GENOGRAM VISUALIZATION USING SOCIAL NETWORK APPROACH FOR MEDICAL FAMILY TREE

SITI FATIMAH BINTI BOKHARE

UNIVERSITI SAINS MALAYSIA

2018

GENOGRAM VISUALIZATION USING SOCIAL NETWORK APPROACH FOR MEDICAL FAMILY TREE

by

SITI FATIMAH BINTI BOKHARE

Thesis submitted in fulfillment of the requirements
for the degree of
Master of Science

August 2018

ACKNOWLEDGEMENT

In the name of Allah, the most gracious and the most merciful. Alhamdulillah, highest praise to Allah for his wills that gave me the strength and patience to complete this thesis and made all things possible

Foremost, I would like to express my sincere gratitude to Dr Wan Mohd Nazmee Wan Zainon as my supervisor and my co-supervisor, Prof. Dr Abdullah Zawawi Talib for their continuous support of my master study and research, for their patience, motivation, enthusiasm, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and co-supervisor for my master study.

I acknowledge, appreciate, and return the love and support of my family, without whom I would be lost, especially to my husband Fadzli and my son Muhammad who have been my emotional anchors through not only during my study years, but my entire life. Both my parents, siblings and their family have also become an important part of my world. I am forever indebted to all my family members for their affection, support, and constant encouragement.

Thanks to many current and former people and all lab members at the School of Computer Sciences USM for their help and friendship over the years.

TABLE OF CONTENTS

Acknowledgement……... ii

Table of Contents……..…….. iii

List of Tables……... vii

List of Figures…….. viii

Abstrak…… xii

Abstract…… xiv

CHAPTER 1 - INTRODUCTION.....1

1.1 Background.....1

1.1.2 Visualization.....3

1.2 Motivation of Research.....5

1.3 Problem Statement.....6

1.4 Objective Research.....8

1.5 Scope of Research.....8

1.6 Contributions.....9

1.7 Thesis Outline.....10

CHAPTER 2 - LITERATURE REVIEW.....12

2.1 Introduction.....12

2.2 Visualization.....12

2.3 Genealogy Visualization....13

2.3.1 Treeplus....16

2.3.2 Treemap....18

2.3.3 Hyperbolic tree....20

2.3.4 Genogram....22

2.4 Genealogy Software....27

2.4.1 Family Tree Software....27

2.4.2 Genogram Software....29

2.4.3 Genealogy Graph Prototype Software....32

2.4 Graph Theory of Kinship Network....34

2.5 Graph Visualization....38

2.5.1 Network Visualization....38

2.6 Discussion....45

CHAPTER 3 - RESEARCH METHODOLOGY....48

3.1 Introduction....48

3.2 Overview of the Methodology....48

3.3 Designing the Framework and its Implementation....50

3.4 Evaluation....53

3.5 Summary 55

CHAPTER 4 - DESIGNING THE FRAMEWORK AND ITS IMPLEMENTATION 56

4.1 Introduction 56

4.2 Designing the Proposed Framework 56

4.3 Designing the Visualization Approach 62

4.3.1 Overview First Technique 62

4.3.2 Zooming and Filtering Technique 66

4.3.3 Detail-on-Demand Technique 68

4.4 Calculating Risk Score using Analysis Method 69

4.5 Implementation 73

4.5.1 Dataset 74

4.5.2 Implementation of "Overview first" Technique 77

4.5.2(a) Layout algorithm, Size and Colour of Node and Label 78

4.5.3 Implementation of "Zooming and filtering" Technique 81

4.5.4 Implementing Detail-on-demand 90

4.5.5 Implementation of Risk Score Calculation 93

4.6 Summary 96

CHAPTER 5 - RESULT AND ANALYSIS 97

5.1 Introduction 97

5.2 Evaluation procedure 98

5.2.1 Case Study Evaluation 98

5.2.2 Sample Selection 99

5.2.3 Questionnaire Design 100

5.3 Analysis and Finding 103

5.3.1 Evaluating the Usefulness 105

5.3.2 Evaluating the Ease of Use 107

5.4 Open-ended strategies evaluation 110

5.5 Discussion 112

5.6 Threats to Validity 113

5.7 Summary 114

CHAPTER 6 - CONCLUSION 115

6.1 Introduction 115

6.2 Conclusion 115

6.3 Future Research 117

6.3.1 Type of visualization 118

6.3.2 Automatic Risk Score Prediction 118

REFERENCES.....110

APPENDICES

LIST OF PUBLICATIONS

LIST OF TABLES

		Page
Table 5.1	Scale items of the usefulness and the ease of use concept	104
Table 5.2	Summary of the comments for the open-ended questions	111

LIST OF FIGURES

		Page
Figure 1.1	Standard symbols for genograms	2
Figure 1.2	Step by step visualization criteria	4
Figure 2.1	Common family tree layout	14
Figure 2.2	Pedigree chart	14
Figure 2.3	Ancestry chart	15
Figure 2.4	Example of treeplus	17
Figure 2.5	Example of treemap	19
Fiqure 2.6	(a) Classical tree layout (b) Treemap layout	20
Fiqure 2.7	Example of hyperbolictree	21
Figure 2.8	BRCA1 pedigree	24
Figure 2.9	Example of medical genogram	26
Figure 2.10	PAF companion	28
Figure 2.11	Myheritage	29
Figure 2.12	Genogram	30
Figure 2.13	Advannce Edraw	31
Figure 2.14	Dual tree genealogy visualization	33
Figure 2.15	Ore-graph	35
Figure 2.16	P-graph	36
Figure 2.17	Bipartite p-graph	37
Figure 2.18	Approaches to deal with large networks	40
Figure 2.19	Visualizing flow through a network	41

Figure 2.20	SocialAction's technique on a subgraph from the global terrorism network	43
Figure 2.21	SocialAction's technique on a subgraph from the global terrorism network	44
Figure 3.1	Research methodology	49
Figure 4.1	Proposed framework design	57
Figure 4.2	Diagram of causes of diabetes, hypertension and heart disease	59
Figure 4.3	Standard gender symbols	60
Figure 4.4	Family relationships key	60
Figure 4.5	Medical genogram symbols (both gender)	60
Figure 4.6	Example of medical genogram	62
Figure 4.7	Example of node-link diagram (disease and relationship)	64
Figure 4.8	Example of node-link diagram filtered using disease link type	67
Figure 4.9	Ordered list for comparing ranking	69
Figure 4.10	Network visualization of eigenvector centrality measure	72
Figure 4.11	Gephi architecture	73
Figure 4.12	Gephi overview	74
Figure 4.13	Medical genogram	75
Figure 4.14	Data imported into gephi data table	76
Figure 4.15	Assigning node with suitable data type (a) Node table (b) Edge table	77
Figure 4.16	Ahmad multi-mode social graph	78
Figure 4.17	Node-link diagram overview	79

Figure 4.18	ForceAtlas 2 layout algorithm	79
Figure 4.19	Data label visualization	80
Figure 4.20	Filter library	82
Figure 4.21	Filtering on gender using the attribute equal filter window	83
Figure 4.22	Ahmad's network visualization filtered on male gender	84
Figure 4.23	Filtering on type using the attribute partition filter window	85
Figure 4.24(a)	Ahmad's network visualization filtered on type of edges (kind) which value is equal to "Hypertension"	86
Figure 4.24(b)	Ahmad's network visualization filtered on type of edges (kind) which value is equal to "Heart Disease"	87
Figure 4.24(c)	Ahmad's network visualization filtered on type of edges (kind) which value is equal to "Diabetes"	88
Figure 4.25	Filtering on age using the attribute range filter	89
Figure 4.26	Ahmad's network visualization filtered on age which value of age in a range of 10 to 69 years old	90
Figure 4.27	Ordered list of network segmentation of node and edge attribute	91
Figure 4.28	Ordered list for degree ranking	93
Figure 4.29	Network visualization of node coloured due to the result of degree ranking in ordered list	93
Figure 4.30	Network visualization using eigenvector centrality measure	95
Figure 5.1	Model of usefulness, ease of use, and self-predicted future usage	101
Figure 5.2	Score mean of usefulness	105
Figure 5.3	Comparison between the score mean for usefulness metric	106
Figure 5.4	Score mean of ease of use	108

Figure 5.5 Comparison between the score mean for ease of use metric 109

VISUALISASI GENOGRAM MENGUNAKAN PENDEKATAN RANGKAIAN SOSIAL BAGI PEPOHON SALASILAH KELUARGA

ABSTRAK

Genogram adalah perwakilan grafik pepohon keluarga yang digunakan untuk menjejaki sejarah dan hubungan keluarga. Perwakilan ini memberikan gambaran keseluruhan tentang sejarah keluarga dan data genealogi yang membolehkan pengguna memahami dengan jelas maklumat umum dan kompleks keluarga mereka selama beberapa generasi. Namun, terdapat beberapa batasan semasa menggunakan genogram, iaitu masalah kebolehskalaan ketika memvisualisasikan kekayaan dan maklumat yang kompleks apabila pepohon keluarga semakin besar. Oleh itu, satu kerangka untuk meneroka data keluarga (genogram) menggunakan pendekatan rangkaian sosial dan kaedah analisis sebagai suatu peningkatan dicadangkan. Kerangka ini mengikuti reka bentuk panduan antara muka pengguna grafik canggih yang merupakan Mantra Maklumat Visual-Pencarian iaitu "Tijauan Keseluruhan Dahulu, Zum dan Tapis, kemudian Perincian-Mengikut Kehendak", yang dicadangkan oleh Shneiderman pada tahun 1996. Kajian ini memberi tumpuan kepada peningkatan interaksi dan pemahaman pengguna kepada data pepohon keluarga perubatan. Satu kajian kes telah dijalankan untuk menilai kegunaan dan kemudahan penggunaan alat prototaip ini. Telah ditemui bahawa alat prototaip yang menggunakan pendekatan rangkaian sosial meningkatkan pemahaman pengguna terhadap keluarga pepohon perubatan dari segi kegunaan dan kemudahan penggunaan. Menurut penilaian kerangka yang dicadangkan, alat ini boleh digunakan untuk kegunaan peribadi atau oleh profesional penjagaan kesihatan dan seterusnya membantu mereka memahami pepohon keluarga mereka dengan lebih baik.

GENOGRAM VISUALIZATION USING SOCIAL NETWORK APPROACH FOR MEDICAL FAMILY TREE

ABSTRACT

Genogram is a graphical representation of a family tree that is used for tracking family history and relationships. It provides an overview of the family history and genealogy data which allow user to clearly understand the general and complex information of their family for several generations. But, some limitations exist while using genogram, namely scalability problems while visualizing the wealth and complex information once the family tree gets bigger. Hence, a framework for exploring medical family tree data (genogram) using social network approach and analysis method as an enhancement is proposed. This framework follows the design of advanced graphical user interface guide which is the Visual Information-Seeking Mantra that is "Overview First, Zoom and Filter, then Details-on Demand", proposed by Shneiderman in 1996. This research focuses on improving user interaction and understanding of medical family tree data. A case study has been conducted to evaluate the usefulness and ease of use of the prototype tool. It was discovered that the prototype tool which employ the social network approach improve user understanding of medical family tree in terms of usefulness and on ease of use. According to the evaluation of the proposed framework, the tool can be utilized for personal use or by healthcare professionals and furthermore it helps them to understand their medical family tree better.

CHAPTER 1

INTRODUCTION

1.1 Background

Genogram is one of the visual representations of genealogy which is similar to that of a family tree. Much like a family tree, the genogram graphically presents demographics with various shaped symbols and colour-coded lines. Genograms can depict a variety of information such as gender, age, ethnic origin, health status, education achievements, as well as different types of relationship such as family, social and emotional relationships. Important dates and any chosen characteristic such as personality, health and vocation of individual family members can also be input in a genogram (McGoldrick et al., 2008; Pendagast and Sherman, 1976).

Currently genograms are used by various groups of people in a variety of fields such as medicine, psychiatry, psychology, social work, genetic research, education and many more. Healthcare professionals have been using genogram as a tool to record important information about their patients and their family histories. The drawing of genogram has to conform to a set of rules in order for users to have the same understanding and obtain a similar interpretation of the genogram. Figure 1.1 shows some basic symbols for genogram.

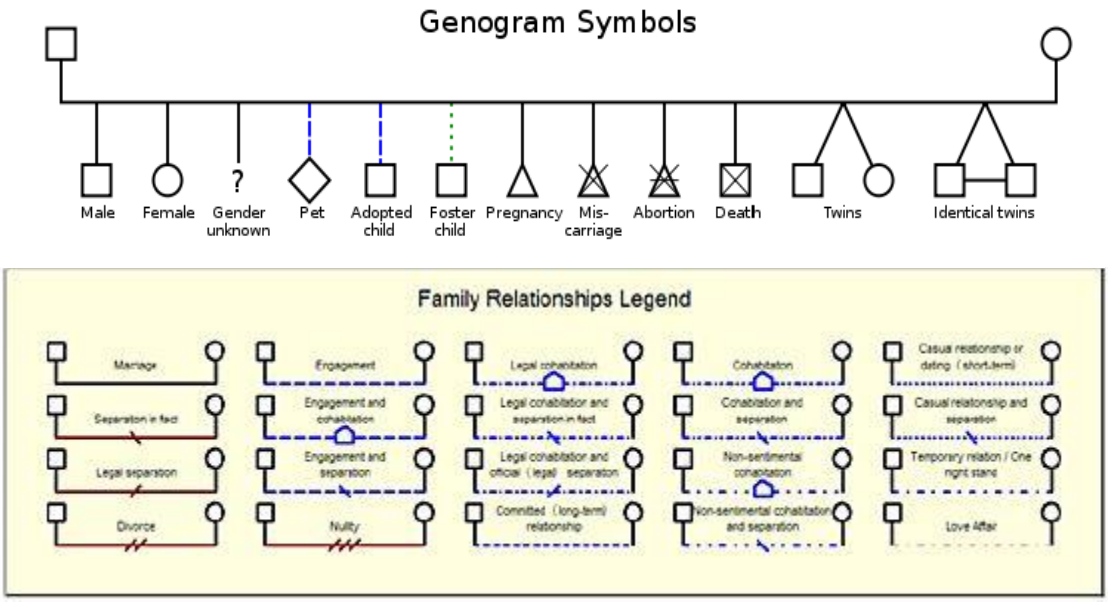

Figure 1.1 Standard symbols for genograms (McCormick et al.,1987)

The drawback of genogram is that it is highly dependent on the size and complexity of the relationships to be portrayed. There are existing software for generating family trees but problems exist due to non-interactive and scalability of trees which may affect the process to comprehend the genogram. This is because the nature of the genogram graphics could be complex and requires some advanced knowledge or detail explanations by those creating the symbol of genogram in order for others to understand the information displayed (Vivienne, 2010).

1.1.2 Visualization

Generally, visualization is the transformation of data and information into pictures. One definition of visualization is to form a mental vision, image or picture of (something not visible or present to sight or of an abstraction); to make visible to the mind or imagination (The Oxford English Dictionary, Third Edition, 2010). The visualization of the data allows the user to gain insight into the data and come up with new hypotheses.

Nowadays the amount of data stored on electronic media is growing exponentially fast. Today's data warehouses dwarf the biggest databases built a decade ago (Kimball and Mer, 2000), and making sense of such data is becoming harder and more challenging.

According to McCormick et al. (1987) visualization is a method of computing which transforms the symbolic into geometric to enable researchers to observe their simulations and computations. Meanwhile Gershon (1994) define that visualization is more than a method of computing, it is about the process of transforming information into a visual form, enabling users to observe the information. In their understanding, visualization technology is based on the integration of older technologies, including computer graphics, image processing, computer vision, computer-aided design, geometric modelling, approximation theory, perceptual psychology, and user interface studies. Figure 2.1 shows where visualization maps the computer representation into images or animation.

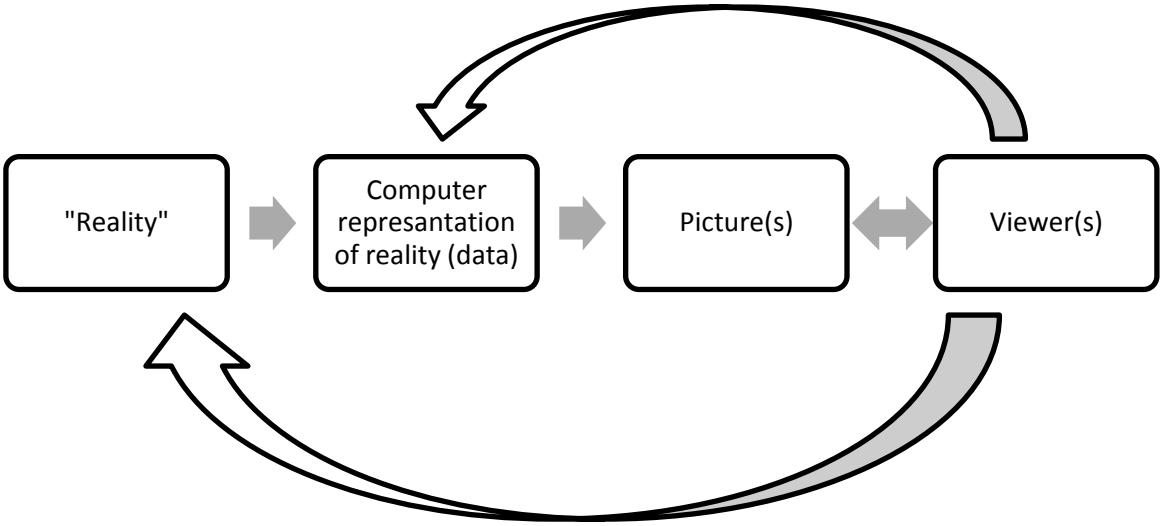

Figure 1.2 Visualization maps the computer representation of reality into perceptual (visual) representation (Domik and Gutkauf, 1994)

In the present day, visualization is used in many scientific areas. Each of these scientific areas has difference data types that need table visualization to be visualized. Keim (2002) has described six data types to be visualized: one-dimensional data (Spence and Apperley, 1982), two dimensional data (Laurini and Thompson, 1992; Egenhofer and Richards, 1993), multi-dimensional data (Inselberg, 1985; Williamson and Shneiderman, 1992;), text and hypertext, hierarchies and graph (Chimera and Shneiderman, 1993; Carriere and Kazman, 1995), algorithm and software.

There are three criteria to be fulfilled in order to visualize data. Firstly, users need to recognize the data type and then try to find key issue for the specific data type. After that, apply a suitable visualization technique that will help to solve the key issue.

This research focuses on visualization interaction and understandability factors in medical family tree data. The object of interest for this research is medical genogram (medical family tree data). The main challenge of this research is how to visualize the medical genogram in order to increase user understanding, through improvement of interaction which leads to a better understanding of medical family tree data especially medical genogram. Hence this research attempts to address the scalability and clustering issue by proposing a visualization technique to improve user interaction and understandability of medical family tree.

1.2 Motivation of Research

Genogram constructs from various symbols and colour-coded lines that represent demographics information, health problem and addiction in individuals and among family members. Thus genogram have a lot of potential to be explored. As someone who is health conscious the researcher believes that medical genogram is very beneficial to be understood because apart from what people know from their family's tree and history, they can also find out about hereditary diseases that could be passed on from their parents to their offspring or descendants. Through this information, people will be more careful and take precautions towards the health problem. The use of different symbols in genogram will confuse the first time users. Furthermore excessive use of colour also contributes to the cluttered look when the family tree gets bigger. According

to Ellis and Dix (2005), "*too much data on too small an area of the display will result in visual clutter, which in turn diminishes the potential usefulness of the visualization, especially when the user is exploring the data*".

The leading benefit of data visualization is that it does not only provide graphical representation of data but also allows changing of form, omitting what is not required and browsing deeper to get further details. This is a great eye catcher and attracts user's attention better and provides better communication. Furthermore, it provides a great advantage over traditional methods. Data visualization gives opportunity to approach huge data and makes it easily comprehensible. Thus to gain the full benefit of genogram data, users require a visualization tool that can present the data in a fully interactive environment conducive to exploration and also allow users to view their own custom data.

1.3 Problem Statement

Genogram contains a wealth of information on the family's history; therefore as a family tree gets bigger, visualizing becomes a more difficult task. Currently medical genograms use symbols, shape and colour-coded lines to represent information such as gender, age, ethnic origin, health status, as well as different types of relationship such as family. GenoPro, Genoware and WinGeno software package uses the symbols shown in Figure 1.1 to encode family's information. However, when the node expands, the genogram will looks confusing and cluttered by the use of different symbols and excessive use of colours (Keller et al., 2011).

Understanding of a medical genogram is helpful in determining patterns of disease or illness within a family. But, the image-map format which is very limited in its ability to support gestural interaction leads to information perception difficulties and the visualization loses its initial purpose to understand the genogram (McGuffin et al., 2005). So, using social network visualization method to visualize the medical genogram reduces the node cluttered (scalability) and supports gestural interaction by using selected visualization technique (McGuffin and Balakrishnan, 2006) which improves user interaction and furthermore increases user understanding of their medical family tree data.

Health problems that run in the family can increase chances of developing the problem. This is because families share their genetics, environment, and habits (Genetic Alliance, 2009). But by knowing only patterns of disease does not help them to know the probabilities of inheritance. Therefore, visualizing statistical results (analysis score) will help user to gain new insight on health problem that affects their family from medical family tree data.

1.4 Research Objectives

The aim of this research is to proposed a better framework for visualising medical family tree data. In more detail, it seeks to fulfil the following research objectives:

1. To design a visualization approach that improves user interaction and understanding of medical family tree data especially genogram in real-time.
2. To apply analysis method which provides user with visualization of analysis score that can improve the tool usability.

1.5 Scope of Research

The scopes of the research are as follow:

(i) Medical Genogram: Medical genogram is used as a tool for mapping family patterns by recording health information (including illnesses and medical conditions) about the patients and their family members. The family health information from three generation of relatives is needed in this research as a complete family history to map pattern of disorder among relatives (Biad, 2015).

(ii) Visualization technique: Visualization techniques can be used to visualize the medical genogram. This research focuses mostly on the visualization techniques that can used to explore the genogram, reduce scalability problem and create interactive environment to user. This research utilizes social network action visualization which implement the design of advanced graphical user interface guide known as Visual Information-Seeking Mantra "Overview first, zoom and filter, then details-on demand" (Shneiderman,

1996). So, three visualization techniques have been used in this prototype tool.

(iii) Novices: This research focuses on beginners and novices as the target users. Users are treated as beginners or novices. This is because, whether they may have some experience or knowledge in tree scturture or medical family tree, the enviroment and experience is different when using the visualization tool. Futhermore, any assitance or guideline provided by using the visualization tool are more meaningful to them.

1.6 Contributions

Generally this research leads to a better understanding of family tree data. This research is expected to makes several contributions to the fields of medical health, genealogy and social network. By using social network visualization method and applying visual seeking mantra, a framework would be established which can present the data in an interactive environment that can assist user in understanding their medical family tree data better.

The analysis method applied in this framework provides statistical analysis of data that gives the user the ability to see its data from a new angle. The approach of Exploratory Data Analysis (Tukey, 1977) emphasizes the importance of curiosity and serendipity (i.e. discoveries made while searching for something else) to data analysis. This method provides a range of metrics for calculating statistics that characterize the data and used statistical results in the visualization by making use of colours and sizes to

create informative visuals and make them particularly suitable for the visual exploration of data.

1.7 Thesis Outline

This thesis has six chapters that are organized in a chronological order consisting of an introduction, literature review, research methodology, the visualization technique, evaluation, and discussion respectively. The organization of this thesis is as follow:

Chapter 1 introduces the overview of family tree and genogram, background of the problems in genogram software. Furthermore, this chapter also explains on the research objectives and problem statement of this research. This chapter also describes the importance of this research, followed by the scope of this research. The theoretical framework for this research is designed and this is followed by a list of research contributions.

Chapter 2 reviews the existing works and some literature survey that are related to this research. Systematic literature review on the aspect of improving interaction and understandability of medical family tree data are explained in detail.

Chapter 3 outlines the research methodology. It provides detail on the procedure taken from the beginning to the end of the conducted research. Research procedures conducted in this study are illustrated and followed by explanations in the chapter.