Missing Migrant Link Analysis with Gephi

Israel Nolazco
Introduction to Data Mining and Analytics
Lewis University
Romeoville, Illinois
israelnolazco@lewisu.edu

***Abstract*— Link Analysis is the method in which two entities are compared to one another in order to find some sort of connection. This approach typically aims to find relatively simple connections within data. Depending on the size of the data, the analysis can be done by hand [1]. This document will utilize the application Gephi in order to conduct Link Analysis regarding the Missing Migrant Project data set with the year of 2018. The data set takes into consideration the whole world but breaks down certain hotspots by regions in order to properly identify its community and the impact on the region. This document will present the analysis between region and month of the year by taking into consideration the number of dead migrants within each analysis. Additionally, this document will present the potential reasoning behind different findings and will ultimately provide an insight regarding the death of migrants while taking their journey.**

Keywords

Link Analysis, Data Analytics, Missing Migrant, Gephi

I. INTRODUCTION

The journey of a migrant can be cruel and painful. Every year thousands of migrants die fleeing their homes, hoping to arrive in a safer area. The Missing Migrant Project is an organization that tracks the death of migrants, including asylum-seekers and refugees [2]. The organization gathers information from government officials, medical examiners, media reports and any other source available at hand. It is crucial to understand that the data analyzed in this document may not be accurate; given the reliance from multiple sources and lack thereof fact-checking. However, the dataset gathered from the Missing Migrant Project provided roughly sixteen thousand entries. The information spams from January to December of 2018. In addition to the data set, the analysis was conducted through the Gephi application, an open-source interactive network exploration and visualization tool for Windows, Linux, and Mac OS X. Ultimately, it is a data analysis tool [3]. The importance of this application is key in order to visualize the results of this data and further comprehend the impact each migrant has on the area they tried to take their journey in.

II. METHODS

In order to understand link analysis and its true potential, there are some keywords needed to explain in order to proceed [4].

Node: A point that represents an object.

Link/Edges: The relationships or connections between nodes. The link may also include associated properties.

Weight: The frequency in which each link is seen in the document.

As previously mentioned, the dataset utilize in this document is provided by the Missing Migrant Project. The dataset contained twenty-two-column and sixteen thousand rows. However, given that we are connecting the regions to the month of the year and factoring in the number of dead; that lead the data set to be trimmed down to only three columns. The nodes are identified as either the month of the year or the region in which a migrant died. The weight is then represented as the total number of the dead by each link (region and month). The source, in this analysis, is considered the region and the target will be considered the month of the year. The reason behind this modification on the data is to further emphasize which region/month of the year is considered the deadliest to migrate and to further investigate the reason behind such discovery. together. Lastly, given the direct connection between the region and month type of link between each node connection is a Direct one. Meaning, the month nodes will not be linking to any other piece of data. Rather, it is best to consider them an endpoint.

Once the data was successfully imported into Gephi, it is then discovered the weight for each edge is the only value at one. Therefore, a separate excel sheet utilize in order to properly filtrate each month and region. Afterward, the dead total will be calculated in order to assign a weight value to the data. Therefore, while analyzing the data a thicker will highlight the link between each node, making the analysis faster.

III. Analysis

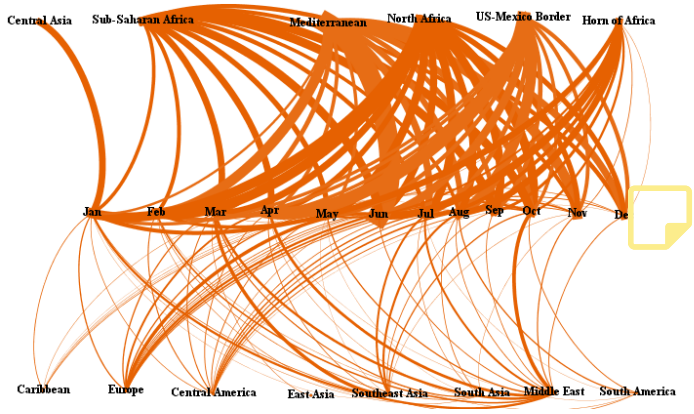

Figure 1. Link Analysis, Missing Migrant Report. This image s hows the initial view between the region and month of the year are associated.

At first glance our image shows a clear pattern between the top regions. Central Asia, Mediterranean, North Africa, etc. Those regions see the death of migrants from any time in the year. H owever, this image also shows a focal point to Central Asia. T he region itself seems to be only interested in migrating in Jan uary and when that occurs it could lead to a high number of de ad migrants.

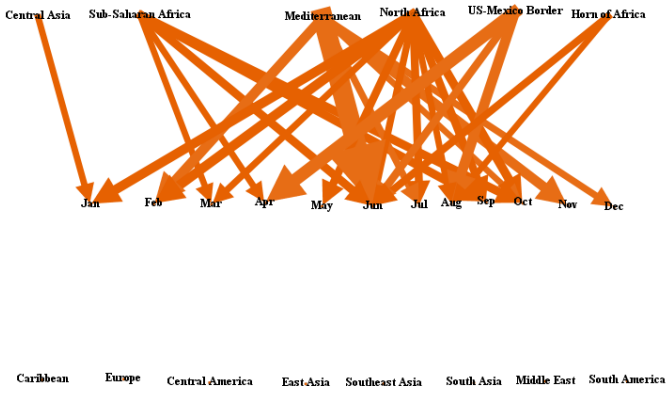

Figure 2. Link Analysis, Missing Migrant Report. This image is created by setting a filter to the edge weight up to an index of fifty. That is to say, the regions point to the time of the month where more than fifty people passed away.

At this point in the analysis, it clear the months of June till October are the months where migrants take on their journey and could potentially face a higher risk of dying. However, it is not yet clear what potential regions foresee a higher risk than others.

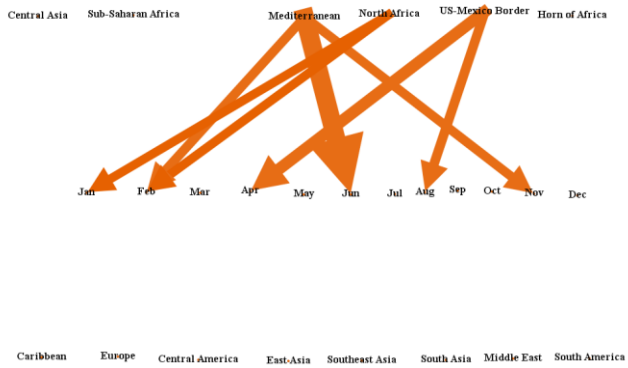

Figure 3. Link Analysis, Missing Migrant Report. This image is created by setting a filter to the edge weight up to an index of seventy-five.

Figure 3 provides an excellent example of an extreme case. The data shows at least seventy-five people have died in the US-Mexico Border. Although, the deadliest locations seem to be North Africa, US – Mexico Border and the Mediterranean. Why?

Figure 4. Google maps image of Northern Africa. https://goo.gl/maps/d8VyMzKbWr132HEd6

As of June 2019, there are about 6.7 million Syria Refugees. Most of which had to cross the Mediterranean Sea. Additionally, during the summertime; the sea waters are relatively warmer than usual. It then possible to assume Syrian Refugees, took advantage of the weather and made their journey to Europe in order to become refugees. However, as seen in Figure 4. The Mediterranean Sea is huge and can be very deadly to go across without the proper equipment.

As for Norther Africa, Figure 4 shows Morocco and Spain being relatively closer to one another. It no surprise for migrants to find Morocco pleasant entry point for migrants. However, Despite the risk and danger of migrating out of Morocco are so intent on leaving that they are willing to leave their families and sometimes risk their lives to change their futures [5].

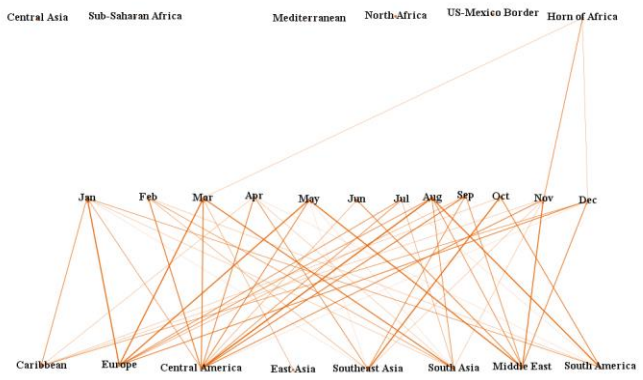

Figure 5. Link Analysis, Missing Migrant Report. This image is created by setting a filter to the edge weight up to an index of ten.

Lastly, the analysis naturally took a closer look at the other side of the spectrum. Figure 5. Shows the regions that see migrants dying in the single digits. However, looking closer at the image; it's noticeable that although the numbers are low Asian regions typically see deaths at almost every month of the years. This indicates that migrants are constantly migrating to Asian regions throughout the year. Additionally, let's keep in mind the regions suggest warm weather. This shows migrants are active throughout the region.

IV. Conclusion

At first glance our image shows a clear pattern between the top regions. Central Asia, Mediterranean, North Africa, etc. Those regions see the death of migrants from any time in the year. However, this image also shows a focal point to Central Asia. The region itself seems to be only interested in migrating in January and when that occurs it could migrants face a platitude of challenges while making their journey. Regardless of legal status a risk carries with every action. As discover in this document, real-world scenarios provide insight to further verify and justify data analysis. Specifically, regarding Link Analysis. This document does not intend to peruse a political agenda or dialog, but rather emphasizes the idea of Link Analysis and its potential tool to analyze real-world scenarios. Additionally, the data analysis conducted only three out of the twenty-one columns. The data could be further broken down to analyze what age groups suffer from migration or further verify specific locations with the region discuss in the document to prove aid or assistance to those migrants. All of these and many other possible options are open given the limitless potential from Link Analysis.

References

[1] McCue, C. (2019). Link Analysis - an overview | ScienceDirect Topics. [online] Sciencedirect.com. Available at: https://www.sciencedirect.com/topics/computer-science/link-analysis [Accessed 10 Oct. 2019]

[2] Missingmigrants.iom.int. (2019). Missing Migrants Project. [online] Available at: https://missingmigrants.iom.int/downloads and https://missingmigrants.iom.int/methodology [Accessed 10 Oct. 2019].

[3] Gephi.org. (2019). Features. [online] Available at: https://gephi.org/features/ [Accessed 10 Oct. 2019].

[4] Doc.arcgis.com. (2019). Link analysis—Insights analyze | Documentation. [online] Available at: https://doc.arcgis.com/en/insights/analyze/link-analysis.htm [Accessed 10 Oct. 2019].

[5] Goldfarb, L. (2019). Fenced Off From the Hopes For a Better Life. [online] us news. Available at: https://www.usnews.com/news/best-countries/articles/2019-10-07/children-in-morocco-eye-migration-to-spain-for-a-better-life [Accessed 10 Oct. 2019].