

Assignment 2 SNA - Island State Communities

Marc Sparhuber

Table of contents

Task 1	2
Task 2	2
Task 3	3
Task 4	5
Task 5	10

Task 1

Formulate theoretical arguments and expectations about interstate relations. Think in network terms. Select the appropriate part of the data for analysis.

In the following, a subset of the territorial change data set will be analyzed. The territorial exchange data set contains all territorial changes involving at least one nation-state for the period 1816-2015. It captures the country which gained, lost and (if applicable), a territorial entity exchanged. Furthermore, other notable variables such as the area, population of the gained/lost territory are captured, as well as dummy variables indicating whether independence was gained, military conflict was involved, and what kind of territorial exchange took place are recorded.

The following subset contains only those entries in the data set in which an island state is recorded as either the gainer, loser or exchanged territory of a territorial change. It is hypothesized that Island states which are bigger in area are more fought for and thus have more ties indicating military conflict. Within a meso context looking at communities within these countries, it is expected that although the communities will to a large extent reproduce the colonial history that many island states endured, that islands which are bigger (and thus, more valuable) had a larger number of territorial exchanges and serve as the nodes between or within more than one community, as they attracted the attention of not just one colonial power. It has to be noted that the ability to test this hypothesis is limited by the fact that islands which are still part of their past colonial power's state are not part of this data set (e.g., Dutch Caribbean).

Task 2

On the selected data, make a table that contains descriptive network statistics for the network and for key node-level variables. Please mention the historical period covered, the nature of tie(s), the number of nodes and ties, average degree, density, the standard deviation of degrees (or outdegrees and of indegrees if network is directed), and average path length. Identify the number of isolates if there are any. Please check if the network is a single component. Which node has the highest degree, and which has the highest betweenness centrality? What do these measures represent in terms of the specific type of relations and data you have selected?

The subset data contains entries ranging from 1819 to 1994. Due to nature of this dataset, all ties are directed. Any dyads in which either the gaining or losing side were not clearly indicated, were removed, thereby reducing the number of all isolates to 0. In total, this leaves 62 nodes and 107 ties within 3 components, the largest of which contains most nodes, at 56 nodes. The single node with the highest betweenness and degree centrality is the United Kingdom, with 473 and 41, respectively. This foreshadows a highly hierarchical graph with the UK at

its core, backed up by its seafaring prowess in previous centuries leading to the colonialization of many island territories (and their later cession).

Table 1: Network Descriptives

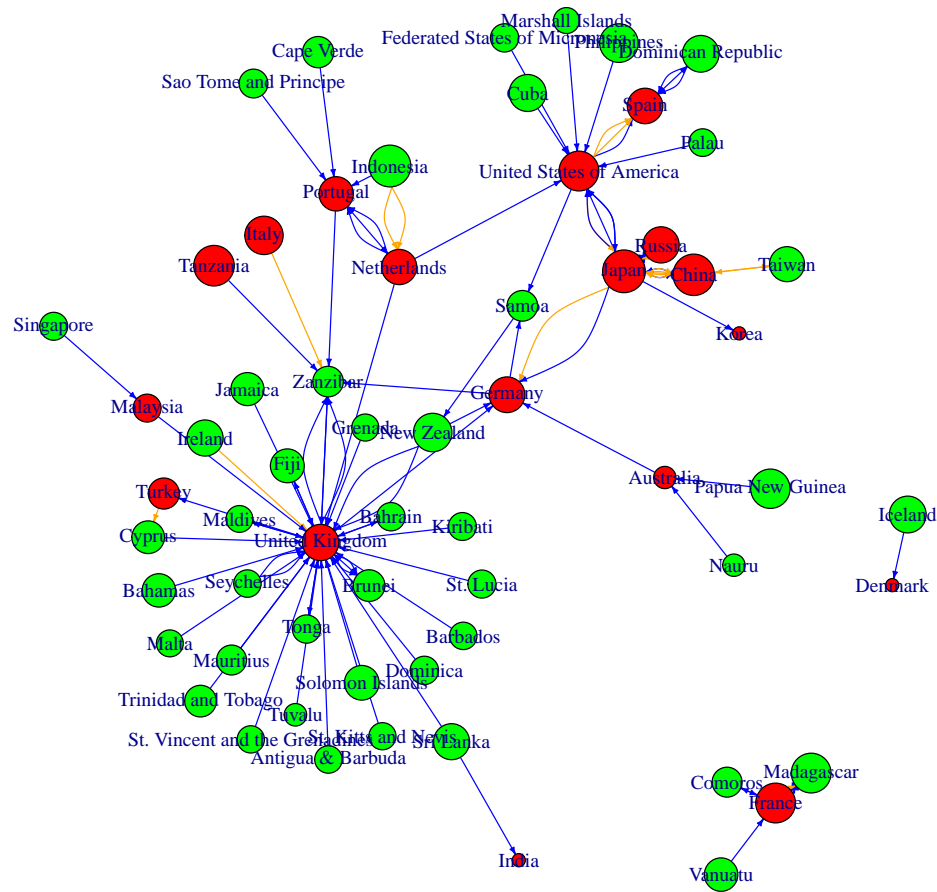
Measure/Variable	Mean/Value	Standard Deviation	Median
Degrees	3.451613e+00	6.004581e+00	1.0
Outdegrees	1.725807e+00	2.240970e+00	1.0
Indegrees	1.725807e+00	4.188859e+00	0.0
Path length	2.841545e+00	NA	NA
Density	2.829190e-02	NA	NA
Reciprocity	3.551402e-01	NA	NA
Transitivity	2.772640e-02	NA	NA
Area gained	1.886729e+05	5.288537e+05	4580.5

Table 1 shows network descriptives and the distribution of gained area in the observed dyads. The large SD and the vastly differing Median and mean indicate the strongly varying sizes in territory gained by regulating the size of node in later visualizations. Meanwhile, the low degree-related measures, specifically the median, indicate that this graph will contain few highly central nodes, which is to be expected, as seafaring and colonizing nations such as the UK, Netherlands and Spain are expected to have the highest centrality with most island states only having a single tie to their previous colonizers. The low density of ~ 0.03 shows that this is a rather sparse network, with many potential but unrealized ties. Regarding reciprocity, 35% of ties are reciprocated. This is larger than anticipated, as most island state ties are expected to only contain the “freeing” tie, rather than the “colonizing” tie, which in many cases falls outside the time frame of this data set. Without interpreting the rest of the measures, these descriptives indicate that many of the island states are unlikely to have ties with one another and largely have “one-way” ties with their previous colonizers that serve as highly central hubs.

Task 3

Visualize the network. Vary degree color and size according to key node attributes. Try to pick a node attribute that relates to an important characteristic (e.g., power, GDP per capita).

Gainers and losers of territory with island states involved 1819 – 1994



In this network visualization, node size scales with the log of total territory gained from corresponding loser nodes, which the color indicates whether this state has obtained independence, with green indicating independence, and the tie color or orange indicating whether military conflict was involved in the territory exchange indicated by the arrow. Importantly, the direction of the arrow indicates the gaining of territory, so that many green-coloured small island states pointing toward the UK represents them gaining their independence - most frequently without military intervention. At first glance, excluding the two smaller components, the main component seems to be roughly divisible into two to four communities (depending

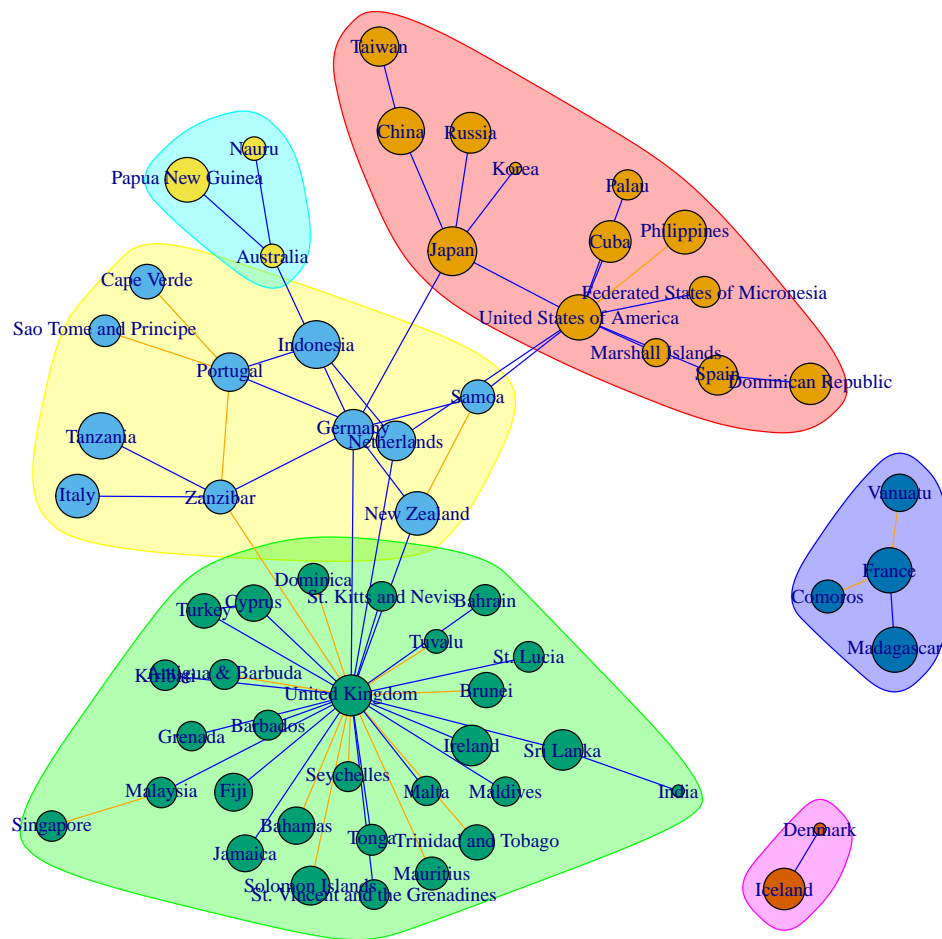
on seed), each centering around a colonizing power with island states contested by more than one colonizing power often serving connectors between the modules.

Task 4

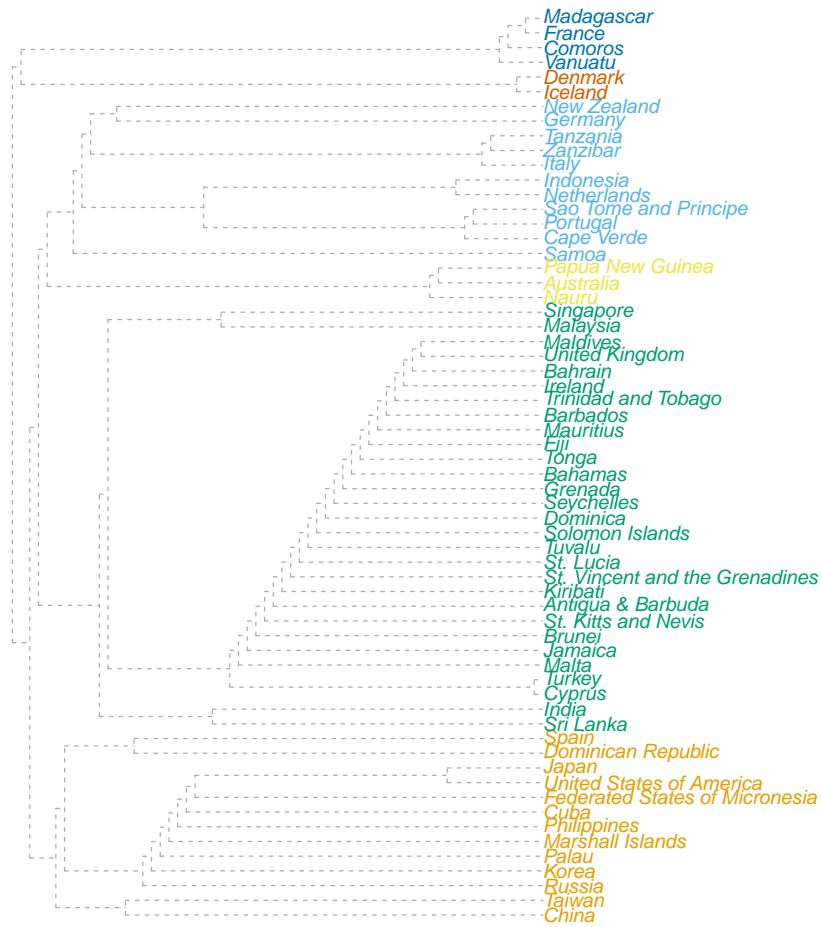
Implement and compare the results (community compositions, their size) of TWO different community detection algorithms to find communities (e.g., Girvan-Newman, Louvain, walktrap, cohesive blocks, Leiden). Please take into account that most of these methods can use undirected graphs only, others (such as Girvan-Newman) might work with weighted graphs. Before carrying out this task, simplify the data if necessary. Create a plot where nodes are either colored based on detected community membership, or you circulate the groups with some color in the background.

Here, three (because the last one is not listed in the instructions) different community detection algorithms are run and plotted, the last of which is meant specifically for directed networks and the earlier two are accompanied by dendrograms. As many community detection algorithms only allow for undirected networks and/or disallow multiple ties between the same nodes, the necessary versions of the graph are calculated. When comparing the first two plots, a clear distinction can be made: the walktrap algorithm leads to much easier to interpret and cogent results, by attempting to find densely connected subgraphs using random walks, while the edge betweenness algorithm (aka. Girvan-Newman) detects a handful of communities consisting of single nodes, by removing edges from the original graph with the still connected nodes being identified as communities. In this case, the prior leads to much clearer interpretability, while the latter seemingly “overdetects” communities. The third algorithm run is called Infomap and is able to take into account directed ties. While this leads to improved results compared to the Girvan-Newman algorithm, there are still some single country communities that admittedly make sense (as these tend to be more central than others, e.g., Zanzibar), yet do not compare to the results of the walktrap algorithm. The dendrograms show the iterative assimilation of countries into their respective communities by the first two algorithms.

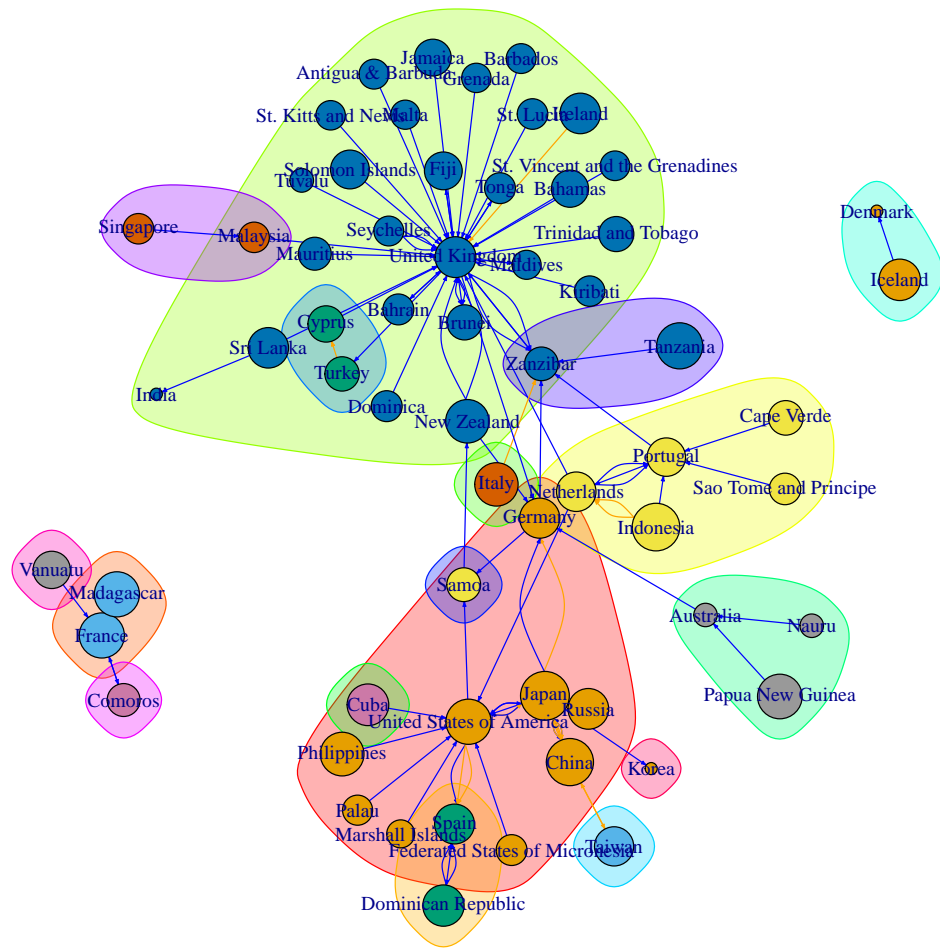
Territory change in island states 1816 – 2015: Walktrap



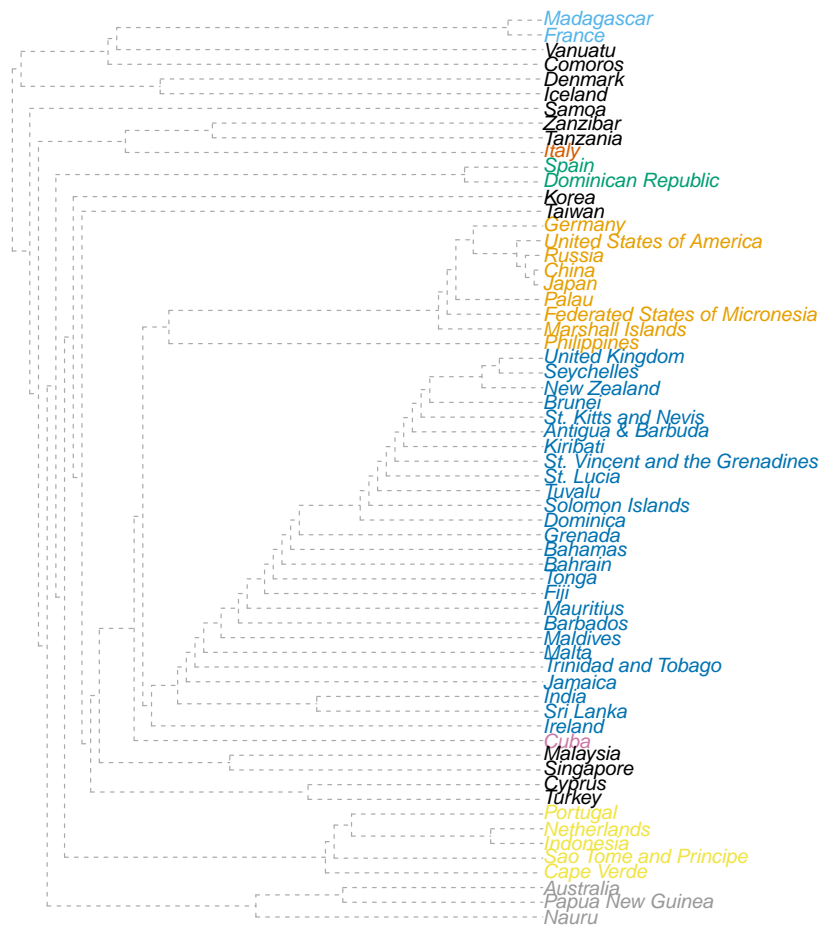
Dendrogram for communities detected with Walktrap algorithm



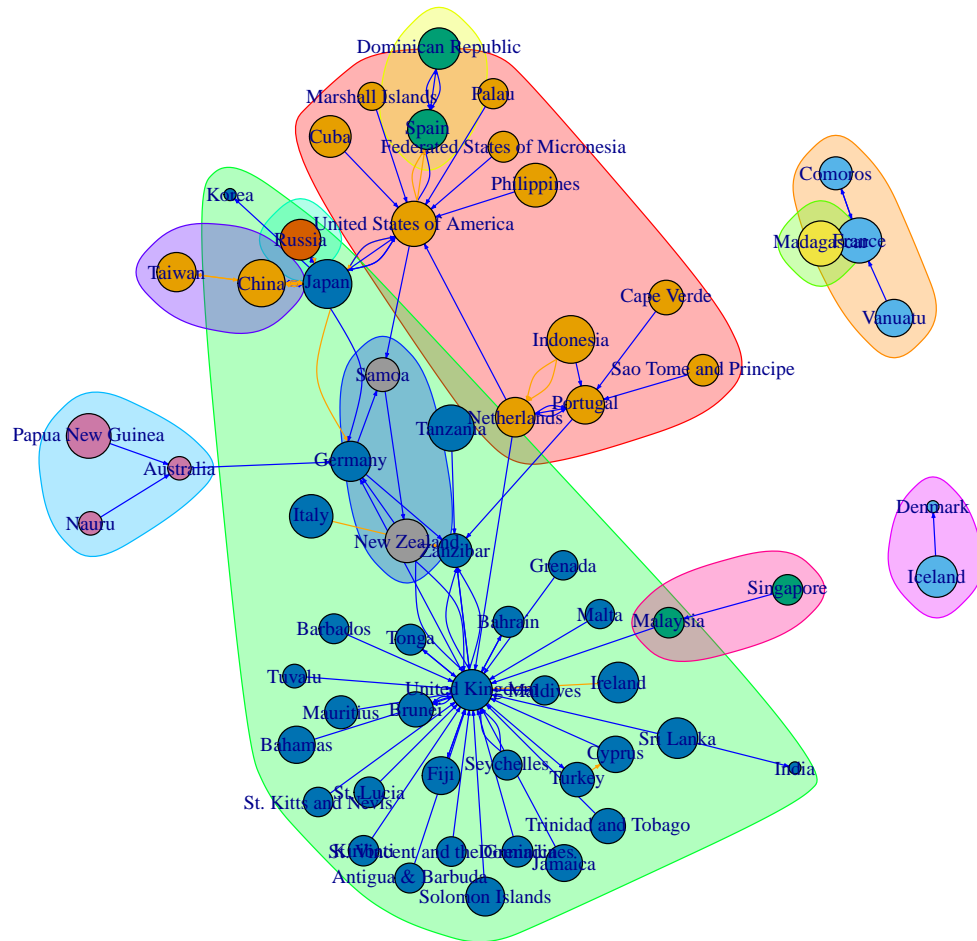
Territory change in island states 1816 – 2015: Edge Betweenness



Dendrogram for communities detected with Edge Betweenness algorithm



Territory change in island states 1816 – 2015: Infomap



Task 5

Make meaningful comparisons between the results of community detection algorithms. Summarize the differences and similarities and try to explain these. Please refer to the relevant literature describing the logic of these algorithms if necessary.

To make meaningful comparisons between community detection strategies past the a visual interpretation, certain measures can be interpreted. One such measure

is modularity. It is close to 1 when there are strongly connected communities, while their connections to other communities are rather weak, indicating clear separations between communities. Meanwhile, values between and around 0 and -1 indicate low modularity, meaning that there are few clear cut communities or that the algorithm doesn't work well with the particular network (Clauset & Newman, 2004). All three community detection algorithms used here result in values around 0.5, indicating moderate modularity. Interestingly, the edge betweenness method scores ~ 0.05 percentage points lower than the other two, which maybe explained by the many one-country communities it detected, thereby decreasing its modularity.

Another method is the variation of information metric, proposed by Marina Meilă (2005). It measures the amount of information lost and gained in changing from the first provided community structure to the second and can take values from 0 to positive infinity. The values the measure produces here indicate a greater dissimilarity of the walktrap algorithm to the edge betweenness algorithm (0.94), than the Infomap community to the edge betweenness community (0.78) and the walktrap algorithm to the Infomap algorithm (0.71). This intuitively makes sense, as the walktrap algorithm creates the least communities, while the edge betweenness algorithm creates the most. Correspondingly, the results of the walktrap algorithm and the Infomap algorithm also do look the most similar when plotted. Notably, these results are the ones produced by just one run and should ideally be averaged across different results of outcomes of community detection algorithms, as they slightly vary from run to run.

Coming back to the initially posed hypothesis it can be said that there is no indication of it being correct, with the exception being cases where it could explain results (i.e., Indonesia or Japan). This has to do with little theoretical groundwork to justify it (and make it worthwhile investigating) and a lack of data to answer it.