Title: Analysis of refugee migration

Done by

19BCE2339-Kabhilan.S
19BCE2313-Kolla Srinivasan Pragathi

AIM:

Perform a social network based analysis on refugee migration between different countries by analysing the given dataset through community detection models and analyse the results.

ABSTRACT:

The present work of our paper is regarding refugee analysis,as migration of refugess is based on community-structured social networks with their individual nodes without having any interactions between them .We implelemt it in adaptive social networks.so to find or to analysis the refugee migration techniques one needs to follow certain methodology to have a clear aspect on refugee migrations to where the people will migrate.

There are different centralites we used for understanding the relationship between the actors.We are going to use some community detection algorithms in this paper to understand the refugee how thedecision making of the refugee migrant to a particular destination is going on. The main usage of this work is to understand the refugee migration in most of the countries. Our analysis would be helpful fo many international organizations for providing help to the individuals for good life in migration.

INTRODUCTION:

The migration of an individual from one society to another society depends on different aspects of the social interaction of the network. The Migration within different societies or countries also affect the markets present in the respective society. This is the reason for discouraging the migrants to enter or exit a current society. Because, the development of an economy depends on the migration across different society.

But, when we come to the topic refugee migration, the migrants are refugees in this case. A refugee is a person who migrates to a different place due to the in-acceptance of the society or the conflict between the public and a particular group of people due to various factors, or also could be the lack of basic sanity in their place of origin. So, in these types of cases there is a possibility of the refugee to be migrated, in some cases the international help is required in order to save the lives of refugees across the world.

So this study will help to get a clear vision where would the refugees migrate ,by this analysis it would be helpful for international or local organizations to understand their migration and help them out.

The data set is taken from a survey called Humanitarian migration by an international survey agency named Migration policy institute and it is available as an open source in online. There are 9 attributes in data set. The nine attributes are year, source, value, target, latitude_dest,longitude_dest, latitude_org, longitude_org, weight. We have eliminated the unwanted data and taken only the useful data that we need i.e. **year -** in which all the migrations happened, **source** of the refugee, **target** or destination the refugee has migrated into, where the total number of refugees who have migrated from source county to destination country is given, its range is given between 0 to 1.

The latitude and longitude are not considered as we are working on a graph irrespective of the geographical positions. The source of the refugee is about the country from which the refugee has left the place, the destination shows to the country to which the refugee has been migrated.

The main theme of this graph is to identify the possibility of a migrant refugee to go to the targets present all around the world. The data set is quite large enough as it contains above 9000 rows where each row shows the data about the year in which the particular number of refugees have migrated from one source country to the particular destination country. We have used simplification technique to simplify the large amount of data, so that we can develop a simple understandable graph to show the different communities identified in the graph.

Each community shows the possibility of particular migrant to go to the counties present in the same community.

To achieve this, we have first started measuring the different centralities by generating a directed graph. The centralities are used to betweenness centrality and closeness centrality, we have also used the cluster coefficient to know the transitive relations and also found shortest paths. After studying the data set we have moved unto the different community detection algorithms. These algorithms detect communities based on its own method of approach to detect the communities where a migrant present in one source country is having a more possibility to migrate within the same community. The algorithms used in this research are fast greedy algorithm, walktrap algorithm, spinglass algorithm, label propagation algorithm and girvan-newman algorithm.

In this paper we have presented different types of community detection methods to understand the refugee migrants. This study so that it could be useful for refugees to get help. There are many refugees in this world who are still waiting to get help. Even though many international organizations are present and helping thousands of refugees every year, it is still not being sufficient as the number of refugees are still increasing in every country all across the globe. There have been refugees waiting more than 10 years and still not getting any help due to the lack of sufficient funds or could be any reason for not being accepted by any society or country. few countries now like America not encouraging any kind of migration into their country. It could be of many reasons as it can affect the development or security of the economy. We hope this research could help the organizations and thus save humanity.

2.Literature Survey:

The study of refugee migration is made with respect to individuals present in the country called Rwanda. In this country, the economy position is very poor with very low GDP due to very low natural resources. This caused many individuals present in the same country to migrate to different countries seeking a new good life. Therefore, a survey was made to know why the particular individual has chosen a particular destination country to migrate. The survey used data from the mobile phone operator present in that country, and construct a social network

based on the connections with the individual. The data is collected before few months from the date when the particular refugee migrated. The method used in this paper is strategic cooperation model. The model is based on the connections made by the mobile by the particular migrant during the survey. In this survey it is found that a person who is having more contacts in the destination is likely to migrate to that place when compared to the refugee who is having relatively less contacts in the particular destination. The total survey is to know the reason behind the refugee has chosen the particular destination based on his social network connections with individuals present in other places.

Social network of a particular individual might reflect the decision of choosing the destination, but there could be other parameters involved in the decision making of the individual, one of such parameters is the gender. The decision of the individual migrant is studied based on their gender, position among his house hold such as mother, father, son, daughter and also the leadership of the household is also considered. The study is conducted in the year 2001 to know whether the gender played a role in making decisions to migrate to a particular country or society. The study has found that the decision does matters according to the gender with respect to few cases or countries. For example, the migrants from Dominican Republic to the US differ in their savings and expenditure behavior due to the matter of fact that men go to work for earning money, while women will take care of their household work. The migration between the rural-urban place of the Thailand had shown that even though men and women go for earning money, women are expected to earn more than that of men. Thus, this paper has provided an insight that how can gender discrimination can affect the decision of the migrant to choose their destination.

But whatever the parameters are involved in the decision of the individual, the main reason they are likely to migrate is due to lack of good conditions in their respective places. By considering this perspective in this paper [3], the migration of an individuals is studied from six regions in Asia, Africa and Europe, based on the poverty relation with migration. The survey is based on the data from the migrants about their household position and their contacts present in the destination place where they have migrated. All types of local and International migrations are present in this data. The survey is made to know what factors made the individual to migrate from the less developed society consisting of poverty in

most of the places. The study made to discover the decision of migration took place based on their gender. The Males took the decision based on his contacts present in the destination region, while the females took decision In the cases where they take decisions about the household wealth. The results showed that the decisions made by the male is a lot different from that of females. Individual interaction with the migrants found that about 75% of the females out of all female migrants have decided to migrate based on the work. Whereas, the 92% of the males have decided to migrate to work.

In other study, from individuals in Bangladesh is done to find the factors made him to migrate. Secondly, the return migration of the individual who have migrated from south Italy to the Germany previously is also studied. The main theme of this paper is to find why the individual are migrating and return migrating from one county to another. In the first case the migration factors are based on lack of wellbeing due to poverty in their area, while in the second case the migration is based on the social per capita, as it had negative impact of the of the Italian migrant in Germany.

In order to help a refugee, costs involved in helping them out is also one of the important and major one to consider for the organizations. The management of migration has also come to a stable state according to a study, where it analyzed to major networks in Asia and Europe. Many restrictive rich Asian countries are helping refugees recently. This study mainly focused on feasibility of flight to help refugees with respect to different governments. In the same way the culture of the society also depends to find whether it is feasible for that refugee or not. A study [6] has found that refugees in India has undergone more psychological distress than refugees in Canada. To reduce such distress conditions among refugees social workers have to study the culture and past experiences of the refugees and migrate in groups. And, also medical treatment should also be considered if the refugees are attacked with any type of cancer, flue or anything else.

3.Community Detection:

Communities are also called as clusters and detecting communities is known to be Clustering. But the main difference is that nodes in this network are connected to other nodes by edges. when it comes to clustering, the data points are not embedded in a network.

Analysis of Communities in Social Networks is very important for many reasons. When it comes to our paper, we found out some communities using various community detection algorithms in igraph. From the date set and communities formed we can easily find out to which countries refugees migrate and the set of countries (a particular community) to which they move.

We choose a Data Set containing the information regarding Refugee Migration from one country to other in this period (2019). Out of Nine attributes (Year, Source, Value, Target, latitude_destination, longitude_destination, latitude_orgin, latitude_origin, weight) in the data set we chose only three(Year, Source, Target) attributes. Generally, when the number of people leaving a particular country indicates that it is not a good Country and people are forced to leave that country. Whereas when there are more and more people entering a particular Country, indicates that it is a good country and also indicates high chances of better living.

Now let us relate the above mentioned with our paper. Consider a Centrality Measure (Degree Centrality) and assume two countries namely Country A and B. If the in degree of a particular country is more Indicates that more and more Refugees are entering that particular country and vice versa if a particular country has higher out degree which indicates bad conditions of living in that country.

Community detection algorithms currently we implemented five Community detection algorithms in our paper using igraph:

Edge betweenness based community detection, Community structure based on statistical mechanics, Community structure based on random walks, Community structure via greedy optimization of modularity**,** Community detection based on label propagation.

Edge betweenness based community detection:

In this method we find communities based on betweenness of edges. M. Girvan and M. Newman invented this algorithm. This community detection algorithm is a hierarchical decomposition process. Firstly, we find betweenness scores of edges and we start removing in the descending order of their betweenness scores. This method gives us out good results, but because of the computational complexity of this algorithm as it find's betweenness for all nodes and betweenness should also be found out again after the removal of a particular node and. This is a slow Process.

Detected Communities based on edge betweenness:

Code for Graph:

```
library(igraph)
set.seed(1492)

l <- layout.fruchterman.reingold(ig, niter=5000, area=vcount(ig)^4*10)
networ <- read.csv(file.choose(),header=FALSE)
g <- graph.data.frame(networ,directed=TRUE)
coords = layout_with_fr(g)
par(mar=c(0,0,0,0))
windows(heigth=50, width=50, record=TRUE, rescale="fit")
g
c3 = cluster_edge_betweenness(g)

modularity(c3)
plot(c3, g, layout=coords,vertex.size=1,edge.arrow.size=0.5,vertex.label.cex=0.75,
     vertex.label.family="Helvetica",
     vertex.label.font=2,
     vertex.shape="circle",
     vertex.label.color="black",
     edge.width=0.5,
      edge.curved=0)
#3dplot for better clarity
rglplot(g,vertex.color=membership(c3),layout=coords,
     edge.arrow.size=0.5,
     vertex.label.cex=0.75,
     vertex.label.family="Helvetica",
     vertex.label.font=2,
     vertex.shape="circle",
     vertex.size=1,
     vertex.label.color="black",
     edge.width=0.5)

plot_dendrogram(c3)
```

Grivan Newman

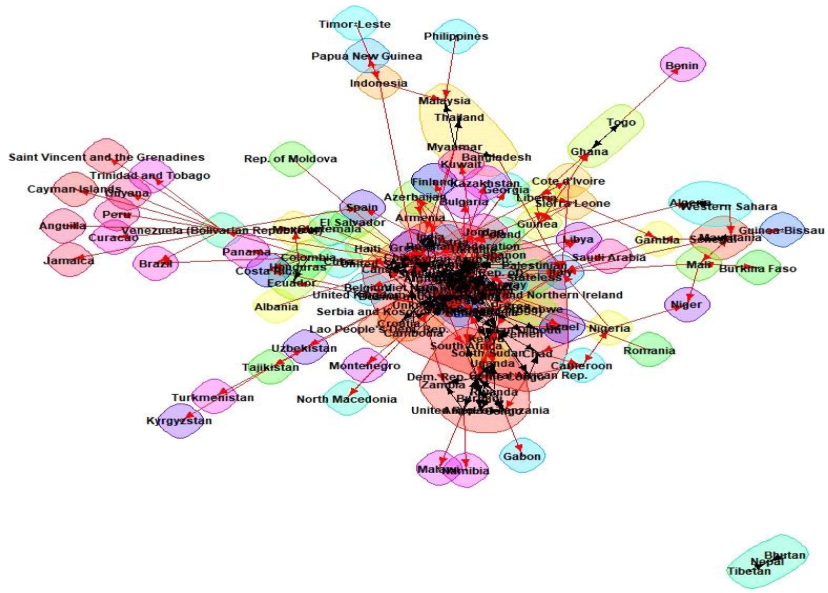

fig1:graph representing communites detected using Grivan Newman algorithm

Dendogram-Grivan Newman

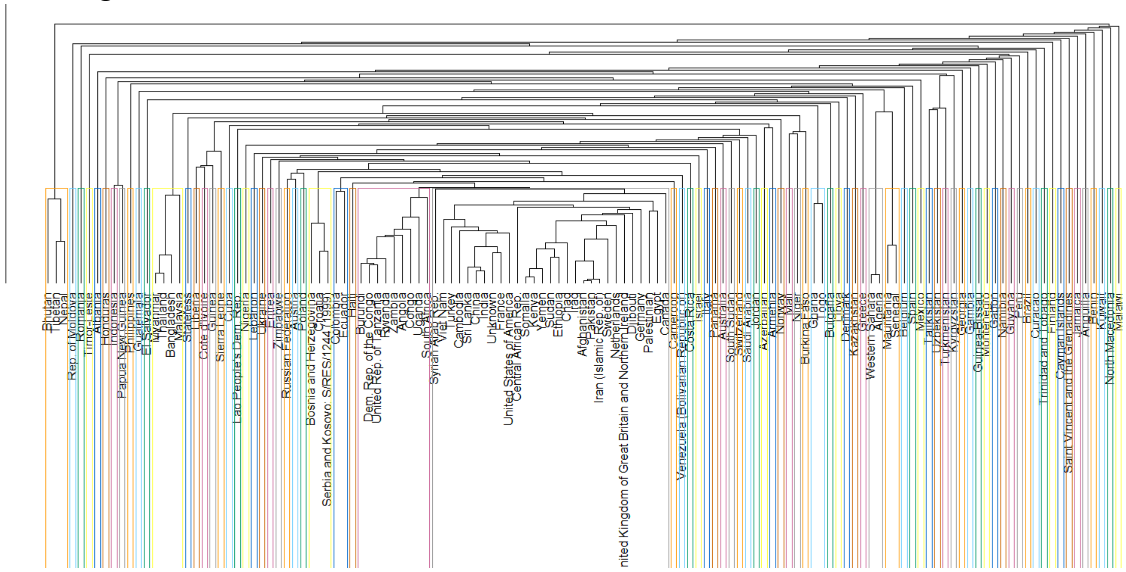

fig2:Dendogram for communities[in Grivan Newman algorithm]

In the above dendogram**(fig2)**, x axis represents the countries and y axis represents the average distance between the two clusters ,from dendogram we can observe that similarly colored lines denote one community[Countries between the similarly coloured lines are of the same community]

From the above **fig1 and fig2** we can see that communities are colour coded for better understanding. observe that larger colour coded community which indicates that people of particular countries in that larger community tend to move within that larger community.

The countries which are geographically closer to each other tend to stay in same communities when we detect communities based on edge betweenness as statistically poverty is a major cause in mass migration of people and refugees in general face difficulty in air travel due to the security precautions in airports and such.

This point is further proven by the graph generated and the dendrogram.

Community structure based on statistical mechanics:

Another algorithm used for community detection is Spin-glass which is based on potts model . In this, each particle can be in any one of spin states c(cytochromes), and the interactions at edges between the particles says which pair of countries would prefer to stay in the same spin state and which ones would rather have different spin states. Based on the length we provided to it will be ended up with communites,at the end we may not find all c communites it would be less than that because some of spins states may become empty. It is not sure that nodes in completely remote parts of the networks have different spin states.

Detected Communities based on spinglass:

Code:

```
library(igraph)
set.seed(1492)


networ <- read.csv(file.choose(),header=FALSE)
g <- graph.data.frame(networ,directed=FALSE)
g2<-decompose.graph(g)
coords = layout_with_fr(g2[[1]])
par(mar=c(0,0,0,0))
windows(heigth=50, width=50, record=TRUE, rescale="fit")
memb <- spinglass.community(g2[[1]])
# computing modularity
mod <- as.character(format(modularity(g,membership(memb)), digits = 2))
# visualizing the result of dividing the network into communities
plot(memb,g2[[1]],edge.arrow.size=0.5,
     vertex.label.cex=0.75,
     vertex.label.family="Helvetica",
     vertex.label.font=2,
     vertex.shape="circle",
     vertex.size=1,
     vertex.label.color="black",
     edge.width=0.5 , layout = coords,edge.curved=0)
```

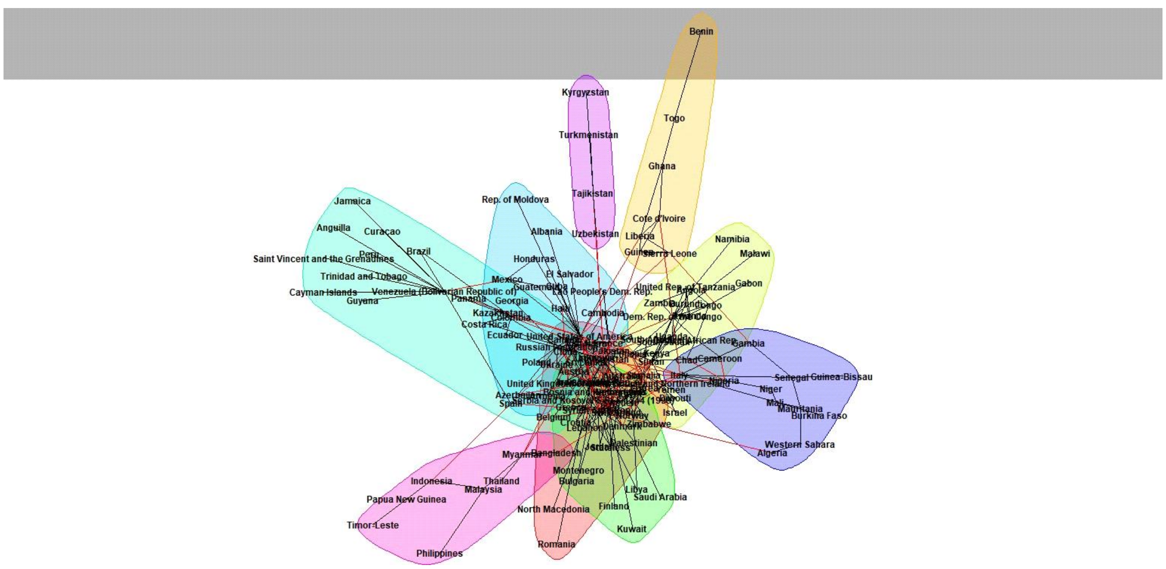

fig3:graph representing communities detected using spinglass alogrithm

From the above graph (**fig3**), we can observe that there are three major communities colour coded which mean that the people in those countries move in that particular larger community.

Community structure based on random walks:

Walktrap is a community detection model based on random walks. If random walks is performed on the graph, then they mostly stay within the same community because as there will be only few edges which are forming community.Walktrap performs short random walks of 3-4-5 steps considering anyone of its paramteres and later on the results that arrived are used to merge separate communities. That is in a bottom-up manner. We use the modularity score to select where to cut the dendrogram. The approach is slower than fastgreedy but it is quite more accurate than fastgreedy.

Detected Communities based on random walks:

Code:

```
library(igraph)
set.seed(1492)
networ <- read.csv(file.choose(),header=FALSE)
g <- graph.data.frame(networ,directed=FALSE)
coords = layout_with_fr(g)
memb <- walktrap.community(g, steps = 4)
# computing modularity
mod <- as.character(format(modularity(g,membership(memb)), digits = 2))
# visualizing the result of dividing the network into communities
plot(memb,g,edge.arrow.size=0.5,
     vertex.label.cex=0.75,
     vertex.label.family="Helvetica",
     vertex.label.font=2,
     vertex.shape="circle",
     vertex.size=1,
     vertex.label.color="black",
     edge.width=0.5 , layout = coords,edge.curved=0)
plot_dendrogram(memb)
text(1,1.2,paste("Modularity = ", mod, sep = ""))
```

GRAPH:

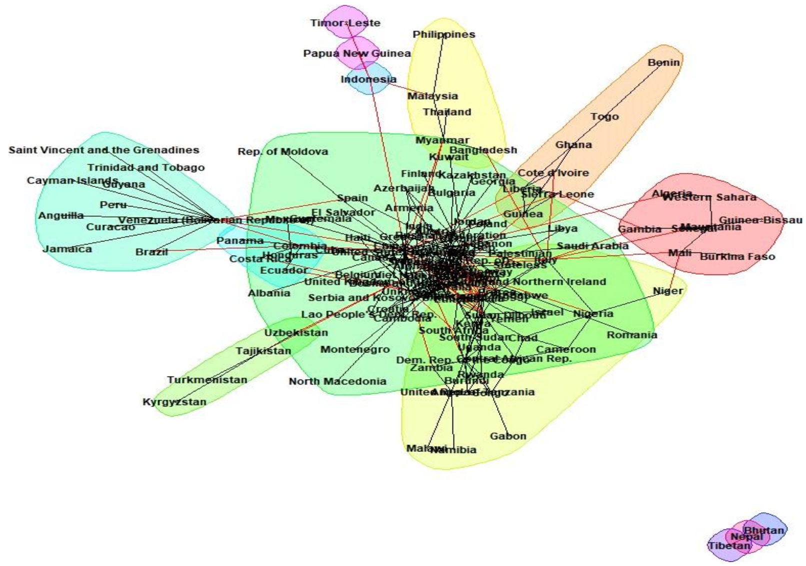

fig4:graph representing communities detected using Walktrap algorithm

Dendogram-Walktrap Algorithm

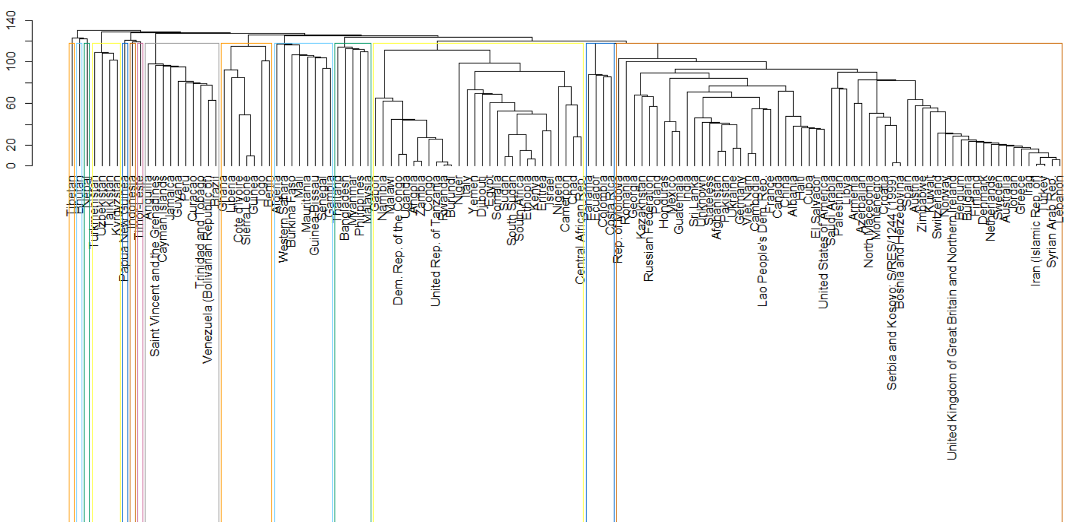

fig5:dendogram for communities [in Walktrap algorithm]

In the above dendogram**(fig5)**, x axis represents the countries and y axis represents the average distance between the two clusters ,from dendogram we can observe that similarly colored lines denote one community[Countries between the similarly coloured lines are of the same community]

From the above **fig4 and fig 5** we can see that, the communities formed were the same for data supplied of different years with consistency. That is because in some counties, due to pre existing conditions like prolonged political instability, long standing wars, geographically deficient in necessary resources ; people tend to migrate from those countries in a serial manner.These type of migration among are detected by walktrap algorithm with minimal which is evident on anlysis of the graph.

Community structure via greedy optimization of modularity

One of the hierarchical approach is fastgreedy, but it is a bottom-up manner instead of a top-down manner. It tries to improve a function modularity in a greedy manner. At first, every vertex here belongs to a separate community, and then communities are merged iteratively such that each merge relents the largest increase in the current value of modularity. When the modularity stops increasing, the algorithm stops, so it gives a grouping and a dendrogram. This method is fast and is tried as a first approximation because it doesn't have any parameters to adapt. Communities below a given size threshold will always be merged with other communities as it suffers from a resolution limit.

Detected Communities based on Fast greedy:

CODE:

```
library(igraph)
set.seed(1492)


networ <- read.csv(file.choose(),header=FALSE)
g <- graph.data.frame(networ,directed=FALSE)
coords = layout_with_fr(g)
par(mar=c(0,0,0,0))
windows(heigth=50, width=50, record=TRUE, rescale="fit")

# fastgreedy community detection
g=simplify(g)
memb <- fastgreedy.community(g)
# computing modularity
mod <- as.character(format(modularity(g,membership(memb)), digits = 2))
# visualizing the result of dividing the network into communities
plot(memb,g,edge.arrow.size=0.5,
     vertex.label.cex=0.75,
     vertex.label.family="Helvetica",
     vertex.label.font=2,
     vertex.shape="circle",
     vertex.size=1,
     vertex.label.color="black",
     edge.width=0.5 , layout = coords,edge.curved=0)
```

GRAPH

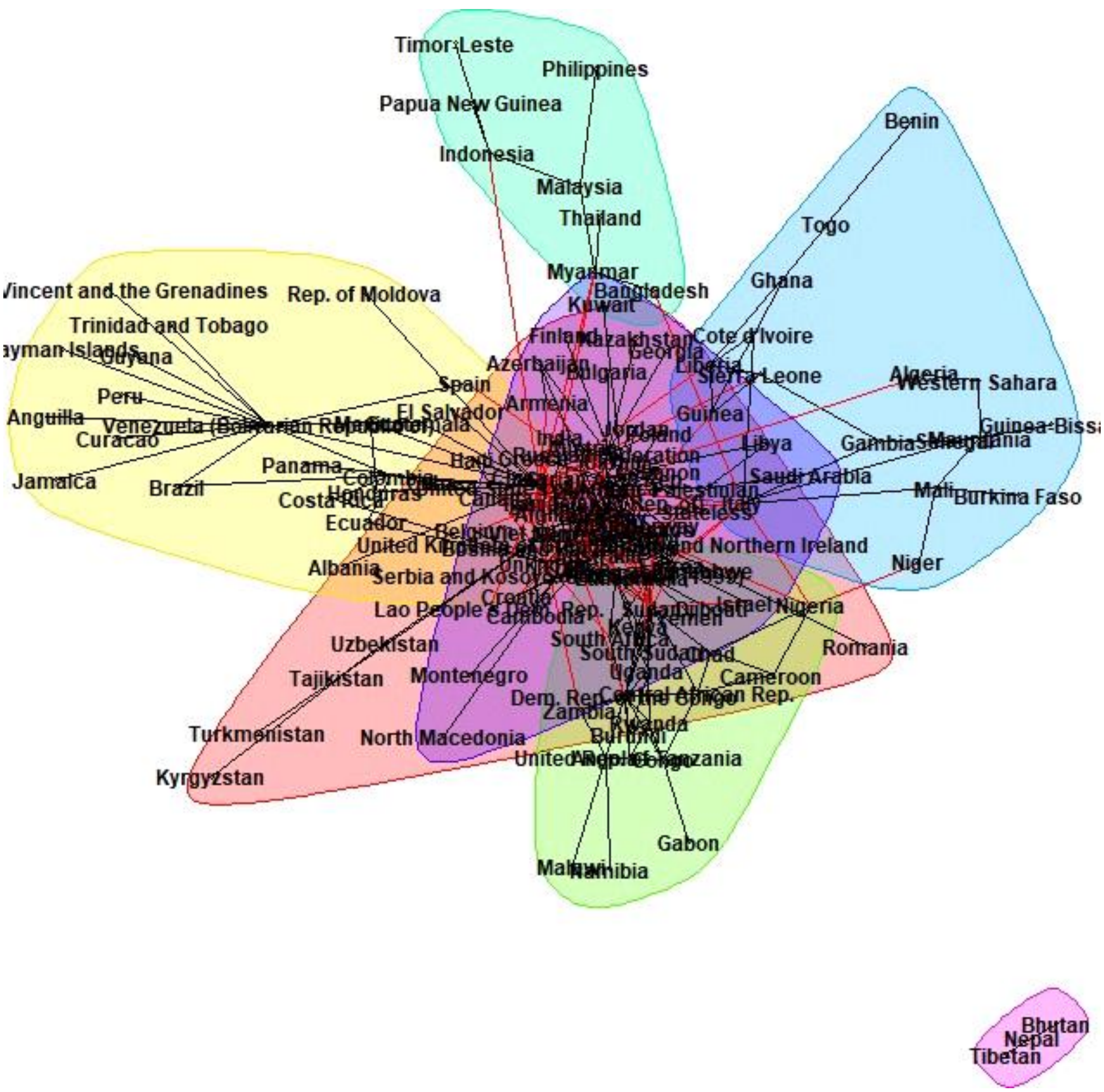

fig6:graph representing communites detected using FastGreedy algorithm

Dendogram-FastGreedy

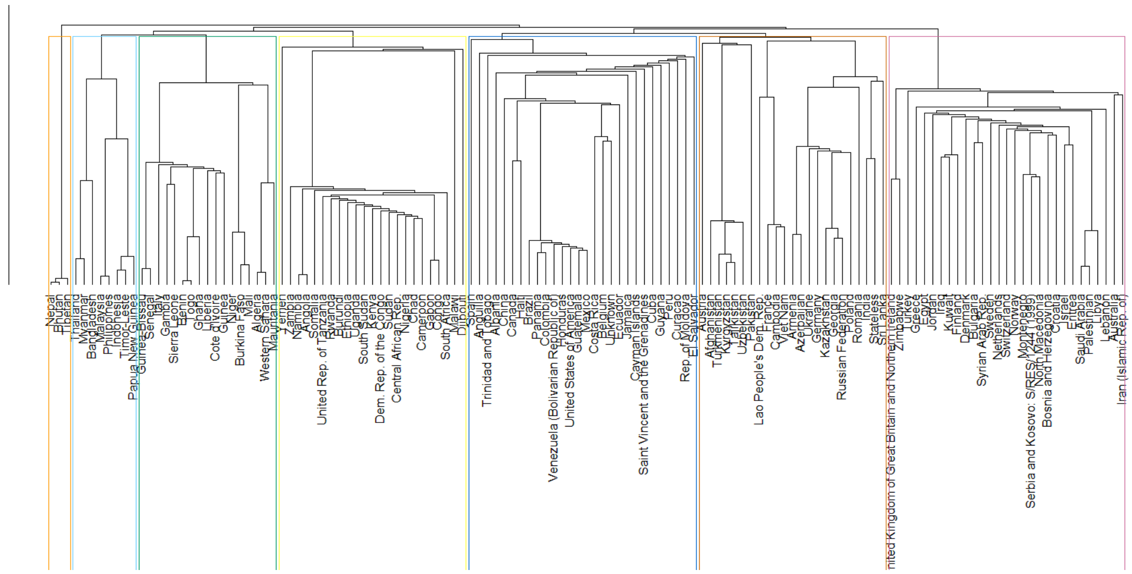

fig7:dendogram for communities[in FastGreedy]

In the above dendogram**(fig7)**, x axis represents the countries and y axis represents the average distance between the two clusters ,from dendogram we can observe that similarly colored lines denote one community[Countries between the similarly coloured lines are of the same community]

From the above **fig6 and fig7**, wee can see that the results of the graph formed by this algorithm is fairly simple as the key aspect of the agglomerative clustering here is the consistent increase of modularity and as modularity is proportional to the degree of the two nodes concerned nodes, i.e.,if a node is to be added to the cluster, the modularity should increase; the countries where either people migrate from or migrate to should be high(they can both be high but it is very unlikely except in case like the countries of India and Pakistan in the year of the partition).

Community detection based on label propagation.

Label propagation community detection is a fast, linear time algorithm for detecting community structure in networks by labelling the vertices with different labels and then updating the names by voting. It's a simple approach, where every node is assigned as one of *k* labels. The method gets going iteratively and changes labels to nodes in a way that each node takes the most frequent label of its neighbours at the same time. This method stops when the label of each node will be the same or maximum of a same neighbourhood node labeled value . It is very fast but sometimes gives different results which is based on initial configuration. Therefore, one should run the method for many times and then build a accurate labelling process

Detected communities based on label propagation:

CODE:

RGui (64-bit) - [C:\Users\HP\Downloads\labelprop.R - R Editor]

File Edit Packages Windows Help

```
library(igraph)
set.seed(1492)


networ <- read.csv(file.choose(),header=FALSE)
g <- graph.data.frame(networ,directed=FALSE)
coords = layout_with_fr(g)
par(mar=c(0,0,0,0))
windows(heigth=50, width=50, record=TRUE, rescale="fit")

# label propagation community detection*
memb <- label.propagation.community(g)
# computing modularity
mod <- as.character(format(modularity(g,membership(memb)), digits = 2))
# visualizing the result of dividing the network into communities
plot(memb,g,edge.arrow.size=0.5,
     vertex.label.cex=0.75,
     vertex.label.family="Helvetica",
     vertex.label.font=2,
     vertex.shape="circle",
     vertex.size=1,
     vertex.label.color="black",
     edge.width=0.5 , layout = coords,edge.curved=0)
```

GRAPH:

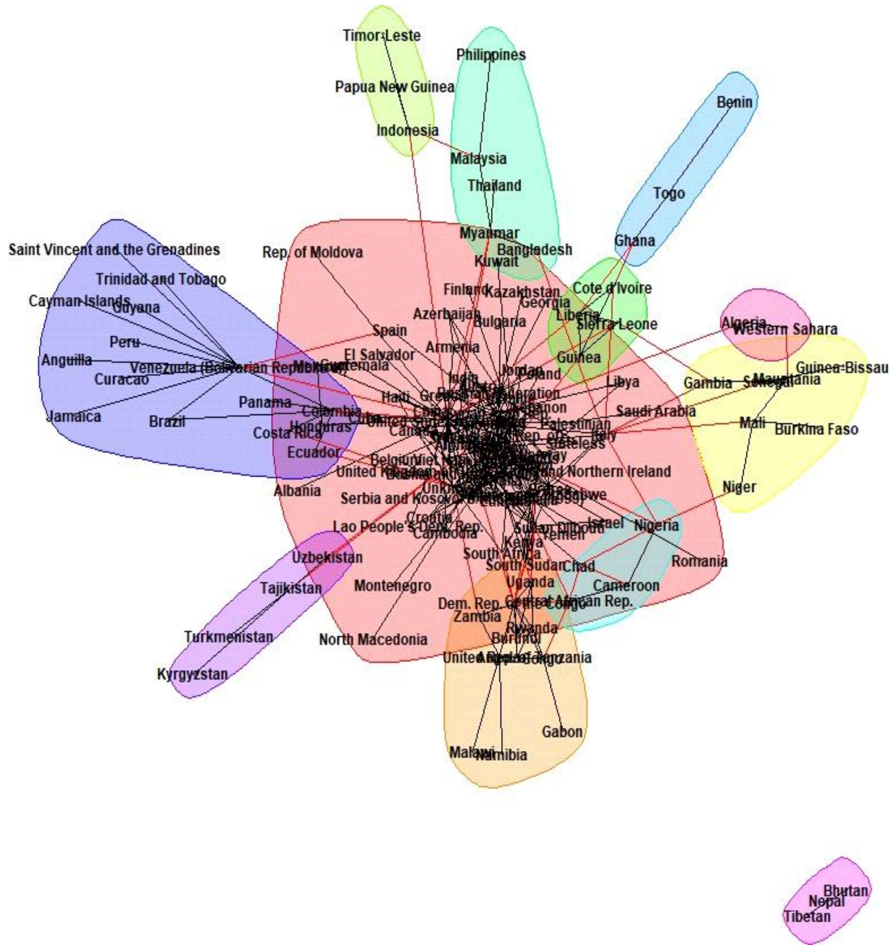

fig8:graph representing communities detected using Label Propogation algorithm

From the above graph **fig8** we can conclude that the countries (vertices) are labled based on their neighbourhood vertex, from above graph we can take best example country that is USA where it has the most in and out degree that is in and out migrants .So by using this algorithm we can get to know where the people migrate mostly in their neighbourhood in a less period of time ,for fast migration ,but optimization with this algorithm is not more accurate.

4.Conclusion:

From the above drawn conclusions, in the community detection algorithms we infer there are many reasons for refugees to move from one country to another country, in search for food, shelter and safety in turn for their very own survival.

Each community detection algorithm depicts a particular reason for a group of refugees to move in that particular community. So, we hope from this study, the drawn conclusions will help international organizations, NGO's, Governments and even United Nations to provide help, raise funds and establish new international laws so that all countries can welcome and provide home to the refugees, and thus save humanity.

References:

- Joshua Blumenstock, Xu Tany "Social Networks and Migration: Theory and Evidence from Rwanda" December 2016
- Sara R. Curran Abigail C. Saguy "Migration and Cultural Change: A Role for Gender and Social Networks?" Journal of International Women's Studies Volume 2 | Issue 3 Article 4 Jun-2001
- Julie Litchfield, Raisul Mahmood, Tasneem Siddiqui, Eva-Maria Egger, and Shayan Ansari "Migration out of poverty" Migration and Social Networks: Evidence from Bangladesh Working Paper 31 September 2015
- Sonja Haug Sonja Haug "Migration Networks and Migration Decision-Making" Journal of Ethnic and Migration Studies. Vol. 34, No. 4, May 2008, pp. 585-605
- Reinhard Biedermann "Towards global management on flight and migration? A comparison of refugee politics in Europe and Asia" Biedermann Asian Journal of German and European Studies (2017) 2:8
- Miriam George & Jennifer Jettner "Migration Stressors, Psychological Distress, and Family—a Sri Lankan Tamil Refugee Analysis" Int. Migration & Integration (2016) 17:341–353
- Miriam George "Migration Traumatic Experiences and Refugee Distress: Implications for Social Work Practice" Clinical Social Work Journal December 2012, Volume 40, Issue 4, pp 429–437
- Jennifer KueEmail, Heather Hanegan, Alai Tan "Perceptions of Cervical Cancer creening, Screening Behavior, and Post-Migration Living Difficulties Among Bhutanese–Nepali Refugee Women in the United States" Journal of Community Health December 2017, Volume 42, Issue 6, pp 1079–1089

- http://igraph.org/r/doc/communities.html
- http://www.pnas.org/content/99/12/7821.full.pdf
- https://arxiv.org/abs/cond-mat/0603718
- https://en.wikipedia.org/wiki/Potts_model
- https://arxiv.org/abs/physics/0512106
- http://igraph.org/r/doc/cluster_fast_greedy.html
- http://igraph.org/r/doc/cluster_label_prop.html