

# A REVIEW ON TOOLS AND TECHNIQUES FOR FAMILY TREE DATA VISUALIZATION

<sup>1</sup>SITI FATIMAH BOKHARE, <sup>1\*</sup>WAN MOHD NAZMEE WAN ZAINON

<sup>1</sup>School of Computer Sciences, Universiti Sains Malaysia, Malaysia

E-mail: <sup>1</sup>[fatimahbokhare@gmail.com](mailto:fatimahbokhare@gmail.com), <sup>2</sup>[nazmee@usm.my](mailto:nazmee@usm.my)

## ABSTRACT

Visualization is an important and helpful ways to support the exploration of large data sets. The leading benefit of 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 paper reviews some of the previous research related to family tree (or sometimes known as genealogy) data visualization. It focuses on existing techniques and applications that are currently available to address family tree visualization issues. The content of this paper is divided into several sections such as visualization usability, family tree visualization, graph theory of kinship network and graph visualization. Visualization gives opportunity to approach huge network type of data and makes it easily comprehensible. In order to gain the full benefit of family tree data, a proper understanding about the current visualization tool or techniques that is used to represent these type of data in a fully interactive environment will be highly beneficial.

**Keywords:** *Data Visualization, Family Tree Visualization, Theory of Kinship Network, Social Network Visualization*

## 1. INTRODUCTION

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. 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; one-dimensional data, two dimensional data, multi-dimensional data, text and hypertext, hierarchies and graph, algorithm and software.

This paper focuses on visualization interaction and understandability factors in family tree data. The main challenge of this research is how to visualize the family tree data in order to increase user understanding, through improvement of interaction. Hence this paper attempts to review current visualization techniques especially those that focus

on the scalability and clustering issue that can improve user interaction and understandability of family tree data.

## 2. FAMILY TREE VISUALIZATION

Family tree also known as Genealogy is the study of family history and relationship which plays an important role in society (e.g. of royal families, and of human migration), genetics, evolutionary biology, and in some cases, religion. Over the years, handcrafted charts have been used to illustrate genealogical relationships of a few dozen individuals in books. So the most widely used method of visualizing family history is to visualize it in a form of a tree often called a family tree (Figure 1).

A family tree is a diagram representing family relationships which depicts a minimum of three generations in a conventional tree structure. There are two types of genealogical relations in every chart which is Parent-child relationships (consanguine relations) which define a hierarchy in genealogical data and Relationships through marriage (conjugal relations) are non-hierarchical and merge family trees [1]. Together these form a network of

relationships complex but simpler than a general graph.

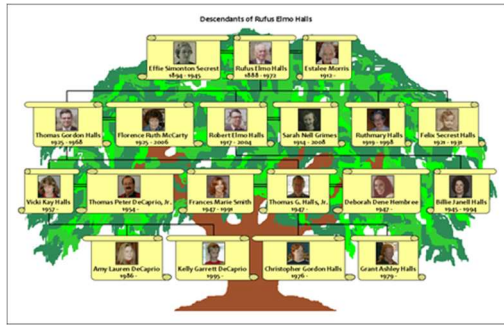


Figure 1: Common family tree layout (taken from cadapplet.com)

The main problem of tree layouts is on handling the size of the tree. Some layouts may effectively handle a tree with a few hundreds of nodes but fail when it has a few thousands of nodes. A big number of nodes and edges lead to a high density of nodes, and it becomes impossible to interact with the tree and get any information from it. Thus careful analysis on layouts is needed to choose the suitable one which can solve a particular problem the best. It is beyond the scope of this paper to describe all the different types of layouts, however, this is the most common ones: Treeplus layout, Tree-map layout, hyperbolic tree. These layouts will be briefly discussed in the following sections, as well as some genealogy software packages.

### 2.1 TreePlus

TreePlus involves converting graphs into trees and shows the missing graph structure with visualization and interaction techniques [2]. TreePlus approach includes allowing preview of adjacent nodes, animating change of the tree structure, and giving visual hints about the graph structure. It allows users to begin with a specific node and incrementally explore the graph in detail and rapidly read labels to analyze the meaning of relationships. According to its author, TreePlus was inspired from the success of their earlier tools, SpaceTree and TaxonTree. Figure 2 shows example of TreePlus with the low density dataset used in the user study.

TreePlus applied pedigree chart layout. This layout shows that every node has its own line and therefore the number of lines needed to visualize the

tree equals to the number of nodes in the tree. As the indentation can be visually perceived only a few lines at once, it becomes difficult to decide if the nodes are aligned or not in a large tree. For example, it might happen that nodes that belong to the same level, e.g. sibling nodes, can be placed on lines that are far away from each other. Therefore, it leads to usability issues when the layout is used to visualize large trees.

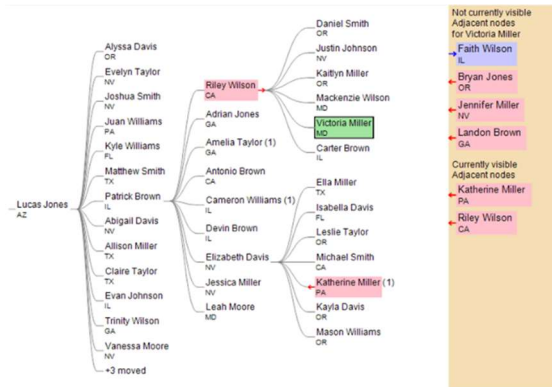


Figure 2: An example of TreePlus [2]

### 2.2 Treemap

Treemap is a space-filling visualization method capable of representing large hierarchical collections of quantitative data [3]. It is formed by using a rectangular display area and recursively subdividing it based on the tree structure, alternating between horizontal and vertical subdivision, and filling the terminal rectangular regions with a color that can be used to represent different types of data. Treemap resulting from the idea to visualize files on a hard drive, have been used practically to a wide variety of domains ranging from financial analysis to sports reporting.

Figure 3 show an example on how classical tree layout being visualize into tree map layout. The weight of each node is the size of the file or folder represented by that node. Every node in that tree has a label and a weight that is the size of that file or folder. The Tree-map layout shown in 3(b) visualizes the same data. The figure clearly shows only the content of the leaf nodes, and it is difficult to extract the content from all the other internal nodes.

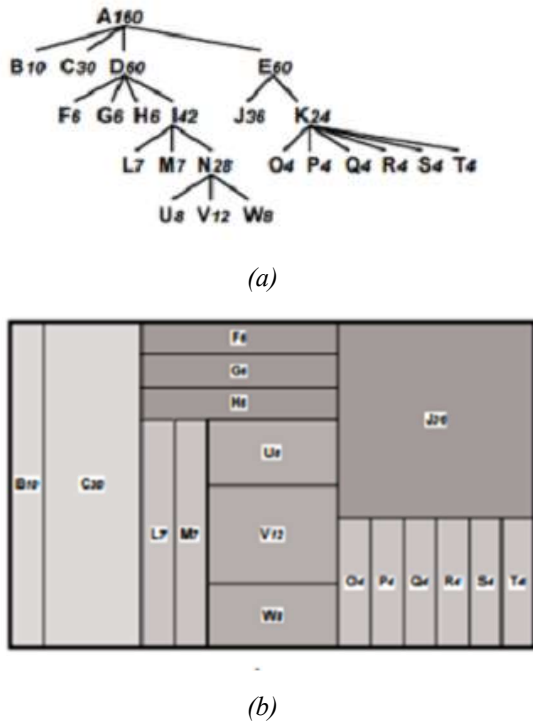


Figure 3 (a) Classical tree layout and (b) Tree-map layout of the same hierarchical data

Information on how to apply the Tree-map layout to genealogy data rarely found. It is most likely because of the mentioned above limitation. Other reason might be in case of genealogy data, ways to assign weights to the nodes of the tree that determine the sizes of the rectangle it not very clear.

### 2.3 Hyperbolic tree

Hyperbolic Tree also called hypertree, employs hyperbolic space in which radial view laid out in it [4]. This view is then mapped to a Euclidean plane so that an arbitrarily large tree fits within an oval-shaped area on the screen (see Figure 4). Hypertree using focus+context (fisheye) technique also known as hyperbolic browser, is used for visualizing and manipulating large hierarchies. The root is placed in the center, and its children are fanned outward. Any part of the tree can be moved to the center with a simple mouse-click or mouse-drag.

A hyperbolic tree follows the same basic principles as a common tree, with a link between a parent and a child. Users can easily grasp the hierarchical structure. However, since nodes are

spread out evenly, users may find it hard in telling exactly how balanced or unbalanced a tree is.

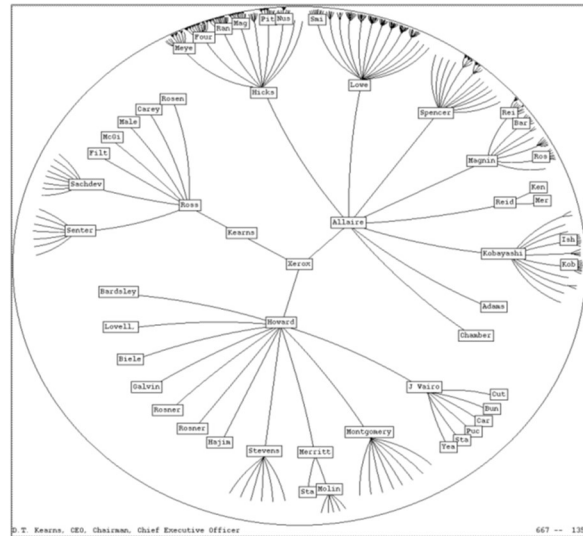


Figure 4: An example of hyperbolic tree [4]

### 2.4 Genogram

Genogram is another visualization of genealogical data that is widely used in the medical field. A genogram is a graphic representation of a family tree that is used for tracking family history and relationships. It provides a detailed pictorial display of a family or groups that goes beyond a family tree in such a way that it contains information about genealogy and different types of relationship [5]. It was first developed and popularized in clinical settings by a family therapist Monica McGoldrick and Randy Gerson by publishing their first book in 1985 [6]. It resembles a family tree but is more sophisticated in its ability to provide particular kinds of information.

Genograms have been used for the last four decades as a tool for mapping family patterns and the psychological factors that interrupt relationships. They allow a practitioner to identify and understand repetitive patterns of behavior and to recognize hereditary tendencies; hence genograms are popular with healthcare professionals in areas such as social work, psychology, psychiatry, genealogy, genetic research, education and many more fields.

Other information that may be contained in a genogram depend on what people want to know; for

example, if the purpose of a genogram is to map the medical history of a family, the genogram might include information such as hereditary patterns, genetics and psychological patterns. Clinical uses of genograms include gaining insight into clients' psyche [7]; as a data gathering device on individuals and families [8]. The development of various types of genogram such as cultural, socio-economic, ethical and career or work genograms is further evidence of the development and broader use of the tool over time [6].

Figure 5, shows an example of a medical genogram. It shows how a deleterious breast and ovarian cancer syndromes 1 (BRCA1) mutated across three generations, including affected family members with breast cancer or ovarian cancer.

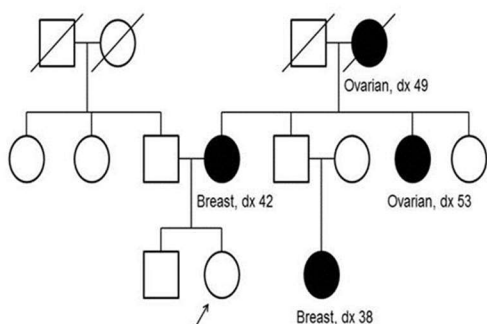


Figure 5: Example of Medical Genogram (Taken from National Cancer Institute)

Medical genogram is a very useful tool used to record health information of the patient and families. It does not only show the relationships among members of the patient's family, but it also includes relevant health information for each person (temporal information). This is because families have many factors in common, including their genes, environment and lifestyle. Together, these factors can give clues to medical conditions that may run in a family [9]. By noticing patterns of disorders among relatives, healthcare professionals can determine whether an individual, other family members or future generations may be at an increased risk of developing a particular condition. For example, people with a higher-than-usual chance of having common disorders, such as heart disease, high blood pressure, stroke, certain cancers, and diabetes [10]. These complex disorders are influenced by a

combination of genetic factors, environmental conditions and lifestyle choices [11]. So it will help the patient and other family member to notices a pattern of disease in their family.

To visually encode disease information various shaped, colored symbols and lines are used. The use of genograms is growing, and the genograms and the symbols are expanded and modified. Different software packages use their own conventions on how to visually encode information with different symbols and colors. Moreover, some of those software packages allow users to add custom symbols or change existing ones to reflect the current needs. Even though it adds flexibility to the tools, it is at the same time a drawback as there is no consistency in the visualization between different software packages. That requires training when changing from one tool to another. Moreover, it can lead to inaccurate or incorrect analysis results. Figure 6, shows an example of medical genogram created for three generation using GenoPro software.

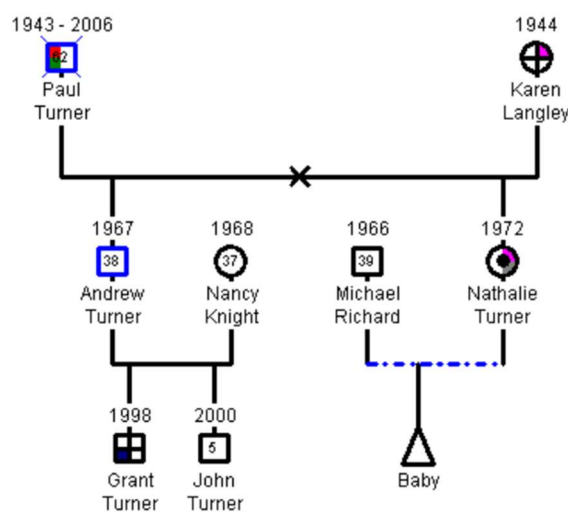


Figure 6: Example of medical genogram (GenoPro)[9]

Based on this figure, we can see that Paul, a smoker, recently died of a heart attack after years of battling with heart disease and diabetes. His wife, Karen has been diagnosed with breast cancer. Nathalie is pregnant, and she is worried that she may have a breast tumor. The anxiety she feels has affected her mental health, and she is also dealing with depression. Meanwhile, Andrew is also a smoker,



and his 8-year-old son has Down syndrome. From the visualization point of view, genograms are just another modification of the standard classical tree layout that visually encodes some additional information by using various symbols. Thus it is similarly does not encode any temporal information except in the form of text labels. It has the same scalability problem when applied to the big genealogy datasets. The big number of individuals in the family will eventually lead to long edges, edge crossings and to exponential crowding. Moreover, it is not strictly standardized and may lead to inaccurate analysis results, and requires additional resources to learn the symbols of each particular software package. Even though it is the most widely used tool in the medical field for visualizing genealogy data, it needs improvement to satisfy the increased needs in the field.

### 3. FAMILY TREE SOFTWARE

There exists numerous commercial along with free software applications for generating family trees. The main purpose of these trees is to allow users to dynamically build a tree by interactively adding nodes. PAF Companion, MyHeritage and GenoPro are examples of software applications for generating family trees.

PAF (Personal Ancestral File) aids users to organize their family history records [12]. It can be generated, either on screen or on paper, and generate pedigree charts, family group records, and other reports that allow users to search for their missing ancestors. The advantages of PAF are that it makes use of colors in order to facilitate users to identify nodes at different levels as the tree gets bigger. Furthermore, each node has a label with information and birth years. Figure 7 shows an example of companion sample chart taken from PAF.

MyHeritage on the other hand, uses the power of the Internet in order to ease people around the world to discover their heritage and strengthen their bonds with family and friends [13]. Unlike PAF, it uses selected colors which are pink and blue to help identify gender. Moreover, it has a label containing qualitative information about the nodes and presents some form of temporal information which is birth

and death years. Figure 8 shows an example of family tree chart taken from MyHeritage.

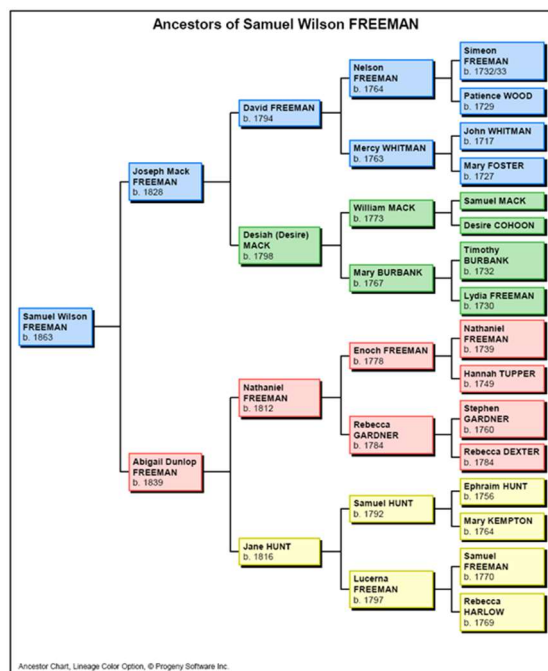


Figure 7: Example of PAF companion [12]

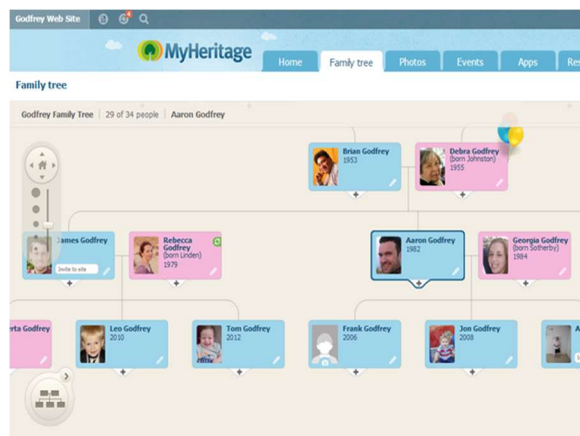


Figure 8: Example of Myheritage [13]

GenoPro is user-friendly comprehensive software that allows one to draw family trees and genograms that can be shared with family [14]. The tree edges encode emotional aspects of relationships well. It allows user to display a node's attribute using color and shape, and the shape represents gender. Meanwhile, the color and the corner of the fill are associated with the disease. Besides, GenoPro also provides user with extra information which is birth

information at the label of each node. Figure 10 shows an example of genogram taken from GenoPro.

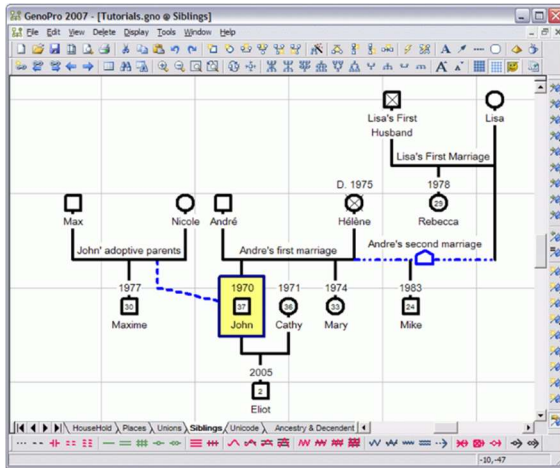


Figure 10: Example of Genogram created using GenoPro [14]

From the examples above, some common problems can be identified. It can be concluded that most software applications create static family trees, without much interaction. Scalability of trees poses problems and temporal or geographical data is often not depicted well, or at all as the software application unable to support exploration of large data set which causes node cluttered and affected user comprehension process. Additionally, searching over tree nodes has not received much attention to overcome the limitation.

It seems that not much can be found in the literature about the graph theoretic properties of genealogical graphs technique [15]. When visualizing genealogical data, it often boils down to the visualization of the tree of ancestors or the tree of descendants of some individual. Usually these trees are combined together to form an hourglass chart. Hourglass charts show only some part of the family, the descendants and ancestors of the central person. Other relationships such as cousins, uncles, aunts etc. are discarded. Moreover, every individual of the ancestor's tree has its own tree of descendants as well as every individual of the descendant's tree has its own tree of ancestors. So this resulted challenging in visualizing all these trees at once.

McGuffin and Balakrishnan [15] made an analysis of genealogical graphs and identified the difficulties of drawing those graphs. So, they proposed a new

visualization approach called the Dual-tree, and made a prototype software to evaluate the proposed approach. Dual-tree was designed to use with genealogical graphs but it can also be applicable in other domains. Figure 11 show example of dual tree visualization.

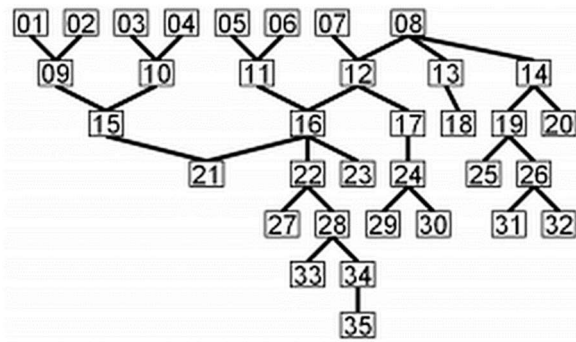


Figure 11: Dual tree genealogy visualization [15]

The Dual-tree supports both layouts: the classical node-link and the indented outline styles. The orientation of the classical node-link style can be chosen from left to right or from top to bottom. The orientation can be chosen depending on how much space is available horizontally and vertically. The area taken by the intended outline tends to be smaller than the area taken by the other two classical node-link styles. It usually does not take much space horizontally. That makes it easier to navigate the tree as the scrolling in only one direction is required in a zoomed-in view.

The Dual-tree technique gives a possibility to visualize big genealogical datasets by showing a subset at a time and allowing to change that subset dynamically. That eventually gives the possibility to explore the whole tree. But it provides no solution proposed to encode temporal information to the Dual-tree except in the form of text labels. The possibility of designing other layout styles to the Dual-tree to visually encode temporal information needs to be explored.

#### 4. GRAPH THEORY OF KINSHIP NETWORK

Genealogies also can be represented as networks in different ways: as Ore-graph, as p-graph, and as bipartite p-graph. Kinship network developed in

British social anthropology is used to describe a social group comprising of parents, parent substitutes and children. It is characterized by an interlocking web of relationships among individuals. Kinship network consists of genealogical data, and this data can be represented as graphs in different ways: as Ore graphs, as p-graphs, or as bipartite p-graphs [16]. In Ore graph of genealogy, every person is represented by a vertex. Meanwhile marriage relationship is represented by edge and the relation between parent and children is represented by arcs pointing from the parents to their children (see Figure 12).

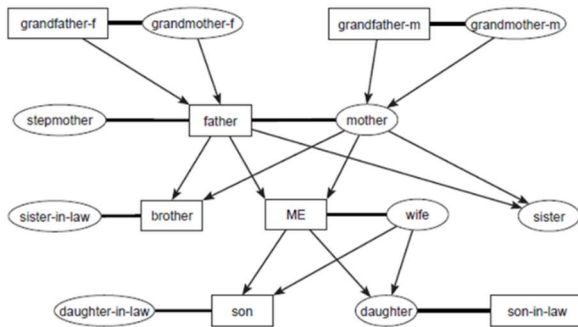


Figure 12: Example of Ore Graph

P-graphs represent couples and uncoupled children as vertices, while parent-child links are the arcs connecting nodes both within and between different nuclear families (see Figure 13). The solid arcs represent the relation with a son of and the dotted arcs represent relation with a daughter of.

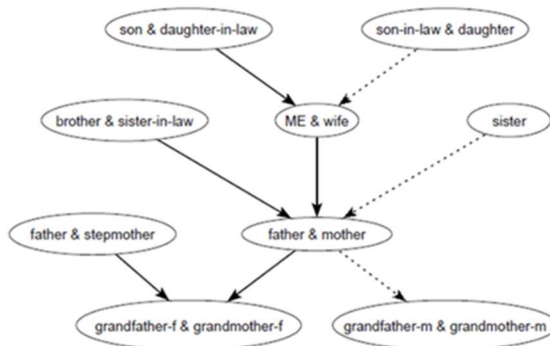


Figure 13: Example of P-Graph

A Bipartite p-graph has two kinds of vertices – vertices representing couples (rectangles) and vertices representing individuals (circles for women and triangles for men). Therefore, each married person is involved in two kinds of vertices (or even

more if he/she is involved in multiple marriages). Arcs again point from children to their parents (see Figure 14).

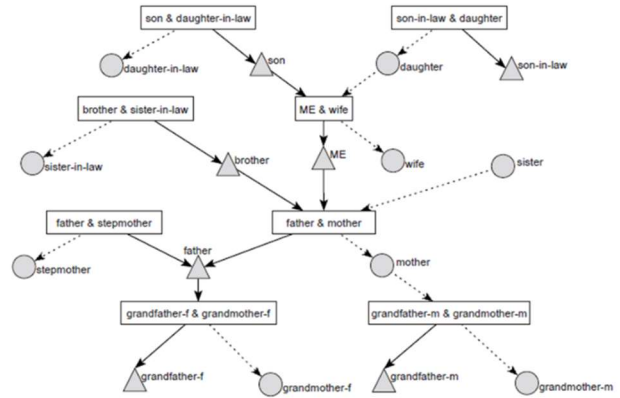


Figure 14: Example of Bipartite P-Graph

P-graphs and bipartite p-graphs have many advantages from ore graphs. This is because P-graphs are directed, acyclic network. So, there are fewer vertices and lines than in corresponding Ore graphs. P-graph is suitable to be used to represent a relinking marriage and more suitable for analyses. Meanwhile, bipartite p-graphs are used to find marriages between half-brothers and half-sisters because of its advantages is in distinguish between a married uncle and a remarriage of a father.

### 5. SOCIAL NETWORK VISUALIZATION

Social Network Visualization is a visual method for supporting characterization, comparison, and classification of large networks. It is most commonly visualized as a graph with individuals as nodes and relations or contacts as edges. Graph theory gives a tools to formally represent social networks as well as quantifying structural properties in the network. Then social network analysis is a methodological approach in the social sciences using graph-theoretic concepts to describe, understand and explain social structure. Furthermore, social network analysis (SNA) has emerged as a powerful method for understanding the importance of relationships in networks. Pajek, Krackplot and Social Action are example of social network analysis application which are been created to handle large network with complex relationship. These techniques are

considered to be some of the major contributors to this area.

Pajek is a powerful tool for analysis and visualization of large networks which having some tens or hundreds of thousands of vertices [17]. It was developed based on experiences gained in development of graph data structure and algorithms libraries (Graph and X-graph), collection of network analysis and visualization programs (STRAN, RelCalc, Draw, Energ) and SGML-based graph description markup language (NetML).

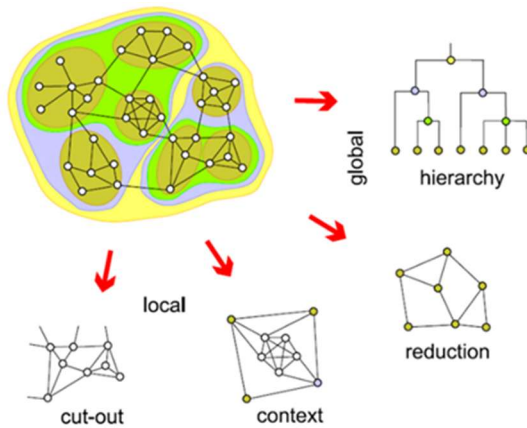


Figure 15: Pajek Approaches to deal with large networks [16]

Pajek approach includes finding clusters (components, neighborhoods of 'important' vertices, cores, etc.) in a network, extracting vertices that belong to the same clusters and showing them separately, possibly with the parts of the context (detailed local view), shrinking vertices in clusters and showing relations among clusters (global view) (see Figure 15). By implementation of a selection of efficient algorithms to deal with large networks, it enables Pajek to support not just ordinary network but also 2-mode networks (bipartite graphs) and temporal networks (dynamic graphs). Furthermore, it provides the option to fix one or more dimensions of the layout space, e.g. by mapping one or more node indices to coordinates.

KrackPlot is a tool for network visualization designed for social network analysts [18]. It allows user to create and delete nodes, move individual nodes or group of nodes and create or delete line. Additionally, user can overview the graph by scroll over large graphs and zoom in on smaller or

subgraph. It is also allowing user to display node attribute visually and the usage of color and/or shape for the node help users in distinguishing nodes with different characteristics such as gender, departmental affiliation or others. There are several algorithm and application that can be integrated with KrackPlot in resulting a better layout, such as Ford-Fulkerson algorithm and UCINET. Figure 16, shows potential flow between a chosen source node in the lower left group and a chosen sink on the right, computed as a max flow with the Ford-Fulkerson algorithm.

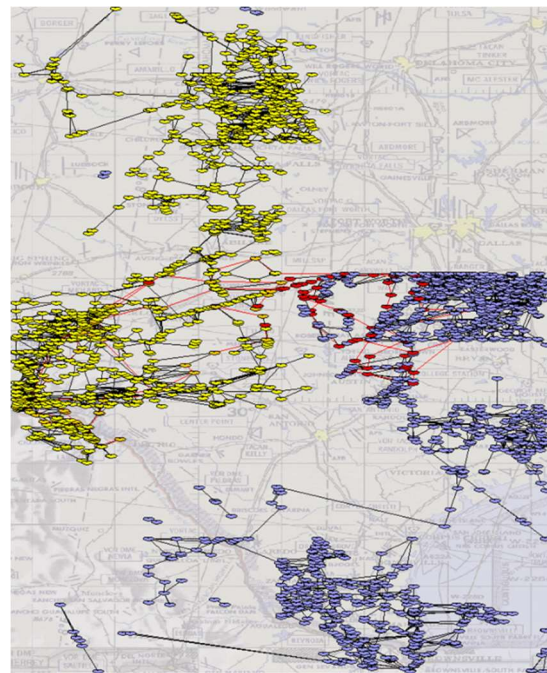


Figure 16: KrackPlot Visualizing flow through a network [18]

Nodes on the flow are shown in red or orange: red means that no more flow could be sent to the node and orange means that flow could potentially be increased. Nodes with no flow are shown in blue or yellow: blue means that no flow could be sent to the node given the current flow. This means that the large yellow section to the left of the blue section highlights a bottleneck in the flow.

**SocialAction** is a tool that integrates visualization and statistics to improve the analytical process [19]. It is inspired by the power of attribute ranking and coordinated views combine principles of information visualization which it allows user to flexibly iterate through visualizations of measures to gain an



overview (order list and Node-link diagram), filter nodes (ranking and color coding), and find outliers (dynamics queries). By using techniques from the social network community, SNA ease user by aggregate networks using link structure, find cohesive subgroups, and focus on communities of interest. Moreover, user can untangle networks by viewing different link types separately, or find patterns across different link. Most importantly SNA approaches follow designing advanced graphical user interfaces guide which is Visual Information-Seeking Mantra “Overview first, zoom and filter, then details-on demand. Figure 17 shows SocialAction’s technique on a subgraph from the global terrorism network.

Rankings		
Betweenness Centrality		
The number of shortest paths between pairs of nodes that pass through a given node.		
Rank	Node	Type
2,516.00	Muslim Militants	Terrorist Group
2,436.50	Corsica (France)	Country
2,413.00	Colombia	Country
2,368.00	Peru	Country
2,280.50	France	Country
2,239.00	Algeria	Country
2,226.00	Rebels	Terrorist Group
2,214.00	GIA (Armed Islamic Group)	Terrorist Group
2,124.00	FARC (Revolutionary Armed For...	Terrorist Group
1,718.00	Bangladesh	Country
1,656.00	U/I	Terrorist Group
1,598.00	India	Country
1,063.00	Pakistan	Country
798.00	Corsican Separatists	Terrorist Group
704.00	FLNC (Corsican National Libera...	Terrorist Group
704.00	Historic Wing FLNC	Terrorist Group
637.00	Indonesia	Country
614.00	Political Activists	Terrorist Group
596.00	Philippines	Country
520.00	Jamaat-E-Islami	Terrorist Group
330.00	Muslim Separatists	Terrorist Group
276.00	ELN (National Liberation Army)	Terrorist Group
187.00	Venezuela	Country
187.00	China	Country
94.00	Egypt	Country
0.00	Dignity for Colombia	Terrorist Group
0.00	Jamie Bateman Canon Front	Terrorist Group
0.00	Sendero Luminoso	Terrorist Group
0.00	Jamiat-ul-Mujahideen	Terrorist Group
0.00	Timorese Students	Terrorist Group

Figure 17: Example of SocialAction technique on a subgraph from the global terrorism network [19]

This network is two-mode, which mean it has two different types of nodes: terrorist groups and countries. In this network, a link exists if a terrorist group committed an attack in that country. In this example, “betweenness” centrality was selected as the ranking criterion. This feature is often used to detect “gatekeepers” between highly connected subgraphs. The nodes are ordered by ranking in Figure 18. In this example, the “Muslim Militants”

group has the highest score. The network visualization, colored according to these rankings, is shown in Figure 18 Groups with high “betweenness” appear red in this network.

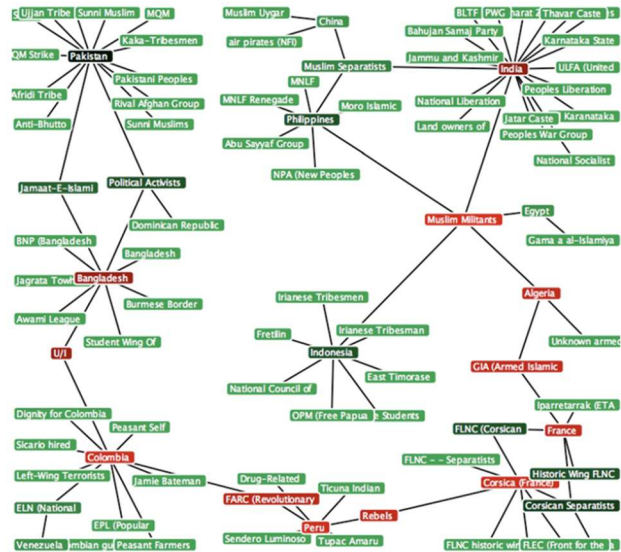


Figure 18: SocialAction example of a subgraph from the global terrorism network [19]

Social network visualization gives a possibility to visualize big genealogical datasets by aggregate networks using link structure, find cohesive subgroups, and focus on communities of interest. That eventually gives the possibility to explore the whole tree. The proposed layout meets the requirements of most of the standard genealogy visualization tasks. Two mode network will allow the node (person) to have more than one edges (relationship) which is the solution to encode temporal information as node attribute and edges. The possibility of designing network layout to visually encode temporal information needs to be explored. In the next section the proposed solution will be describe according to the problem.

## 6. DISCUSSIONS AND CONCLUSIONS

From the literature study, one of the main problems that family tree visualization’s face is the scalability problem which leads to visual clutter and affect user understandability on the data that has been visualized [20]. Besides, the image-map format also would be the other issue which distracts user exploration tool experience that slow down process of understanding medical family tree data [15].

Most of the existing family tree data visualization tools only focuses on visualizing family relationship which is believed to be the most important data in genealogy research. So, there are many layout approach introduced by the researcher, for example TreePlus, Treemap and Hyperbolic tree to solve scalability issue. This approach has successfully worked, but it only works when applied to a relatively small dataset. However, when the size of a dataset increases, the visualization task with the existing tools turns to be a difficult.

The available genealogy information becomes more precise and includes many additional details besides family relationships. The additional detail is called temporal information which plays an important role for genealogy researchers since multifactorial inheritance disorders caused by a combination of environmental factors and mutations in multiple genes [21]. The temporal information such as gender, age and disease, gives additional value for the genealogy specialists and helps to make more precise analysis. So, in order to get full advantage of that information, it needs to be visually encoded as well. Then, GenoPro which use genogram are used to visually encoded all the information needed. Genogram is the most widely used tool in the medical field for visualizing genealogy data because it can visualize all recorded data in just one diagram. But, the image map format results in the same scalability problem when applied to big genealogy datasets [15]. Moreover, it encodes temporal information in the form of symbol which is required sometime to learn about the symbol in order to have the same interpretation with the researcher. Having a big number of individuals in the family will eventually lead to long edges, edge crossings and to exponential crowding.

One of the other existing solutions is to show only a portion of a family at a time instead of showing the whole family. But, how to choose the right part of the family and still get the maximum information from it. Some visualization tools show only the tree of ancestors, some can show both the tree of ancestors and the tree of descendants. Sometimes it is not enough, and information about a more extended family is required. The dual-tree solution that was discussed earlier, proposed a new way of

choosing the subset of a family to visualize. By using the interaction technique, it allows the process of exploring the whole family node by dynamically changing the subset. Hereby, user can dynamically choose the focal person and the ancestor whose descendant tree is displayed. By choosing the offset between these individuals, the chosen subset represents the family in greater or less details. In that way, it handles the scalability problem of the genealogy data visualization. But dual tree fails to properly encode the temporal information which is important in visualizing medical family tree.

The social network proposed a way of reducing scalability problem by characterization, comparison, and classification of large networks. Moreover, the proposed layout avoids another genealogy data visualization problem which is edge crossings. The visualization technique used allows exploring the whole family by dynamically changing the subset using zoom technique. User is also able to filter the node according to the interest since users' perceptual clarity improves when the number of visualized elements is limited [19]. The temporal information can also be encoded by using two mode network visualization and node attribute method which a solution for dual tree limitation.

The data type is a key issue in visualization. A suitable visualization type had to be applied for specific data type. This paper review several work consists of combination of multiple layouts in order to represent medical family tree data such as node-link diagram and. There are a lot more visualization types that can be explored. Further study should look into other types of visualization that deal with different type of data that have inherent relation among the data elements which can improve display strategies of family generations' structure.

As a conclusion, the social network visual method seems to be the most suitable approach to visualize family tree data as it reduces scalability problem, improves user interaction and helps user to understand family tree data better. Furthermore, social network analysis (SNA) has emerged as a powerful method for understanding the importance of relationships in networks which will improve the tool usability.

**ACKNOWLEDGMENTS**

This work was supported by Universiti Sains Malaysia under RUI Grant Scheme (1001/PKOMP/8012206).

**REFERENCES:**

- [1]. Nam WK, Stuart KC, and Jeffrey H. (2010). Tracing genealogical data with TimeNets. In Proceedings of the International Conference on Advanced Visual Interfaces, pp. 241–248. ACM
- [2]. Lee, B., Parr, C. S., Plaisant, C., Bederson, B. B., Vekslar, V. D., Gray, W. D., & Kotfila, C. (2006). Treeplus: Interactive exploration of networks with enhanced tree layouts. *IEEE Transactions on Visualization and Computer Graphics*, 12(6), pp. 1414-1426.
- [3]. Shneiderman B (1992). Tree visualization with tree-maps: A 2-D space-filling approach. *ACM Transactions on Graphics* 11(1), pp. 92-99.
- [4]. Lamping J & Rao R (1995). The hyperbolic browser: A focus + context technique for visualizing large hierarchies. *Journal of Visual Languages and Computing* 7(1), pp 33-55.
- [5]. Gill S.S. (2003) Evolutionary Computing For Cancer Genetics Risk Assessment, MSc IMM dissertation, MACS, Heriot Watt University.
- [6]. McGoldrick, M., Gerson, R., & Petry, S. S. (2008). *Genograms: Assessment and Intervention* (3rd ed.). New York: W.W. Norton & Co.
- [7]. Kennedy V, (2010), *Genograms*, Mai Review, (3) pp. 1 - 12
- [8]. Berthold, M., Dill, F., Kötter, T., & Thiel, K. (2008). Supporting creativity: Towards associative discovery of new insights. *Advances in Knowledge Discovery and Data Mining*, 14-25
- [9]. Bokhare, S.F., Zainon, W.M.N.W. & Talib, A.Z., (2014) A Study on Visual Representation of Medical Family Tree Using Genograms, *Advances in Intelligent Systems Research*, , pp. 1 - 6.
- [10]. Genetics Home Reference (2018). Inheriting Genetic Conditions Available from: <https://ghr.nlm.nih.gov>
- [11]. Hernandez LM, Blazer DG, (2006). *Genes, Behavior, and the Social Environment: Moving Beyond the Nature/Nurture Debate*. Washington (DC): National Academies Press (US)
- [12]. Personal Ancestral File software, available from: <https://www.familysearch.org/paf> (2016)
- [13]. MyHeritage Software, available from: <https://www.myheritage.com/> (2017)
- [14]. Genepro-Geneology Software, available from: <http://www.genopro.com/> (2017).
- [15]. McGuffin MJ and Balakrishnan R.( 2005). Interactive visualization of genealogical graphs. In Proceedings of IEEE Symposium on Information Visualization, INFOVIS, pages 16–23.
- [16]. Batagelj, V., Mrvar, A. (2002) Pajek — Analysis and Visualization of Large Networks. In: P. Mutzel, M. Jünger, S. Leipert (eds.) *Graph Drawing '01, Lecture Notes in Computer Science 2265*, Springer-Verlag, 477–478
- [17]. Batagelj, V., Mrvar, A (1998): Pajek-A Program for Large Network Analysis. *Connections* 21 (2), 47–57.
- [18]. Krackhardt, David, (1994). Graph theoretical dimensions of informal organization. In Kathleen Carley and Michael Prietula, editors, *Computational Organization Theory*. Lawrence Erlbaum Associates, Inc., eds
- [19]. Perer, A., & Shneiderman, B. (2006). Balancing Systematic and Flexible Exploration of Social Networks, 12(5).
- [20]. Bokhare, S.F., Zainon, W.M.N.W. (2015), Visual Information Framework for Medical Family Tree Data (Genogram). *Advances in Visual Informatics*, H. Badio Zaman et al. (Eds.), LNCS 9429, pp. 3-13.
- [21]. Stöppler, M.C. (2014), *Genetic Diseases* Available from: [http://www.medicinenet.com/genetic\\_disease/article.htm](http://www.medicinenet.com/genetic_disease/article.htm) (2016)
- [22]. Chi, E. H. H. (2000). A taxonomy of visualization techniques using the data state reference model. In *IEEE Symposium on Information Visualization, 2000. InfoVis 2000*. pp. 69-75.
- [23]. Gratzl, S., Lex, A., Gehlenborg, N., Pfister, H., & Streit, M. (2013). Lineup: Visual analysis of multi-attribute rankings. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12), 2277–2286
- [24]. Keim DA, Mansmann F, Schneidewind J, Thomas J, & Ziegler H (2008). Visual analytics: Scope and challenges, *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics*. LNCS 4404, pp.76–90.
- [25]. Von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J. J., Fekete, J. D., & Fellner, D. W. (2011). Visual analysis of large graphs: state-of-the-art and future research challenges. In *Computer Graphics Forum*, Vol. 30, No. 6, pp. 1719-1749.
- [26]. Pendagast, E. G., and Sherman, C. O. (1976).

- A guide to the genogram family systems training, *The Family*, vol. 5, No. 1
- [27]. Gershon, N. D. (1994). From perception to visualization. *Scientific Visualization*, 129-139.
- [28]. Hadlak, S., Tominski, C., Schulz, H. J., & Schumann, H. (2010). Visualization of hierarchies in space and time. In *Workshop on Geospatial Visual Analytics: Focus on Time at the AGILE International Conference on Geographic Information Science*.
- [29]. Craft, B., & Cairns, P. (2005). Beyond guidelines: what can we learn from the visual information seeking mantra? In *Proceedings. Ninth International Conference on Information Visualisation*, pp. 110–118.
- [30]. Abello, J., Van Ham, F., & Krishnan, N. (2006). Ask-graphview: A large scale graph visualization system. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 669-676
- [31]. Herman I, Melancon G & Marshall MS (2000). Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics* 6(1), pp. 24-43.