**LUIS FELIPE VILLOTA**

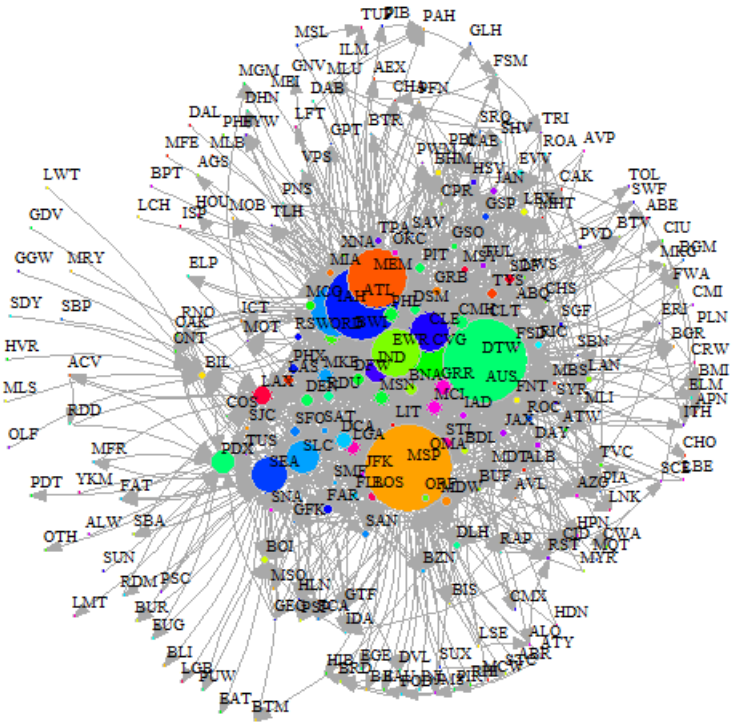
LUND UNIVERSITY
USING SOCIAL THEORY – SIMP56

# PART 2B ASSIGNMENT RE-EXAM: SOCIAL NETWORK ANALYSIS

## Part A (40 points)

## AIRLINES DATASET:

*Figure 1. Plot of the network of routes(flights) made between airports by degree centrality*




1. **Is the network directed or undirected?**

   The network is directed.

2. **How many airport and flights routes there are?**

   We have 235 unique airports (nodes) and 1297 different routes (edges). The total number of edges are distinctive from-to combinations; implying that there are no observations in our data that have more than one connection between the same two airports. On other hand, there were 210 unique "sources" and 193 unique "targets", meaning that we have more variety of places of departure than actual destinations.

3. **What is the density of the network?**

   Density for our directed network: *N*(N-1)* = 0.0235

4. **What is the average degree of the network?**

   *Total edges / total nodes* or *(N-1) *density* = 5.519

5. **What is the average shortest path of the network?**

   The mean of the shortest path between each pair of airports = 3.006053

6. **Which are the 5 most central airports in this network? Use 3 different centrality measures and present the top airports in a nice table. Describe and interpret the results.**

   *Table i. Top 5 most central airports (by descending degree centrality)*

| | Id | Label | City | Code | Degree | Eigenvector | Closenness | Betweenness | Hub | Authority |
|---|---|---|---|---|---|---|---|---|---|---|
| *1* | 136 | Minneapolis - St. Paul Intl | Minneapolis | MSP | 130 | 0.9974053 | 0.002958580 | 8006.087 | 1.0000000 | 0.5067683 |
| *2* | 50 | Detroit Metropolitan Wayne County | Detroit, MI | DTW | 125 | 1.0000000 | 0.002832861 | 6121.558 | 0.3464955 | 1.0000000 |
| *3* | 80 | Hartsfield-jackson Atlanta International | Atlanta, GA | ATL | 112 | 0.9489819 | 0.002808989 | 5794.573 | 0.5674164 | 0.7965608 |
| *4* | 130 | Memphis International | Memphis | MEM | 90 | 0.8252992 | 0.002624672 | 5717.460 | 0.7779475 | 0.4572270 |
| *5* | 70 | George Bush Intercontinental | Houston | IAH | 81 | 0.8081088 | 0.002538071 | 2928.226 | 0.3248180 | 0.7570478 |

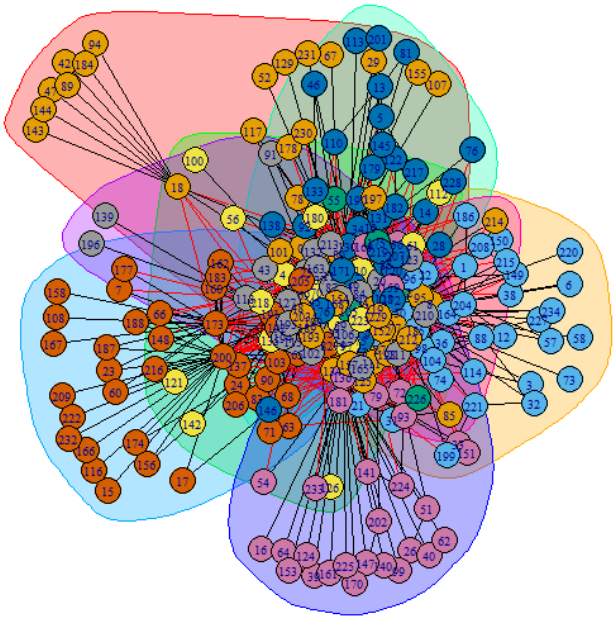
   The most central vertex in the network is St. Paul International airport in Minneapolis, -in terms of connections concurrent upon it (130 routes, of which 49 were incoming flights and 81 were departures from this place). We used 6 centrality measures which corroborate its salient position. Although, we found out that Detroit Metropolitan – Wayne County airport in Detroit, MI is almost as central (with 5 incident links less), an actually scored higher for the Eigenvector indicator, meaning also it is better connected to other highly influential (central) nodes.

7. **Are there different communities in the network (use Louvain algorithm)? Describe, interpret and visualize the results.**

   Yes. First of all, we have a connected network (all nodes are connected somehow). Then, using the Louvain algorithm, we have found that its *modularity* = 0.242, which reflects that the number of edges in a given cluster is superior to the number of edges

expected by chance within that module. There were 9 communities identified, and the biggest module had 42 nodes (airports). Finally, the network is not hierarchical.

*Figure 2. Visualization of the 9 communities in the Airports dataset (using the Louvain algorithm)*



**IRON MAN MOVIE (2008) DATASET:**

8. **Is the network directed or undirected?**

   The network is directed.

9. **How many actors and ties there are?**

   We have 18 vertices (characters) and 54 connections.

10. **What is the density of the network?**

    Density for our directed network: *N*(N-1)* = 0.1764

11. **What is the average degree of the network?**

    *(N-1) *density* = 3

12. **What is the average shortest path of the network?**

    The mean of the shortest path between each pair of characters = 1.98366

13. **Who are the 3 most important characters in this movie in terms of speaking themselves?**

*Table ii. Top 3 most important characters (by descending "nlines"= total lines spoken)*

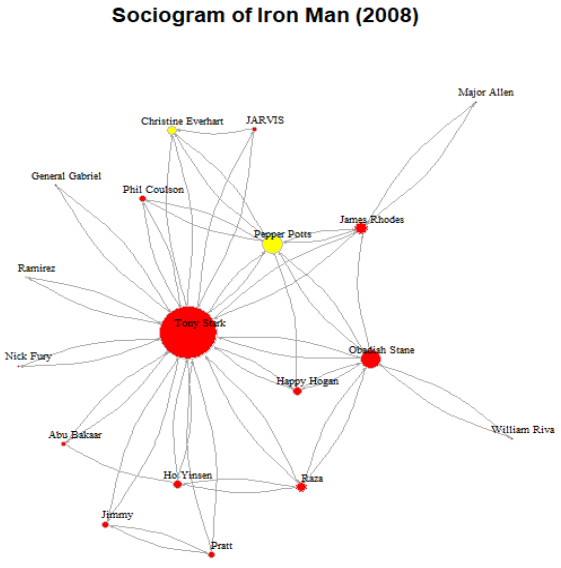
| | char_ID | char_name | char_female | nlines | linesin |
|---|---|---|---|---|---|
| *1* | 1 | Tony Stark | 0 | 353 | 368 |
| *2* | 8 | Pepper Potts | 1 | 123 | 135 |
| *3* | 4 | Obadiah Stane | 0 | 80 | 62 |

**14. Who does James Rhodes interact with in the movie? Are there any characters who he speaks to, but they do not speak back, or vice versa?**

James Rhodes speaks to 3 characters: Tony Stark (ID: 1), Major Allen (ID: 15) and Pepper Potts (ID: 8). All the above characters speak back to him, plus Obadiah Stane (ID: 4), so in total 4 characters speak back to him. In conclusion, James does not speak to Obadiah.

**15. Visualize the network so that node size depends on some centrality measure and node color corresponds to the sex of the character.**

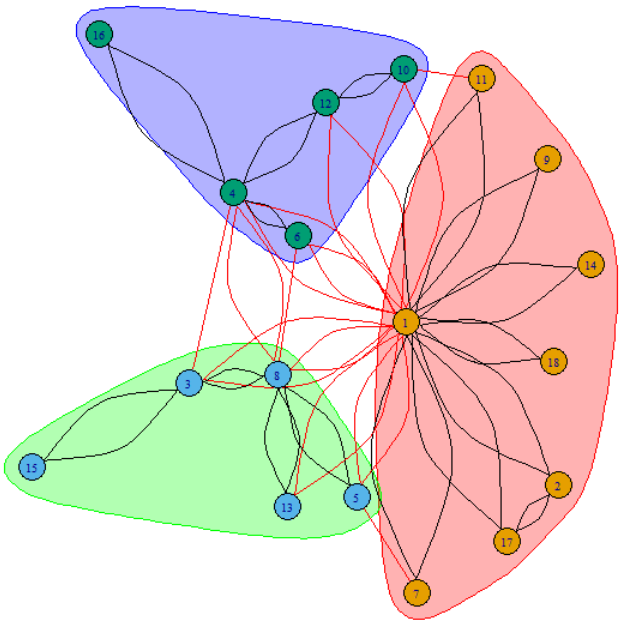
*Figure 3. Network of characters' interactions (node size by degree and colored by sex)*



16. **Are there different communities in the network (use Louvain algorithm)? Describe, interpret and visualize the results.**

Yes. First of all, we have a connected network (all nodes are connected somehow). Then, using the Louvain algorithm, we have found that its *modularity* = 0.234, which reflects that the number of edges in a given cluster is superior to the number of edges expected by chance within that module. There were 3 communities identified, and the biggest module had 8 nodes (characters). Finally, the network is not hierarchical.

*Figure 4. Visualization of the 3 communities in Iron Man Movie (2008) (using the Louvain algorithm)*




17. **Perform clustering based on edge betweenness for an undirected network (set seed = 2). How closely do the clustering results match with the community algorithm results? Tabulate the two memberships and calculate the association between the two memberships (χ-squared)). Describe, interpret and visualize the results.**

*Figure 5. Dendogram (clustering based on edge betweenness as an undirected network)*

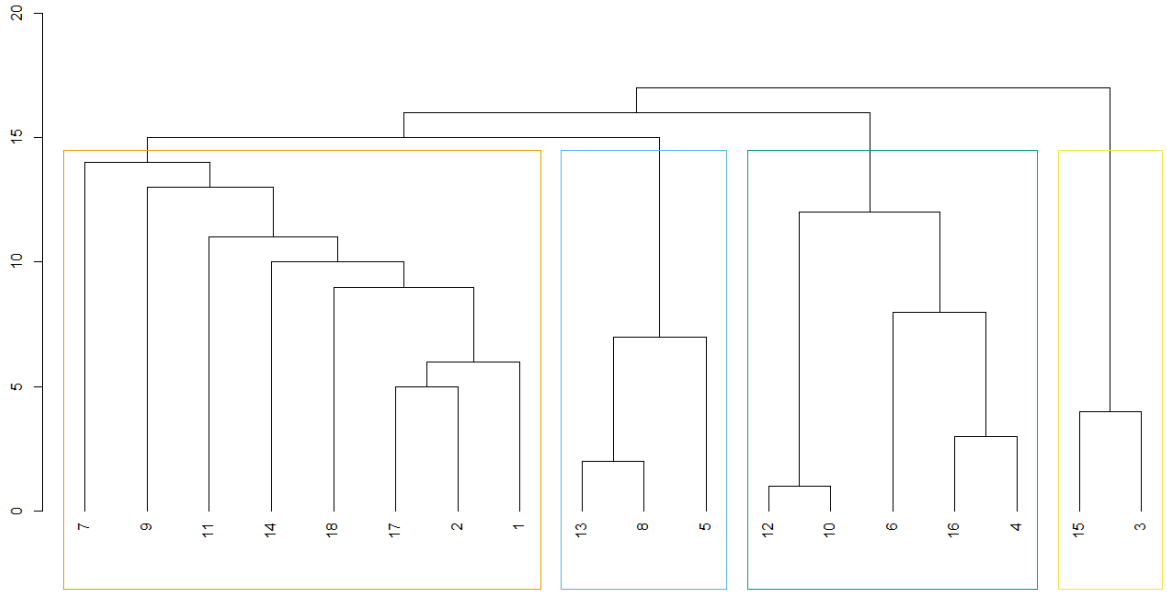



*Figure 6. Visualization of the clusters in Iron Man Movie (2008) (clustering by the edge betweenness)*

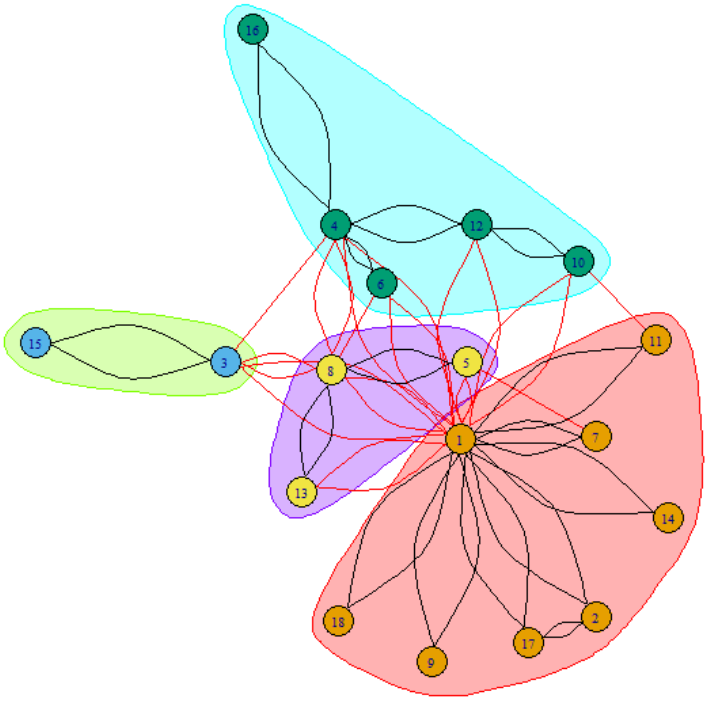


*Figure 7. Visualization of the network with nodes labelled and colored by clustering membership*

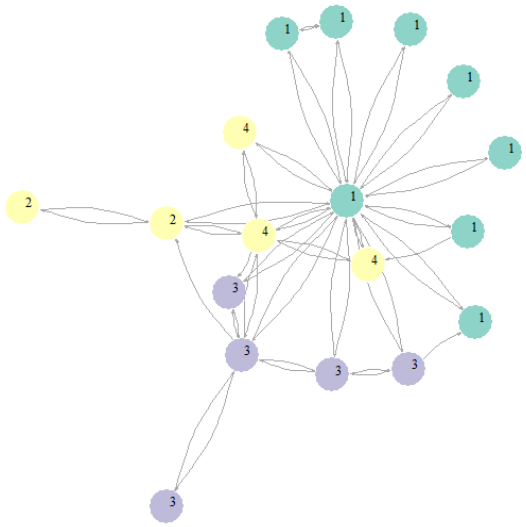


*Table iii. Tabulation of community membership (Louvain algorithm) vs Clustering (edge betweenness)*

| Communities \ Clustering | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 8 | 0 | 0 | 0 |
| 2 | 0 | 2 | 0 | 3 |
| 3 | 0 | 0 | 5 | 0 |

We observe that the clustering by edge betweenness "opened" a new group of 3 nodes (airport ID's: 5, 8 13). Moreover, their association (using Pearson's Chi-squared test) shows: $X\text{-squared} = 36$, $df = 6$, $p\text{-value} < 0.05$. The two "memberships", are hence significantly associated.

**18. Briefly compare the two networks to other real-life networks (use the examples from the last slide in Lecture 2). Include a table with the main descriptives of networks that you compare and interpret the results.**

*Table iv. Comparison of the two studied networks (with the indicators used)*

| Network | Type | Nodes | Edges | Density | Av. Degree | Sh. Path | Recip. | Comm. | Top Node by Deg. C. |
|---|---|---|---|---|---|---|---|---|---|
| Airlines | D | 235 | 1297 | 0.023 | 5.519 | 3 | 0 | 9 | MSP |
| Iron Man Movie (2008) | D | 18 | 54 | 0.176 | 3 | 1.983 | 0.925 | 3 | Tony Stark |

Both networks had more edges than nodes and turned out to be directed in our analysis, meaning that not each and every one of their vertices had "a mutual" connection. However, the majority of the interactions between characters in the Iron Man movie were reciprocal. Hence, this graph is almost the same as the example of the Slide 2 of film actors (undirected, with a low average of distances between nodes, e.g.). The network of flights between airports can be compared to the technological (train routes) network, with the difference that the transportation means (airplanes flying) allow for a more diversified "flow", - allowing our graph to be directed.

## Part B (10 points)

**Create a social network (e.g. friends, relatives, classmates etc.) with at least 16 actors.**

19. **Compare your network to a random network and small world network of the same size (also set seed). Provide a brief description of network, including a table with the main descriptives and figures of degree distribution for all 3 networks.**

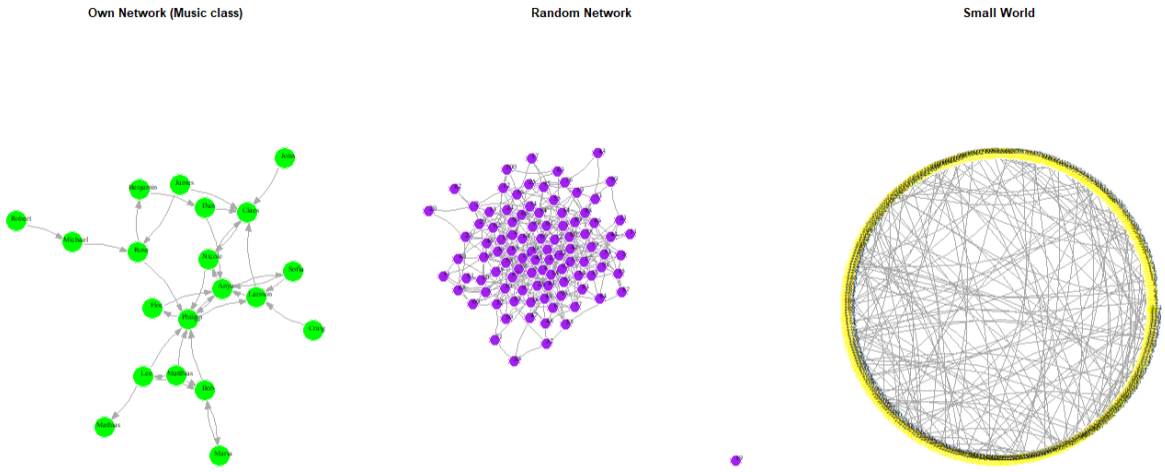
*Table v. Comparison of own network, a random network and a small world network (basic indicators)*

| Network | Type | Nodes | Edges | Density | Av. Degree | Sh. Path | Recip. |
|---|---|---|---|---|---|---|---|
| Own | D | 20 | 35 | 0.092 | 1.75 | 2.378 | 0.22 |
| Random | UD | 100 | 340 | 0.068 | 3.4 | 2.598 | 1 |
| Small World* | UD | 333 | 333 | 0.006 | 1 | 10.165 | 1 |

*In Milgram's "small-world" (1967) average degree of separation is 6 degrees.

My network is composed of 20 individuals who are in the music classes I attend to in Malmö. It turned out to be a directed network, maybe because of the types of instruments that are taught to play, - but not "mingled" in a musical ensemble yet, or people who have heard others playing one-way, e.g.

**20. Present a very basic visualization of all 3 networks (just to capture the basic structure of connections).**




**21. Create a list of top 5 members by 3 centrality measures for your network.**

*Table vi. Top 5 members in own network (by decreasing degree centrality)*

| | Name | Degree | Eigenvector | Closenness |
|---|---|---|---|---|
| *1* | Philipp | 10 | 0.9978295 | 0.03125000 |
| *2* | Anya | 9 | 1.0000000 | 0.02564103 |
| *3* | Claes | 6 | 0.3469663 | 0.02272727 |
| *4* | Larsson | 5 | 0.5230953 | 0.02564103 |
| *5* | Bob | 5 | 0.3317373 | 0.02222222 |

**In your network:**

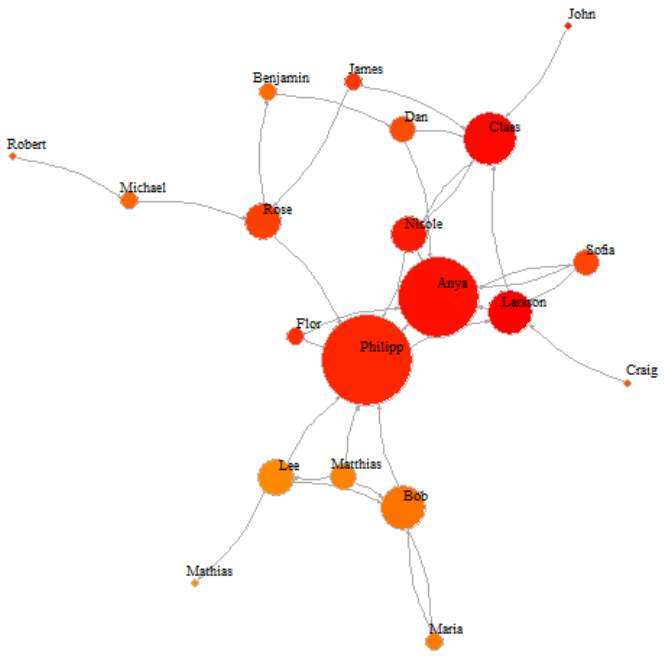
a. **Who you consider the key figure to contact for distributing information? Why?** Anya. She scored the highest in the Eigenvector

centrality and has meaningful connections, -as she is a teacher, actually.

b. **Who should get vaccinated first to avoid quick spreading of contagious disease? Why?** Philipp. Is the most central individual with 10 link (7 incoming and 3 out).

**22. Create a visualization of your social network (use labels, colors etc.).**

*Figure 8. Visualization of own network with color and sizes by degree centrality*




R code: https://github.com/FelipeVillota/SIMP56---Using-Social-Theory/blob/main/sna_reexam.R