National Research University Higher School of Economics

Faculty of Social Sciences

Introduction to Network Analytics

Homework 2

Prepared by:

Podolskaya Anastasia

(aapodolskaya_4@edu.hse.ru)

Shadrunov Aleksey

(asshadrunov@edu.hse.ru)

Emelyanenko Maksim

(mvemelyanenko@edu.hse.ru)

Leading tutor:

Daria Maltseva

Moscow, 2022

Contents

1 Task 1

In this task, we were to transform the Network 8 ("Positive and negative choices in a football team") to Pajek [**?**]. The given network is **unimodal** (vertices represent only football players), **signed** (positive and negative relationships are observed between players), and both **arcs and edges** are present in the network. The original network is displayed in the figure below.
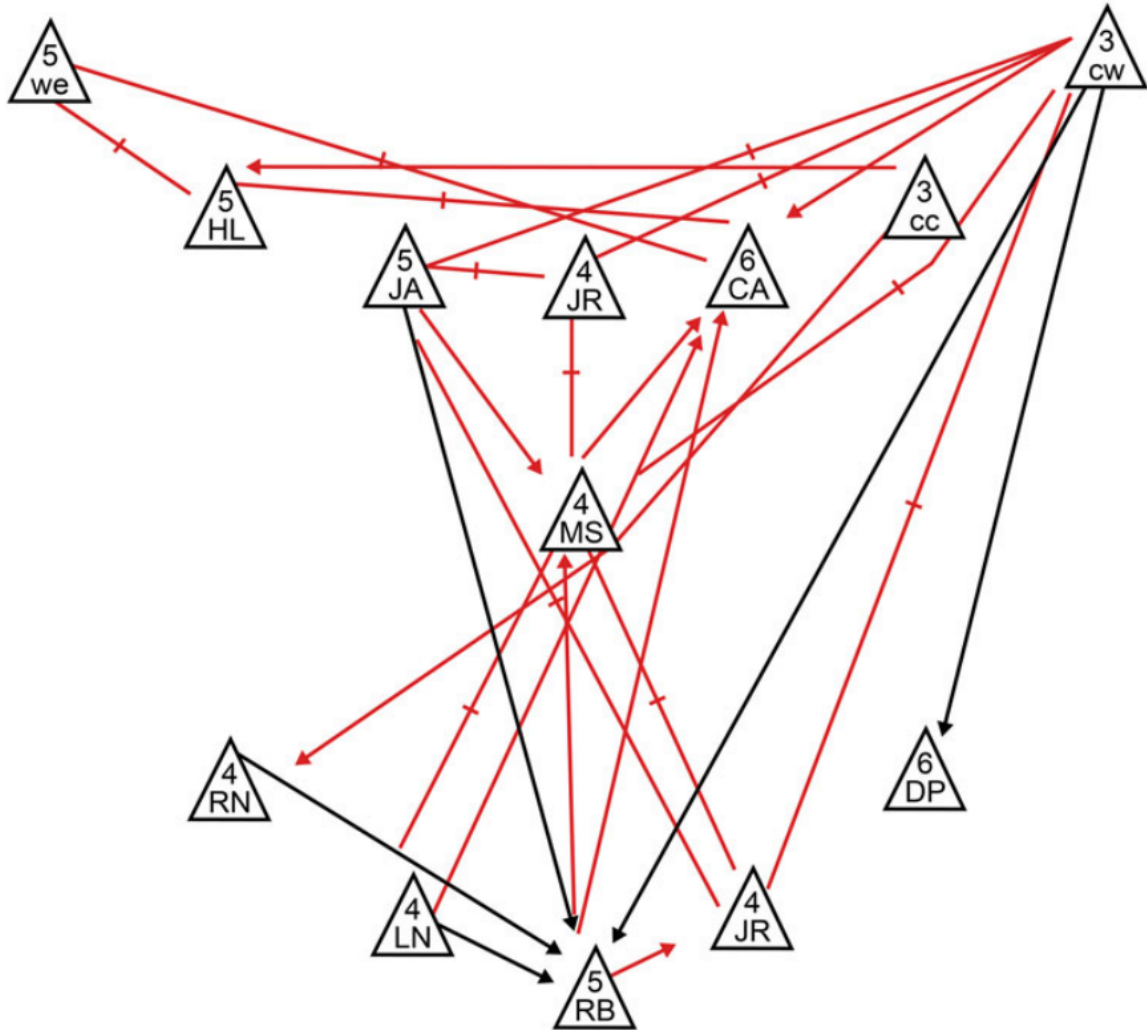

Figure 1 – Original network

At first, we created a Pajek's network data format (**.net**) file with all vertices, arcs and edges of the network [**?**]. In addition, the colours of lines were specified [**?**]. The structure of the file is listed below.

```
1  *Vertices 13
2      1 "5 WE"
3      2 "5 HL"
4      3 "5 JA"
```

```
 5     4 "4 JR"
 6     5 "6 CA"
 7     6 "3 CC"
 8     7 "3 CW"
 9     8 "4 MS"
10     9 "4 RN"
11     10 "4 LN"
12     11 "5 RB"
13     12 "4 JR"
14     13 "6 DP"
15 *Arcs
16     3 8 1 c Dandelion
17     3 11 -1 c RedOrange
18     6 2 1 c Dandelion
19     6 9 1 c Dandelion
20     7 11 -1 c RedOrange
21     7 13 -1 c RedOrange
22     8 5 1 c Dandelion
23     9 11 -1 c RedOrange
24     10 5 1 c Dandelion
25     10 11 -1 c RedOrange
26     11 5 1 c Dandelion
27     11 12 1 c Dandelion
28 *Edges
29     1 2 1 c RedViolet
30     1 5 1 c RedViolet
31     2 5 1 c RedViolet
32     3 7 1 c RedViolet
33     3 4 1 c RedViolet
34     3 12 1 c RedViolet
35     4 7 1 c RedViolet
36     4 8 1 c RedViolet
37     7 8 1 c RedViolet
38     7 12 1 c RedViolet
39     8 10 1 c RedViolet
40     8 12 1 c RedViolet
```

Listing 1 – Pajek net file

Next, we proceeded with visualisation of the network [**?**]. After selecting **Draw** command, the default circular layout of the network appeared (Figure 2).
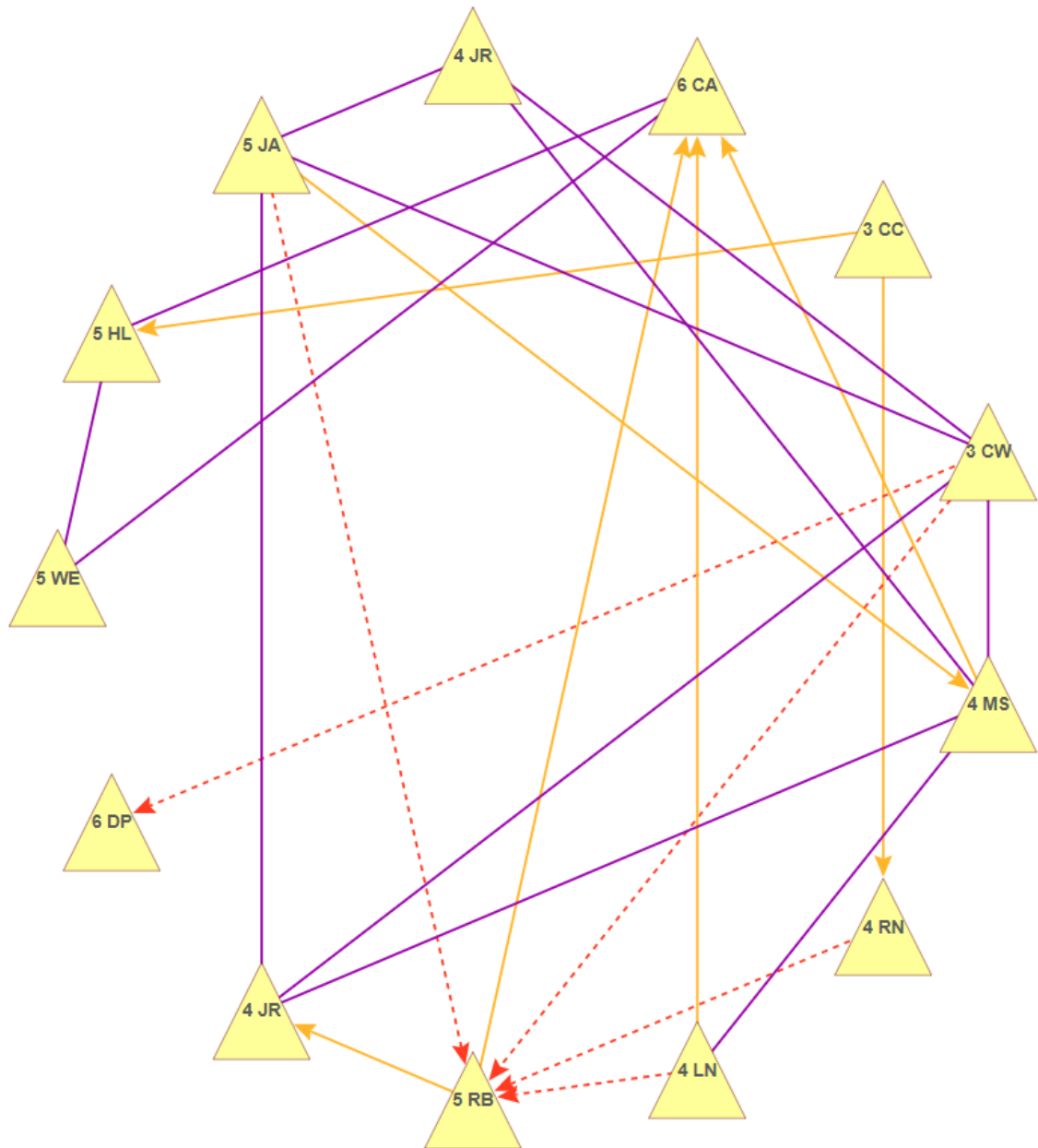
Figure 2 – Default layout

To restore the resemblance, we need to manually move vertices, as on the original network. The layout might be not very neat, however, it represents the **actual positions** of a football team on a stadium. This way we understand how the observed network describes the actual situation in a real football game.

The result is in the Figure 3.

As it is common in network analysis, negative links are drawn with **dotted lines**, and here we additionally marked them with red colour. Arcs have arrows on the ends and they are orange or red, and edges are violet (all edges are positive).

We may try another layout — **Kamada-Kawai**. This layout puts important ver-
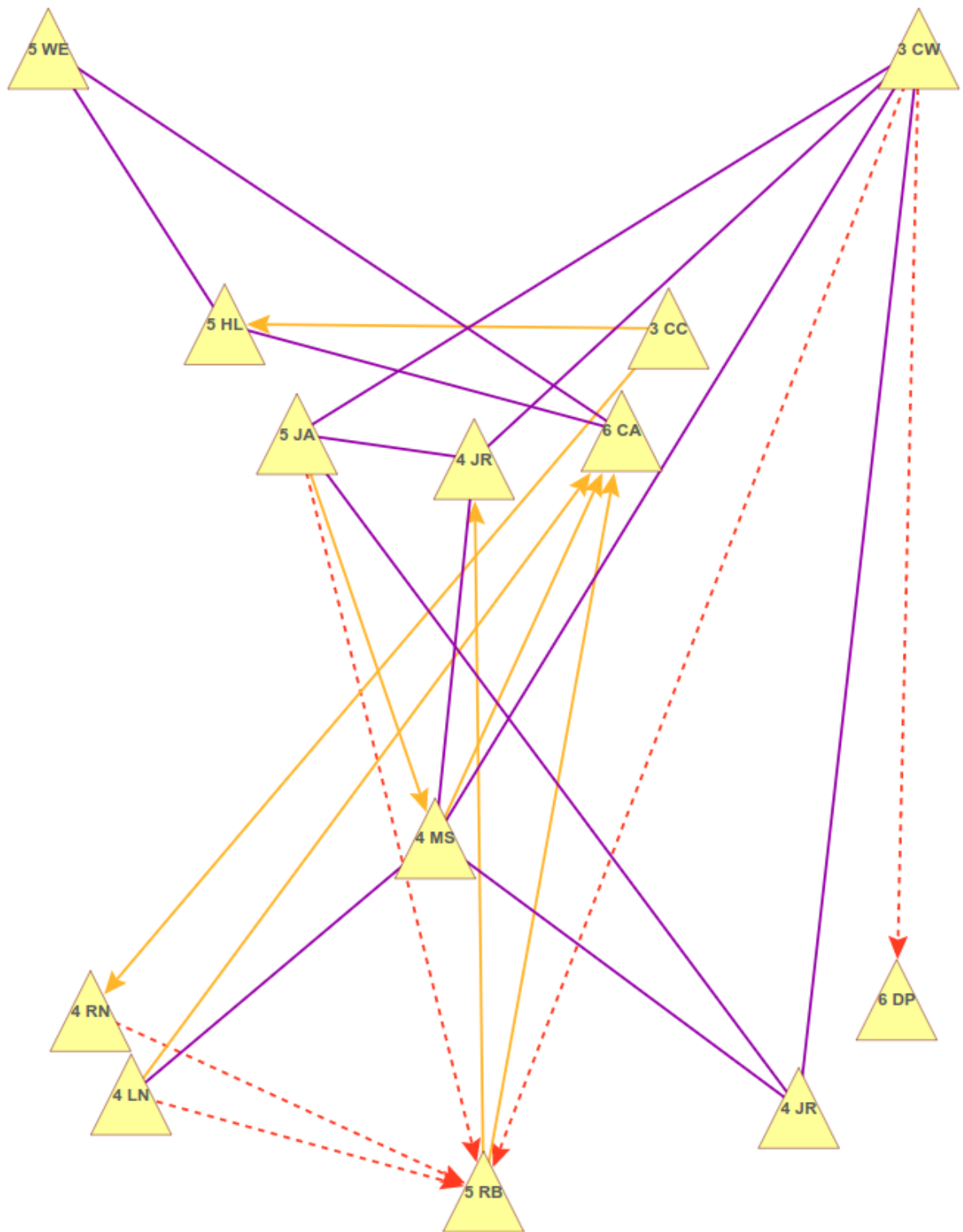
Figure 3 – The layout similar to the given network

tices closer to the centre of the diagram. The result looks satisfying (Figure 4)

Figure 4 – The layout similar to the given network

2 Task 2

In this task, we are to analyse the Mexican polite elite network, which is displayed in the picture below. At first glance, one can notice that the given network is undirected and unweighted. Let us examine the network more closely.
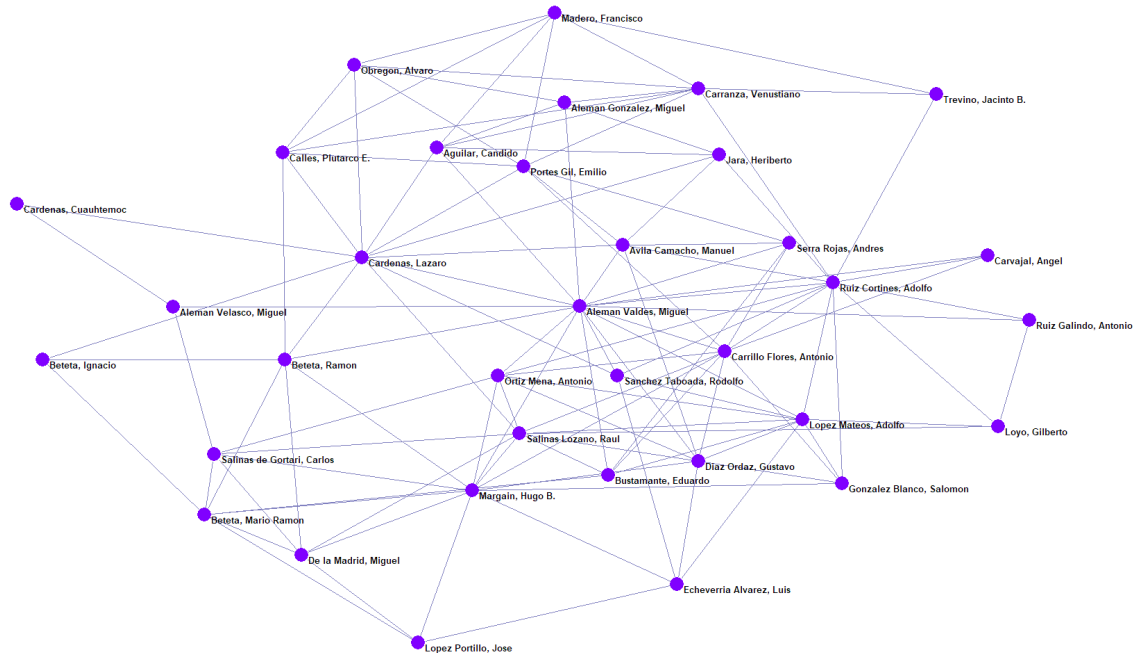
Figure 5 – Mexican polite elite network

2.1 Subtask 1

Overall, the network has **35 vertices**, which represent Mexican presidents and close collaborators, and **117 edges** that indicate the presence of political, kinship, friendship, or business ties between the actors (Figure 6). Also, our network has zero arcs (due to its **undirected nature**), no loops. Given that all of the network's lines are equal to 1, we can confirm that the network is **unweighted**. Since the observed network does not differentiate between the types of actor interaction, it is **unimodal** and **does not contain multiple lines**. The density of our network, which is basically the ratio of links to nodes, is quite low, reaching 0.196, which means that only 19.6 percent of all possible edges are present. This implies that the network is not as connected as it could be. The network's **average degree** equals **6.69**, which suggests that on average, each node in the network has 6.69 links.

```
kilobytes of free virtual address space:    137 434 464 676 kB

================================================================
1. mexican_power.net (35)
================================================================
Number of vertices (n): 35
----------------------------------------------------------------
                                       Arcs          Edges
----------------------------------------------------------------
Total number of lines                    0            117
----------------------------------------------------------------
Number of loops                          0              0
Number of multiple lines                 0              0
----------------------------------------------------------------

Density1 [loops allowed]    = 0.19102041
Density2 [no loops allowed] = 0.19663866
Average Degree = 6.68571429
```

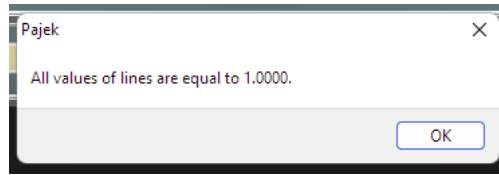Figure 6 – General network information

```
Pajek                                                    ×

All values of lines are equal to 1.0000.

                                            OK
```

Figure 7 – The network is unweighted

As can be noticed from Figure 8, the network's **lowest degree** equals 2 (Cuauhté-
moc Cárdenas), whereas the largest amounts to 17 (Miguel Alemán Valdés). Thus, we
can conclude that having direct contact with many other actors, Miguel Alemán Valdés is
the **most influential node** in our network. The network's plot with the vertex sizes and
colours adjusted by their degree values can be viewed in the Figure 16.

```
================================================================
5. All Degree of N1 (35)
================================================================
Dimension: 35
The lowest value:                              2.0000
The highest value:                            17.0000
----------------------------------------------------------------
Sum (all values):                            234.0000

Arithmetic mean:                               6.6857
Median:                                        6.0000
Standard deviation:                            3.2669
 2.5% Quantile:                                2.8500
 5.0% Quantile:                                3.0000
95.0% Quantile:                               12.3000
97.5% Quantile:                               13.6000
```

Figure 8 – Degree

The network's **diameter** amounts to 4, which refers to the length of the longest
shortest path — in our case, it is the path from Madero Francisco to José López Portillo
(Figure 9).

```
=====================================================================
Searching the longest shortest path in 1. mexican_power.net (35)
=====================================================================
 Working...

         Result:
 The longest shortest path from Madero, Francisco (1) to Lopez Portillo, Jose (29). Diameter is 4.
```

Figure 9 – Degree

Our network has only **one component**, consisting of the whole graph (Figure 10). As is known, an undirected graph is called connected if there is a path between every pair of distinct vertices of the graph. Since a component of a graph is defined as a maximal subgraph in which a path exists from every node to every other, we can conclude that the given network is **connected**.

```
Strong Components
=====================================================================
 Working...
 Number of components: 1
 Size of the largest component: 35 vertices (100.000%).
```

Figure 10 – Number of components

Looking at the **triad census** of our network, one can not help but notice that only 4 types of triads are present (a characteristic of undirected graphs): an **empty triad** (003), a triad with a **reciprocated connection** between two vertices (102), a triad with **two mutual relations and one null relation** (201), a **complete triad** (300). So, it is obvious that there are many cases of empty triads, or triads with one reciprocated connection (Figure 11).

Meanwhile, triads with two relations, or the ones in which all three dyads have a relationship are not as common. Thus, since the majority of triads are concentrated in the left side of the triad distribution, we can assume that the network is **not quite connected**. This corresponds to the fact that only 19.6 percent of all possible edges are present.

```
Triadic Census 1. mexican_power.net (35)
=================================================================
Working...
-----------------------------------------------------------------
  Type    Number of triads (ni)        Expected (ei)        (ni-ei)/ei    Model
-----------------------------------------------------------------
  3 - 102          2460                    316.24                 6.78    Balance
 16 - 300           101                      0.38               265.93    Balance
-----------------------------------------------------------------
  1 - 003          3435                   1759.44                 0.95    Clusterability
-----------------------------------------------------------------
  4 - 021D            0                    316.24                -1.00    Ranked Clusters
  5 - 021U            0                    316.24                -1.00    Ranked Clusters
  9 - 030T            0                    154.81                -1.00    Ranked Clusters
 12 - 120D            0                     18.95                -1.00    Ranked Clusters
 13 - 120U            0                     18.95                -1.00    Ranked Clusters
-----------------------------------------------------------------
  2 - 012             0                   2583.95                -1.00    Transitivity
-----------------------------------------------------------------
 14 - 120C            0                     37.89                -1.00    Hierarchical Clusters
 15 - 210             0                      9.28                -1.00    Hierarchical Clusters
-----------------------------------------------------------------
  6 - 021C            0                    632.47                -1.00    Forbidden
  7 - 111D            0                    154.81                -1.00    Forbidden
  8 - 111U            0                    154.81                -1.00    Forbidden
 10 - 030C            0                     51.60                -1.00    Forbidden
 11 - 201           549                     18.95                27.98    Forbidden
-----------------------------------------------------------------
Transitive          101                    193.08
Intransitive        549                   1059.81
=================================================================
Chi-Square: 62165.5364***
1 cells (6.25%) have expected frequencies less than 5.
The minimum expected cell frequency is 0.38.
```

Figure 11 – Triadic Census

2.2   Subtask 2

As expected, the largest **(weak) component** is the whole graph. The result corresponds to the one we have established in the Subtask 1 because there is no distinction between weak and strong components in undirected graphs.

```
3. Weak Components of N1 [>=1] (35)
================================================================================
Dimension: 35
The lowest value:  1
The highest value: 1

Frequency distribution of cluster values:

   Cluster       Freq      Freq%   CumFreq  CumFreq% Representative
  --------------------------------------------------------------
         1         35  100.0000        35  100.0000 Madero, Francisco
  --------------------------------------------------------------
       Sum         35  100.0000
```

Figure 12 – Triadic Census

Now, we are to compute the **standard importance measures** of the largest component — the network itself — and rank nodes in accordance with their centrality values.

High **betweenness** indicates an actor who is on many paths between other actors. The higher the value, the more powerful the actor is. In our case, with the betweenness value amounting to 0.2303, the most powerful actor is Miguel Alemán Valdés. The second best is Lázaro Cárdenas, whose score is lower by 0.0732 points (Figure 13).

The **closeness centrality** of a node measures its average farness (inverse distance) to all other nodes. The highest value equals 0.6667, while the lowest amounts to 0.3864. Thus, as expected, the node with the shortest distances to all other nodes is Miguel Alemán Valdés, whereas Lázaro Cárdenas is a few points behind him (Figure 14).

Freeman's degree centrality shows the number of connections that the political elites have. Having direct contact with 17 actors, Miguel Alemán Valdés is the most central node. Following Miguel Alemán Valdés with 13 ties, Adolfo Ruiz Cortines ranks second.

```
1. Betweenness centrality in N1 (35)
=======================================================================
Dimension: 35
The lowest value:                      0.0000
The highest value:                     0.2302

Highest values:

      Rank    Vertex                   Value   Id
-----------------------------------------------------------
         1       12                    0.2302  Aleman Valdes, Miguel
         2       10                    0.1570  Cardenas, Lazaro
         3       18                    0.1316  Ruiz Cortines, Adolfo
         4       20                    0.0896  Margain, Hugo B.
         5       31                    0.0649  Salinas Lozano, Raul
         6       23                    0.0598  Carrillo Flores, Antonio
         7       14                    0.0582  Beteta, Ramon
         8        7                    0.0389  Portes Gil, Emilio
         9       19                    0.0353  Lopez Mateos, Adolfo
        10        2                    0.0309  Carranza, Venustiano
-----------------------------------------------------------
Sum (all values):                      1.1729

Arithmetic mean:                       0.0335
Median:                                0.0160
Standard deviation:                    0.0487
 2.5% Quantile:                        0.0019
 5.0% Quantile:                        0.0026
95.0% Quantile:                        0.1392
97.5% Quantile:                        0.1680
```

Vector Values			Frequency	Freq%	CumFreq	CumFreq%
(	...	0.0000]	1	2.8571	1	2.8571
(	0.0000 ...	0.0767]	30	85.7143	31	88.5714
(	0.0767 ...	0.1535]	2	5.7143	33	94.2857
(	0.1535 ...	0.2302]	2	5.7143	35	100.0000
Total			35	100.0000		

Figure 13 – Betweenness Centrality

```
4. All closeness centrality in N1 (35)
=================================================================================
Dimension: 35
The lowest value:                      0.3864
The highest value:                     0.6667

Highest values:

        Rank     Vertex                Value     Id
    -------------------------------------------------------
            1        12                0.6667    Aleman Valdes, Miguel
            2        10                0.5862    Cardenas, Lazaro
            3        23                0.5763    Carrillo Flores, Antonio
            4        18                0.5667    Ruiz Cortines, Adolfo
            5        31                0.5484    Salinas Lozano, Raul
            6        20                0.5484    Margain, Hugo B.
            7        14                0.5397    Beteta, Ramon
            8        19                0.5313    Lopez Mateos, Adolfo
            9        28                0.5231    Ortiz Mena, Antonio
           10        11                0.5231    Avila Camacho, Manuel
    -------------------------------------------------------
Sum (all values):                      16.8776

Arithmetic mean:                       0.4822
Median:                                0.4722
Standard deviation:                    0.0615
 2.5% Quantile:                        0.3864
 5.0% Quantile:                        0.3959
95.0% Quantile:                        0.5793
97.5% Quantile:                        0.5983

        Vector Values                          Frequency      Freq%      CumFreq   CumFreq%
    ---------------------------------------------------------------------------------------
    (                   ...               0.3864]           0     0.0000         0     0.0000
    (          0.3864 ...                0.4798]          20    57.1429        20    57.1429
    (          0.4798 ...                0.5732]          12    34.2857        32    91.4286
    (          0.5732 ...                0.6667]           3     8.5714        35   100.0000
    ---------------------------------------------------------------------------------------
     Total                                                 35   100.0000
```

Figure 14 – Closeness Centrality

```
5. All Degree of N1 (35)
==========================================================================
Dimension: 35
The lowest value:                        2.0000
The highest value:                       17.0000

Highest values:

     Rank     Vertex                  Value   Id
   -----------------------------------------------------------
        1       12                  17.0000   Aleman Valdes, Miguel
        2       18                  13.0000   Ruiz Cortines, Adolfo
        3       10                  12.0000   Cardenas, Lazaro
        4       20                  12.0000   Margain, Hugo B.
        5       23                  11.0000   Carrillo Flores, Antonio
        6       31                  10.0000   Salinas Lozano, Raul
        7       19                  10.0000   Lopez Mateos, Adolfo
        8        7                   8.0000   Portes Gil, Emilio
        9       28                   8.0000   Ortiz Mena, Antonio
       10        2                   8.0000   Carranza, Venustiano
   -----------------------------------------------------------
Sum (all values):                       234.0000

Arithmetic mean:                          6.6857
Median:                                   6.0000
Standard deviation:                       3.2669
 2.5% Quantile:                           2.8500
 5.0% Quantile:                           3.0000
95.0% Quantile:                          12.3000
97.5% Quantile:                          13.6000
```

Vector Values		Frequency	Freq%	CumFreq	CumFreq%
(...	2.0000]	1	2.8571	1	2.8571
(2.0000 ...	3.0000]	5	14.2857	6	17.1429
(3.0000 ...	4.0000]	2	5.7143	8	22.8571
(4.0000 ...	5.0000]	6	17.1429	14	40.0000
(5.0000 ...	6.0000]	7	20.0000	21	60.0000
(6.0000 ...	7.0000]	4	11.4286	25	71.4286
(7.0000 ...	8.0000]	3	8.5714	28	80.0000
(8.0000 ...	9.0000]	0	0.0000	28	80.0000
(9.0000 ...	10.0000]	2	5.7143	30	85.7143
(10.0000 ...	11.0000]	1	2.8571	31	88.5714
(11.0000 ...	12.0000]	2	5.7143	33	94.2857
(12.0000 ...	13.0000]	1	2.8571	34	97.1429
(13.0000 ...	14.0000]	0	0.0000	34	97.1429
(14.0000 ...	15.0000]	0	0.0000	34	97.1429
(15.0000 ...	16.0000]	0	0.0000	34	97.1429
(16.0000 ...	17.0000]	1	2.8571	35	100.0000
Total		35	100.0000		

Figure 15 – All Degree Centrality

Let us plot the network, partitioned by the degree centrality measure (Figure 16).
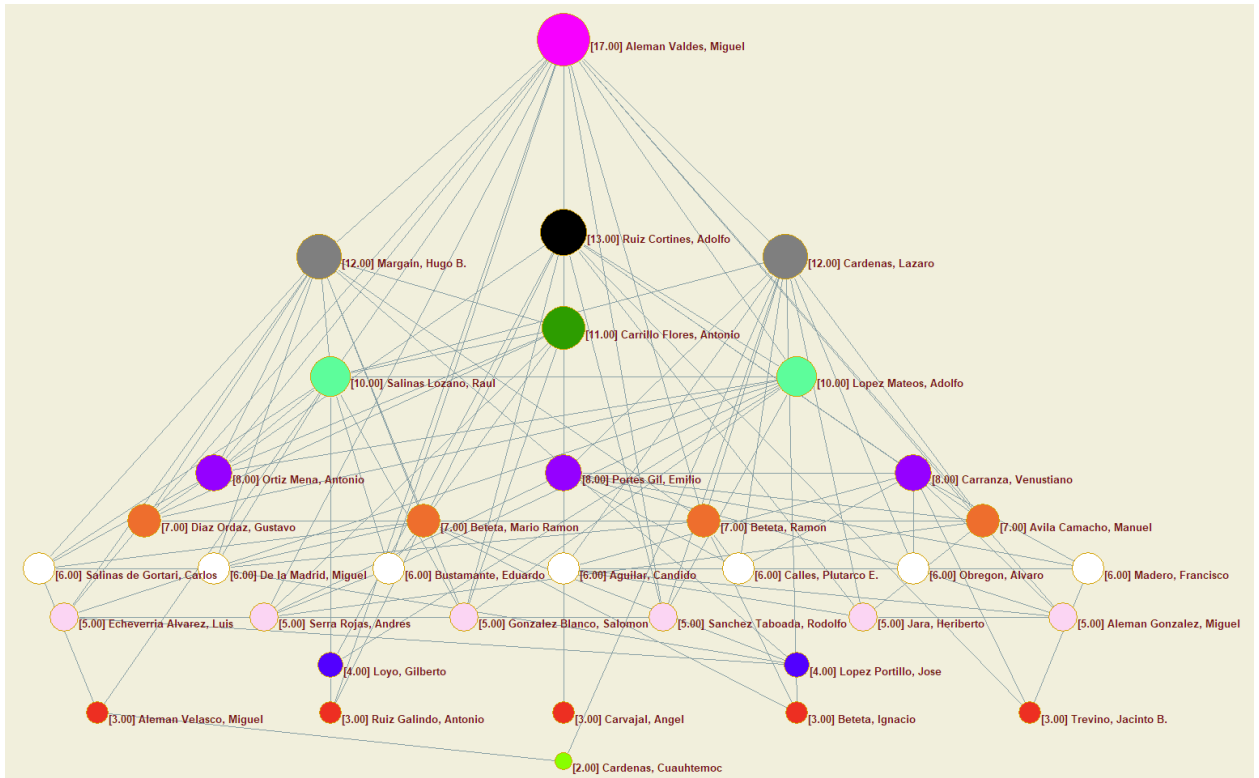
Figure 16 – The network partitioned by the degree

In this layout, the vertex with the highest degree value is at the top, whereas the least central ones are positioned below. The nodes are coloured according to their degree values as well.

As can be seen, the most common degree value is 6 (7 actors), while the most unique ones are 2, 11, 13, 17. With only two connections present, Cuauhtémoc Cárdenas is the least central node, whereas positioned at the very top, Miguel Alemán Valdés is the most powerful one.

2.3 Subtask 3

Now, let us determine the cores in our network (Figure 17).

```
6. All core partition of N1 (35)
============================================================================
Dimension: 35
The lowest value:  2
The highest value: 5

Frequency distribution of cluster values:

   Cluster      Freq      Freq%    CumFreq   CumFreq% Representative
   ----------------------------------------------------------------
         2         2     5.7143          2    5.7143 Aleman Velasco, Miguel
         3         5    14.2857          7   20.0000 Trevino, Jacinto B.
         4         3     8.5714         10   28.5714 Sanchez Taboada, Rodolfo
         5        25    71.4286         35  100.0000 Madero, Francisco
   ----------------------------------------------------------------
       Sum        35   100.0000
```

Figure 17 – All core partition

Coreness is a measure that can help identify tightly interlinked groups within a network. A **k-core** is a maximal group of entities, all of which are connected to at least k other entities in the group. In the k-core, each actor is connected to at least k other actors. The given network contains a large 5-core (25 vertices). In addition, there is a 2-core (2 vertices), a 3-core (5 vertices), and a 4-core (3 vertices).

Given that the 5-core is the largest one, we can conclude that the actors in our network entertain wide-ranging social ties: 71.42 per cent of the members of the elites socialise with at least 5 other political actors.

The plot of the largest core is shown below.
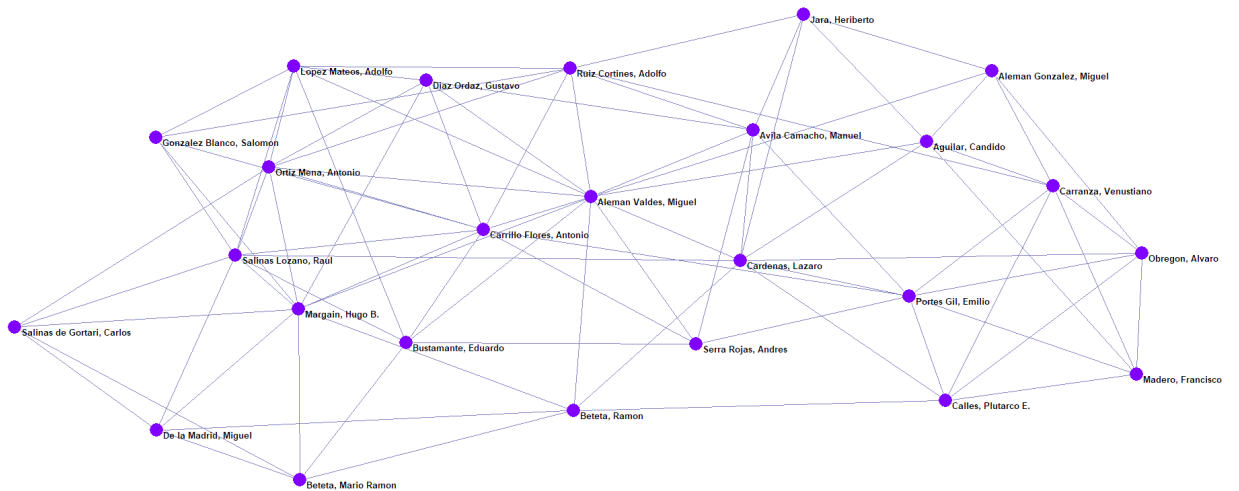
Figure 18 – The plot of the largest core

2.4   Subtask 4

Now we have to use an **island approach** to determine some link islands in the network.

By definition, an **island** is a subnetwork of vertices connected directly or indirectly by lines with a value greater than the lines to vertices outside the subnetwork [**?**]. So, to successfully apply this approach we need to measure the values of links in the network. To do so, we may use the 3-rings method.

3-rings method means that the ring counts are stored as line values — how many times each line belongs to 3-rings. In Pajek, we select `Create New Network > with Ring Counts stored as Line Values > 3-Rings > Undirected`. The result is in the figure below.

Figure 19 – Weighted network

Next, we should try and determine some link islands in the network. To select islands, we have to specify the **maximum size** of it. Depending on this, the number of islands varies. First, we select 10, and then the number will be increased to 18.

18

2.4.1 Maximum Size 18

Firstly, we create a weighted network (**Create New Network > with Ring Counts stored as Line Values > 3-Rings > Undirected**). Then we enable **Generate Network with Islands** and select **Network > Create Partition > Islands > Line Weights, Maximum Size — 18, Minimum Size — 2** (Figure 20-21). Next, we remove isolated vertices (**Network > Create Partition > Degree > All, Operations > Network + Partition > Extract > SubNetwork Induced by Union of Selected Clusters**).
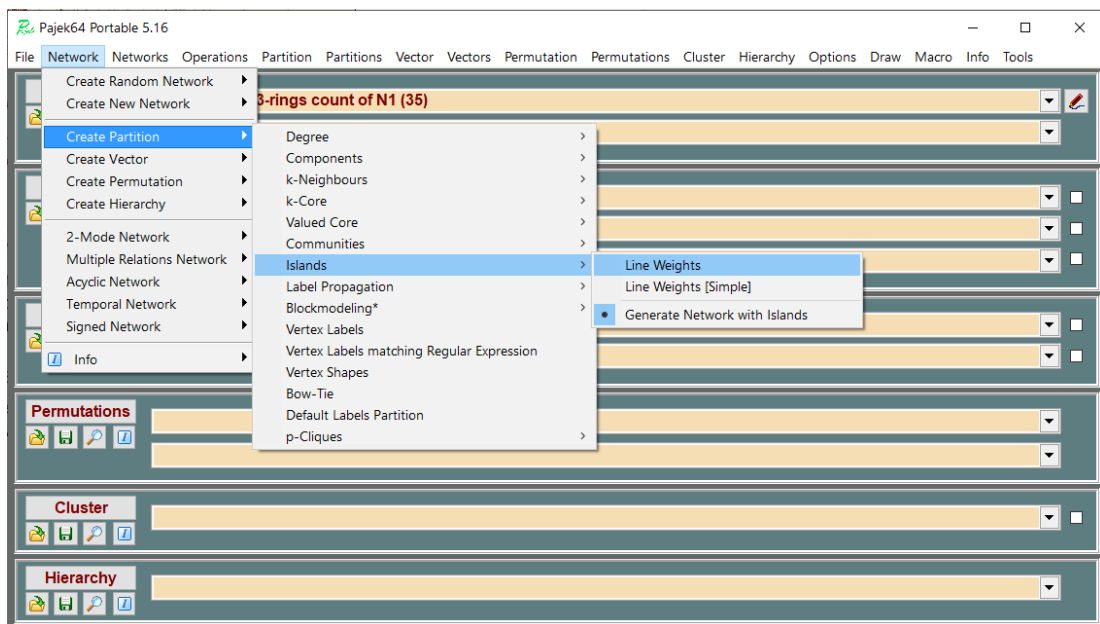
Figure 20 – Generate Islands
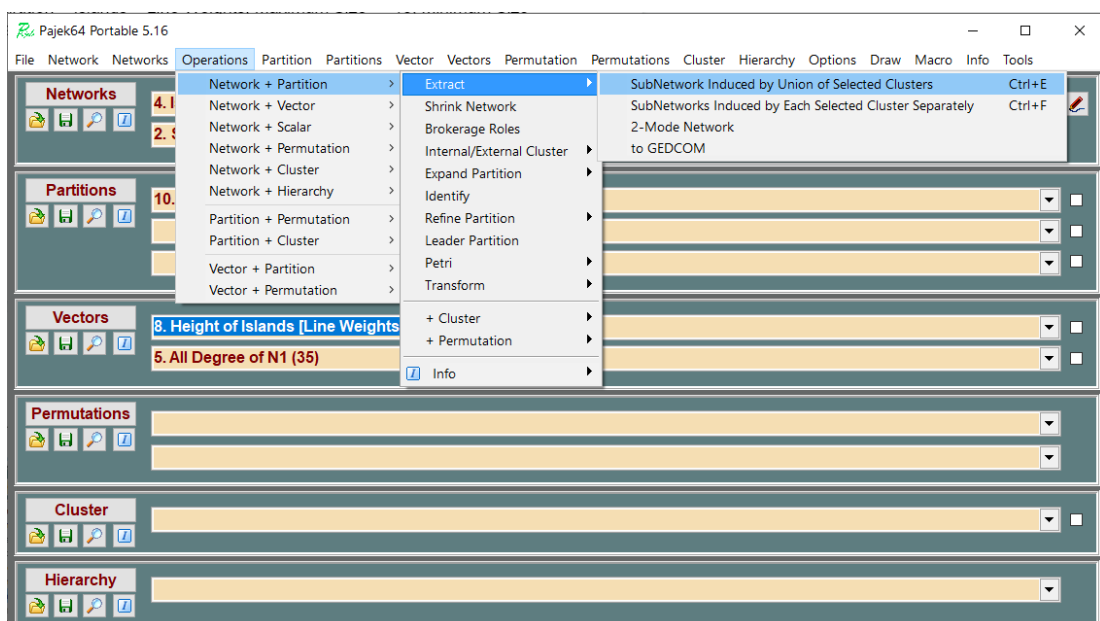
Figure 21 – Remove isolated vertices

After all, we can get a very nice diagram by drawing **Network + First Partition** and selecting **Separate Components**.
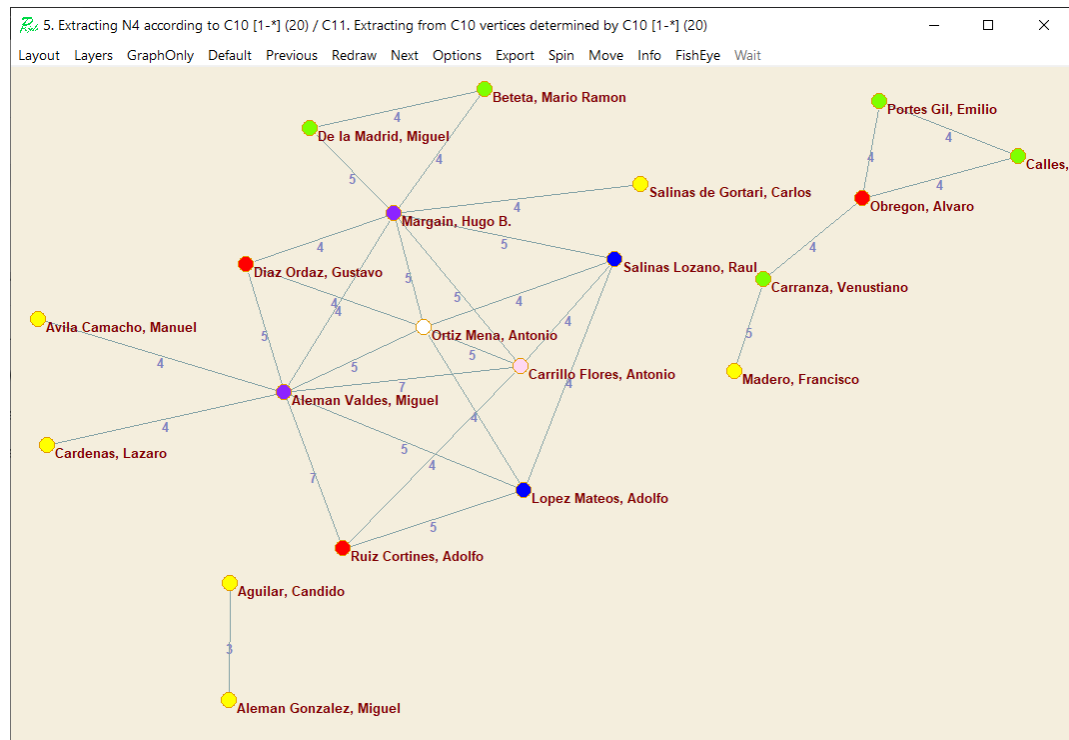
Figure 22 – Generated Islands

What we see is **three islands**, and colours in this case represent the **height of each vertex** (for example, white one is the highest point of the island). The big island contains **13 vertices**, and two others — **2 and 5 vertices** (Figure 23).

Figure 23 – Islands

In order to interpret the picture, we should apply given **attributes** to the network (military/civilians and year). For doing that, we have to **cut unused items** from **mexican_year.clu** and **mexican_military.clu** partitions (add islands partition as the second one and select **Partitions > Extract SubPartition (Second from**

First). Also we copy the year partition to a vector, thus we can add these values to vertex labels: **Options > Mark Vertices Using > Vector Values**).
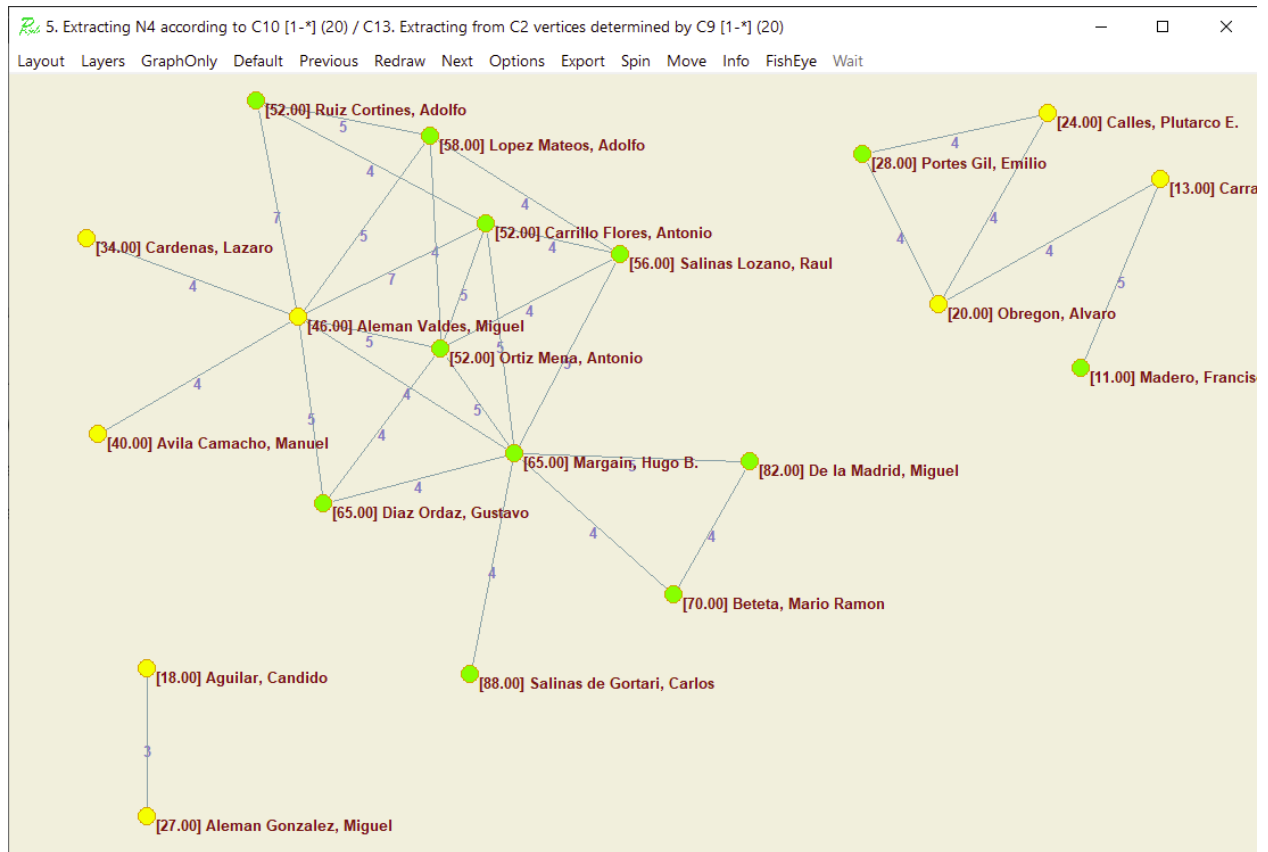
Figure 24 – Islands with attributes

Here we see that the big island consists **mostly of civilians** (green colour), so those civilian actors form a cohesive group with links of significant values. Also this island is mostly constructed of people who were active in the **second half of the century**, as can be observed from the labels.

The island of size 5 consists of three military and two civilian actors, all of whom were active **between 1911 and 1928**. So it may be suggested that during that time such a mixed group was active.

The island of size 2 is not particularly informative.

2.4.2   Maximum Size 10

Let's reduce the maximum size of an island to 10. Create new islands with **Network > Create Partition > Islands > Line Weights, Maximum Size — 10, Minimum Size — 2**. Also we remove isolated vertices (**Network > Create Partition > Degree > All, Operations > Network + Partition > Extract > SubNetwork Induced by Union of Selected Clusters**). The result is in the figure below.
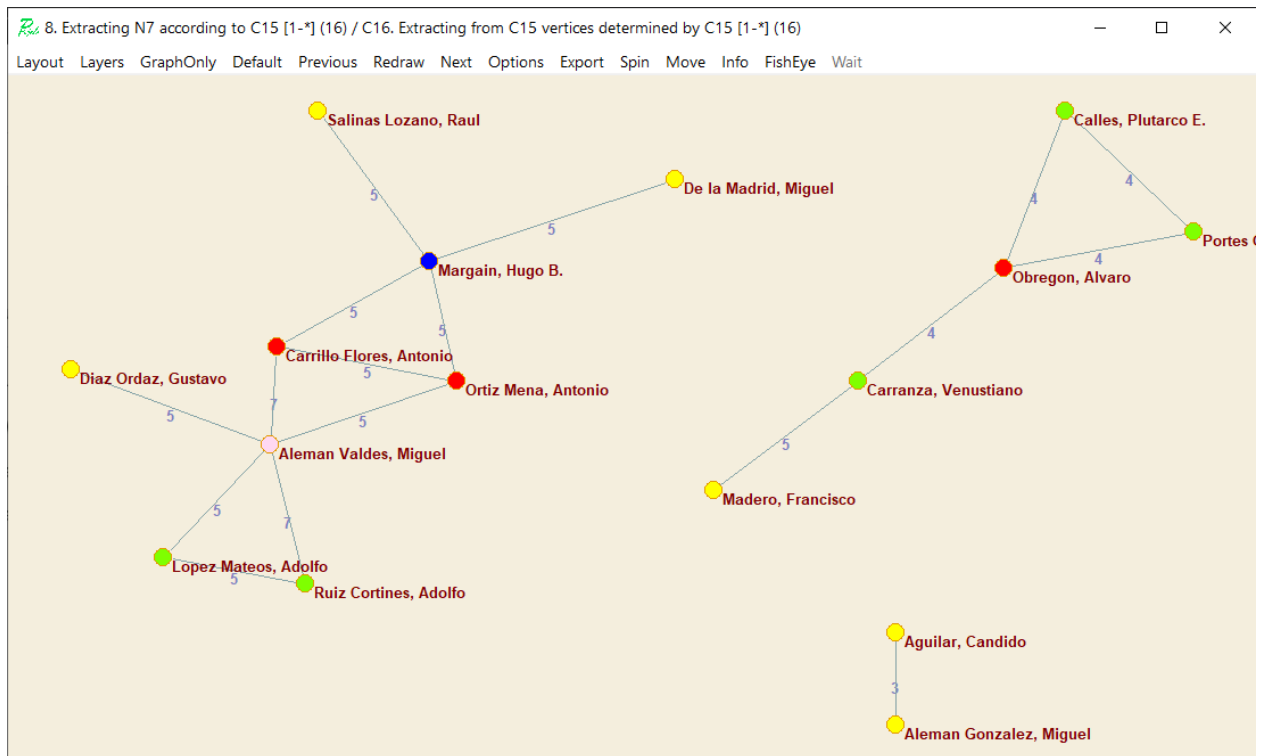

Figure 25 – Generated Islands

Again, we should apply attributes to the islands (Figure 26).
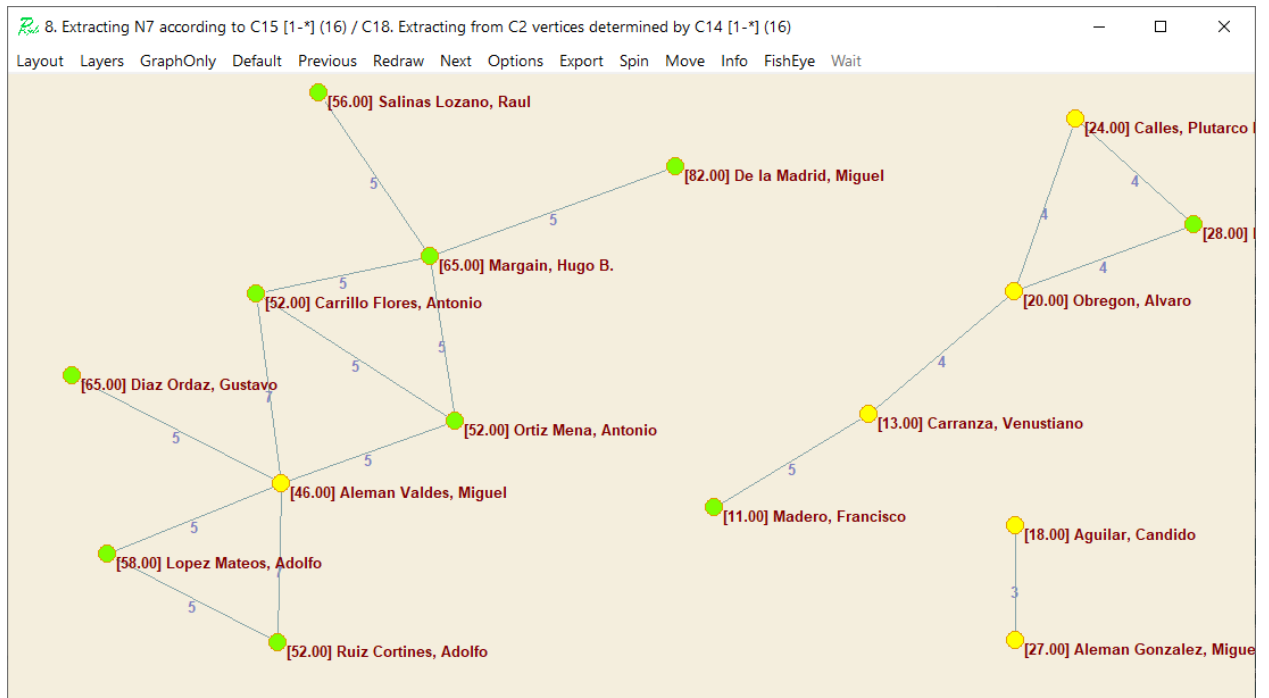
Figure 26 – Islands with attributes

As might be seen, two small islands did not change, and the biggest island now contains only **9 vertices**. They are **mostly civilians**, except the highest one, who is a military person.

This version of the island is **more coherent** than in the previous case, and here we may see that all actors have been active **since 1946** (after World War II). Furthermore, this island might have been "started" with the military person *Aleman Valdes* (1946), as he is in the middle of the island, and all other actors who performed later are directly or indirectly connected to him.

2.5 Subtask 5

Now we need to make a **line-cut** of the network. Line-cut is another useful technique for **extracting parts of a network** where only lines (and their vertices) above a certain value are retained [?]. In other words, we select only those parts of the network where values of lines are above a certain threshold, and those parts are probably of more importance than others.

To do so in Pajek, we start with the **weighted network** from the previous step and select **Network > Create New Network > Transform > Remove > Lines with Value > lower than**. But before we start, it is useful to analyse what **line values** are present in the network (**Network > Info > Line Values**, Figure 27).

```
================================================================================
3. Undirected 3-rings count of N1 (35)
================================================================================
Lowest value of line:          0.00000000
Highest value of line:         7.00000000

      Line Values                         Frequency      Freq%    CumFreq   CumFreq%
      ----------------------------------------------------------------------------------
(                    ...        0.0000]         6       5.1282         6      5.1282
(            0.0000 ...         1.0000]        19      16.2393        25     21.3675
(            1.0000 ...         2.0000]        37      31.6239        62     52.9915
(            2.0000 ...         3.0000]        26      22.2222        88     75.2137
(            3.0000 ...         4.0000]        17      14.5299       105     89.7436
(            4.0000 ...         5.0000]        10       8.5470       115     98.2906
(            5.0000 ...         6.0000]         0       0.0000       115     98.2906
(            6.0000 ...         7.0000]         2       1.7094       117    100.0000
      ----------------------------------------------------------------------------------
  Total                                       117     100.0000
```

Figure 27 – Line Values

As we see, lines with values from 0 to 2 form 52% of the network, so to get more important half of it, we may cut starting with value 3.
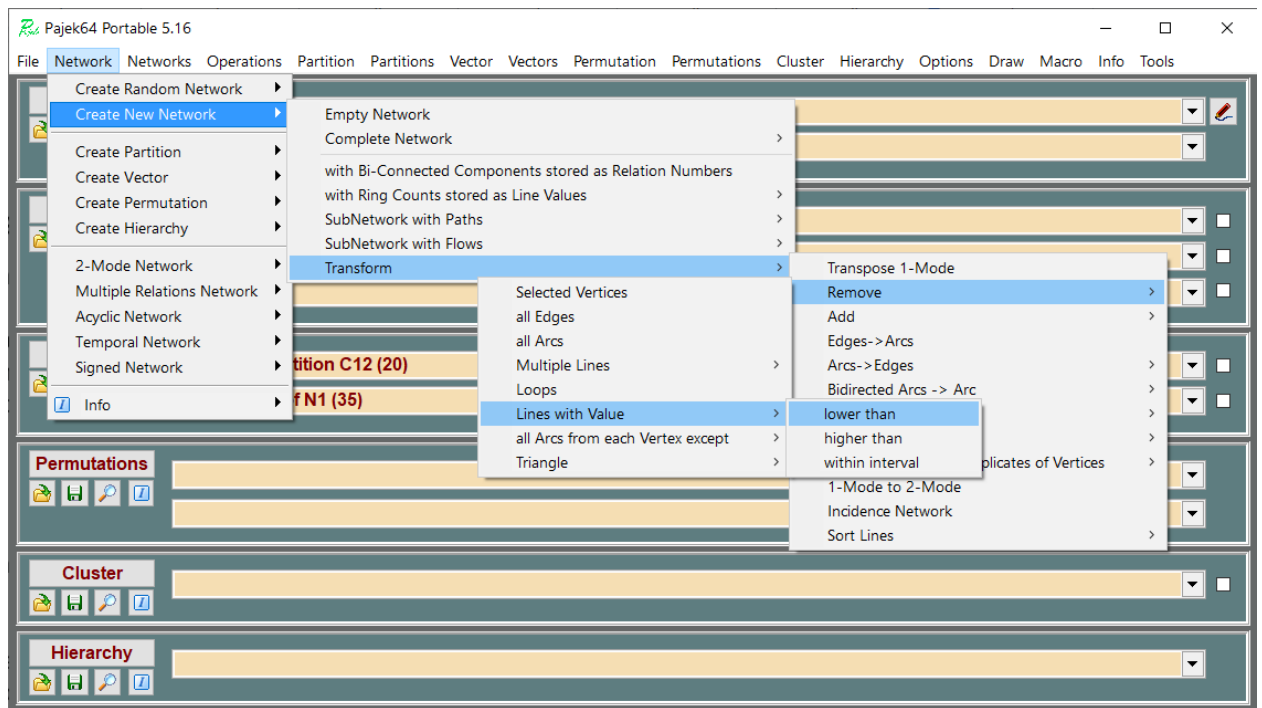
Figure 28 – Line-cut

After applying attributes, we see that the separated part of the network is rather massive. It contains mostly civilians, however, all military actors are connected with each other. It also could be noticed that years are increasing from right to left, specifically on the left there is a well-connected group of vertices (in a star shape) with year varying between 1911 and 1928. In fact, this group has already been observed before, as it is an island.
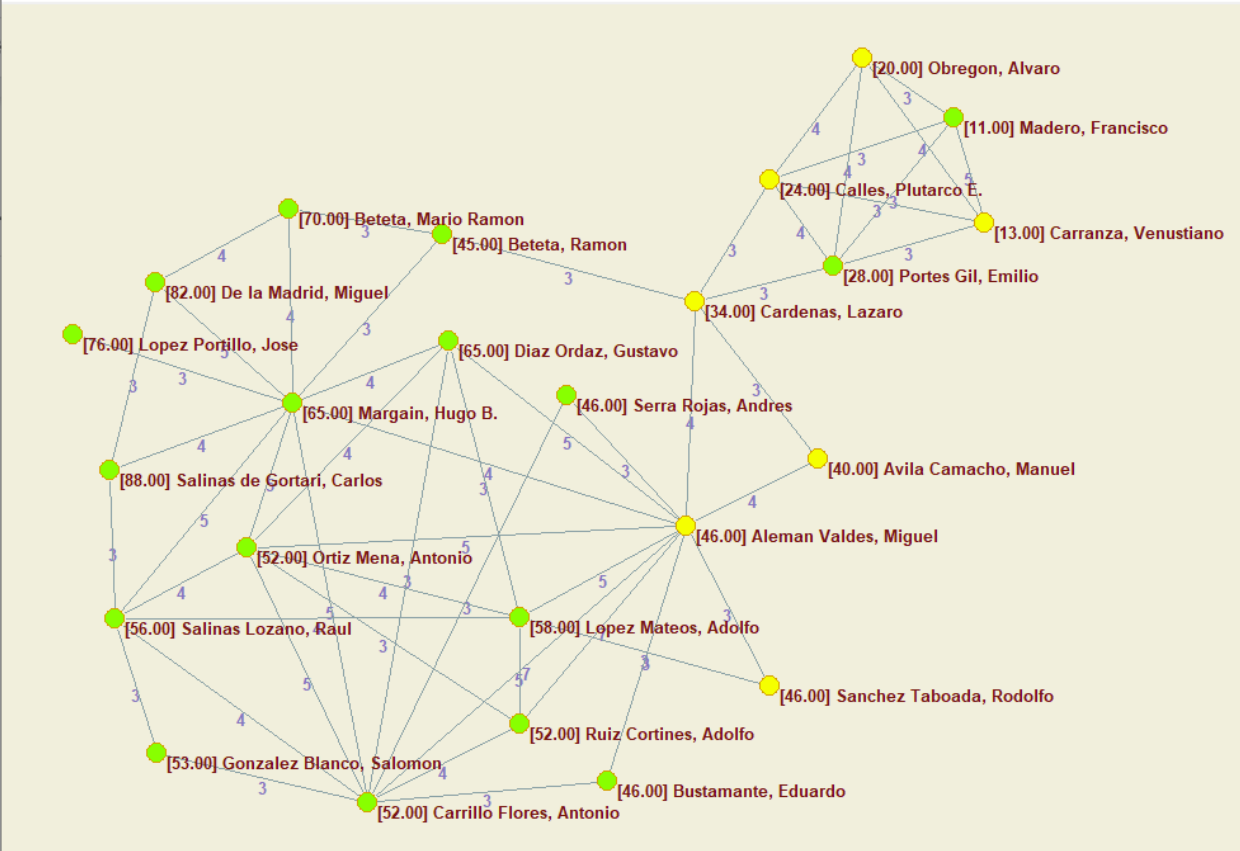
Figure 29 – The result of line-cut of 3

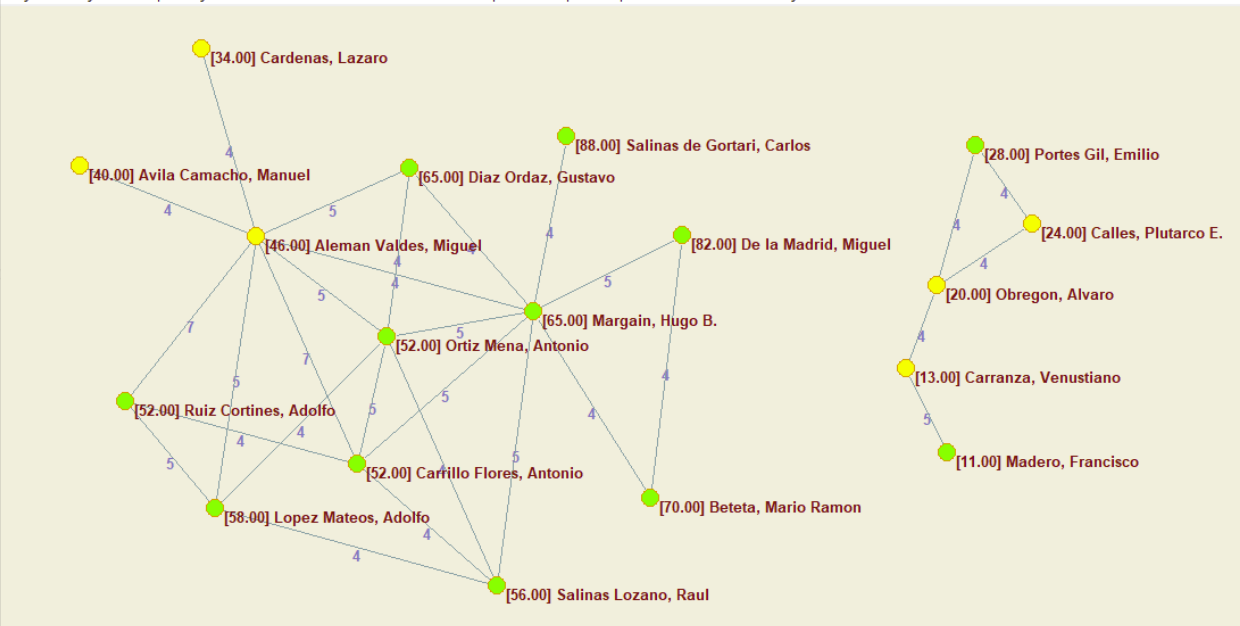Let's cut further and select the threshold of 4.

Figure 30 – The result of line-cut of 4

Now the smaller island is separated from the bigger one, and the number of vertices decreased as well. In fact, the big one is similar to the big island we discussed earlier. So we may conclude that in our case **line-cut works similarly to the island approach**.

Additionally, if we remove isolated vertices from the line-cut network, we can limit the property partitions by doing the following. `Select All degree partition, then Partition > Binarize partition`. Then Extract SubPartition and copy the result to vector.

2.6 Subtask 6