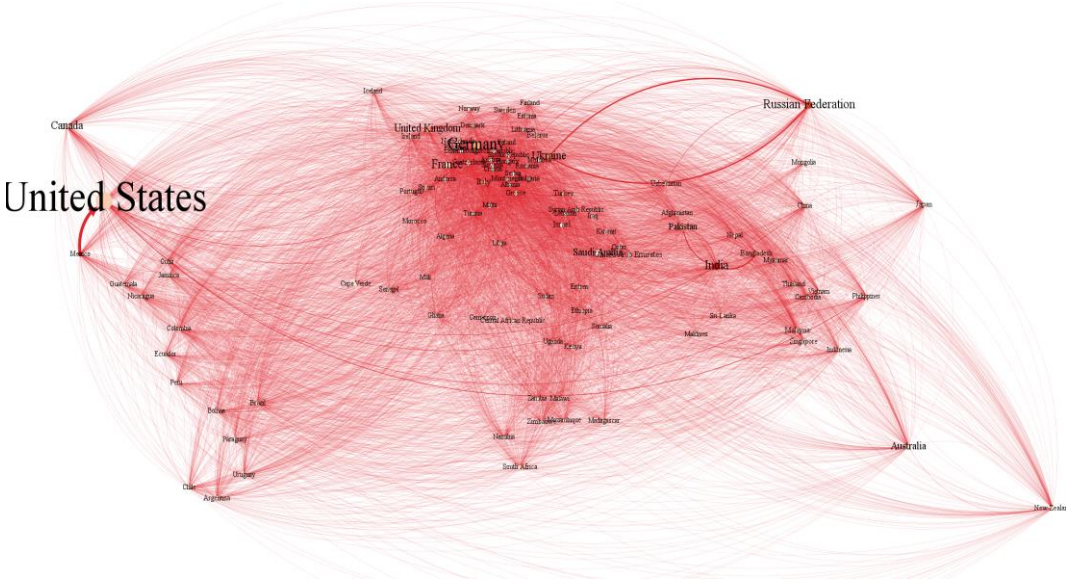

Figure 5. Migration stocks 2000. Order: 106, size: 7889.

The countries with the highest weighted in-degree are, in order, the United States, Germany, the Russian federation, India and France. Those with the highest out-degree are Mexico, India, the Russian Federation, Ukraine and Poland. The country hosting the greatest minority is the USA, counting 9367910 migrants from Mexico. The total amount of international migrants is 113511192, about 1.8% of the global population.

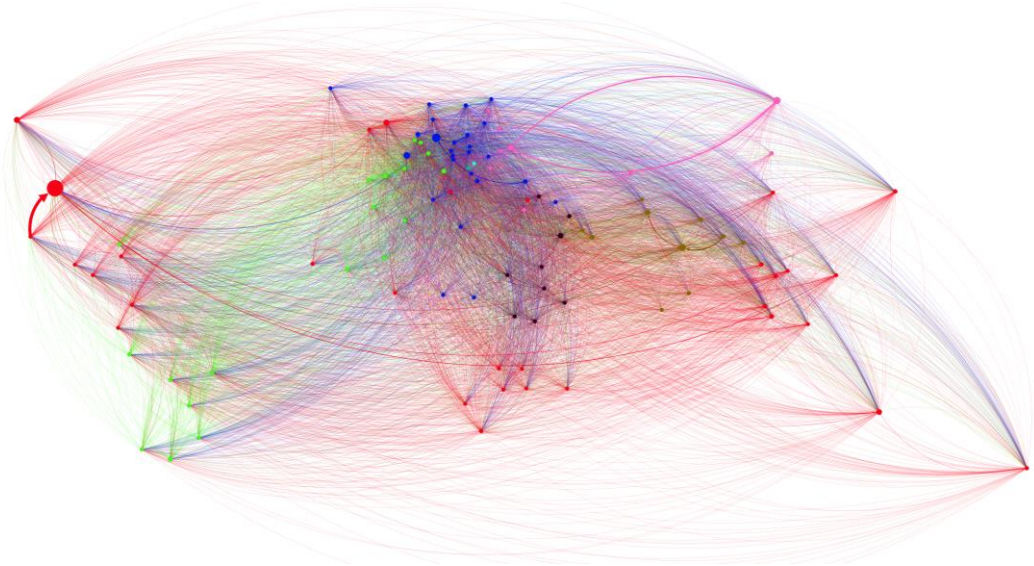

Figure 6. Communities detection, 8 clusters.

The partitioning in this case does not differentiate much from the very last.

A reasonable improvement to the analysis could be given by considering the number of migrants in relative terms, proportionally to the population size of the country of destination, or of origin, depending on the information of interest. That would be more meaningful in order to better capture the evolution of the international migration phenomena.