Report- Assignment 02

Social Network Analysis

Written By: Mishu Dhar.

Task 01:

Creating new variables

To enhance the analysis, I extracted a subset of data from the original dataframe, comprising nodes (SideA and SideB) and additional edge attributes. A new variable, "total_death," is derived from the "Batdeath" and "NonstateDeath" variables. This variable is added as edge attributes by indicating the total number of casualties, providing insight into the intensity of connections within the graph object. In cases where information was missing (-8 or -9), the resulting variable may lack data for certain events. Nonetheless, for the sake of analytical simplicity, I did not bring any change here.

Another variable, "duration," was generated from the "startYear" and "endYear" variables. Instances where wars initiated and concluded within the same year were considered to have a duration of one year. Notably, -7 was identified as a placeholder for missing information regarding ongoing wars. Given that the dataset was compiled in 2006, I assumed that such conflicts had concluded by that time. For instance, if a war was labeled "Afghan Resistance," -7 was replaced with 2022 to reflect the withdrawal of the American army from Afghanistan. Similarly, for wars labeled "Iraqi Resistance," -7 was substituted with the end year 2013. These are information I collected from internet. However, for missing information (-8) pertaining to Side A and Side B, no adjustments are made. Given the historical context of colonial and imperial wars spanning over 200 years, it was deemed impractical to accurately replace missing data with specific participants. Therefore, these values remained unchanged (-8, indicating missing information).

THEORETICAL ARGUMENT

I set up the network as an undirected one. Each edge between the vertices represents their involvement in a war. Since some parties (vertices or nodes) were involved in multiple wars, there might be multiple edges between them.

Given that the dataset consists of colonial and imperial war data, I expect to see high connectivity among countries that were imperialist, such as the United Kingdom, France, Portugal, and the Netherlands. These countries are likely to be the most influential nodes in the network.

In more recent times, countries like the USA, Afghanistan, and Iraq have been involved in numerous conflicts. Therefore, I anticipate that these countries will also emerge as highly connected nodes in the network.

TASK 02 HISTORICAL PERIOD, NATURE AND DESCRIPTIVE STATISTICS

Historical Period: The Extra-State War Data from the Correlates of War (COW) project covers historical periods from the early 19th century to the present day. It includes data on conflicts involving non-state actors, such as rebel groups, insurgencies, and militias, as well as conflicts between states and non-state actors. The specific historical periods covered are from 1816 to 2004 (starting year) and ending year 2004. However, I added 2022

and 2013 as the ending year for Afghan and Iraq year for the simplicity of network analysis. I have mentioned the logic of doing this in the data preparation section.

Nature of Ties: The ties in the network represent instances where parties were involved in war at least once. It is assumed that there may be multiple ties between the same two vertices. This is because countries like the United Kingdom and France, which had colonies, were involved in multiple battles with the same entities. For example, the United Kingdom had several conflicts with various Indian states(several times), while France had engagements with Algeria, among others.

number_of_nodes	number_of_ties	graph_density	avg_path_length
163	198	0.01499659	3.48072

Interpretation:

There are total 163 Nodes and 198 ties in this network.

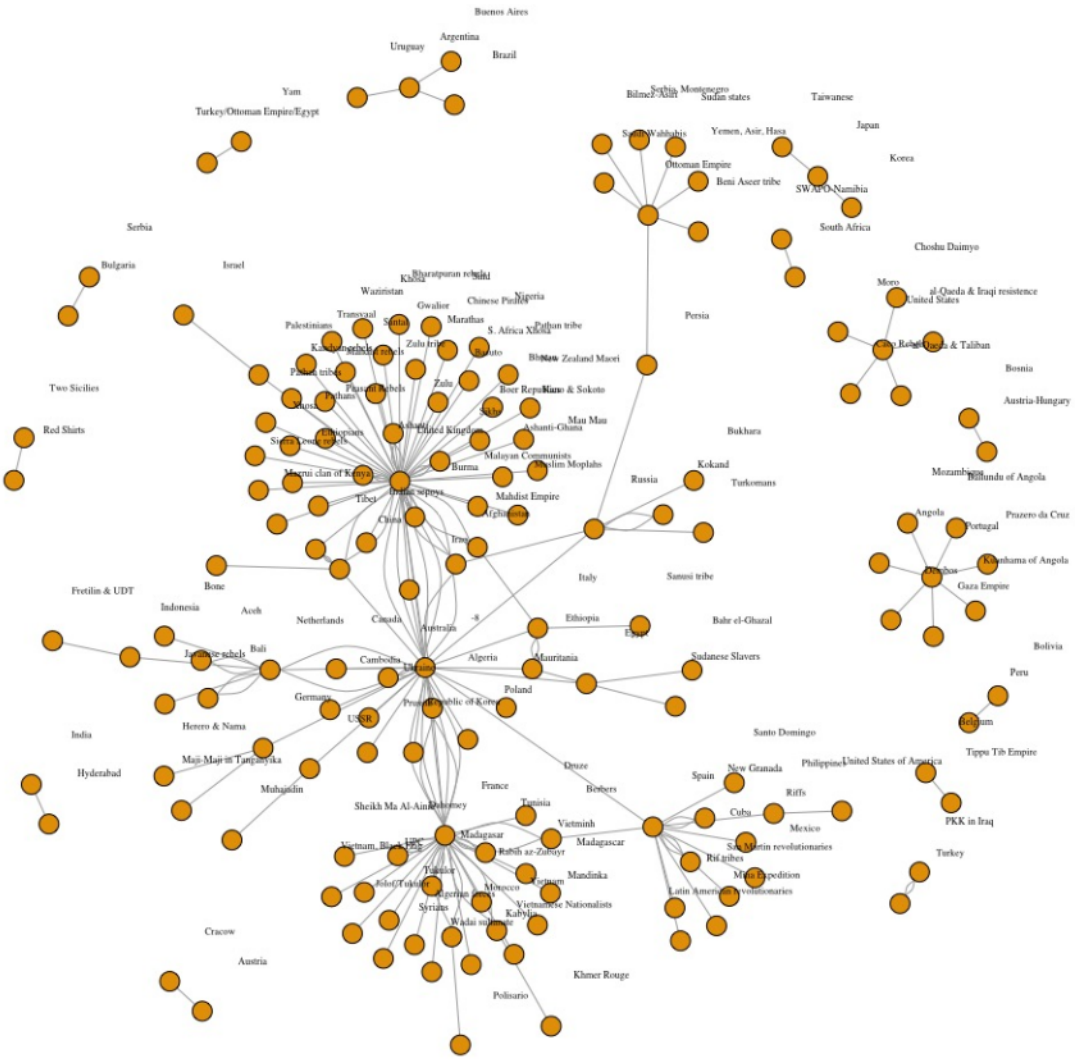
Density measures how interconnected the nodes are in the network, representing the proportion of potential edges that are actually present in the network. It ranges from 0 to 1. In this, graph density is approximately 0.015, indicating a relatively sparse network. This suggests that only about 1.5% of all possible connections between nodes are realized in the network.

The average path length average shortest path length between all pairs of nodes in the network. Here, avg_path_length is approximately 3.481. It signifies that, on average, it takes about 3.481 steps to traverse from one node to another in the network, considering all possible paths.

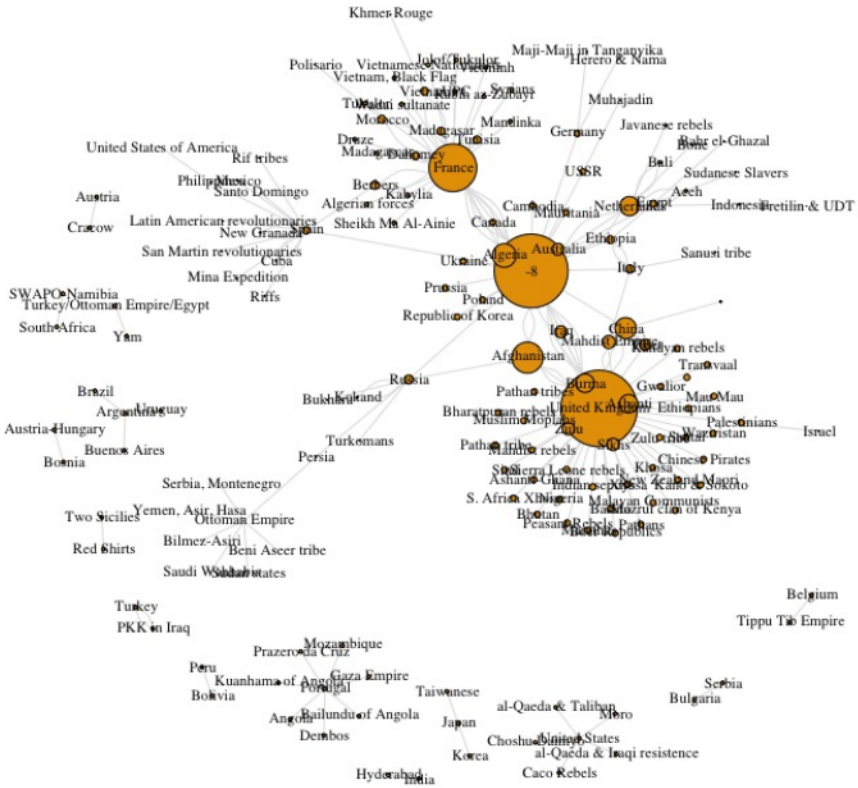
Isolates: as each edge between the vertices at least one engagement in the war, from the nature of the data it is evident that there is no isolate (a node which is not connected to any of the nodes) in the network.

TASK 03: PLOTTING WITH VARIOUS ATTRIBUTES

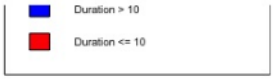
Extra State War Network

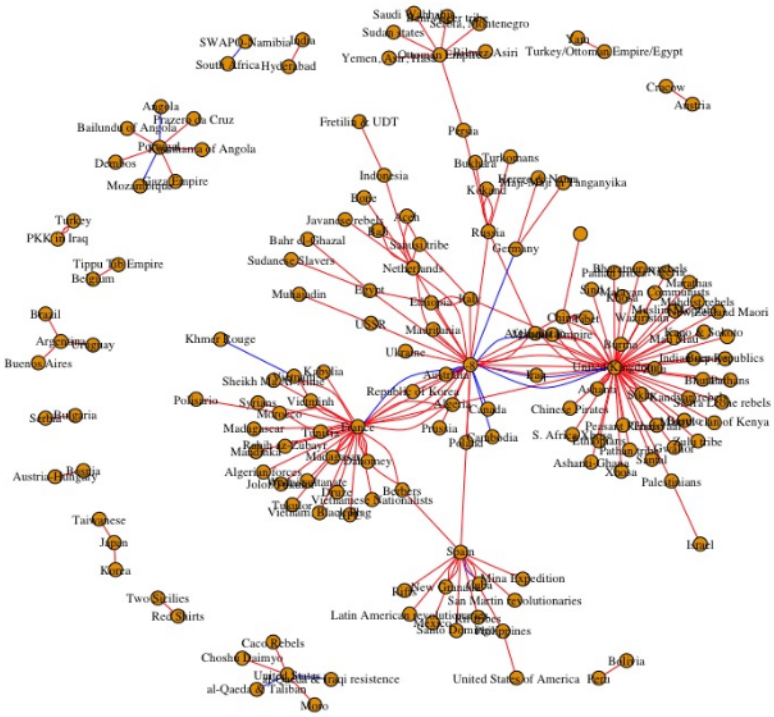


Extra State War Network- Influential Nodes

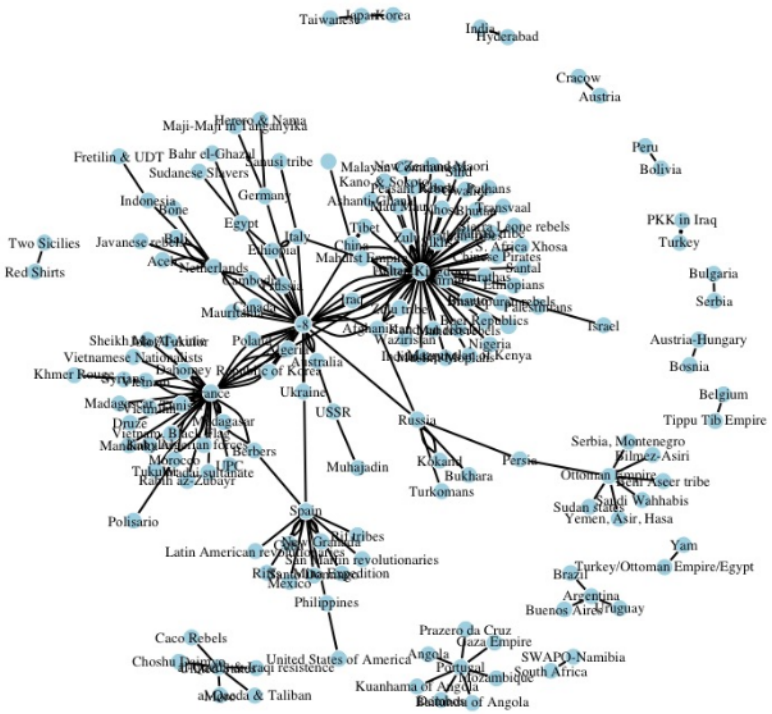


War Duration

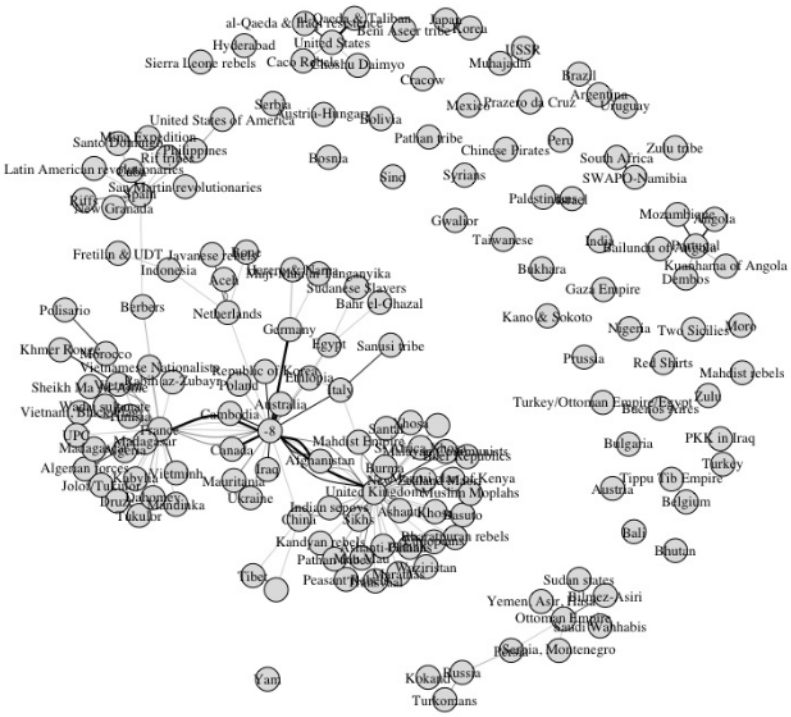



Saudi
Sudan
Montenegro
SWAPO
Namibia
South Africa
Hyderabad
Yemen
Asiri
Turkey/Ottoman Empire/Egypt
Croatia
Algeria
Angola
Frazer da Cruz
Bailundo of Angola
Dembos
Mozambique
Fretilin & UDT
Persia
Turkmans
Indonesia
Kokand
Tanganyika
Turkey
PKK in Iraq
Tippu Tib
Empire
Belgium
Java rebels
Aceh
Bahr el Ghazal
Sahel tribe
Russia
Germany
Sudanese Slavers
Netherlands
Muhammadin
Egypt
England
Boers
USSR
Mauritania
Ukraine
Brazil
Argentina
Uruguay
Buenos Aires
Khmer Rouge
Korea
Sheikh Mahdi
Polisario
Syrians
Morocco
Madagascar
Tunisia
France
Algerian forces
Jolo
Vietnamese
Nationalists
Vietnam
Bloc
Austria
Hungary
Taiwanese
Japan
Korea
Two Sicilies
Red Shirts
Cacos
Rebels
Choshu
Daimyo
United States
Iraqi resistance
al-Qaeda
Taliban
Moro
Spain
Mina Expedition
New Granada
San Martin
revolutionaries
Latin America
Mexico
Santo Domingo
Philippines
United States of America
Berlin
Bulgaria
Israel
Mauritius
Maori
Ashanti
Ghana
Senegal
Palestinians
Chinese
Pirates
Peasants
S. Africa
Kenya
Pathans
Sikh
Tribe
Kingdom
Prussia
Portugal
Cambodia
Canada
Austria
Republic of Korea
Ottoman Empire
Russia
Mali
Tuareg rebels
Mauritania
Mozambique
Kenya
Indian Republics
Brahmans
Muslim rebels
Wahhabis
Mahdi
Sudan
Ethiopia
Libya
Algeria
Tunisia
Morocco
Egypt
Iraq
Kuwait
Saudi Arabia
Yemen
Oman
Bahrain
Qatar
Iran
Afghanistan
Pakistan
Turkey
Russia
Germany
Italy
Austria
Hungary
Poland
Belgium
Netherlands
Denmark
Sweden
Finland
Norway
Switzerland
Greece
Romania
Serbia
Bulgaria
Albania
Croatia
Slovenia
Bosnia
Montenegro
Ireland
Portugal
Spain
Cuba
Mexico
Brazil
Argentina
Uruguay
Paraguay
Chile
Peru
Bolivia
Colombia
Venezuela
Ecuador
Panama
Costa Rica
Nicaragua
Honduras
El Salvador
Guatemala
Haiti
Dominican Republic
Jamaica
Trinidad
Guyana
Suriname
Cuba
Philippines
Japan
Korea
Taiwan
Vietnam
Laos
Cambodia
Thailand
Burma
Malaysia
Singapore
Indonesia
Philippines
Papua New Guinea
Australia
New Zealand
Fiji
Samoa
Tonga
Vanuatu
Solomon Islands
Kiribati
Tuvalu
Nauru
Micronesia
Marshall Islands
Palau
Guam
Hawaii
Alaska
Canada
United States
Mexico
Guatemala
Belize
Honduras
El Salvador
Nicaragua
Costa Rica
Panama
Colombia
Venezuela
Guyana
Suriname
Ecuador
Peru
Bolivia
Paraguay
Chile
Argentina
Uruguay
Brazil
Portugal
Spain
France
Italy
Germany
Austria
Hungary
Poland
Belgium
Netherlands
Denmark
Sweden
Finland
Norway
Switzerland
Greece
Romania
Serbia
Bulgaria
Albania
Croatia
Slovenia
Bosnia
Montenegro
Ireland
Israel
Palestinians
Ashanti
Ghana
Senegal
Kenya
Pathans
Sikh
Tribe
Kingdom
Prussia
Portugal
Cambodia
Canada
Austria
Republic of Korea
Ottoman Empire
Russia
Mali
Tuareg rebels
Mauritania
Mozambique
Kenya
Indian Republics
Brahmans
Muslim rebels
Wahhabis
Mahdi
Sudan
Ethiopia
Libya
Algeria
Tunisia
Morocco
Egypt
Iraq
Kuwait
Saudi Arabia
Yemen
Oman
Bahrain
Qatar
Iran
Afghanistan
Pakistan
Turkey
Russia
Germany
Italy

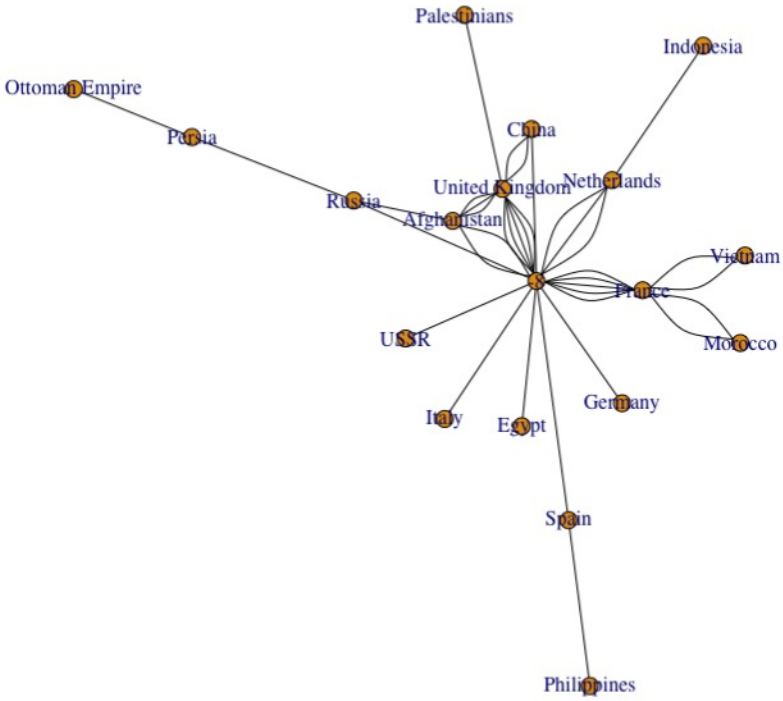
Thicker Edge- Longer Duratiom



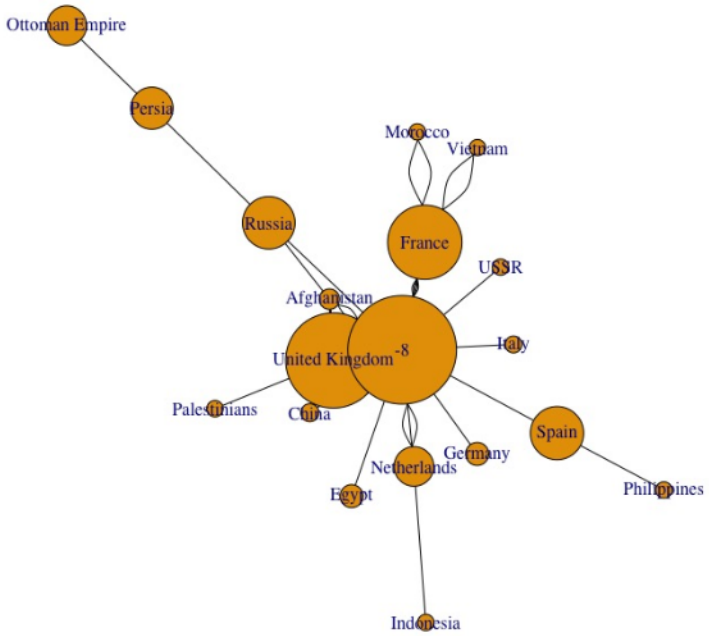
War Longer than Two Years


Hyderabad
United States
Sierra Leone rebels
Caco Rebellion
Daimyo
Cracow
Japan
Korea
USSR
Mujahaddin
Brazil
Argentina
Uruguay
Prazero da Cruz
Mexico
Serbia
Austria-Hungary
Bolivia
United States of America
Pathan tribe
Chinese Pirates
Peru
Zulu tribe
South Africa
SWAPO-Namibia
Bosnia
Sind
Syrians
Palestinians
Gwalior
Mozambique
Angola
Taiwanese
India
Bukhara
Gaza Empire
Dembos
Kano & Sokoto
Nigeria
Two Sicilies
Moro
Prussia
Red Shirts
Mahdist rebels
Zulu
Turkey/Ottoman Empire
Egypt
Bulgaria
PKK in Iraq
Turkey
Tippu Tib Empire
Austria
Belgium
Bali
Bhutan
Sudan states
Yemen
Asir, Hejaz-Asiri
Ottoman Empire
Saudi Wahabis
Serbia, Montenegro
Russia
Kokand
Turkomans
Yam
Tibet
Peasant Rebellion
Wazirdtan
Kandyan rebels
Sikhs
Indian sepoys
Chinese
United Kingdom
Muslim Moplahs
Basuto
Mauritania
Ukraine
Iraq
Canada
Afghanistan
Mahdist Uprisings
Burma
Kenya
Algerian forces
Jolo Tulkar
Druze
Tukulor
Madagascar
UPC
France
Cambodia
Australia
Italy
Poland
Ethiopia
Sanusi tribe
Germany
Netherlands
Bahr el-Ghazal
Sudanese Slayers
Berbers
Polisario
Morocco
Khmer Rouge
Vietnamese Nationalists
Sheikh Ma'ruf
Vietnam, Black Flags
Indonesia
Aceh
Fretilin & UDT
Javanese rebels
Belize
New Granada
Latin American revolutionaries
Philippines
Santo Domingo
Mau Expedition
Rif tribe
San Martin revolutionaries
Spain
Rufiji
Kuashama of Angola
Dembos
Thailand
Burma
Mujahaddin
Tanganyika
Bukhara
Gaza Empire

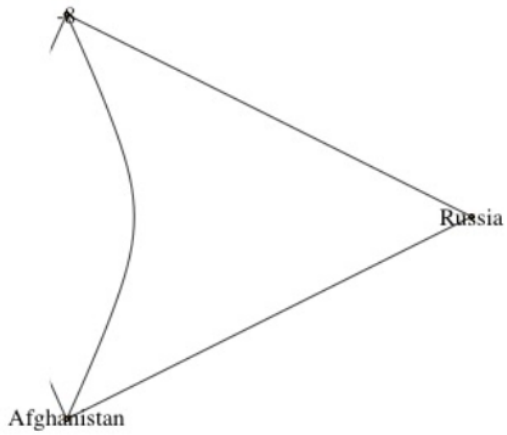
Influential Nodes- Base on Betweenness Centrality



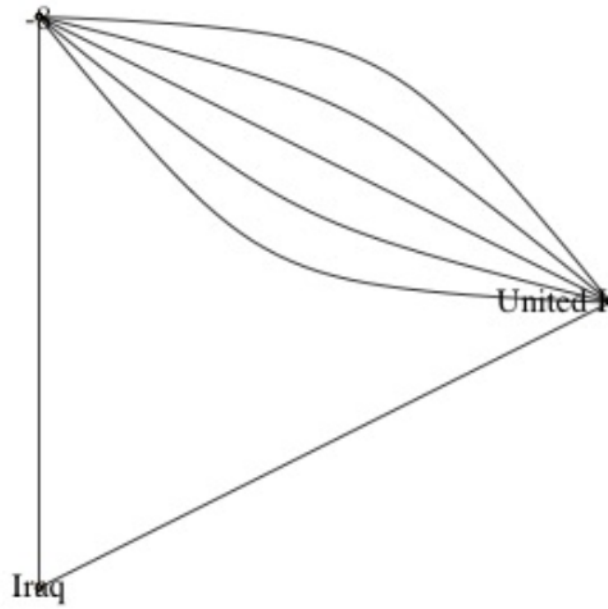
Influential Nodes- War Network



Largest Clique 1



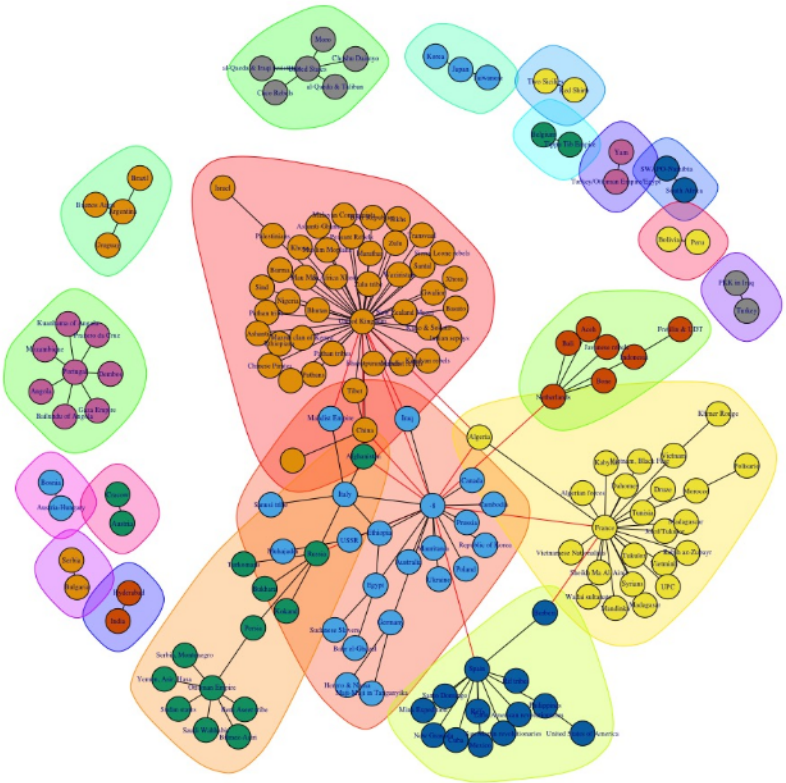
Largest Clique 2



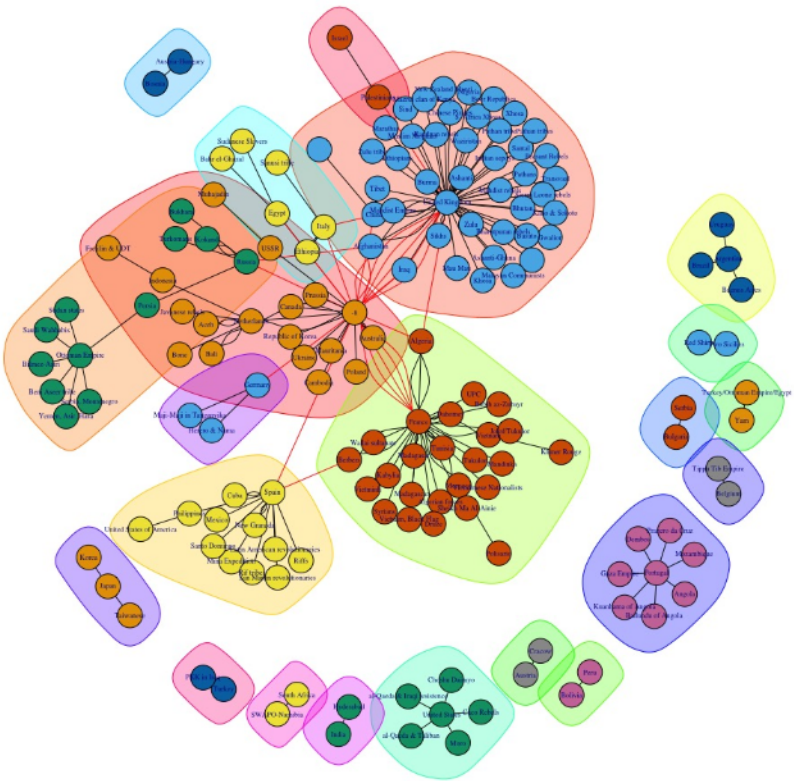
Task 04- Community Detection

Identifying cliques and calculating assortativity and reciprocity provide us with some information as to patterns of vertex association in a network. A more formal way of achieving this is through community detection. If a network is said to have a community structure then it is possible to assign vertices to unique sets. Within each set of vertices, the connections between members will be more dense than the connections between different sets of vertices

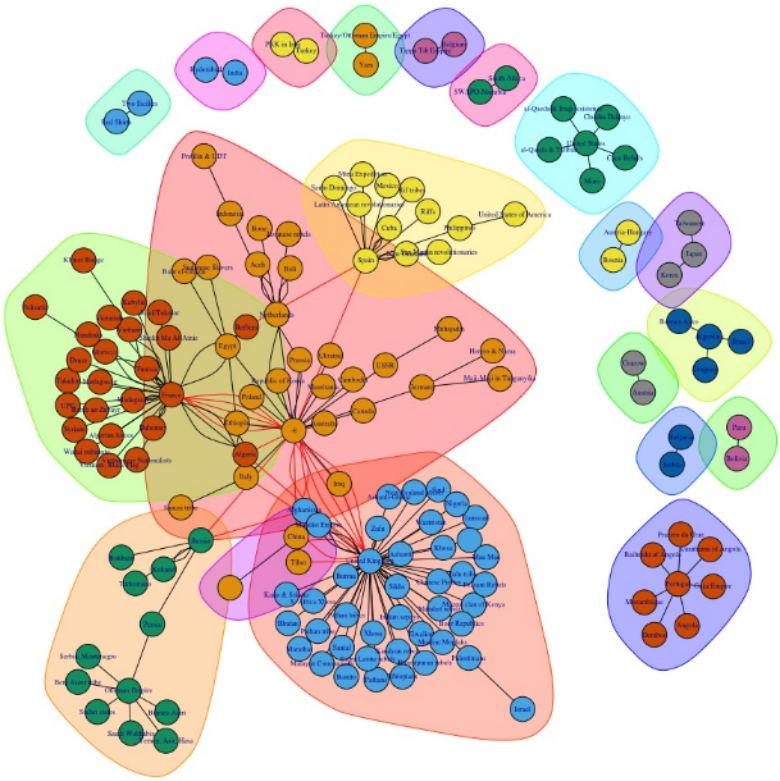
Community Detection- Fast Greedy



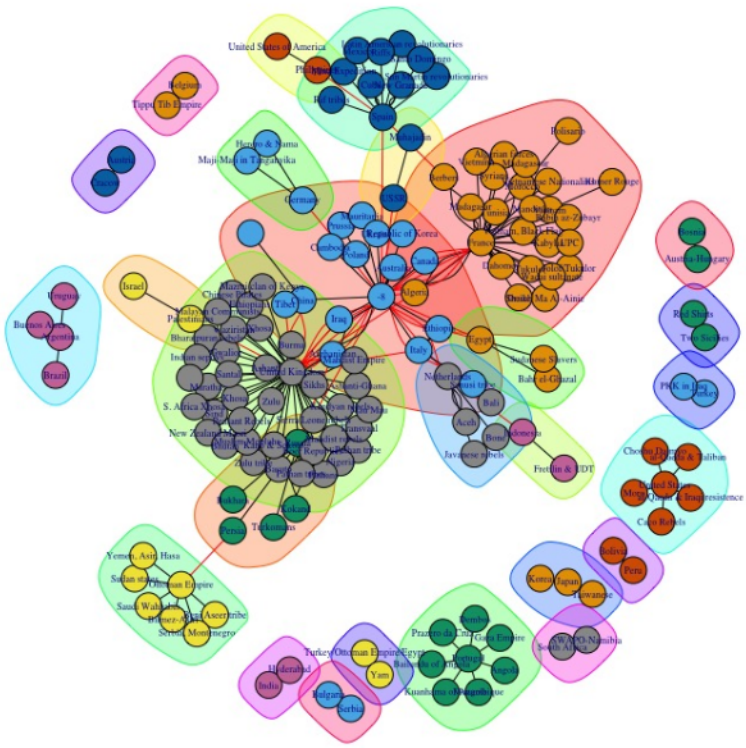
Community Detection- Edge Betweenness



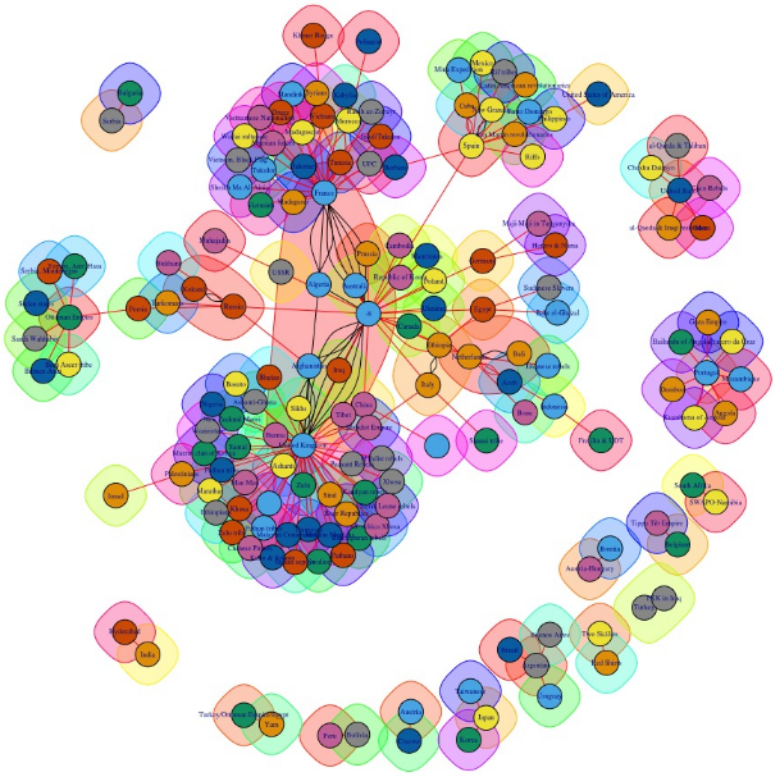
Community Detection- Louvain



Community Detection- Walktrap



Community Detection- Leiden



TASK 05- COMPARISON BETWEEN THE COMMUNITY DETECTION ALGORITHMS

Leiden Community Length: 153
Louvain Communities Length: 20
Walktrap Community Length: 27
Edge Betweenness Community Length: 22
Fast Greedy Community Length: 20

From the figure and from the length it is shown that the Leiden community detection algorithm identified the largest number of communities, indicating that it may have subdivided the network into smaller, more fine-grained communities compared to other algorithms. Choosing a community detection algorithm may vary according to the context or the purpose of the analysis. In my point of view, for this project (assignment) considering the size of the network Louvain and Walktrap performed better than Leiden. As the Leiden algorithm detected a very large number of communities (153), the figure is very complex to perceive. We are keen to know here which are the most war-prone states, so for that particular purpose, Louvain and Walktrap algorithms detected communities in a better way around the influential nodes such United Kingdom, France, Portugal, Spain, etc.

Findings:

Following the hypothesis, from the statistics and figures above it is evident that countries such as the United Kingdom, France, Portugal, Spain, etc that established colonies overseas are the most influential nodes in this war network and were involved in more wars than others.

Limitation of the Analysis:

As I did not modify the missing information -8 from the participants (sideA and sideB) it is considered a whole state entity, which is clearly misleading. But I don't have the exact historical information to replace it.

For some unknown reason, the arguments degree( ) and betweenness( ), and isolates( ) are not working, when I restarted my R studio today. But, yesterday it worked and I stored the value of betweenness as a variable object_betweenness. I uninstalled and reinstalled the R studio again but still, it is not working.

```
> which.max(object_betweenness)

-8

10
```