COMP5048 - Visual Analytics

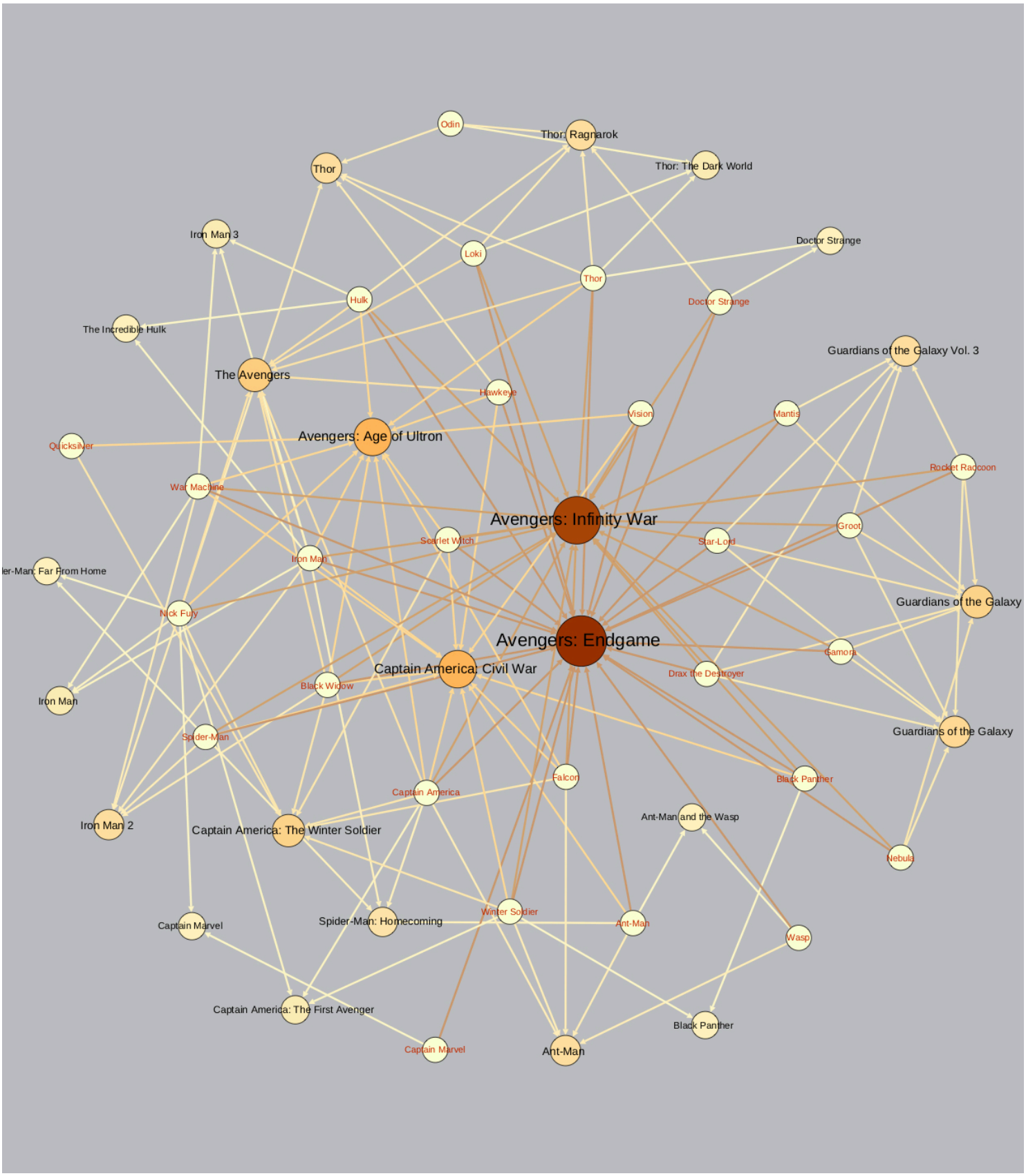
Assignment #1: Individual Work

Zhecheng Zhong (490319299)
zzho7727@uni.sydney.edu.au

Semester 2, 2019

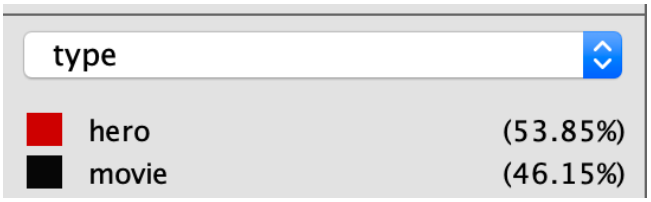
1. A1 - Marvel Cinematic Universe

1.1. Visualization of graph



(Figure 1.1: Marvel Cinematic Universe)

1.2. Graph Description



(Figure 1.2: Hero node and movie node are discriminated by the color of label name)

Based on comic books, the marvel cinematic universe is a series of superhero film. The graph shown above depicts the relationship between 28 characters and 24 movies. As illustrated in (Figure 1.2), hero names are represented by red label name while movie names are highlighted by the black color.

This is a directed graph. Nodes with deeper color and bigger size, have more inter-connections, which is the total degree of nodes. For instance, Avengers: Endgame and Avengers: Infinity War have the most number of characters appeared in the movies.

1.3. Tools & Algorithm

Python and Gephi are the main tools to implemented the graph where Python is utilized for preprocessing given data to a .gml data set. Then Gephi is used to draw the graph.

The graph employs Fruchterman and Reingold algorithm which is a standard force-directed layout. It has two principles: 1. Vertices connected by an edge should be drawn near each other. 2. Vertices should not be drawn too close to each other[1]. Fruchterman and Reingold algorithm assumes vertices to be point-shaped and defines two forces for influencing vertices[2]:

f_{attr}: an attractive force that pushes connected vertices towards each other.
f_{rep}: a repulsive force that disperses connected vertices against each other.

Let u, v being two connected nodes, the absolute value of the forces is computed as:

$$fattr(u,v) = k^2/distance(u,v)$$
$$frep(u,v) = distance(u,v)^2/k$$

The positions of the two-dimensional vertices determine the directions of the forces. In the above equation, K describes the distance between two connected vertices whose f_{attr} = f_{rep}.

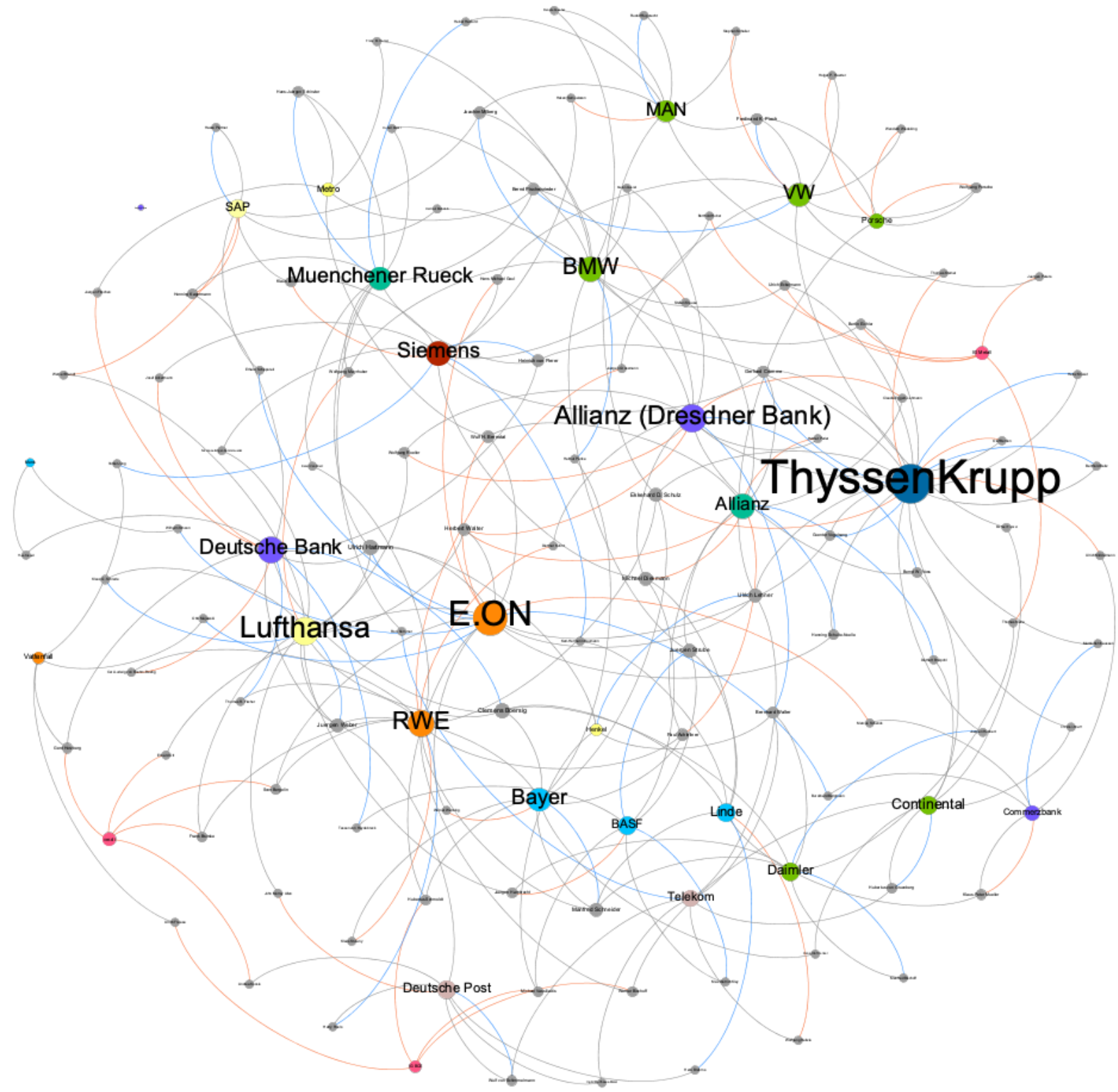
1.4. Strength and Weakness

This approach works fine for point-shaped vertices. It clearly specify the edge relationship between two nodes. However, it cannot cluster related nodes and it cannot deal with two-dimensional vertices. It cannot ensure that vertices do not overlap, but only prevents their centers from touching.

2. B2 - Network from Social Science

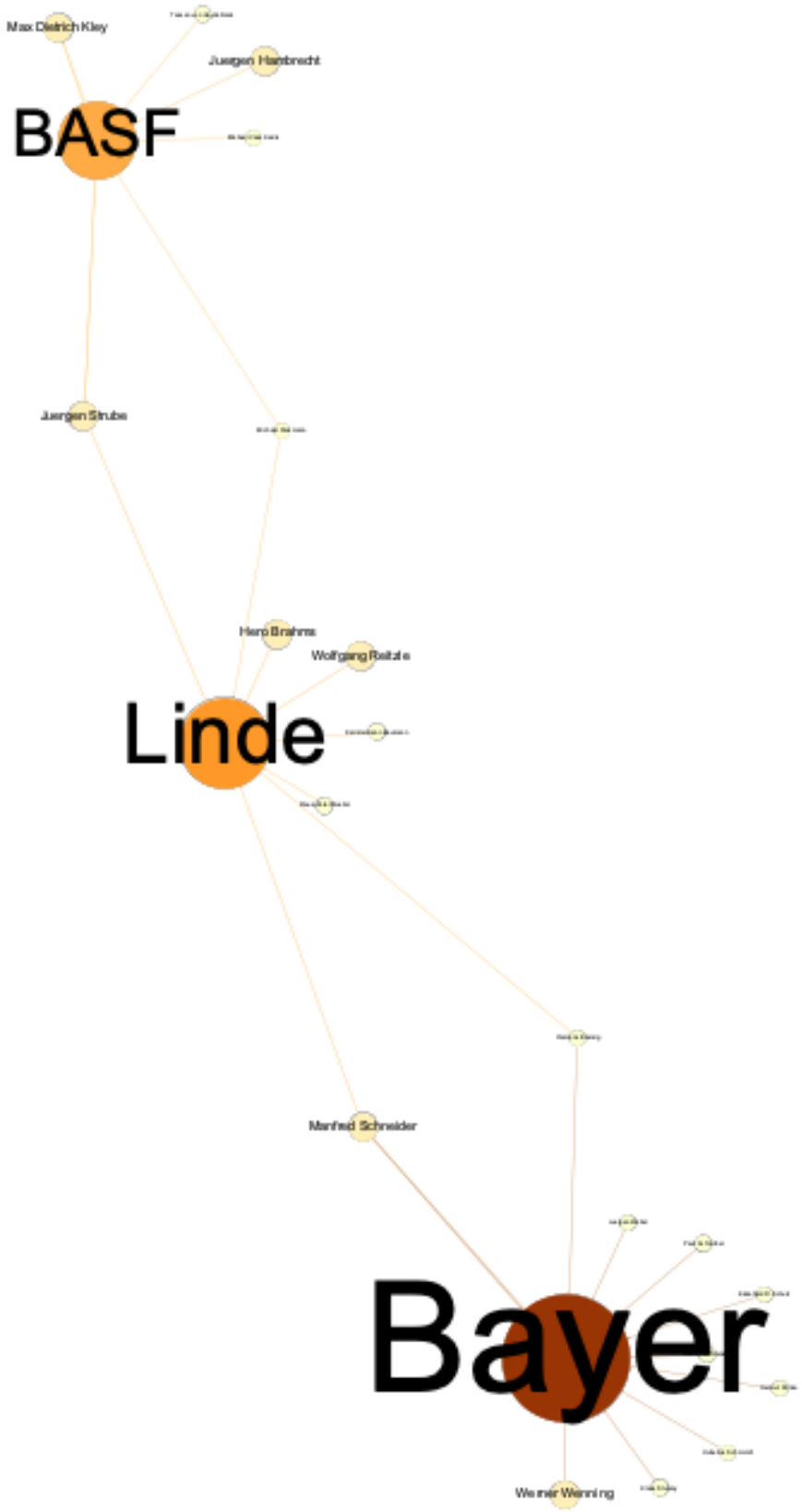
2.1. Visualization of graph

2.1.1. Visualization of whole graph - Fruchterman-Reingold



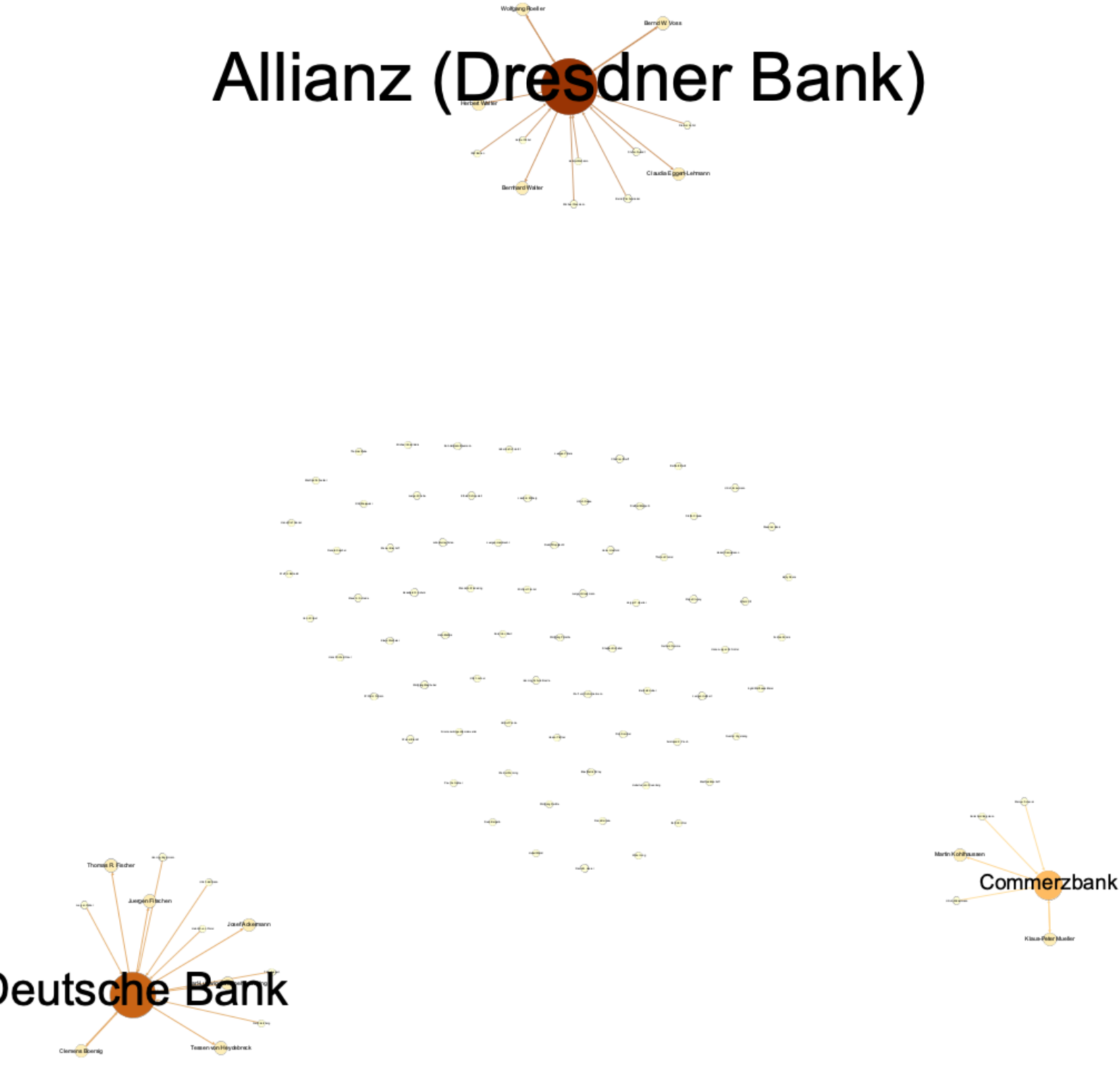
(Figure 2.1.1: relationships between companies and various persons in Germany)

2.1.2.Visualization of chemistry industry - Force Atlas



(Figure 2.1.2: relationships between chemistry companies and various persons in Germany)

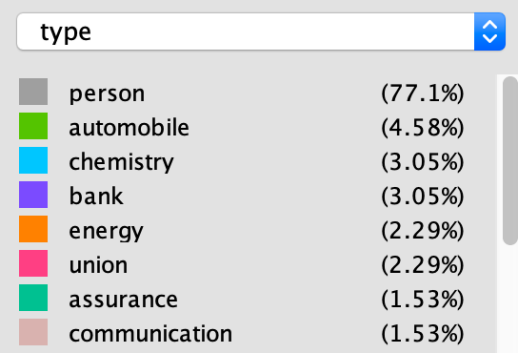
2.1.3.Visualization of bank industry - Force Atlas



(Figure 2.1.1: relationships between banks and various persons in Germany)

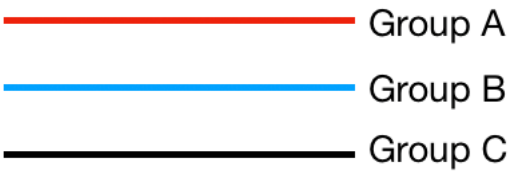
2.2. Graph Description

2.2.1. Visualization of whole graph



(Figure 2.2.1: Node colors are cataloged and differentiated by industries)

Figure 2.1.1 depicts the network of employment between person and companies in various industries. It consists of 129 nodes and 271 edges. Nodes are explained in Figure 2.2.1 that different colors stand for different industries. And those nodes with deeper color and bigger size, have more inter-connections, which is the total degree of nodes.



(Figure 2.2.2: edge colors are cataloged and differentiated by group)

As shown in Figure 2.2.2, there are three edge colors representing different meaning of connections. Group A connects from a company to a person, meaning that this person is employed by that company. Group B connects from a company to a person meaning that this person was recently employed by that company. Group C connects from a person to a company meaning that this person serves in the supervisory board of the company.

2.2.2. Visualization of chemistry company and bank

Figure 2.1.2 and Figure 2.1.3 draw the employment network between person and chemistry companies & banks. It clearly shows the employment relationship among person and companies, especially when a person has relations with several different companies in same industry.

2.3. Tools and Algorithm

Python and Gephi are the main tools to implemented the graph where Python is utilized for preprocessing given data to a .gml data format. Then Gephi is used to draw the graph.

Figure 2.1.1 is drawn based on the Fruchterman and Reingold algorithm. Details about the algorithm have been illustrated in section 1.3.

Figure 2.1.2 and figure 2.1.3 are drawn based on Force Atlas algorithm, which is a spatial layout algorithm for real-world networks. It has different way of computing repulsion and attraction forces, which is also called as “energy model”. Besides, Force Atlas algorithm continuously estimates the situation to maximize speed and minimize the swinging[3].

Force Atlas algorithm has a classical attraction force which is expressed by :

$$fattr(u, v) = d(u, v)$$

The repulsion part is inspired to bring poorly connected nodes closer to very connected nodes[3]. The formula of repulsion force is as follows:

$$F_r(n_1, n_2) = k_r \frac{(deg(n_1) + 1)(deg(n_2) + 1)}{d(n_1, n_2)}$$

Where the coefficient k_r is defined by the settings. The +1 is important as it ensures that even nodes with degree zero still have some repulsion force.

2.4. Strength and Weakness

	Strength	Weakness
Figure 2.1.1 Fruchterman and Reingold	Clearly specify the differences of nodes and edges.	This algorithm cannot cluster nodes with common characteristics.
	Layout of graph is easy to read.	
	Correlated companies and people are placed in adjacent areas.	
Figure 2.1.2 & 2.1.3 Force Atlas	Clearly specify the employment relations between person and companies.	It focuses mainly on the company node, ignoring the information of person.
	Easy to find nodes cluster (employees of a company) and connection node(an employee engaging in two or more companies).	

Reference

[1] Fruchterman T M J, Reingold E M. Graph drawing by force-directed placement[M]. 1991.

[2] Drawing Graphs with Circular Vertices. *https://schneide.blog/tag/fruchterman-reingold/*

[3] Mathieu Jacomy, Tommaso Venturini, ForceAtlas2, A Graph Layout Algorithm for Handy Network Visualization