Linnæus University
School of Computer Science, Physics and Mathematics

Evaluation of Network Comparison Approaches

Author: Hailelul Lakew Teshome
Supervisors: Prof. Dr. Andreas Kerren and Ilir Jusufi.

Subject: Computer Science
Level: Master
Date: June, 2013

Abstract

Network visualizations have been used for quit long time. Different disciplines use this visualization to compare a given dataset. Identifying better comparison approach that is used for information visualization is indispensable both for the people who are using it and for developers who are looking for a better way of visualizing huge data. In this thesis a task based approach has been used to analyze two different network comparison approaches namely Juxtaposition (showing different objects compared in separate space or time) and Superposition (overlaying objects in the same space). Thirty students at Linnaeus University have participated in the questionnaire to evaluate the usability of the two approaches. SPSS tool is used to analyze the data collected from the participants and the result explicitly indicates that there is no significant variation between Juxtaposition and Superposition comparison approaches. The result can be used as a recommendation for domain specific professionals and developers in their quest for better network comparison for their audience.

Key words: Network/graph visualization, Juxtaposition, Superposition, Explicit Encoding.

Acknowledgement

I would like to express my deepest gratitude and appreciation to all those who have helped me complete this thesis. Specially, my supervisor Prof. Dr. Andreas Kerren and Ilir Jusufi for your unreserved support and advice. Many thanks go to Mathias Hedenborg and Hans Frisk for arranging their students participate in the study. Further, I would like to thank all students who gave me their share of time to participate voluntarily in the research. In addition, I would like to thank my family, without a doubt, if had it not been for your help, everything would not have been a reality.

Contents

1 Introduction 1

1.1 Problems 1

1.2 Motivation 2

1.3 Goal 3

1.4 Thesis Outline 3

2 Background 4

2.1 Graph Drawing 4

2.1.1 Drawing Conventions 5

2.2 Graph Drawing Approaches 6

2.2.1 Node-Link Diagrams 6

2.2.2 Matrix Representation for Graphs 7

2.3 Aesthetical Criteria 8

2.4 Graph Drawing Algorithms 9

2.5 Perception in Visualization and Gestalt Laws 10

2.6 State-of-the-Art in Graph Visualization 11

2.7 Evaluation of Information Visualization Techniques 14

2.7.1 Evaluation Methods 14

2.7.2 Challenges of Evaluation of Information Visualization 15

2.7.3 Evaluation in Network Visualization 16

3. Network Comparison 18

3.1 Introduction to Network Comparison 18

3.2 Juxtaposition 21

3.2.1 Juxtaposition using Node-Link Diagram 25

3.3 Superposition 25

3.3.1 Superposition using Node-Link Diagram 27

3.4 Explicit Encoding 28

3.5 Hybrid Approach 30

4 Evaluation of Network Comparison Approach 34

4.1 Questionnaire 34
4.2 Method 35
4.3 Procedure 37
4.4 Participants 37
4.5 Analysis of Result 38
4.5.1 Participants' Comments 43
5 Discussion and Conclusion 45
5.1 Future Work 46
Bibliography 47
Appendices 53
Appendix A: Questionnaire 54
Appendix B: Participants Personal Information 64
Appendix C: Statistical data of participants 66
Appendix D: Independent Samples T-Test 72

List of Figures

Figure 2:1 Drawing of the same graph: 5
Figure 2:2 Drawing of the same digraph: 6
Figure 2:3 A simple node-link diagram 7
Figure 2:4 Matrix representation of graphs 7
Figure 2:5 Nodes to the right are in a state of equilibrium 10
Figure 2:6 A regular network representing the yeast life cycle. 12
Figure 2:7 Visual representation of the GAL4 protein interaction 12
Figure 2:8 With whom do you discuss issues important to your work? 13
Figure 2:9 Types of methodologies in relation to precision, ... 15
Figure 3:1 Screenshot of a reconciled phylogenetic tree displayed 19
Figure 3:2 shows how primary categories 21
Figure 3:3 Dendogram matrix view (Jin Chen et al., 2009). 22
Figure 3:4 Toot tip showing the information that matches between 22
Figure 3:5 AdaptiveTree to show tournament features 23
Figure 3:6 ManyNets, rows represent networks of call in 5-hours 24
Figure 3:7 Pop-up window that you can get by right click on specific cell 24
Figure 3:8 A simple node-link graph representation 25
Figure 3:9 The orbit-based comparison method of Protein-Protein 26
Figure 3:10 Visualization of Movies in orange and people in blue 27
Figure 3:11 A simple superimposed node-link diagram 27
Figure 3:12 EMDialog a selected statement and the corresponding tree 29
Figure 3:13 The screenshot of DAG visualization for multiple classification, 30
Figure 3:14 The Semantic Graph Visualizer (SGV) comparing two 31
Figure 3:15 A screenshot of the BGplay system (Pier Francesco et al., 2006) 32
Figure 3:16 Protein network comparison simultaneously using 33
Figure 4:1 shows distribution of educational level of participants 38
Figure 4:2 Shows the percentage distribution based on the 39
Figure 4:3 Shows the percentage distribution among the participants 39
Figure 4:4 shows the percentage distribution among the participants 40
Figure 4:5 shows the percentage distribution among the participants 40
Figure 4:6 Shows how participants detect the difference between 42
Figure 4:7 Shows how participants detect the difference between 42
Figure 4:8 Shows us that the number of participants who favor 43

List of Tables

Table 3:1 M. musculus PPI-network rank of correlation coefficient 26
Table 4:1 Shows order of questions in the questionnaire................................. 35
Table 4:2 shows the number of differences between two paired graphs.......... 36

1 Introduction

Currently due to the advancement of information communication technology and its easy access, people are exposed to huge data. During their daily activities, tremendous amount of data are being produced. To get useful information from these data, we need to show them graphically so that people can get the information easily by using the visual understanding of human being. The unique capability of human visual system detects patterns and features within very short period of time (Fekete et al., 2008). We need to have efficient visualization methods to show the textual description visually by exploiting the perceptual ability of people towards the collection of large datasets (Tarawneh et al., 2011). Traditionally, visualization tools were used to compare objects and to give visual figure that can be understood easily by humans. It enables people to easily understand complex datasets and developing insight (Fekete et al., 2008). But, using traditional visualization to represent large data often result in loss of data and even some of the information would be hidden (Krempel, 2009). Currently, there are lots of improvement in the technology combining visual abstraction and data mining to get the best out of it (Novotny, 2004).

Groundbreaking work of network visualization was done by Moreno in the context of the analysis of social networks (Moreno, 1953). With the growing use of networks in many disciplines, computer programs are available to allow networks to be represented visually. Researchers in the past decades are highly interested visualizing relational datasets (Carnecky et al., 2012). Many application areas use network visualization to depict their domain specific structures. Often, networks are visualized so that they can be compared to other networks. There are three fundamental and well known visual designs for comparison in general: Juxtaposition (showing different objects compared in separate space or time), Superposition (overlaying objects in the same space) and Explicit Encodings (Gleicher et al., 2011).

1.1 Problems

Data objects can be depicted using network visualization. Different field of science use network visualization to visualize large datasets for the audience. The issue of comparison has to address the demand of connectivity within and between objects by user. Currently, we can see a number of visualization designs for comparison. Amongst, Juxtaposition, Superposition and Explicit Encoding are the prominent (Gleicher et al., 2011). We can use these and their combination for comparison. So far, to the best of our knowledge there is no known study as to how users perceive these approaches. Therefore; the aim of this thesis is to come up which network comparison is best and why.

1.2 Motivation

Network visualization is very important for analyzing relational data that can be observed in different fields (Kerren et al., 2007). Comparing objects can be done with different visualization systems. Comparative visualization systems compare complex objects by simplifying the objects using abstractions (Gleicher et al., 2011). Visualizing complex conceptual structures is very important for many different fields of science. A lot of information visualization system use graph drawing to realize and depict the connection between objects. Graph drawing is vital to show graphs, networks of complex conceptual structures for many applications (Di Battista et al., 1999). We can represent facebook users and their interaction among themselves using this technique where the vertices represent entities and edges representing their relationships (Leeuw, 2009). In biology, the technique can be used to show the phylogenic maps of individuals or species. It is used to show the relationships between different individuals along their phylogenic root and to show if they are highly related evolutionary or if they are distantly related. In biochemistry or bioinformatics, the technique is used to show the arrangements of different molecules and chemical reactions that are going on in certain metabolic reactions. In software engineering, network visualization is used to show the complexity of different systems and also to depict the internal behavior or state of a compiler. In object oriented field, it used to show the relation between different components such as classes and UML of a given system (Tarawneh et al., 2011). Researchers in geography and cartography use visualization to analyze and explore spatiotemporal data (Andrienko et al., 2003).

In all these fields of science above the network comparison could be used during visualization of a given set of data. Network comparison is vital for comparing two groups of data using graph or network for better visualization and understanding of the dataset in question. For instance, network comparison can enable us compare social networks such as telephone call networks and get some insight about the networks behavior which can be used as an input for stakeholders (Freire et al., 2010). The three fundamental network comparison approaches can mixes to form a hybrid. We have Juxtaposition + Explicit Encoding, Superposition + Explicit Encoding, Juxtaposition + Superposition and the combination of all three. Each and every network graph that we are using of visualization can be grouped into the one of the above categories. Knowing which one is better in providing easy comparison for the audience is the key for people who are engaged in data visualizations. Understanding the different methods of network visualizations categories which provide the best comparison approaches is very important. Knowing the suitable visualization techniques for giving people the best perception of visual information is indispensable. People favor one visualization technique than the other for their purpose of comparison and get information easily and comfortably, knowing which one of the techniques that are favored by the audience could be one step ahead for further research in the area.

1.3 Goal

Audience has their own way of favoring one network comparison approach from the other. Better knowledge of different comparison approaches is important for researchers and develpers to develp comparision tools or come up with specific visualization solutions (Gleicher et al., 2011). *The goal of this thesis is to identify explicitly which network comparison approach namely, Juxtaposition or Superposition is better for the purpose of comparison of a given dataset by analyzing data collected from the task based usability study.*

1.4 Thesis Outline

In the above discussions we have described: Problems, Motivation and Goal of the thesis. Detailed literature review presented in Chapter 2, Background, which contains: Graph Drawing, Graph Drawing Approaches, Aesthetical Criteria, Graph Drawing Algorithms, Perception in Visualization and Gestalt Laws, State-of-the-Art in Graph Visualization, Evaluation of Information Visualization Techniques. Chapter 3 discusses introduction of network comparison where three fundamental of visual comparison designs described. Chapter 4 describes the evaluation of the network comparison approach using the data collected from the participants and Chapter 5 allocated for discussion and conclusion.

2 Background

This chapter is divided into five parts. Section 2.1 (Graph Drawing) discusses about graph drawing, where general information about graph drawing concept and terminologies are discussed. Section 2.2 (Graph Drawing Approaches), discusses about graph drawing approaches where prominent drawing approaches presented and dicussed. Section 2.3 (Aesthetical Criteria) part goes to the criteria that make a given diagram appealing for the audience and Section 2.4 (Graph Drawing Algorithm) discusses graph drawing algorithms in relation to graphs, Section 2.5 (Perception in Visualization and Gestalts law) discusses about perception and Gestalts law and Section 2.6 (State-of-the-art in graph visualization) deals with the state-of-the-art in network visualization.

2.1 Graph Drawing

Graph is an abstract structure that is used to model information and present it to the user. The study of graphs came to the scene in the 18th century when the Konigsberg bridge problem was realized by Euler. Since then the graphs have become an interest for many researchers. Graph drawing is an established field that focuses on how to draw network using different algorithms by enhancing its aesthetics (Chaomei Chen, 2004). The entities of a graph are represented by vertices and the relationships by edges. Many information visualization systems use graphs to represent a given information content with easy and understandable manner. Mathematicians have been identifying geometric representation of graphs for centuries for the purposed of visualization and intuitions. In the 1960s, pioneer work of graph drawing as a diagram is presented to help and understand software come into the picture (William T. Tutte, 1963).

A graph *G* can be represented mathematically: $G = \{V, E\}$ consists of a set of vertices *V* (also called nodes) and a set of edges *E* and are sometimes called links, arcs or connections. The set of nodes in the graph may represent an object in question and the edges may represent relation between the two nodes depending on what type of data that we are dealing with. These nodes and edges could be used to visualize complex data. Human perception of these sets of nodes and edges vary from individual to individual. The different approach that we follow to draw a graph can affect the way we visualize the information in questions. In this regards, it is worth mentioning the importance of graph drawing concepts related to drawing conventions (Di Battista et al., 1999).

2.1.1 Drawing Conventions

Drawing conventions are very important to illustrate graphs that can be acceptable for representation. Some of the drawing conventions that are widely used while we are drawing graphs are (Di Battista et al., 1999):

Polyline Drawing: in this convention, each edge represented as chain of edges in the polygons, it is illustrated in Figure 2.1 a.

Straight-line Drawing: edges are represented as straight lines, it is illustrated in Figure 2.1 b.

Orthogonal Drawing: each edge is drawn as a polygonal chain of alternating horizontal and vertical segments of having 90 degrees between them (Di Battista et al., 1999). It is illustrated in Figure 2.1 c.

Grid Drawing: vertices, crossings, and edge ends have integer coordinates.

Planar Drawing: no two edges cross.

Upward (resp. downward) Drawing: each edge drawn as a curve monotonically non-decreasing or non-increasing. If the edges are pointing upward, then we call it strictly upward and if the edges are all pointing downward, then we call it strictly downward (Di Battista et al., 1999).

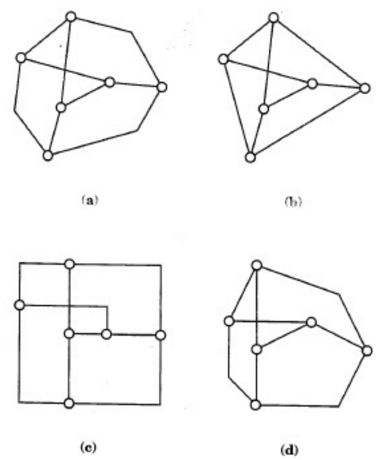


Figure 2:1 Drawing of the same graph: (a) polyline (Di Battista et al., 1999); (b) straight-line (Di Battista et al., 1999); (c) orthogonal (Di Battista et al., 1999); (d) polyline grid (Di Battista et al., 1999).

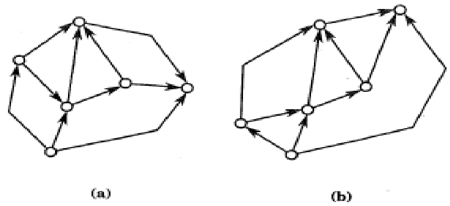


Figure 2:2 Drawing of the same digraph: (a) planar polyline; (b) strictly upward planar polyline (Di Battista et al., 1999).

2.2 Graph Drawing Approaches

The main idea of information visualization is to represent a given dataset visually so that people can easily get the information fast and clear. So far, there has been lots of network visualization system developed (Di Battista et al., 1999). Internet connectivity, telephone calls , telephone system, world wide web, social network such as facebook etc.. are some of the areas where the visualization system engaged (Freire et al., 2010; Pier Francesco et al., 2006; Staša Milojević et al., 2012).

There are two different graph drawing approaches. These are: Node-link diagrams and Matrix display (Matthew O. Ward et al., 2010).

2.2.1 Node-Link Diagrams

Graphs are often represented by node-link diagrams. Currently there are different graph drawing methods that enable us to produce node-link diagrams. Node-link diagrams are widely used in depicting graphs and networks and also hierarchical structure and parent child relationships can be easily represted by node-link diagram (Micheal Burch et al., 2011). Freeman in his survey and history of social network visualization has described it well (L. Freeman, 2000). Nested structure can be clearly depicted when we are using node link diagram. But this comes with its own drawbacks. It uses screen inefficiently and do not scale well when we are dealing with large datasets (Shengdong Zhao, et al., 2005). Figure 2:3 below is a simple node-link diagram, the nodes are data objects (for expame people) and the edges could be relationships etc.

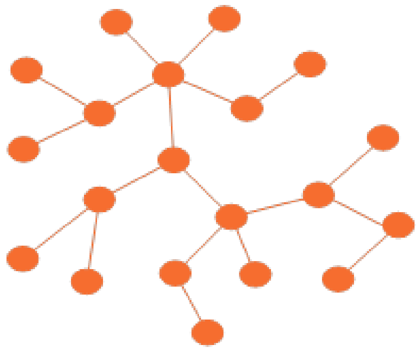
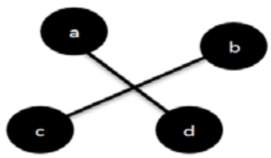


Figure 2:3 A simple node-link diagram.

Planar graph drawing technique is another. The main purposed of this graph drawing technique is to avoid edge crossing. This algorithm is well known because planar graph drawing techniques is practised for quite long peroid of time and we can get a lot of information from literature (Matthew O. Ward et al., 2010). Second, edges crossing in the visualization of graph could result in bad aesthetics. Therefore, it is always a good idea to minimize or get rid of edge crossing (Matthew O. Ward et al., 2010).

2.2.2 Matrix Representation for Graphs

The other representation of graph is using adjacency matrix, which is an *N* by *N* grid, where *N* is the number of nodes and the position*(i, j)* corresponds to the existence (or not) of links between nodes *i* and *j*. This method is very important because it is used to alleviate the problem of edge crossing that is observed in node-link diagrams (Matthew O. Ward et al., 2010).



	a	b	c	d
a	0	0	0	1
b	0	0	1	0
c	0	1	0	0
d	1	0	0	0

Figure 2:4 Matrix representation of graphs.

Graph visualization plays important role in depicting very large relational dataset visually so that people can easily get the general picture of the dataset in question. In this regards, it is worth mentioning the importance of graph drawing concepts related to aesthetic values (Di Battista et al., 1999).

2.3 Aesthetical Criteria

The main purpose of visualization is to show information that can satisfy the audience. In this regard, the graph layout should be easy to read, understand, and remember. Moreover, it should have aesthetics. It is logical to think as the complexity of the information that we are dealing increases, easy and clear visualization of a graph could be hampered. In order to tackle such shortcomings, researchers and developer start to give due attention to aesthetics of graphs. The common aesthetics are (Di Battista et al., 1999):

Edge Crossings: reducing the number of crossing between the edges is very important to increase the aesthetic value of a graph. Ideally, planar drawing is what we want to achieve. A graph with a lot of edge crossing could make a graph clatter and congested.

Area: reducing the area of a graph in question could be used to achieve its aesthetics. Having very large size could make the graph hard to follow.

Total Edge Length: reducing the sum of the length of the edges could help to maximize the graph aesthetics. A graph with long edges could be difficult to get insight easily.

Maximum Edge Length: reducing the maximum length of an edge is could help to insure aesthetics of a graph.

Uniform Edge Length: having a reduced variance of the lengths of the edges could be helpful to make sure the graph has aesthetics. Uniform edge length could be followed easily than graph with large variations among the edges.

Total Bends: reducing bends along the edges could maintain its aesthetics. A graph with quite many edge bends is hard to follow and to get insight.

Maximum Bends: reduction of the maximum number of bends on an edge could increase the aesthetic value of a graph.

Uniform Bends: the variance of number of bends in an edge should be reduced to ensure its aesthetic. Reducing the number of bends in the edges of a graph could make the graph easy to follow.

Symmetry: increasing the symmetry of a graph could maximize its aesthetics. Dividing symmetrical objects evenly could be perceptually appealing. When we have unconnected symmetrical elements, we tend to connected them and have a coherent shape out of it (Soegaard, 2005).

2.4 Graph Drawing Algorithms

In the late 1960's hand drawing circuit design became too complicated because of the increased number of elements in the circuit. This prompted the need of algorithms to assist the process of circuit design. The algorithm overview can be found in the book (T. Lengauer, 1990). W.T. Tutte (1917-2002) was the inventor of the first graph drawing algorithms (William T. Tutte, 1963).
Graph drawing algorithms play vital role for visualization of network or graphs. Most graph drawing algorithms meet common criteria that make graphs/network look attractive visually and efficiently display the information intended to show. Spring-embedder algorithm is most widely known algorithm. The idea behind this algorithm is, each node in the graph have forces. The edge that exists between two nodes can have attraction force and the nodes between the end points of the edges can exert repulsion force (Di Battista et al., 1999). The force that exists within this works like a spring, that is the edge between the two nodes can attract the nodes depending on the position and the nodes can repel to each other depending the distance they have until they achieve their state of equilibrium. As we compress the spring it tries to force us apart and as we pull the two ends of the spring it tries to drag us in. We can see, to the right of Figure 2:5 that nodes are in a stable state and no force acting on the other node and becomes in the state of equilibrium. There is a tendency for a graph to be aesthetically pleasing using this algorithms. The layout is good so that people can easily get the information without the nodes being congested. But this comes with problem of scalability. The forced-directed approach does not work perfect for large graphs (Stephen G. Kobourov, 2012).

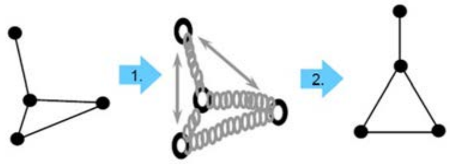


Figure 2:5 Nodes to the right are in a state of equilibrium (graph drawing course given by P. Eades and S. Hong).

Detailed discussion and annotation bibliography of a number of graph drawing algorithms could be referred in the literature (DiBattista, 1998; Di Battista et al., 1994).

There are many algorithms for tree drawing, such as space-filling radial drawing algorithm and layered drawing algorithm. In the first algorithm the appropriate root is chosen and from that root each subtree is drawn inside a wedge (pie shaped structure) by keeping the angle of it to be proportional with the number of leaves in that subtrees. In the latter, the graph is plotted with set of vertices having their own specific layer, with the dominat vertices hold the top and further move down the layer progressively. Such type of arrangement can be used to easily represent flow charts etc. (wolfram research, 2008).

There are also algorithms for Hierarchical graphs. Hierarchical graphs are directed graphs where nodes are placed into layers. Hierarchical graphs can be witnessed in many applications such as software visualization, CASE (Computer-Aided Software Engineering) tools etc. The idea behind this algorithm is to put vertices on the same horizontal line, and the edges pointing to the Y- direction (Seok-Hee Hong and Hiroshi Nagamochi, 2006). All the above mentioned algorithms could achieve their target if the final outputs that are used for the purpose of visualizing a given dataset is pleased audience. In oder to display a given dataset visually for people understanding of human perception is key.

2.5 Perception in Visualization and Gestalt Laws

Perception is very important for visualization. In order to have effective visualization of data we have to consider the capability of human perception and power of recognition. During visualization process we need to have a pattern at the end that can give us insight and the first thing happens during the interpretation of visuals pattern is its perception (Dastani M, 1998). According to (Ware, 2000) we have three stage model of perception. The first stage is parallel processing, i.e. information has to be processed parallel in order to get insight about the environment, the second stage is pattern perception to visualize our scene in different perspective such as colour, texture, and motion pattern and the final stage objects in our memory by retaining our visual thinking. Perception and cognitive power of human is important for the understanding and analyzing of data visualization.

Human perception has a significant role to play in the area of information visualization. In order to harness the benefit of information visualization and provide best visual figure to the audience, thorough understanding of visual perception in human is very important to detect meaningful patterns from data represented visually (Christopher G. Healey et al., 2012).

Pattern perception for the first time got attention by the group of German psychologists in 1912 when they founded Gestalt School of Psychology. The contribution of these German psychologists has still solid ground. This is because many of the perceptual phenomena were described clearly. Gestalt laws of pattern perception was produced by them. The laws describe how we see patterns in visual display. Proximity, similarity, connectivity, continuity, symmetry, closure, relative size, and common fate are eight Gestalt laws (Ware, 2000). These laws are very important still and their contribution is immense. It is wise to consider these laws while we are trying to generate visual picture of a dataset for the audience and during graph drawing as well.

2.6 State-of-the-Art in Graph Visualization

Visual representation is becoming very important and key in the world where numerous amounts of data being produced in a second. This makes information visualization to be an important research topic. In the ancient time image was used to represent our thoughts. But using information visualization to understand the pattern and structure in a given data is a more recent phenomenon. Currently information visualization is being used in many different disciplines, this includes: capturing the changing dynamics of different system, mapping timelines, science, literature, and blogosphere (Staša Milojević et al., 2012).

In science, information visualization used widely. In the Figure 2:6 below, we can see Gene-regulatory and signal-transduction networks. A grid layout of the yeast cell cycle regulatory network visualized by CADLIVE. The rounded rectangles are proteins, the rectangles metabolites, the parallelograms mRNAs and the ovals represent events (Weijiang Li and Hiroyuki Kurata, 2005). Here the biological network use directed edges to show the flow of information.

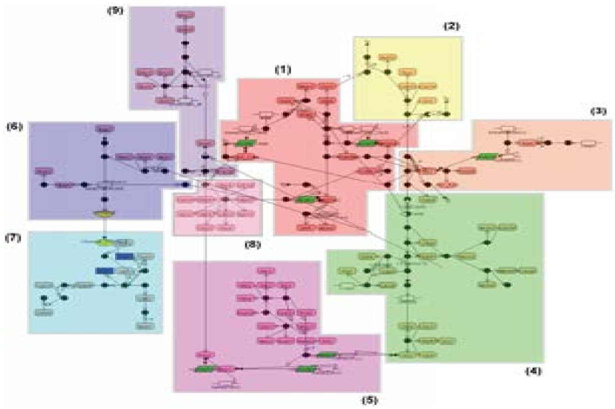


Figure 2:6 A regular network representing the yeast life cycle.

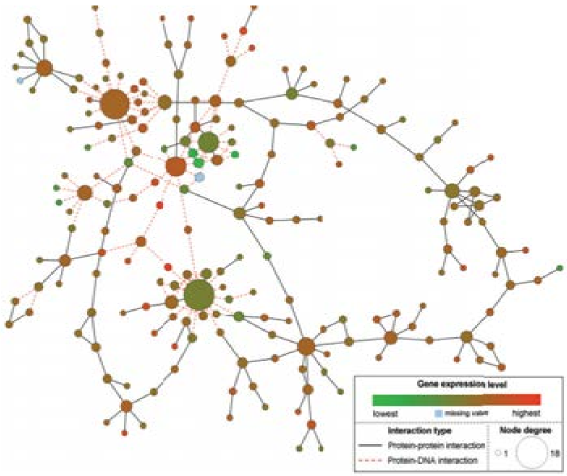


Figure 2:7 Visual representation of the GAL4 protein interaction subnetwork in yeast.

The Figure 2:7 shows the visual representation of the GAL4 protein interaction subnetwork in yeast (M. Albrecht et al., 2009). This network is used to represent protein interactions or between other genetical components such as DNA or RNA. The nodes in the network represent proteins or set of proteins. As we can see the edges in the network are normally undirected, but could be directed to show interactions such as activation. This leads to have mixed graphs. As you can see above the node-link representation of graphs has been used to visualize this complex biological data for the audience.

Another area where we can apply the benefit of graph visualization is on social networks. With the advancement of technology and internet currently there are many social network sites. Visualizing these social networks is very important. Social network visualization is simply a geometric representation of the abstract information of the social network (Ing-Xiang Chen and C. -Z. Y., 2010). People who are in the managing posts may need abstracted information how the social network is being used and how its development through time is and see improvements in organizational performance or to have a policy intervention for any change in the organization.

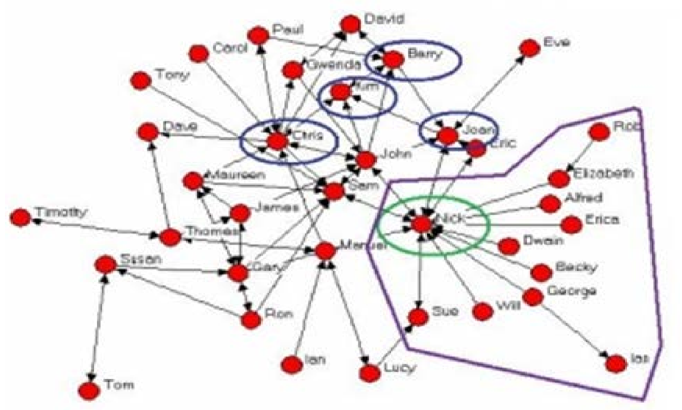


Figure 2:8 With whom do you discuss issues important to your work? (Paola Tubaro, 2012).

From the visualization of people's interaction in the Figure 2:8, we can easily see that Nick is the busiest guy talking to important issues related to the work in the company. This could easily give clear information for the policy makers or managers to do something before they make any decision.

Since late 1990's weblogs, commonly known as blogs are becoming cyber culture for today's society. The blogs are reaching 150 million currently. The interconnection of these millions of blogs, result in Blogosphere. Currently there are good user interaction mechanisms that enhance the exploitation of information from the blogosphere using BlogConnect (Justus Broß et al., 2011).

2.7 Evaluation of Information Visualization Techniques

Information visualization is involved in many fields to help users gain fast and clear understanding of sets of large data. Researcher's interest towards the advancing of information visualization technology was observed throughout 1990s (Ware, 2000). These advanced visualization technique has been used by the society at large. These capture the interest of researchers and other stakeholders in the field (Chen and Czerwinsk, 2000). Currently we are witnessing more and more techniques of information visualization in the academic and business world. Identifying which visualization technique is better than the other is crucial in certain cases. To realize such situation practically we need to have evaluation methods. Today, we have lots of information visualization evaluation methods.

2.7.1 Evaluation Methods

For the purpose of any research to be successful, choosing the right type of evaluation method is important. In fact, choosing the best research method for our situation is one the most important part of empirical research. All methodologies have their own advantage and disadvantage. But, they have some commonalities; they all start with question that should be investigated in relation with the current idea, theories and findings. The result that we get may consolidate the existing concept or may contradict totally and raise questions on the existing ideas (Carpendale S., 2008).

Figure 2:9 (McGrath, 1995) is adapted and used by (Carpendale S., 2008) shows eight methodologies for evaluation.

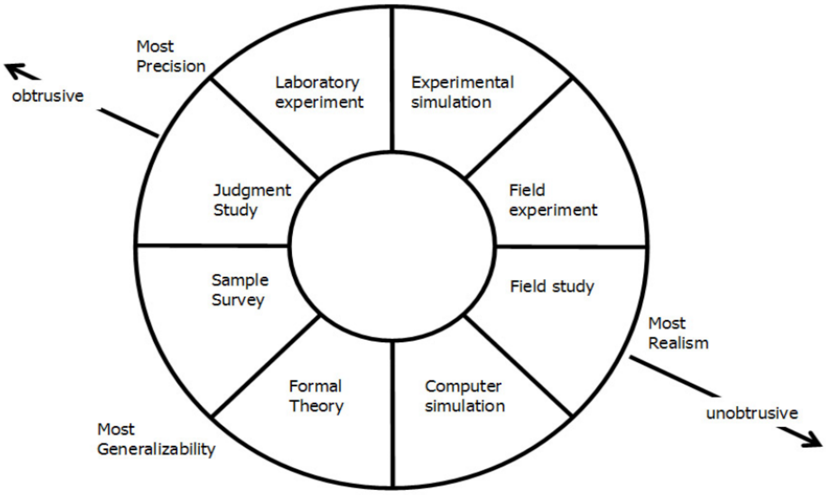


Figure 2:9 Types of methodologies in relation to precision, generalizability and realism (McGrath, 1995).

According to (Carpendale S. 2008) we have two major evaluation methods. These are Qualitative Evaluation and Quantitative evaluation methods. Qualitative evaluation includes interviews, field studies, observation. This evaluation method has no statistical data to manipulate. The Quantitative evaluation includes laboratory experiment survey or questionnaire and has statically manipulation tasks (Carpendale S. 2008). Our research is done using quantitative evaluation method by manipulating data collected from the distributed questionnaires for students at Linnaeus University.

2.7.2 Challenges of Evaluation of Information Visualization

The availability of different information visualization techniques in different discipline has raised the issue of having appropriate evaluation of these techniques (Carpendale S. 2008). But these evaluation techniques are not without challenges. In empirical research, it is difficult to target the right focus and ask questions specific to that focus area, for interesting question that we have it is difficult to have the right methodology to deal with it and get as much data collection as we can. The above bottleneck is not only for information visualization research but many other fields are also facing the same problems (Carpendale S. 2008). The paper singled out three evaluation challenges with respect to human computer interaction (HCI) empirical research, perceptual psychology empirical research and cognitive reasoning empirical research (Carpendale S. 2008).

It is usual that interface interaction tasks such as zooming, filtering and accessing data details tasks are prominent task of interest in HCI empirical research (Shneiderman, 1996). These interactive tasks are very important to get detailed information about a dataset and it is strongly related with the functionality of the system (Carpendale S. 2008). Getting the appropriate size of participant when we are doing HCI empirical research is another challenge especially if the participants are domain specific experts then it is difficult to get their time. The other is if our HCI empirical research is about the complex visualization software we are not sure that if the result is due to specific functionality that we singled out or the overall system solution. Moreover, if we want to compare the interactive visualization technique of new information visualization software using the existing software as a benchmark, then the user may be familiar with the old system and this could biase the evaluation result (Carpendale S. 2008). Human perception is very important to perceive a visualization of a given dataset. It used to assess the readability of visuals features. According to the data type and character we have, there are different cognitive reasoning task in information visualization. Not all of these tasks are defined clearly especially those that leads to new insight into the data are challenging to test them empirically (Ware, 2000).

2.7.3 Evaluation in Network Visualization

Graph comparison is widely known problem in computer science and there are different algorithms solution to compare the similarity and difference between graphs (Alan G. Melville et al., 2011). According to the paper mentioned above, an experiment has been made to compare two visualizations of multiple graphs using matrix format. These are Juxtaposition and Superposition in matrix visualization and super bowl game data was used during the experiment process. The questionnaire was automated in such a way that teams are represented by vertices and edges used to represent the winner and loser, the edge directed out from the winner to the loser. Then different questions were presented that has to be answered in the system. The task based result was analyzed using statistical tools and presented according to the mean and standard deviation. The result indicates that superimposed visualization has better accuracy than the Juxtaposition. Tester of the system have found it difficult to associate where the difference are related in Juxtaposition than superimposed matrices which were more correct (Alan G. Melville et al., 2011).

Another recent evaluation of network visualization is done by domain specific versus generic on metabolic network. According to the paper (Romain Bourqui, 2011) before this particular research there has been no efficiency evaluation studies done. The idea was to identify if there is efficiency difference between the automatic graph layout algorithms and any algorithms that specifically considers the currently practiced convention among biologists. For the experiment three layouts have been used. Two of the layouts were generated using the generic graph layout algorithms, specifically forced directed and Hierarchical layout algorithms, and the third was produced using MetaViz (Romain Bourqui et al., 2007), which produced the layout according to the convention of biologists. Evaluation software was implemented to collect data from the questionnaire. All the three graphs which are produced using Forced-directed algorithm, Hierarchical and MetaVis were scrutinized. The task was to identify motif (is unordered set of reactions were each reactant or product of it shared with at least one other motif). The result indicates that the hierarchical layout produces worst time performance than both the MetaVis and GEM (Frick A. et al., 1994) (using forced graph algorithm), and no statistical difference observed between GEM layout and MeatVis (Romain Bourqui, 2011).

There was also an experiment which is performed on the graph drawing algorithms. The idea of the experiment is to investigate if aesthetics of graphs have effect on the understanding of graph drawing. Three aesthetic values were selected in the experiment. These are: minimizing arc crossing, minimize arc bends, maximize symmetries. The experiment was done on the undirected graphs. The three hypotheses during the commence of the experiment were: minimizing arc crossing increase aesthetics, minimizing arc bends increase the aesthetic of a graph and maximizing symmetry increase aesthetics. Nine drawings of graphs were used each with sparse and dense with the three aesthetic features. During the experiment controlled timing were used. The result indicates that indeed minimizing arc crossing and minimizing arc bends enhance the aesthetics of a graph whereas the maximizing symmetry has no effect on the aesthetics of a graph (Helen C. Purchase et al., 1995).

3. Network Comparison

Network visualization is one of the techniques that enable people to show complex data in simple and recognizable pattern. Comparing different network visualization is currently important in many fields of science. Understanding network comparison thoroughly is vital to bring the best network visualization on the table that can be used for comparison of complex objects. In this regards, we have visualization categories for network comparison that we need to study thoroughly to identify which comparison technique could be favored among the other and to suggest a further input for the future research and development in the field. The three fundamental taxonomies of object comparisons approaches are: Juxtaposition, Superposition and Explicit Encoding (Gleicher et al., 2011). These three fundamental of comparison approaches can be used to produce hybrids that are used for visualization as well, as we shall see in the next sections.

3.1 Introduction to Network Comparison

Information visualization has played significant role in the past decades for understanding complex information. In recent decades the role of computers was significant for better utilization of information visualization. The emerging of new technology, computers, has also brought a challenge for the information visualization. Large data that we are visualizing for the people using computers could not guarantee passing the message equal to audiences with different level of education, background, profession, age etc. (Mikko Berg, 2012).

Currently because of the technological advancement in the information technology there are a lot of data being produced as well as processed. These data are becoming complex in time. Information visualization has to deal with complex data. In order to get the information from complex data, visual comparison is very important (Gleicher et al., 2011). Comparing individual objects was the past trend of information visualization but currently because of the complexity of data and the advancement of science, comparison of complex objects is becoming a key to get information easily from this complex data. The comparison applies to the sequences or graphs not for the individual objects. Biology could be one example, we compare the different phylogenic trees of animals or plants or we may compare the genetic make-up of DNA or RNA and find out if they have something in common or not (Gleicher et al., 2011). Understanding the comparison task is very important to alleviate some of the potential challenges that are faced by different tools used for comparison. The comparison of objects has its own challenge. When we are comparing objects the viewer needs a connection within and between objects. Scalability is another challenge during the process of comparison, which depends on the level of complexity of object that we are dealing with and the number of objects (Gleicher et al., 2011).

There are different systems in the past that used to compare single objects. The Figure 3:1 below shows a screenshot of a reconciled phylogenetic tree displayed with ETE (a python Environment for Tree Exploration (Huerta-Cepas J et al., 2010)).

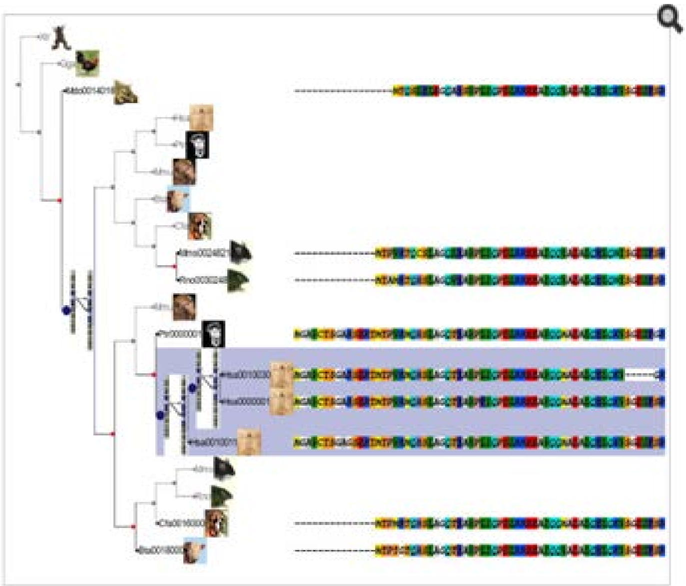


Figure 3:1 Screenshot of a reconciled phylogenetic tree displayed with ETE (Huerta-Cepas J et al., 2010).

The reconciled phylogeny tree, Figure 3:1, is generated using ETE's strict reconciliation algorithm to show a portion of a reconciled tree. As you can see different aspects of a tree has been highlighted using a custom ETE layout function such as inferred gene losses represented by grey dashed lines, duplication events by blue nodes, speciation events by red nodes (Huerta-Cepas J et al., 2010).

Though visualization of complex objects is difficult, usually visualization systems have their way out. The first is, abstracting the object so that we can get rid of unnecessary details and make the objects simpler (Amenta N, Klingner J, 2002). The other alternative is not to use any explicit comparison of objects in the visualization system, so that the user uses his/her own memory for comparison or use separate comparison in space (Gleicher et al., 2011).

The paper (Gleicher et al., 2011) proposes the designs used to compare objects into three general categories. The design category depends on the way how each comparison approach shows the object that they are dealing with. These categories are: Juxtaposition, Superposition and Explicit Encoding. The above three categories are the building blocks. This is because all other categories are derived from them. These three categories can be combined to form hybrids. Generally there is a possibility of seven categories. These are: Juxtaposition, Superposition, Explicit Encoding, Juxtaposition + Superposition, Juxtaposition + Explicit Encoding, Superposition + Explicit Encoding and the combination all three.

According to the paper (Gleicher et al., 2011) the above taxonomies are useful and most comparative visualization can be grouped into one of the categories described above. This could be because of a number of reasons. First, categorizing the space of designs enable related methods to be grouped by design, so that we can see the pros and cons of each form. This can make similar issues and solutions can be transferable between related designs even if the data that we are dealing with is from different domains. It also bridges the gap between the different categories of comparison and resource used for comparison such as cognition and perceptual ability of human beings.

Figure 3:2 shows how the three primary categories namely: Juxtaposition, Superposition and Explicit Encoding and their hybrids used for the purpose of comparison of complex object in the database system that collects different information visualization literature during their study (Gleicher et al., 2011). The literatures which are collected during their research are show in the Figure 3:2. You can see most of the literatures are using the three fundamentals of network comparison approaches namely: Juxtaposition, Superposition and Explicit Representation for the purpose of comparison. In addition, you can see that there are literatures which use the hybrids approach namely Explicit Encoding + Juxtaposition, Explicit Encoding + Superposition and Juxtaposition + Superposition.

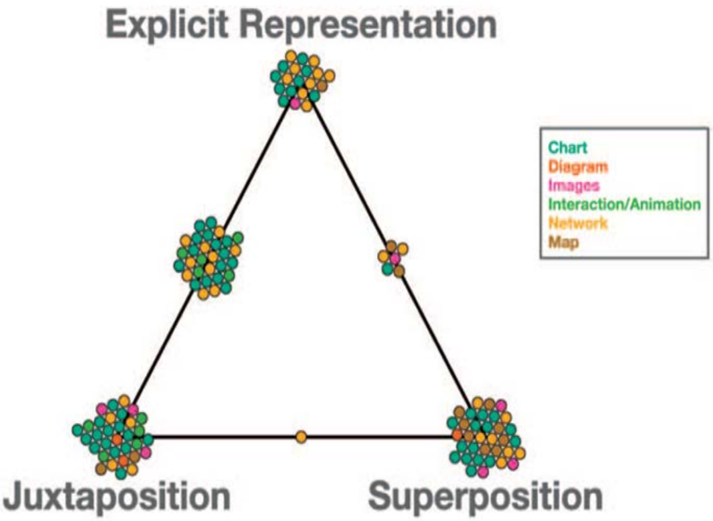


Figure 3:2 shows how primary categories (Juxtaposition, Superposition and Explicit Encoding) and their hybrid used in the literature collected for research purpose of a paper, Visual comparison for information visualization, (Gleicher et al., 2011).

At this point it is vital for us to consider those network comparison approaches that have been used in the past for the purpose of visualization in the different disciplines out there.

3.2 Juxtaposition

The idea behind Juxtaposition is to show individual objects independently or separately (Gleicher et al., 2011). This type of design is usually occurs in space, that is putting individual objects next to each other. In this category the comparison is totally dependent on the memory of a person who is engaged in comparison. Animation film is best example of showing how the comparison is effective and user memory of the present and past is very important to get the whole picture (Gleicher et al., 2011). This comparison is effective when the difference between the object is significant. The Figure 3:3 (Jin Chen et al., 2009) below shows screenshot of interactive dendogram matrix view, as we can see the overview is on the top and the detail view is on the bottom. The overview on the top is represented by the leaf node and detail view in the bottom is by sub-trees. If we select a specific leaf node on the overview, then we can see it on the detail view as a sub-tree and the information is exactly the same between the two as you can see in the Figure 3:4, that is between the leaf node on the overview and sub-tree on the detail view (Jin Chen et al., 2009).

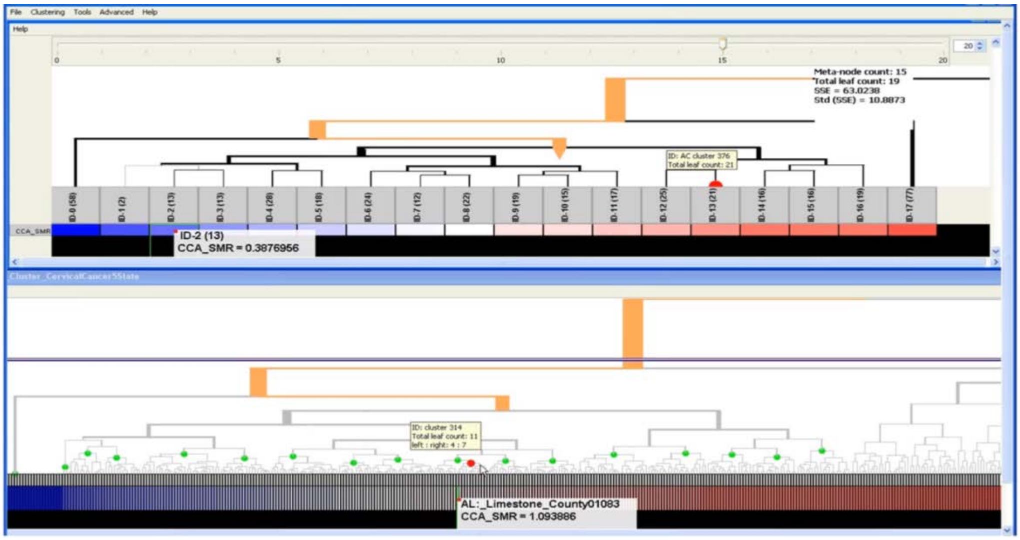


Figure 3:3 Dendogram matrix view (Jin Chen et al., 2009).

We can explore by mouse hover on the sub-trees as they are side by side. In the picture all the detail view is not shown because of the screen size problem. As you can clearly see those data related to the meta-node selected on the top are juxtaposed side by side which is highlighted with green circle on the bottom in the detail view and the user can get detail information by comparing each of them.

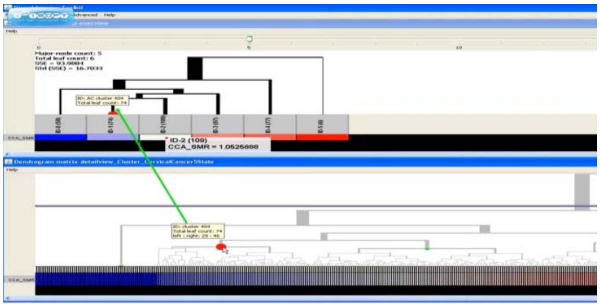


Figure 3:4 Toot tip showing the information that matches between the leaf node of the overview and sub-tree of the corresponding sub-tree (Jin Chen et al., 2009).

Letting the audience see the relation between the object is a challenge in this comparison design (Gleicher et al., 2011). But, by putting the objects close to each other and with a space comfortable to make comparison, we can let people see the difference easily. However, if the difference between the objects is not big, it puts some burden on the user to identify the difference (Shen et al., 1998). As we can see from the Figure 3:5 (Tan D. et al., 2007), the AdaptiveTree there are 20 games have played. The first round of the game is represented by a short edge coming out of every team. The winner of the two teams given bold line (dark) and the loser takes grey colour. Games which are yet to be held are given broken lines. From this diagram we can clearly see the winner and the loser between the two categories.

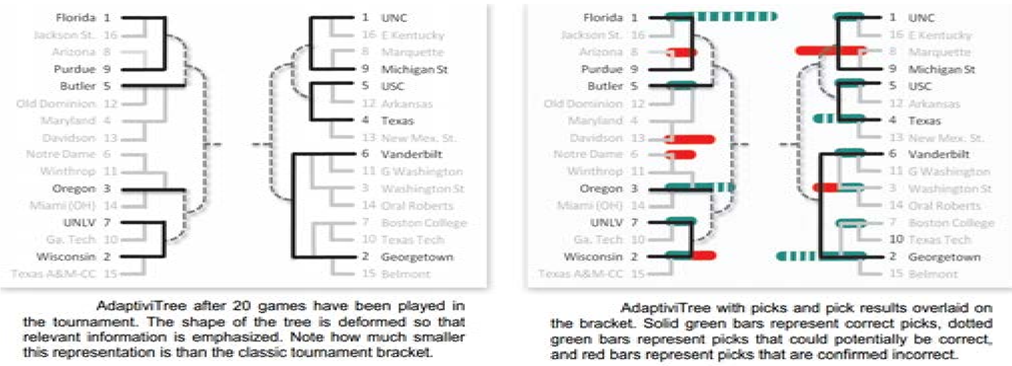


Figure 3:5 AdaptiveTree to show tournament features (Tan D. et al., 2007).

Figure 3:6 (Freire et al., 2010) below shows the interface of ManyNets which is used to visualize social networks and compare them in the rows of a table. The table is used to compare networks using rows and the column of each row used to hold their statistics. We have the overview of each column on the top row and the detail on the lower rows of each specific column. As you can see in some columns such as Vertex count is displayed using horizontal histogram and that make a user easily compare and get insight about each network across the rows. As you can witness from the Figure 3:6 visualization of data is juxtaposed and user simply needs to compare and if there is interesting row they need to pay attention then the user can examine it in detail (Freire et al., 2010).

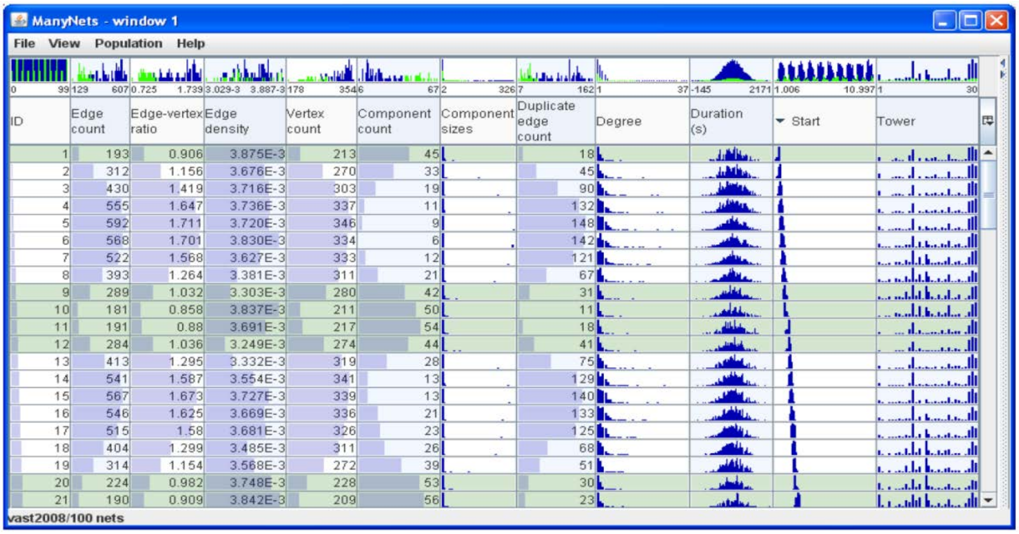


Figure 3:6 ManyNets, rows represent networks of call in 5-hours period, with 50% overlap with pervious row, therefore 10 rows cove the whole day (Freire et al., 2010).

By mouse hovering in a cell we can get small description about a value on tool tip, but if we want more detail information we can right click on a specific cell or column which enable us to get pop-up window with detail information about that specific cell or column as you can in the figure 3:7 (Freire et al., 2010) and compare it with other attributes of the network in question.

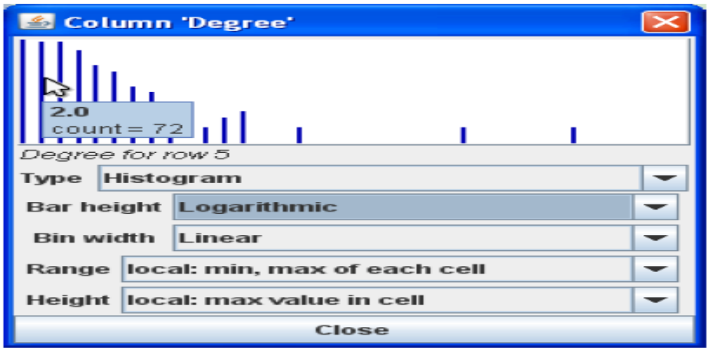


Figure 3:7 Pop-up window that you can get by right click on specific cell to get detail information (Freire et al., 2010).

3.2.1 Juxtaposition using Node-Link Diagram

Figure 3:8 below is the node-link representation of a simple network (Gleicher et al., 2011).

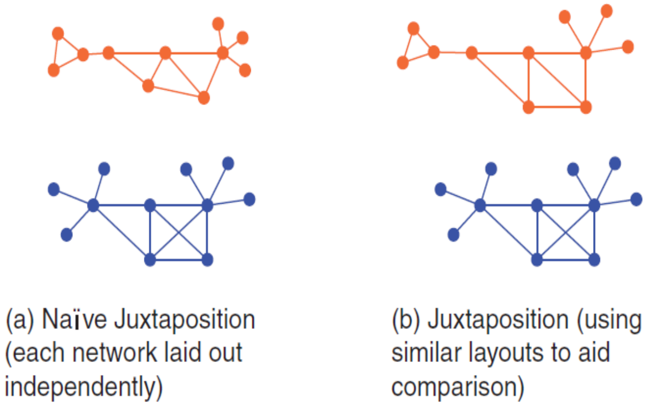


Figure 3:8 A simple node-link graph representation.

Figure 3:8 shows the different way of comparison using Juxtaposition. In figure (a), we can see that the comparison is done using different layout between the two networks and in part (b) you can see that the two networks are compared using similar layout.

3.3 Superposition

This type of comparison is used to show objects one on top of the other and presenting them in the same time and space or place. The overlays of objects can be done by making one object semi-transparent or allowing one object oppress the other visually. This taxonomy of comparison is used to show several objects in the same space. But this has a drawback especially if the objects are dense, for example image. Blending or making the object semi-transparent may be the solution, but this may result in clutter. Superposition usually used for objects which are similar and can be put the same space so that the user can easily see the similarity and difference (Gleicher et al., 2011).
Table 3:1 (Tim D. et al., 2006) below shows the centrality ranks for M. musculus PPI (Protein-Protein-Interaction) using eccentricity, closeness, eigenvector, degree and rwbetweenness centrality measures (Tim D. et al., 2006).

	$\mathcal{C}_e$ eccentricity	$\mathcal{C}_c$ closeness	$\mathcal{C}_\lambda$ eigen-vector	$\mathcal{C}_d$ degree	$\mathcal{C}_r$ rwbet-weenness
$\mathcal{C}_e$	1.00000	0.85319	0.52707	0.35398	0.36017
$\mathcal{C}_c$	0.85319	1.00000	0.58717	0.42786	0.46054
$\mathcal{C}_\lambda$	0.52707	0.58717	1.00000	0.44396	0.43527
$\mathcal{C}_d$	0.35398	0.42786	0.44396	1.00000	0.74693
$\mathcal{C}_r$	0.36017	0.46054	0.43527	0.74693	1.00000

Table 3:1 M. musculus PPI-network rank of correlation coefficient (Tim D., et al., 2006).

We can see in the Figure 3:9 (Tim D. et al., 2006), the comparison of vertices which are in the PPI-network using orbit-based comparison with their centrality measures of the above five methods. As you can see all the vertices of the graph *Gi* lie on the orbits which depends on the centrality value of the vertex, that is each vertex *v* gets value of an orbital constrain with radius *r = f(C(u)) where the function f derived from the centrality(C) value of each vertex u.* Interested readers can get full information from the source (Tim D. et al., 2006). But the idea we are trying to emphasis is on the left bottom part of the visualization, two dimensional representation, you can see that the whole visualization is superimposed so that we can clearly get the insight about the centrality of the each vertex in the orbits.

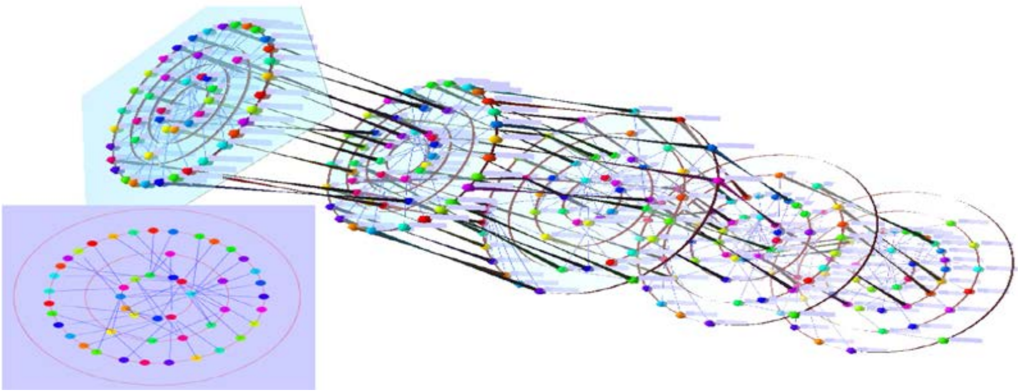


Figure 3:9 The orbit-based comparison method of Protein-Protein Interaction (PPI) network (Tim D., et al., 2006).

Below you can see the visualization of a social network visual analytic tool called Ontovis (Zeqian Shen et al., 2006). Large heterogeneous social network is analyzed by using information visualization technique and supporting the analysis using an auxiliary graph called ontology (Zeqian Shen et al., 2006). Figure 3:10 shows visualization of a movie which is orange nodes and the people in blue colours. As you can see the central biggest blue nodes corresponds to Woody Allen, a film actor. And three medium sizes correspond to Maria “Mia” Farrow, Louise Lasser and Diane Keaton who are worked most often with him and who have relation with Woody Allen, film actor (Zeqian Shen et al., 2006). The two dimensional view shows us that people who are participating in a given film (orange node) with Woody Allen, that is to mean superimposed view give as a clear view about people who participated in a particular film with the famous actor, as you can see in the Figure 3:10.

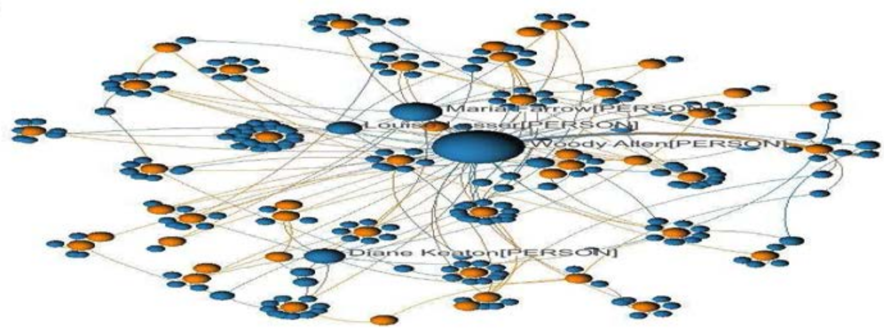


Figure 3:10 Visualization of Movies in orange and people in blue related to Woody Allen (Zeqian Shen et al., 2006).

3.3.1 Superposition using Node-Link Diagram

As we can see in the Figure 3:11 the pink and blue node-link diagram are shown. They are superimposed one on the top of the other.

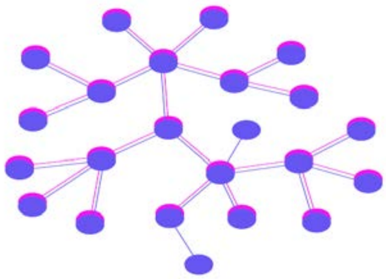


Figure 3:11 A simple superimposed node-link diagram.

3.4 Explicit Encoding

This category of visualization for comparison depicts the relationship between objects by encoding visually (Gleicher et al., 2011). This type of visualization tries to show explicitly the relationships between objects so that users do not need to stress themselves to get the relation between objects they are visualizing. But, we have to make sure that, we have identified the proper relationship that exists between objects before we visually encode them (Gleicher et al., 2011). So, we can clearly see that there are two tasks to consider. First, the proper relationships should be correctly identified and second, we need to depict this identified relationship visually. These make the technique to come up with new objects which are completely different from the old ones. This certainly raises the issue of decontextualization, the user loses the connection between the old objects from the newly represent objects (Gleicher et al., 2011).

EMDialog visualization (Hinrichs et al., 2008) could be taken as an example of Explicit Encoding. The EMDialog was used to show the work of Emily Carr (December 13, 1871 – March 2, 1945), well known Canadian painter and writer. To the left of Figure 3:12 the temporal dimension is represented by tree cut section and to the right the context is represented by upright tree. The timeline to the left ranges from 1890-2010 where every data related to Carr's painting and works written by different authors represented by small circles within a specific tree cut. If a finger touches a particular circle it automatically displays the content so that the user can get the information and to the left it shows others data that contain that specific keyword using upright tree. We can clearly see the issue of decontextualization here. What is touched and depicted is totally decontextualized to the left where multiple unrelated data reviled.

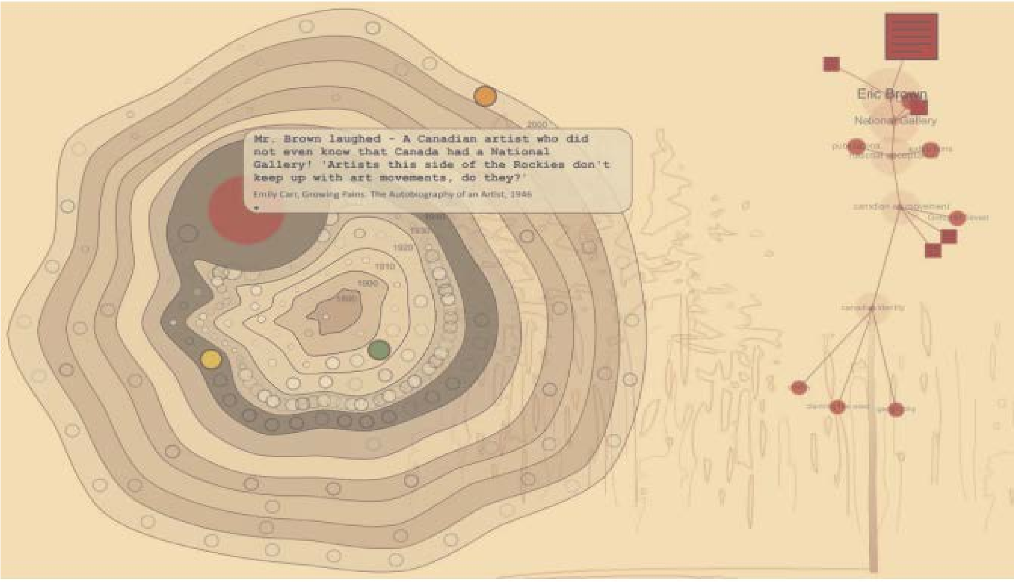


Figure 3:12 EMDialog a selected statement and the corresponding tree diagram (Hinrichs et al., 2008).

In the Figure 3:12, we can see that what is displayed to the right depends on the keyword touched on specific tree cut to the left. The displayed content may not be related to the users interest which result in decontextualiazation, which is something that we face when we are using pure Explicit Encoding. Because of the above reason, we rarely use pure Explicit Encoding (Gleicher et al., 2011).

If we see the DAG (Directed Acyclic Graph) representation in the Figure 3:13 (Martin Graham and Jessie Kennedy, 2007) below, we can see the Apiaceae taxonomic family using eight different hierarchical classifications. As you can see there are layers on the DAG separated by horizontal band, where the nodes are represented by small rectangles with colour that match the trees in which the node found. In the middle you can see Ammieae has been selected and we see five nodes are blue in colour indicating that this current classification is found in them where as three nodes have no blue colour indicating these nodes do not found in that classification. If we move up to the Family level we can see that all nodes are in the classification, but only five nodes are colored blue indicating they are in the classification but the three nodes are colored grey to show that they are in the classification but does not have relationships with the current selection (Martin Graham and Jessie Kennedy, 2007). In this diagram what we select and what you get lack context. If you select a given sub-family you can get lots of children which are in their Genus level for that particular Sub-family of organism which is difficult to trace back and get full insight about the displayed information, this what we referred in the previous chapter as decontexualization.

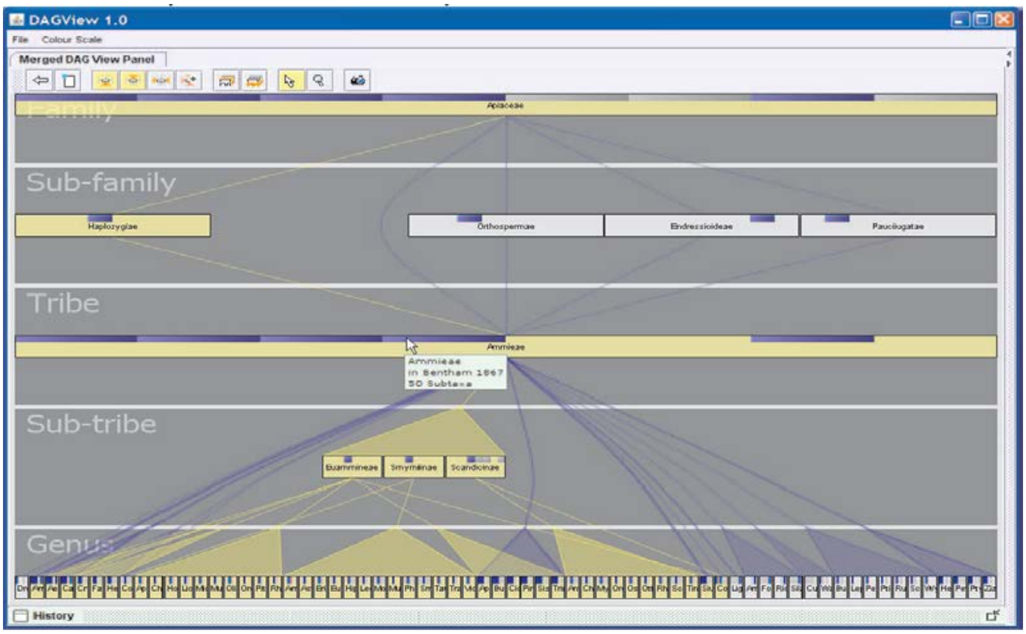


Figure 3:13 The screenshot of DAG visualization for multiple classification, in this instance centered on Ammieae (Martin Graham and Jessie Kennedy, 2007).

3.5 Hybrid Approach

To alleviate the problem of decontextualization of the Explicit Encoding, we need to use Explicit Encoding together with Superposition representation and /or Juxtaposition. In the Figure 3:14 we can see how Explicit Encoding can be used together with Juxtaposition.

Figure 3:14 shows two networks are being shown independently to the left and to the right. And the two graphs merged and shown in the middle, and colour coding are used to show the similarity and difference in the merged graph (Andrews K. et al., 2009).

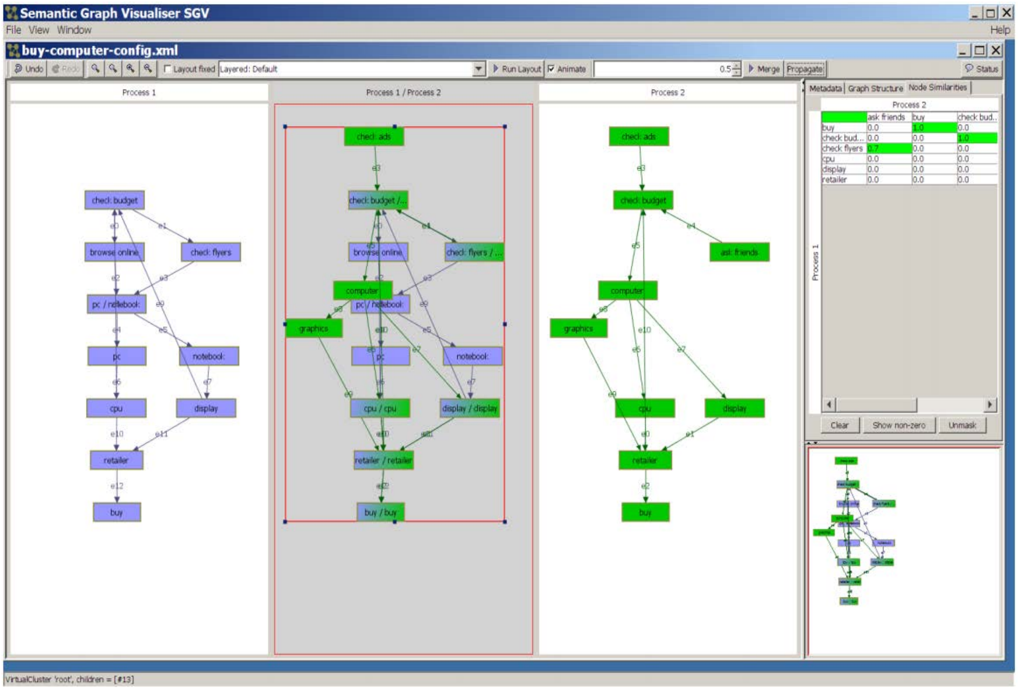


Figure 3:14 The Semantic Graph Visualizer (SGV) comparing two process graphs representing workflows involved in buying a computer (Andrews K. et al., 2009).

Superimposing of an encoded visualization could be used together with Explicit Encoding as we can see in the Figure 3:15 below. It shows screenshot of BGPlay. Here topographic map is used to visualize an internet hierarchy. AS (Autonomous System) within the same hierarchy level are confined using contour lines. The autonomous systems inside the center are the top levels. As we move down the hill the height decrease and also the hierarchy. This visualization is very important to detect those paths which are not using the higher hierarchy or the backbones, this is because those paths are efficient and less expensive (Pier Francesco Cortese et al., 2006).

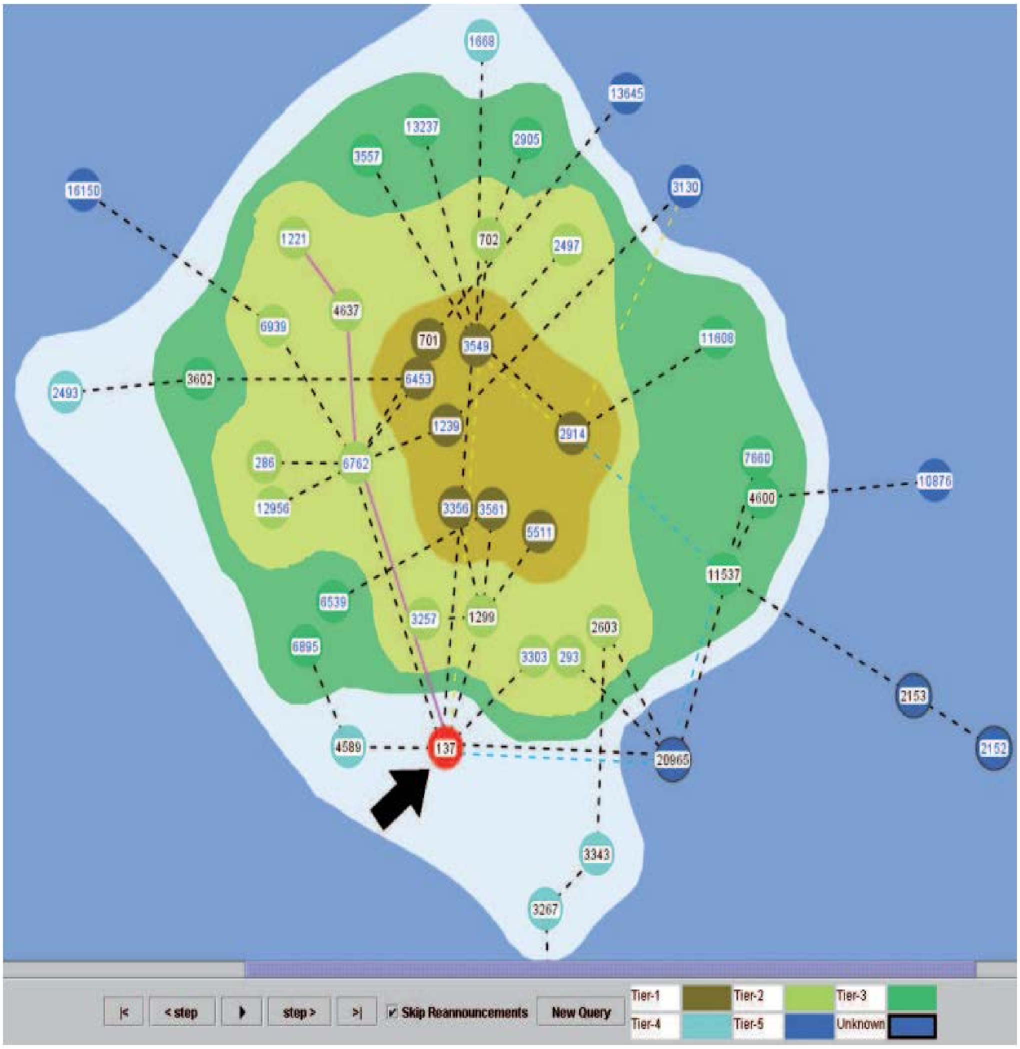


Figure 3:15 A screenshot of the BGplay system (Pier Francesco Cortese et al., 2006).

Figure 3:16 shows a hybrid design of Juxtaposition and Superposition (Gleicher et al., 2011). The design of these two contradicts because the former uses separate space whereas the latter uses same space. But still many information visualization systems use this view.

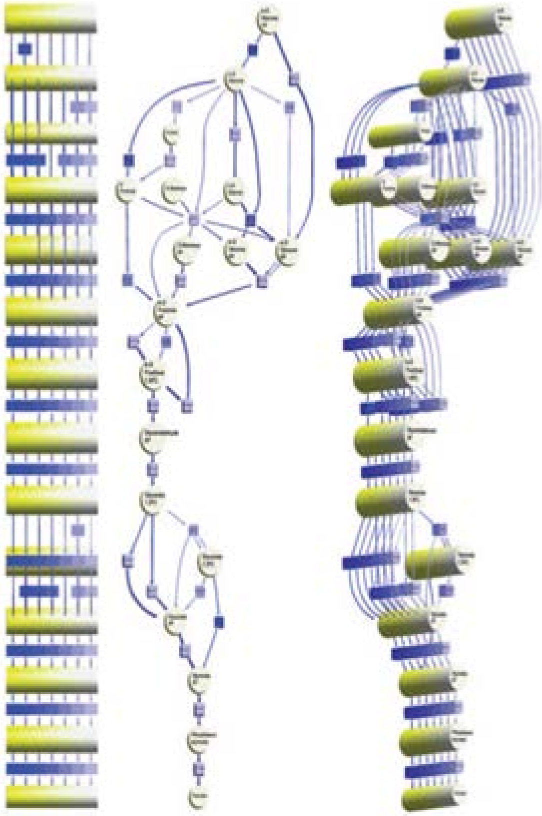


Figure 3:16 Protein network comparison simultaneously using Juxtaposition and Superposition (Brandes U. et al., 2003).

4 Evaluation of Network Comparison Approach

In this thesis three main network comparison approaches were discussed in detail. These are Juxtaposition, in which comparison is made between two objects separated in space and time; Superposition, in which comparison is made between two objects by overlapping one on the top of the other; and Explicit Encoding. The first two network comparison approaches are considered in this analysis. The aim of this thesis is to identify explicitly which network comparison approach namely, Juxtaposition or Superposition is better for the purpose of comparison of a give dataset by analyzing data collected from the questionnaires, which are collected from the students at Linnaeus University.

The comparison of data objects using Juxtaposition has to depend on the memory of the viewers or audience. They have to have attention and compare the two objects and look for the difference to get the insight, whereas the comparison of data objects using Superposition do not need much attention and rely on the visual system of the users. All the objects are in a single place and can be easy to compare objects. Therefore, the experiment hypothesis for our experiment was*: Superposition comparison approach is better for comparison than Juxtaposition comparison approach*.

4.1 Questionnaire

The questionnaire was designed to collect some data from participants as well as collect their answers to the questions presented to them. They were asked to fill their personal information: their Age Group, Gender, and Department, Major Study, Education Level and a question that asks participants if they have done network analysis before. Then, each participant performed twelve tasks of networks compared using the Juxtaposition and Superposition comparison approaches. The graphs in the questionnaire were presented in such a way that to consider the following:

- Participant need to identify node(s) which is/are present in one graph and not present in the other
- Participant need to identify edge(s) which is/are present in one graph and not in the other
- Participant need to identify node(s) and edge(s) which is/are present in one graph and not present in the other.

You can get full information about the questionnaire in Appendix A.

4.2 Method

In the questionnaire twelve questions were asked with some order to make the participant face the comparisons approaches alternately after having one approach two times. In the Table 4:1 S is stands for sparse graph and D stands for dense graph. The table shows how the order of the questions were in the questionnaires.

S1-Super	D1-Juxt	S2-Juxt	D2-Super	S3-Super	D3-Juxt	S1-Juxt	D1-Super	S2-Super	D2-Juxt	S3-Juxt	D3-Super

Table 4:1 Shows order of questions in the questionnaire.

As you can see in the Table 4:1, we deliberately change or alternate the Superposition, Juxtaposition and sparse and dense graphs as well. The idea behind this is to make the user not used to a single approach so that he/she could answer the question without past experience. The other important point to make here is between the two consecutive comparison approaches we have changed the number of nodes and edges. We have dense graphs with edges 32 in average and sparse graphs with 24 edges in average. We have a maximum of 26 nodes and 36 edges in the dense graphs whereas among the sparse graphs we have a maximum of 23 nodes and 27 edges. We slightly differ the number of nodes and edges just to make the user faces something different from the previous questions and we believe that the person who is participating in the questionnaire could not use the past memory at least in the beginning. We labeled those pair of graphs which are relatively dense with prefix D followed by a number 1,2 or 3 to show that we have three different types of dense paired graphs and we explicitly put the comparison approach used separated by hyphen. The same thing applies for the paired graphs which are sparse. We used prefix S followed by a number 1, 2 or 3 to show that we have three different types of sparse paired graphs and a hyphen followed by the comparison approach used . For instance, when we write S1-Super it means a sparse pair of graph one compared using Superposition approach. Below the table shows the differences that we put in each pair of graph to be identified by the questionnaire participants.

Detail information about the questionnaire is shown in Appendix A.

Type of networks	Meaning	Total difference between them
S1-Super	Sparse pair of graphs one with superposition approach	2
D1-Juxt	Dense pair of graphs one with juxtaposition approach	3
S2-Juxt	Sparse pair of graphs two with juxtaposition approach	2
D2-Super	Dense pair of graphs two with superposition approach	2
S3-Super	Sparse pair of graphs three with superposition approach	1
D3-Juxt	Dense pair of graphs three with juxtaposition approach	2
S1-Juxt	Sparse pair of graphs one with juxtaposition approach	2
D1-Super	Dense pair of graphs one with superposition approach	3
S2-Super	Sparse pair of graphs two with superposition approach	2
D2-Juxt	Dense pair of graph two with juxtaposition approach	2
S3-Juxt	Sparse pair of graphs three with juxtaposition approach	1
D3-Super	Dense pair of graphs three with superposition approach	2

Table 4:2 shows the number of differences between two paired graphs.

In the data analysis only those differences that are detected were our interest. We take the average value of detected value. The weight (the number of detected difference divided by the total difference that the graph has) of each detected value was taken. This is because each pair of graph has their own variation in the total difference they have so their weigh has to be considered rather than taking simply the detected number. You can get the full data analyzed in Appendix C.

4.3 Procedure

Some procedures were followed while the participants engaged in the questionnaire. Each and every individual was told to use only 30 seconds (we believe that if the time allocated for each questionnaire is long, then there is a possibility that participants could identify all the differences and on the other hand if we allocate short time, then the participant may overlook the differences, therefore, we choose 30 seconds to be appropriate for the experiment) for each question and we tried to make them follow by guiding them in such a way that every individual do answer the questionnaires questions at the same time. And also we tried to make participants aware that it is necessary for the research that they have to follow strictly the procedure for the success of the study. We have written that they have to use 30 seconds for each question in the paper too.

4.4 Participants

We have got thirty participants in the questionnaire. From this group only two were instructors and the remaining twenty eight were students from different departments and level of studies. You can get full information about the participant statistics in Appendix B.

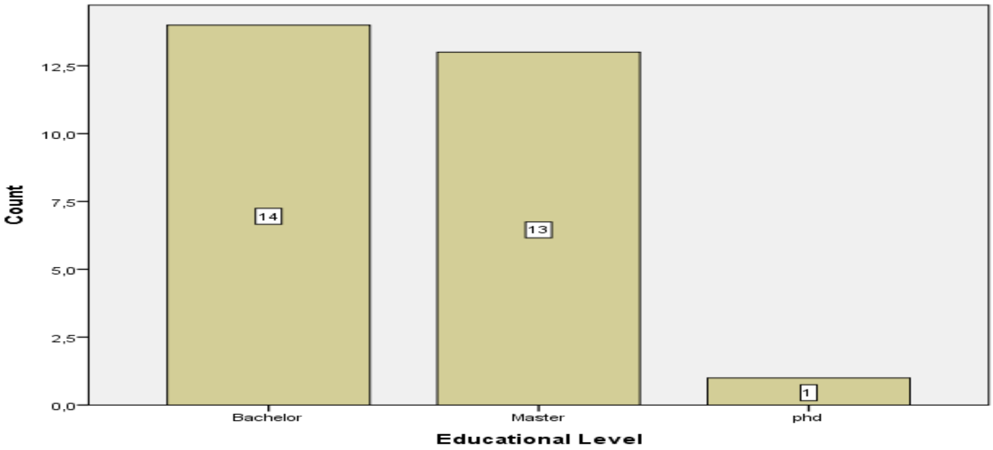


Figure 4:1 shows distribution of educational level of participants.

4.5 Analysis of Result

There are two groups of graphs used, these are Sparse and Dense graphs. We have sparse graph one, two and three and also we have dense graph one, two and three. Each graph is juxtaposed and also superimposed using pair of graphs. These gives as a total of 12 pairs of graphs which are the total number of pair of graphs presented in the questionnaire. Sparse graph one has two difference, sparse graph two has two difference to detect and sparse graph three has one difference to detecte. We have a total of 5 differences to be detected. This applies for both Juxtaposition and Superposition approaches. The percentage detected by each user is calculated by adding all the difference detected by each user and divide it by the total number of difference, which is five. Dense graph one has three differences to be detected, dense graph two has two differences and dense graph three has also two differences. We have a total of 7 differences to be detected. This again applies for both Juxtaposition and Superposition approaches. The percentage detected by each user is calculated by adding all the difference detected by each user and divide it by the total number of difference, which is seven. In the Figure 4:2 below we summarized the percentage of difference detected when it comes to graphs which are sparse using juxtaposed comparison approach.

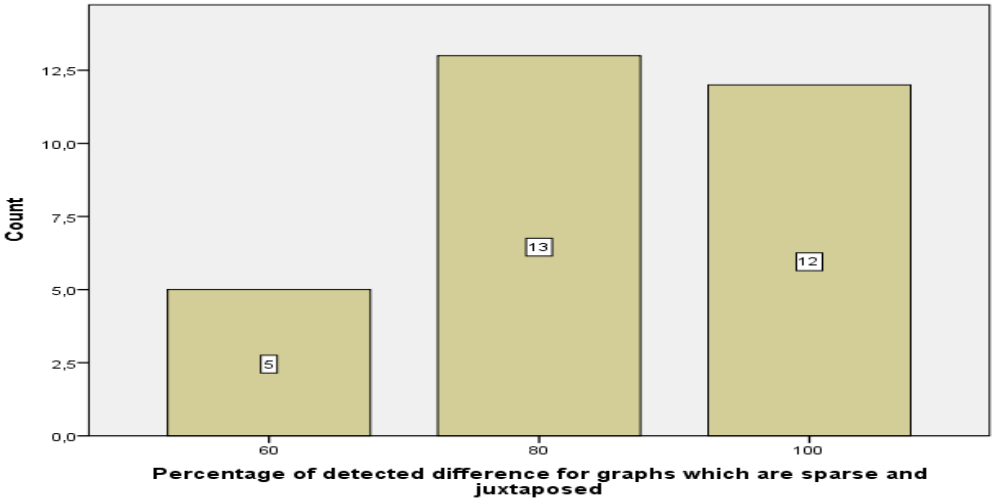


Figure 4:2 Shows the percentage distribution based on the number of differences detected using graphs which are sparse and juxtaposed.

As we can see in the Figure 4:2 above from the total of 30 participants, 5 of the them detected 60% of the difference, 13 of them detected 80% and 12 of them detected 100% of the difference that exist between two graphs who are sparse and juxtaposed.

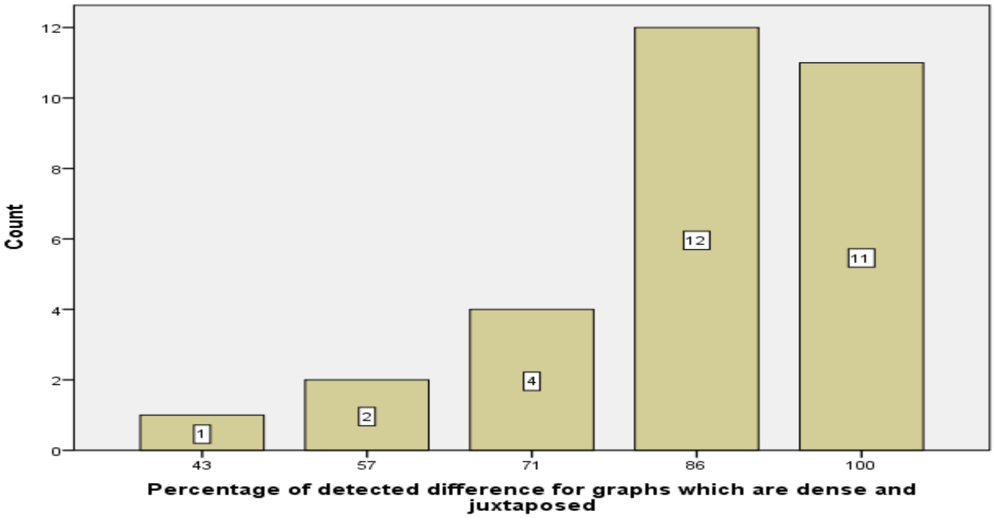


Figure 4:3 Shows the percentage distribution among the participants who detected the difference in the pair of graphs which are dense and juxtaposed.

As we can see in the Figure 4:3 above the percentage distribution for the graphs are 1 participant detected 43%, 2 participants detected 57%, 4 participants detected 71%, 12 participants detected 86% and 11 participants detected 100%.

Figure 4:4 shows the percentage distribution among the participants who detected the difference in the pair of graphs which are sparse and superimposed.

From the above Figure 4:4 it is clear to see that 2 participants detected 60% of the difference, 14 participants detected 80% and 14 participants detected 100%.

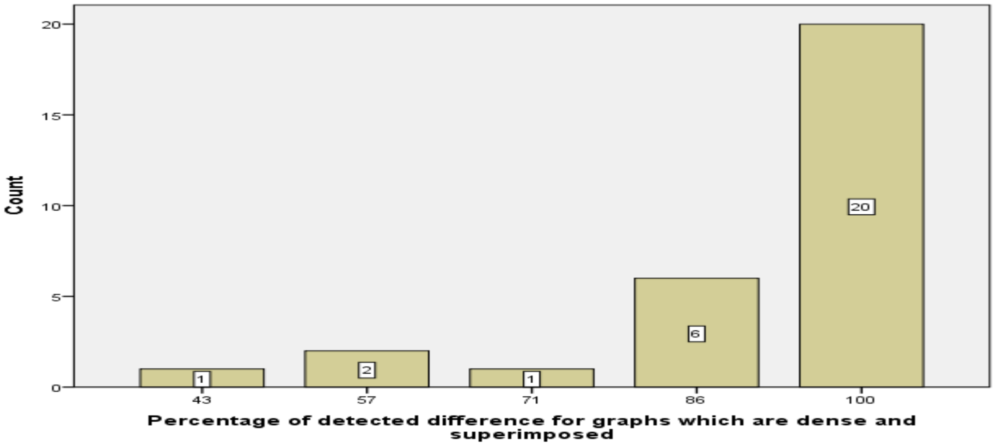


Figure 4:5 shows the percentage distribution among the participants who detected the difference in the pair of graphs which are dense and superimposed

Only one participant detected 43% of the difference between two pair of graphs, 2 participants detected 57%, 1 participant detected 71%, 6 participants detected 86% and 20 participants detected 100%.

The independent samples T-Test between dense Juxtaposition and dense Superposition graphs have no statistically significant variations between the two groups. The mean for the graphs, which is dense and presented using Juxtaposition approach, is 85.77% meaning on average 30 participants have detected 85.77% of the difference in such type of graphs and 91.47% of the difference has been detected using graphs which is dense and using Superposition approach. And also “Sig.” is equals to 0.777 which is well above α which is 0.05 indicating we have no difference statistically. Please do visit Appendix D to see the independent samples T-test table.

Again from the independent samples T-Test between sparse Juxtaposed and sparse Superimposed graphs we get no difference statistically. The mean differences between the two groups are close. That is when the sparse and juxtaposed approach is used the participants identified 84.67% of the difference and when sparse and superimposed graphs used the participants identified 88% of the difference. Moreover, the "Sig." value is 0.519 which is well above 0.05 indicating that we do not have statistically significance variation between the two groups. From the two T-tests we note that there are subtle variations in detecting difference between the two approaches. That is to say, 91.47 % to 85.77% between superimposed dense graph and juxtaposed dense graph, and 88% to 84.67% between superimposed sparse graph and juxtaposed sparse graphs. But these minute variations could not result differences statistically as we can see in the t-test. You can get full informalton about the T-test in Appendix D.

The other important point that we can observe is, the first question in the questionnaire was a graph which is sparse and compared using Superposition approach. And there were two differences to be identified by the questionnaire participants. Let see how the participant answered it below in the Figure 4:8

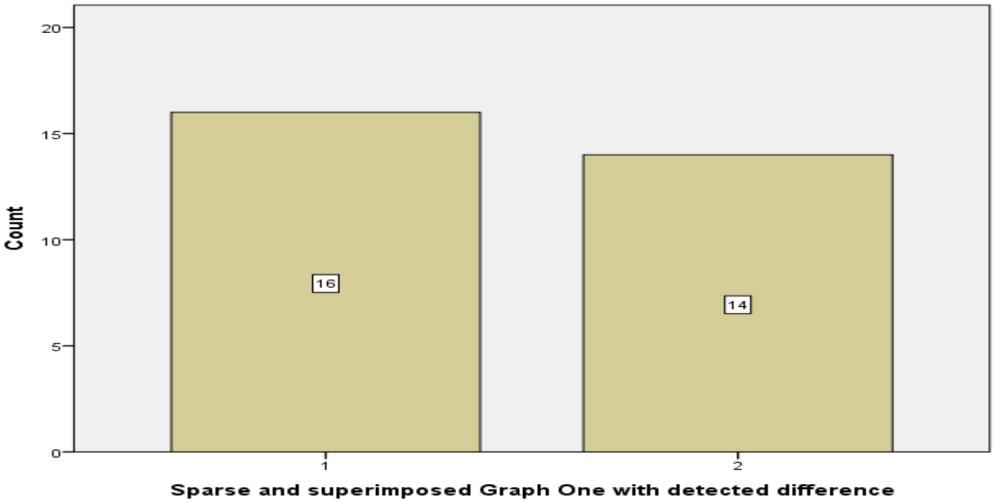


Figure 4:6 Shows how participants detect the difference between a pair of graphs which are sparse and superimposed one another.

From the Figure 4:6 above we can see that 16 individuals detected only one difference and 14 individuals detected both.

The last question in the questionnaire was a graph which is dense and compared using superimposed, which is something more complicated than the first question. Now let see how the participants detected the two differences between two graphs which are superimposed.

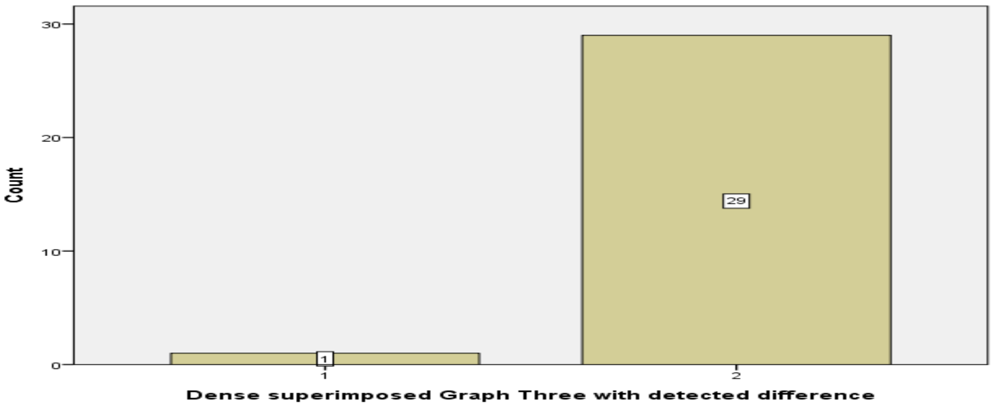


Figure 4:7 Shows how participants detect the difference between a pair of graphs which are dense and superimposed one another.

As you can see in the Figure 4:7 above except one individual all detected both differences correctly. This clearly shows that participants might have gained experience which have could have an effect on the answer given by the questionnaire participants.

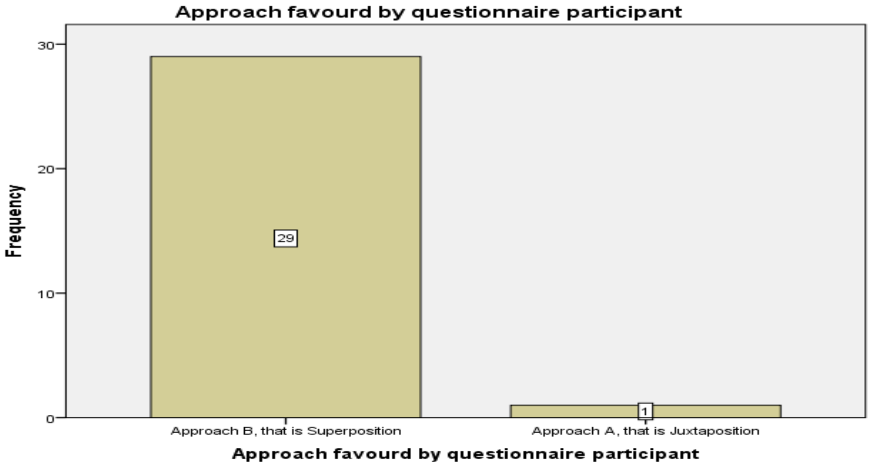


Figure 4:8 Shows us that the number of participants who favor Superposition approach and Juxtaposition approach of network comparisons approaches.

As you see above in the Figure 4:8 the overwhelming majority i.e. 29 of the questionnaire participants have found out graphs which are compared using superimposed comparison approach is better to singled out difference than graphs which are using Juxtaposition comparison approach.

4.5.1 Participants' Comments

We can summarize the reasons given by participants as to why they favor Superposition comparison approach than Juxtaposition comparison approach into four categories. These are as follows:

First, using Superposition comparison approach it is possible to compare the difference easily and fast. Eight participants in the questionnaire have answered that Superposition is better than Juxtaposition for the above reason. The second is Superposition is better than Juxtaposition because the colour difference plays important role to identify the missing line from a pair of edges or nodes. Seven individuals of the participants favor Superposition for the reason mentioned above. The third reason why Superposition is favored is focusing. That is in Superposition we only focus at one place and get the difference just concentrating at a single area. Twelve participants in the questionnaire choose Superposition to show difference better than that of

Juxtaposition for the above reason. The last reason why Superposition is favored than Juxtaposition is the memory issue. In Juxtaposition we have to see here and there to identify the difference and put some information in our memory so that to get the difference correctly. Whereas, using Superposition we do not need to play around with our memory because we found everything in one place. Four participants of the questionnaire have favored Superposition than Juxtaposition for the reason mentioned above. In addition, two participants reasoned out that the reason they choose Superposition than Juxtaposition is that, both colour difference to identify the difference and focusing at one place to look for the difference give Superposition approach an edge.

5 Discussion and Conclusion

Two network comparison approach, namely Juxtaposition (where objects compared using different space and time) and Superposition (where objects compared by overlapping one on the top of the other) were particularly undergone through analysis. Questionnaire was prepared to collect data and to know how individuals detects difference within a pair of graphs using the two approaches for the purpose of comparison. The graphs in the questionnaires were mainly of two types, one group of graphs is sparse and the other is dense. These two groups were designed using three different graphs and each graph is compared using Juxtaposition and Superposition approaches. Therefore, we get twelve pairs of comparisons, six from Juxtaposition and six from Superposition. In section 4 Table 4:1 you can see the 12 graphs used in the questionnaires.

In addition, we have tried participants to face different graphs type while they are attempting the questions in the questionnaire to avoid the effect of past experience or memory effect.

From the total participants who identify difference using sparse and dense juxtaposed graph 12 and 11 of them detected 100% respectively, but using sparse and dense superposition graph 12 and 20 of them detected the difference 100%

Although we have not got statistically difference using independent samples T-test, due to subtle variation, participants have detected slightly more difference using superimposed design approach than juxtaposed as you could get it from the above discussion. The other point is, almost all participants favor overwhelmingly Superposition approach to be better approach in showing the difference between the two graphs.

The above facts tend to favor Superposition (i.e. overlapping one object on the top of the other using same space and time) than Juxtaposition (which puts two graphs in a separate place and time) but we have no statistical support to accept our hypothesis. Therefore, we can only conclude that we have no statistically significant variation between Superposition and Juxtaposition comparision approach to detect differences between two pair of graphs, which rejects our initial hypothesis. The result sounds interesting because almost all our participants got the feeling that Superposition to be better than Juxtaposition in identifying the difference in pair of graphs but the statistical analysis shows that we have no significance variation between the two approaches.

5.1 Future Work

We strongly believe that we have moved one step ahead regarding the network comparison approach. We believe the result is interesting and would pave the way for study and investigation on the two approachs further. In this thesis we do not have any timing mechanism that can force individual use exactly 30 seconds to answer a question. We believe, this could influence the answer each and every participants of the questionnaire gave. Creating a web-based tool could insure the time restriction capability and we can access more people. And any related piece of work that can include such capability could produce even more interesting results. The other point that we would like to make is, having more version of survey with alternate questions order that could avoid the problem of memory effect and any further studies should put these points into consideration. And also making the graphs between dense and sparse more diverse than what has been used here, to see if we could get significance variation in the analysis at the end. The other point is, this thesis has not considered algorithms that produce such layouts. It would be interesting to find out which algorithms are efficient in providing better network comparison approaches.

Bibliography

Alan G. Melville, Martin Graham, and Jessie B. Kennedy (2011). *Combined vs. Separate Views in Matrix-based Graph Analysis and Comparison*. In Proceedings of the 15th International Conference on Information Visualization, School of Computing, Edinburgh Napier University, Edinburgh, United Kingdom. pp. 53 – 58.

Albert Park. (2006). *Intended Use Evaluation Approach for Information Visualization*, Maters thesis. Faculty of the Virginia Polytechnic Institute and State University Blacksburg, Virginia.

Amenta Nina and Klingner Jeff (2002). *Case study: Visualizing sets of evolutionary trees*. Proceedings of the IEEE Symposium on Information Visualization, INFOVIS ’02. IEEE Computer Society: Washington, DC. pp. 71–74.

Anne M. Treisman and Garry Gelade (1980). *A feature-integration theory of attention*. Cognitive Psychology 12, 97–136, Academic Press, Inc.

Anne Treisman (1985). Preattentive processing in vision. Published in Journal: Computer Vision, Graphics, and Image Processing. Academic Press Professional, Inc. San Diego, CA, USA.

Arne Frick, Andreas Ludwig, and Heiko Mehldau (1994): *A Fast Adaptive Layout Algorithm for Undirected Graphs*. Proc. Graph Drawing (GD’94), pp. 399-403.

Ben Shneiderman (1996). *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*. In Proceedings of the IEEE Symposium on Visual Languages, pp. 336–343, Los Alamitos.

Björn Zimmer. Dennie Ackermann, Manfred Schröder, Andreas Kerren, and Volker Ahlers (2010). *Comparative Visualization of User Flows in Voice ortals*. In Proceedings of the 18th International Symposium on Graph Drawing (GD '10), poster paper, pp. 404-405, volume 6502 of LNCS, Konstanz, Germany. Springer.

Carnecky R., Fuchs R., Mehl S., Jang Y., and Peikert R. (2012). *Smart Transparency for Illustrative Visualization of Complex Flow Surfaces*. IEEE transactions on visualization and computer graphics. 19(5):838-51.

Carpendale Sheelagh (2008). *Evaluating Information Visualizations.* Department of Computer Science, University of Calgary. Andreas Kerren et al. (Eds.): Information Visualization, 2008. Springer-Verlag Berlin Heidelberg. LNCS 4950:19 – 45.

Chaomei Chen (2004). *Information visualization: Beyond the Horizon*, 65-87. Springer London, London.

Chaomei Chen and Mary Czerwinski. (2000). *Empirical evaluation of information visualization: an introduction*. International Journal of Human-Computer Studies, John Wiley and Sons, Inc. Vol. 53:631 - 635.

Chaomei Chen and Yue Yu. (2000). *Empirical studies of information visualization: a meta analysis*. International Journal of Human-Computer Studies, Academic Press . Vol. 53:851 - 866.

Christopher Collins, Gerald Penn, and Sheelagh Carpendale (2009). *Bubble Sets: Revealing Set Relations with Isocontours over Existing visualizations*. IEEE transactions on visualization and computer graphics. 15(6):1009-1016.

Christopher D. Shaw, James A. Hall, David S. Ebert, and D. Aaron Roberts. (1999). *Interactive lens visualization techniques*. Proceedings of the Conference on Visualization ’99: Celebrating Ten Years, VIS ’99. IEEEComputer Society Press: Los Alamitos, CA; 155–160.

Christopher G. Healey, Senior Member, IEEE, and James. T. Enns (2012). *Attention and Visual Memory in Visualization and Computer Graphics*. IEEE transactions on visualization and computer graphics. 18(7):1170-88.

Colin Ware (2000). *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers, Inc., San Francisco, California.

Dastani M. (1998). *Languages of Perception*. ILLC Dissertation Series 1998-05, University of Amsterdam.

Desney Tan, Greg Smith, Bongshin Lee, and George Robertson (2007). *AdaptiviTree: Adaptive tree visualization for tournament-style brackets*. IEEE transactions on visualization and computer graphics. 13(6):1113-20.

Giuseppe Di Battista, Peter Eades, Roberto Tamassia, and Ioannis G.Tollis (1998). *Graph Drawing: Algorithms for the visualization of Graphs,* vii (397), Prentice Hall.

Helen C. Purchase, David Carrington , Jo-Anne Allder (2002). *Empirical Evaluation of Aesthetics based Graph Layout.* School of Information Technology and Electrical Engineering, The University of Queensland, Australia. Empirical Software Engineering, 7:233–255. Kluwer Academic Publishers, The Netherlands.

Helen C. Purchase, R. F. Cohen, M. I. James (1995). *An Experimental Study of the Basis for Graph Drawing Algorithms.* Springer.

Hinrichs U., Schmidt H., and Carpendale S. (2008). *Emdialog: Bringing information visualization into the museum*. IEEE Trans Visual Comput Graph; 14:1181–1188.

Ing-Xiang Chen, C. -Z. Y. (2010). *Visualizaion of Social Networks Handbook of Social Network Technologies and Application*. Springer-Verlag New York Inc. pp. 585-610.

J. L. Moreno (1953). *Who Shall Survive: Foundations of Sociometry, Group Psychotherapy, and Sociodrama*. Beacon House, Beacon, N.Y.

Jaime Huerta-Cepas, Joaquín Dopazo, and Toni Gabaldón (2010). *ETE: a python Environment for Tree Exploration*, licensee BioMed Central Ltd, Huerta-Cepas et al.

Jan de Leeuw (2009). *Journal of Statistical Software.* John Wiley & Sons, Inc.

Jakob Nielsen and Rolf Molich (1990). *Heuristic Evaluation of User Interfaces*, presented at proceedings of ACM CHI '90, Seattle, VA.

Jean-Daniel Fekete, Jarke J. van Wijk, John T. Stasko, and Chris North (2008). *The value of information visualization*. Information Visualization, Springer-Verlag Berlin Heidelberg. LNCS 4950:1–18.

Jin Chen, Alan M. MacEachren, and Donna J. Peuquet (2009). *Constructing Overview + Detail Dendrogram-Matrix Views*. IEEE transactions on visualization and computer graphics. 15(6):889-96 and http://www.youtube.com/watch?v=IRhKPpluwQ4.

Justus Broß, Patrick Schilf, Maximilian Jenders, and Christoph Meinel (2011). *Visualizing the Blogosphere with BlogConnect*. Institute at the University of Potsdam, IEEE CS, MIT, Boston, USA. pp. 651 - 656 .

Keith Andrews, Martin Wohlfahrt, and Gerhard Wurzinge (2009). *Visual graph comparison*. In Proceedings of the 2009, 13th International Conference InformationVisualization. IEEE Computer Society. Washington, DC. pp. 62–67.

Linton C. Freeman (2000). *Visualizing social networks*, Journal of Social Structure Article, University of California, Irvine.

Lothar Krempel (2009). *Network Visualization*, 1–31, Sage, London/New Delhi.

Manuel Freire, Catherine Plaisant, Ben Shneiderman, Jen Golbeck (2010) ManyNets: *An Interface for Multiple Network Analysis and Visualization*, Atlanta, GA, USA.

Mario Albrecht, Andreas Kerren, Karsten Klein, Oliver Kohlbacher, Petra Mutzel, Wolfgang Paul, Falk Schreiber, and Micheal Wybrow (2009). *On Open Problems in Biological Network Visualization*. In Proceedings of the 17th International Symposium on Graph Drawing, Chicago, USA. Springer. Volume 5849 of LNCS:256-267.

Martin Graham and Jessie Kennedy (2007). *Exploring Multiple Trees through DAG Representations*. IEEE transactions on visualization and computer graphics. 13(6):1294-301.

Matej Novotny (2004). *Visually effective information visualization of large data*. Proceedings of the 8th Central European Seminar on Computer Graphics.

Matthew O. Ward, Georges G. Grinstein, Daniel Keim (2010). *Interactive Data Visualization, Foundation, Techniques and Applications*, Taylor & Francis Inc, United States.

Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Hansen, Charles D. Hansen, and Jonathan C. Roberts (2011). *Visual comparison for information visualization*. Information Visualization, 10 (4):289–309, sage.

Mikko Berg (2012), *Human abilities to perceive, understand, and manage multidimensionalinformation with visualization,* phd thesis, Department of Media Technology, Aalto Univeristy, Espoo, Finland.

Natalia Andrienko, Gennady Andrienko, and Peter Gatalsky (2003). *Exploratory spatio-temporal visualization: an analytical review*. Journal of Visual Languages & Computing, ELSEVIER. *14*(6):503–541.

Paola Tubaro (2012). *Introduction to social network analysis*. University of Greenwich, London.

Pier Francesco Cortese, Giuseppe Di Battista, Antonello Moneta,Maurizio Patrignani, and Maurizio Pizzonia (2006). *Topographic Visualization of Prefix Propagation in the Internet*. IEEE transactions on visualization and computer graphics. 12(5):725-32.

Purvi Saraiya, Chris North, and Karen Dunca (2004). *An Evaluation of Microarray Visualization Tools for Biological Insight*. Presented at proceedings of IEEE Symposium on Information Visualization, Austin, Texas, USA.

Qin Shen, Alex Pang, and Sam Uselton (1998). *Data Level Comparison of Wind Tunnel and Computational Fluid Dynamics Data*. Proceedings of the IEEE Visualization 98 (VIZ'98). pp. 415-418.

R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. (2002). *Network Motifs: Simple Building Blocks of Complex Networks*. Science 298, no. 5594: 824|827.

Raga'ad M. Tarawneh, Patric Keller, and Achim Ebert (2011). *A General Introduction To Graph Visualization Techniques*. Proceedings of IRTG 1131 – Visualization of Large and Unstructured Data Sets Workshop. pp. 151–164.

Romain Bourqui, Helen Purchase, Fabien Jourdan (2011). *Domain Specific Vs Generic Network Visualization: an evaluation with metabolic networks*, Australasian Conference on User Interface. Perth, Australia.

Romain Bourqui, Ludovic Cottret, Vincent Lacroix, David Auber, Patrick Mary, Marie-France Sagot, and Fabien Jourdan (2007). *Metabolic network visualization eliminating node redundance and preserving metabolic pathways*. BMC Systems Biology, 1(29).

Seok-Hee Hong, Hiroshi Nagamochi (2006). *Convex Drawings of Hierarchical Plane Graphs*. In Proceedings of the 17th Australasian Workshop on Combinatorial Algorithms, AWOCA.

Shengdong Zhao, Michael J. McGuffin, Mark H. Chignell (2005). *Elastic Hierarchies: Combining Treemaps and Node-Link Diagrams*. University of Toronto. IEEE Symposium on Information Visualization (InfoVis). pp. 57 - 64.

Soegaard, Mads (2005). *Gestalt principles of form perception*. Retrieved 14 July 2013 from http://www.interaction-design.org/encyclopedia/gestalt_principles_of_form_perception.html

Staša Milojević, Bradley M. Hemminger , Chaomei Chen, Loet Leydesdorff , Jason Priem, Scott B. Weingart (2012). *Information Visualization State of the Art and Future Directions*. ASIST, Baltimore, MD, USA.

Stephen G. Kobourov (2012), *Spring Embedders and force directed Graph Drawing Algorithms*, university of Arizona.

T. Lengauer John Wiley and Sons (1990). *Combinatorial Algorithms for Integrated Circuit Layout*, John Wiley & Sons, Inc. Chichester, UK.

Tim Dwyer, Seok-Hee Hong, Dirk Koschutzki, Falk Schreiber, Kai Xu (2006). Visual Analysis of Network Centralities, NICTA VALACON (Visualisation and Analysis of Large and Complex Networks) Workshop on Biological Networks organised by Seok-Hee Hong on November, 2004 in Sydney.

Ulrik Brandes, Tim Dwyer, and Falk Schreiber (2003). *Visualizing related metabolic pathways in two and a half dimensions*. Graph Drawing , Springer-Verlag Berlin Heidelberg. pp. 111–122.

Weijiang Li and Hiroyuki Kurata (2005). *A grid layout algorithm for automatic drawing of biochemical networks*, Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan.

William T. Tutte. (1963). *How to draw a graph*. Proceedings of the London Mathematical Society, 13 (52):743 -768.

Wolfram Research (2008). *Introduction to Graph Drawing*, Wolfram Research, Inc.

Ying Tu and Han-Wei Shen (2007). *Visualizing Changes of Hierarchical Data using Treemaps*. IEEE transactions on visualization and computer graphics. 13(6):1286-93.

Zeqian Shen, Kwan-liu Ma, and Tina Eliassi-Rad (2006). *Visual Analysis of Large Heterogeneous Social Networks by Semantic and Structural Abstraction*. IEEE transactions on visualization and computer graphics, special issue on visual analytics. 12(6):1427-39.

Appendices

Appendix A: This appendix contains information about questionnaires used in the research.

Appendix B: This appendix contains personal information of participants who are participated in the study.

Appendix C: This appendix contains the row data which is collected from each questionnaire and displayed analysis of some information

Appendix D: Independent Samples T-Test

Appendix A: Questionnaire

Remark**:** Please don’t write your name

Age Group: 14-17, 18-24, 25-34, 35-44, 45-55, over 55 (please circle)

Gender: ________________________

Department: ________________________

Major Study: ________________________

Educational Level: Bachelor, Master, PhD (please circle)

Have you ever done network/graph analysis before? Yes, No (please circle)

If your answer is Yes, please explain us how you did it?

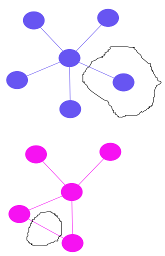
____
____
____
____
____
____
____
____

The purpose of this questionnaire is to get information about the two basic network/graph comparisons namely **Juxtaposition** and **Superposition**, and to identify which comparison is better in providing information easily and clearly to the audience**.**

How to answer the questions in the questionnaire?

Below we have two approaches: **Approach A** and **Approach B**

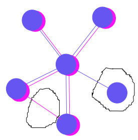
Approach A: What difference can you observe between the two diagrams below?



Juxtaposition

As you can see in the diagram above (blue) has six nodes but the lower (red) has only five. To show the difference we need to circle on the node that makes them different. And also the lower graph (red) has additional edge between the bottom nodes which is not found in the graph (blue) above it. Likewise we circle on the edge to show the difference.

Approach B: What difference can you observe between the two diagrams below?



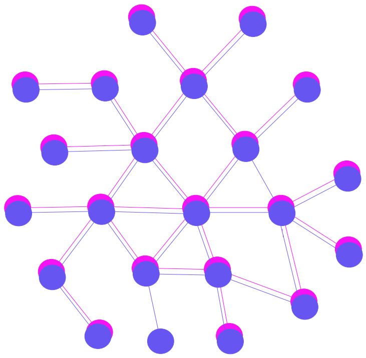
Superposition

As you can see in the graph above with blue colour (overlaying) has six nodes and the graph with red colour (underlying) have only five and has additional edge at the bottom. To show their difference we need to circle the node (nodes) that makes them different as well as the edge(s) which is only found in the underlying graph (red).

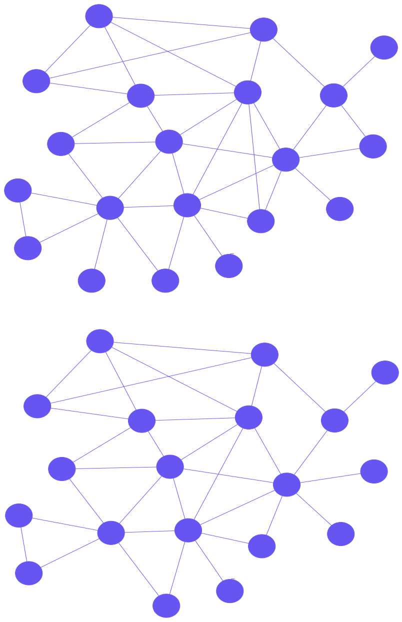
This page intentionally left blank

<u>Note</u>: Please answer each question within 30 seconds. This is very important for the research. We need your kindness to follow our request for the success of our study.

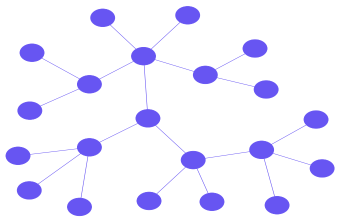
1. Please circle the difference that you see between these two overlapped diagrams? (Answer it within 30 seconds please)

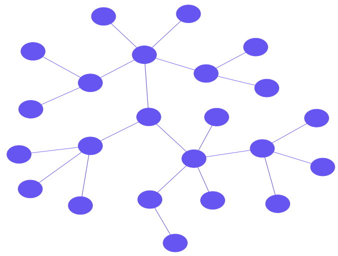


2. Please circle the difference that you see between these two diagrams? (Answer it within 30 seconds please)

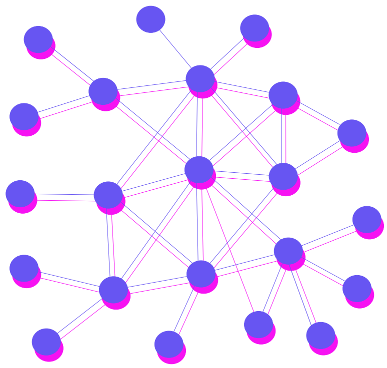


3. Please circle the difference that you see between these two diagrams? (Answer it within 30 seconds please)

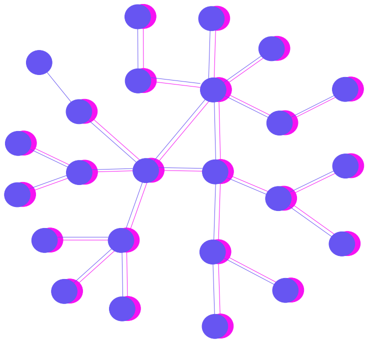




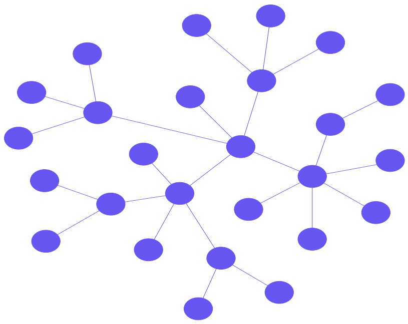
4. Please circle the difference that you see between these two overlapped diagrams? (Answer it within 30 seconds please)

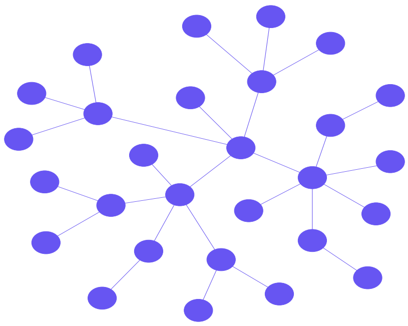


5. Please circle the difference that you see between these two overlapped diagrams? (Answer it within 30 seconds please)

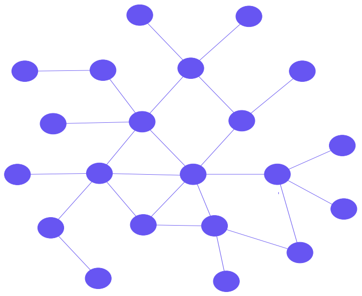


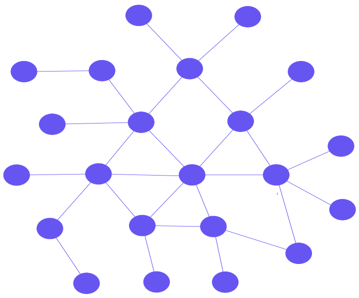
6. Please circle the difference that you see between these two diagrams? (Answer it within 30 seconds please)



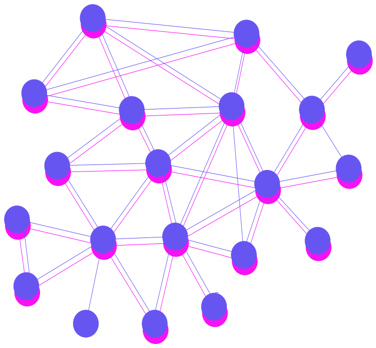


7. Please circle the difference that you see between these two diagrams? (Answer it within 30 seconds please)

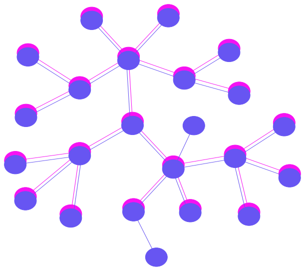




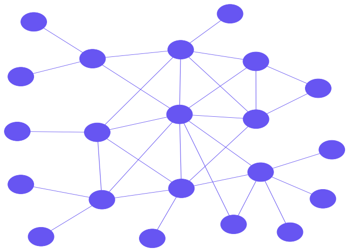
8. Please circle the difference that you see between these two overlapped diagrams? (Answer it within 30 seconds please)

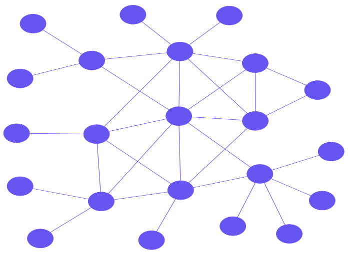


9. Please circle the difference that you see between these two overlapped diagrams? (Answer it within 30 seconds please)

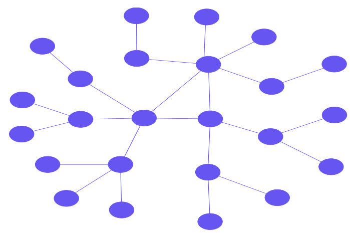


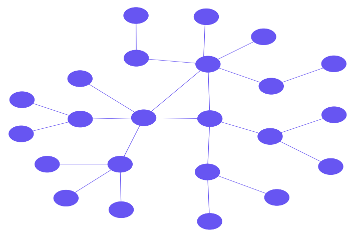
10.Please circle the difference that you see between these two diagrams? (Answer it within 30 seconds please)



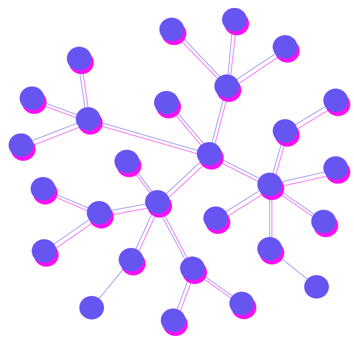


11. Please circle the difference that you see between these two diagrams? (Answer it within 30 seconds please)





12. Please circle the difference that you see between these two overlapped diagrams? (Answer it within 30 seconds please)



In your opinion, which graph/network comparison technique allows an easier identification of the differences between two graphs/networks?

A. **Approach A**(i.e. showing two graphs separately)
B. **Approach B** (i.e. overlapping one graph on the other)

Why ?

Finished !!!

Thank you very much for your effort !!!

Appendix B: Participants Personal Information

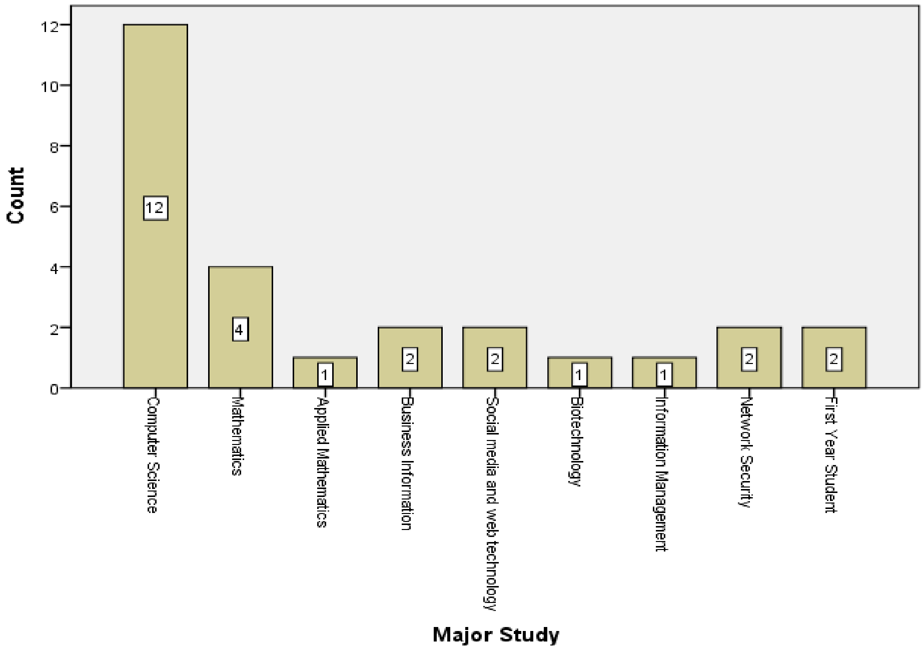
The table below shows the demography of 30 individuals who are participating in the questionnarie.

P	Age Group	Gender	Department	Major Study	Educational Level	Done Network Analysis Before
1	25-34	M	Computer Science	Network Security	Bachelor	No
2	-	M	Computer Science	Software Technology	Master	No
3	25-34	F	DFM	Mathematics	Master	No
4	-	M	Computer Science	Computer Science	Bachelor	No
5	25-34	M	-	Biotechnology	Master	No
6	-	F	Mathematics	-	Bachelor	No
7	18-24	M	Mathematics	First year student	Bachelor	No
8	-	-	-	-	-	-
9	Over 55	M	Math	-	PhD	Yes. As a teacher
10	18-24	F	DFM	Mathematics	Bachelor	No
11	18-24	F	DFM	Mathematics	First year student	-
12	18-24	M	Mathematics	First year student	Bachelor	No
13	18-24	F	Math	Math	Bachelor	No
14	25-34	M	Math	Applied Math	Bachelor	No
15	18-24	M	Computer Science	Computer Science	Bachelor	No
16	25-34	M	Software Technology	Computer software	-	Yes. Yes I have analysed a project by a built in tool. We have analysed how many classes, methods and other total number of lines of code. These

						are different matrix which do it
17	18-24	F	Social media and web technology	Social media and web technology	Master	No
18	25-34	M	Computer Science	Software Technology	Master	No
19	25-34	M	Computer Science	Information Management	Master	No
20	25-34	M	Computer Science	Business Information	Master	No
21	18-24	F	Computer Science	Business Information System	Master	No
22	25-34	M	DFM	Computer Science	Bachelor	No
23	18-24	M	Computer Science	Computer Science	Master	No
24	Over 55	M	Computer Science	Computer Science	Master	No
25	-	M	DFM	Computer Science	Bachelor	No
P	Age Group	Gender	Department	Major Study	Educational Level	Done Network Analysis Before
26	25-34	M	Computer Science	Computer Science	Bachelor	No
27	25-34	M	Computer Science	Network Security	Bachelor	Yes, ISP backbone designs , MPLS and BGP node topologies
28	18-24	M	Computer Science	Information Technology	Master	No
29	18-24	M	DFM	Computer Science	Master	No
30	18-24	F	Media Technology	M	Master	Yes, I have had Graph theory mathematical course and also understanding graphs on Algorithms and Data structure course on Bachelor studies

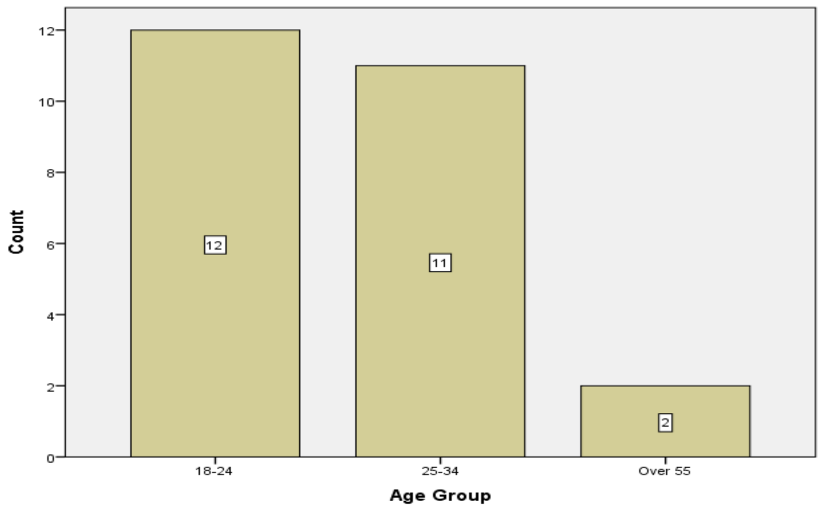
Appendix C: Statistical data of participants

The following shows the major studies that the 30 individuals participating in the questionnaire



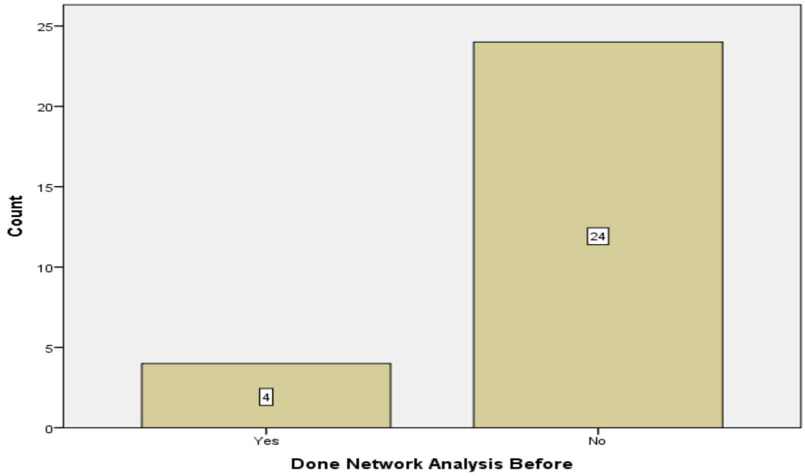
Below we can see the distribution of questionnaire participants according to their department.

Below you can see that the sex distribution among the 30 individuals who are participating in the questionnaire



Below we can see the gender distribution among the 30 individuals who participated in the questionnaire.

In the Figure below you can see that 24 of the individuals out of 30 them have not done any network analysis. And four of the participants have said they have done network analysis. The first participant who is PhD holder said he worked network analysis "As a teacher" which vague and did not explain it further. The second participant wrote that he has done network analysis to determine how many classes, methods, and number of lines a given software has, if it is for the purpose of knowing the above facts I do not think it is relevant to the question asked. The third participants answer yes and wrote, ISP backbone designs, MPLS (Multiprotocol Label) and BGP (Border Gateway Protocol) node topologies but failed to explain how is it was done. And the last participant answered yes and wrote he/she understands graph theory for mathematics course and also knows about graph from data structure and algorithms course, which is irrelevant for the question asked.



Here I summarized the data collected from the questionnaire participants. A total of 30 students participate in the study. In order to minimize space for the table I used abbreviation for some words that are described below:

Abbreviation	Meaning		Abbreviation	Meaning
P	Participant		S1-Juxt	Sparse graph one - Juxtaposition
M	Male		S2-Juxt	Sparse graph two - Juxtaposition
F	Female		S3-Juxt	Sparse graph three -Juxtaposition
Dif	No of difference between two graphs/network		D1-Super	Dense graph one-Superposition
D	Number of difference detected		D2-Super	Dense graph two-Superposition
ND	Number of difference not detected		D3-Super	Dense graph three - Superposition
S1-Super	Sparse graph one - Superposition		D1-Juxt	Dense graph one-Juxtaposition
S2-Super	Sparse graph two-Superposition		D2-Juxt	Dense graph two - Juxtaposition
S3-Super	Sparse graph three - Superposition		D3-Juxt	Dense graph three-Juxtaposition

P	S1-Super			D1-Juxt			S2-Juxt			D2-Super			S3-Super			D3-Juxt			S1-Juxt			D1-Super			S2-Super			D2-Juxt			S3-Juxt			D3-Super		
	Dif	D	ND	Dif	D	ND	Dif	D	ND	Dif	D	ND	Dif	D	ND	Dif	D	ND	Dif	D	ND	Dif	D	ND	Dif	D	ND	Dif	D	ND	Dif	D	ND	Dif	D	ND
1	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: Easy to see																																				
2	2	1	1	3	1	2	2	1	1	2	1	1	1	1	0	2	2	0	2	2	0	2	1	1	2	2	0	2	1	1	1	1	0	2	2	0
Which approach is best? B: It was easy to observe the difference in approach B because of colours make it easy to see the difference node																																				
3	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: Faster to see the difference																																				
4	2	2	0	3	3	0	2	1	1	2	2	0	1	1	0	2	1	1	2	1	1	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: No difference for me																																				
5	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: Easier to compare																																				
6	2	1	1	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: Easy to see because it is close																																				
7	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0	1	1	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: Easier to see different many points. Easy to follow the lines and the structure																																				
8	2	1	1	3	1	2	2	2	0	2	2	0	1	1	0	2	2	0	2	2	0	3	2	1	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: Faster to detect, less time to compare																																				
9	2	1	1	3	2	1	2	2	0	2	1	1	1	1	0	2	2	0	2	1	1	3	1	2	2	2	0	2	1	1	1	1	0	2	2	0
Which approach is best? B: The vision is not splited easier to miss a difference with approach A I think																																				
10	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0	2	1	1	3	3	0	2	2	0	2	1	1	1	1	0	2	2	0
Which approach is best? B: Since you can easily see the differences. And fast																																				
11	2	1	1	3	2	1	2	2	0	2	2	0	1	1	0	2	2	0	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: Because then you see the difference immediately																																				
12	2	2	0	3	2	1	2	2	0	2	2	0	1	1	0	2	2	0	2	2	0	3	2	1	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: Everything is focused in one place it is easy to spot if a graph is missing a part since it really sticks out																																				
13	2	1	1	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0	2	1	1	3	3	0	2	2	0	2	1	1	1	1	0	2	2	0

Which approach is best? B: It was much easier to compare the 2 graphs you didn’t have 2 constantly scan back and forth between the 2 images																																				
14	2	1	1	3	2	1	2	2	0	2	2	0	1	1	0	2	2	0	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: Faster identification of the differences																																				
15	2	1	1	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0	2	1	1	3	1	2	2	1	1	2	2	0	1	1	0	2	2	0
Which approach is best? B: I don’t have to work with my memory of one graph when I look at another																																				
16	2	1	1	3	2	1	2	1	1	2	2	0	1	1	0	2	2	0	2	1	1	3	2	1	2	2	0	2	1	1	1	1	0	2	2	0
Which approach is best? A: If the two graphs are drawn separately we can concentrate easily. If two graph more concentrate and deep eye is required you can compare two graphs easily and fast if they are shown separately as compare on one another																																				
P	S1-Super			D1-Juxt			S2-Juxt																													
17	2	2	0	3	2	1	2	1	1	2	2	0	1	1	0	2	2	0	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: Because it is easier to find the missing colour in one place																																				
18	2	1	1	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0	2	2	0	3	3	0	2	2	0	2	2	0	1	0	1	2	2	1
Which approach is best? B: In my opinion the “Approach B” is easy to identify the differences between the two graphs. It is quite easy to compare the two graphs and differences are vivid and easy to identify																																				
19	2	2	0	3	3	0	2	2	0	2	1	1	1	1	0	2	2	0	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: The colours play a big role. But it was often hard to find mostly connections																																				
20	2	2	0	3	2	1	2	2	0	2	2	0	1	1	0	2	2	0	2	1	1	3	2	1	2	2	0	2	1	1	1	1	0	2	2	0
Which approach is best? B: Eye movement is not that much more focus, less irritation																																				
21	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: In the example you also use different colour, it makes easier to compare the 2 graphs, besides you directly have the graphs on one sight																																				
22	2	1	1	3	1	2	2	1	1	2	1	1	1	1	0	2	1	1	2	1	1	3	1	2	2	1	1	2	1	1	1	1	0	2	1	1
Which approach is best? B: Personally, I choose B because the space between the two graphs are dense. It is very easy to be identified. However, I think this may be more like a psychological question. Some people may think A is easier																																				
23	2	2	0	3	1	2	2	1	1	2	2	0	1	1	0	2	2	0	2	1	1	3	3	0	2	2	0	2	1	1	1	1	0	2	2	0
Which approach is best? B: You can focus on both images at the same time instead of trying to remember the previous image																																				
24	2	1	1	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0	2	2	0	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0
Which approach is best? B: It is easier to compare when you do not need to shift between different figures																																				
25	2	1	1	3	3	0	2	2	0	2	1	1	1	1	0	2	1	1	2	1	1	3	3	0	2	2	0	2	2	0	1	1	0	2	2	0

<table>
<tr><td colspan="37">Which approach is best? B: It is easy to see missing edges and modes due to the colouring and overlap</td></tr>
<tr><td>26</td><td>2</td><td>1</td><td>1</td><td>3</td><td>2</td><td>1</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>1</td><td>1</td><td>3</td><td>3</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td><td>2</td><td>0</td></tr>
<tr><td colspan="37">Which approach is best? B: Easier to identify when the visible overlap is missing</td></tr>
<tr><td>27</td><td>2</td><td>1</td><td>1</td><td>3</td><td>2</td><td>1</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>3</td><td>3</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td><td>2</td><td>0</td></tr>
<tr><td colspan="37">Which approach is best? B: With different colour at one scene, one can easily identify which colour is missing or surplus</td></tr>
<tr><td>28</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>3</td><td>3</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td><td>2</td><td>0</td></tr>
<tr><td colspan="37">Which approach is best? B: Using approach A, I have to follow two separate graphs which means that I have to inspect 1st one remember it inspect 2nd graph. Compare them some data could be lost/ forgotten during the workflow. With approach B I don’t have to remember anything , I just have to detect single objects within a set of duplicated ones</td></tr>
<tr><td>29</td><td>2</td><td>1</td><td>1</td><td>3</td><td>3</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>3</td><td>3</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td><td>2</td><td>0</td></tr>
<tr><td colspan="37">Which approach is best? B: With approach B you don’t have to pay too much attention as compared to the other approach</td></tr>
<tr><td>30</td><td>2</td><td>2</td><td>0</td><td>3</td><td>2</td><td>1</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>1</td><td>1</td><td>3</td><td>3</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>2</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td><td>2</td><td>0</td></tr>
<tr><td colspan="37">Which approach is best? B: Because we had the graphs very close and colours are easier to identify. When I compare two graphs (as in option A) I have to check each node and their connection. On second case I just check for missing colour in the pattern</td></tr>
</table>

Appendix D: Independent Samples T-Test

Shows the T-test between the two comparisons approach namely dense Juxtaposition and dense Superposition graphs.

T-Test

Group Statistics

	Comparision	N	Mean	Std. Deviation	Std. Error Mean
Difference detected in graphs	Percentage of Detected Difference in Dense and Juxtaposed Graphs	30	85,77	15,069	2,751
	Percentage of Detected Difference in Dense and Superimpsed Graphs	30	91,47	15,290	2,792

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Difference detected in graphs	Equal variances assumed	,081	,777	-1,454	58	,151	-5,700	3,919	-13,546	2,146
	Equal variances not assumed			-1,454	57,988	,151	-5,700	3,919	-13,546	2,146

Shows the T-test between the two comparisons approach namely sparse Juxtaposition and sparse Superposition graphs.

T-Test

Group Statistics

	Comparision	N	Mean	Std. Deviation	Std. Error Mean
Difference detected in graphs	PercentageOf Detected Difference In Sparse And Juxtaposed Graphs	30	84,67	14,559	2,658
	Percentage Of Detected Difference In Sparse And Superimposed Graphs	30	88,00	12,429	2,269

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Difference detected in graphs	Equal variances assumed	,421	,519	-,954	58	,344	-3,333	3,495	-10,329	3,663
	Equal variances not assumed			-,954	56,608	,344	-3,333	3,495	-10,333	3,666

Linnæus University
School of Computer Science, Physics and Mathematics

SE-391 82 Kalmar / SE-351 95 Växjö
Tel +46 (0)772-28 80 00
dfm@lnu.se
Lnu.se/dfm