A Family-Centric Genealogy Visualization Paradigm

Robert Ball and **David Cook**
Stephen F. Austin State University
Department of Computer Science
{ballrg | cookda}@sfasu.edu

ABSTRACT
Genealogy is a way to tell a story. Stories tell us who we are and where we came from, and also serve as a foundation for future behavior. Genealogical visualizations are used to help understand and share the story of the past with others. Modern research for genealogy emphasizes the stories of families and communities. Most genealogical visualizations are focused on the individual and have a number of limitations, including not giving accurate stories of families and timelines. This paper presents a new visualization paradigm that focuses on families and communities. It is unique in that it shows all the family relationships of individuals while at the same time putting the people into a correct historical and temporal context. Our new paradigm allows people to understand their stories of their families and the communities that they came from in order to create a whole story with context and details.

Author Keywords
Genealogy, family history, information visualization

INTRODUCTION

Genealogy is as much about pedigree charts, birth dates, death certificates, and marriage dates as computers are about zeroes and ones.

Genealogy is about people, their lives, and about how their stories shape who we are. We present a new visualization paradigm about making familiar units the central part of the story.

First, what is genealogy and genealogical data?

According to Webster's dictionary [9]: Genealogy (noun):

1. The study of family history.

2. The history of a particular family showing how the different members of the family are related to each other.

Therefore, genealogical data is data that one gathers from the study of family history trying to piece together how different people are related. In essence it is history, or as explained in another way, it is a narrative of a family [2].

In one sense, it is how the entire human race is related. However, in practice, genealogy is usually a highly personal, individualized history of one's ancestors (and sometimes descendants) [10].

Genealogical visualization is a way to see – visualize – the data in a way that helps the user understand the geological data. Its purpose it to provide insight into the data. However, current visualization paradigms focus primarily on an individual's past and often leave out important context.

Current genealogical visualizations, such as pedigree charts and fan charts, fail to adequately describe complex family situations. For example, important historical personages such as King Henry VIII, King of England (who had six marriages) or King David from the Bible (seven concurrent wives and an unknown number of concubines) can only be shown with one wife/concubine at a time.

Imagine being present in King David's court and trying to understand all the family connections of the people present. For example, one might see several fathers-in-law, cousins, a number of half-brothers (i.e. same father, different mother), uncles, aunts, etc. How are the different people related to each? By understanding the different relationships that people would have had helps better understand the narrative of the people, thus creating better understanding of the genealogical data and the people that it represents.

Our research question is as follows: **What is an efficient visualization technique that shows the families and relations of a particular person in regard to temporal data? In other words, how can we see the contemporaries and *all* possible family dynamics of a person?**

In this paper we first explain the psychological and social reasons for doing genealogy and history research. We then briefly review a number of more popular genealogical visualizations. Finally, we present a new paradigm that incorporates families and communities into a new genealogical visualization paradigm.

WHY DO GENEALOGY RESEARCH?

Genealogy is a way of writing history. Genealogy helps create a feeling of belonging and an explanation of who you are and where you came from as an individual [10].

Other academic research concurs that genealogical research is often performed for finding out the answers to who we are and where we came from. It is argued that in modern life people that perform genealogical research do it because they "find themselves living in a 'new country', separated

from the 'old', the question of *where* ones comes from is not always so easily answered." In other words, genealogy provides personal meaning to individuals [1].

Although health history, human biology evolution, the human genome project, etc. are often something that news media pick up on as being important in terms of family history research, Anne-Marie Kramer did a thorough study of how and why people perform genealogy research from 2000 to 2008 and found that for most people health history is not the primary reason for doing genealogical research. She states, "the past is deployed here to shape respondents' relationships to themselves in the present, as well as to shape dispositions towards their future" [7].

We find that genealogists are storytellers, historians, and detectives: finding people that are not found, uncovering that one's ancestors were thieves, common people, farmers, nobility, and the like are all very important and interesting in shaping the view of where one came from. Common words that describe genealogists are the following:

- Storyteller
- Narrator
- Detective
- Historian

Understanding who we are and where we came from also influences our social behavior. Recently, a news story that went viral and spread very quickly throughout media was about a white supremacist that found through a DNA test that he was 14% African [5]. Understandably, the white supremacist was shocked.

Karla Hackstaff suggests that genealogy is used to link personal family histories to social-historical contexts. She suggests that individuals construct current identities based on genealogical research by looking at how their family's past can be seen within "social memories." This is turn contributes to diverse stories from new standpoints in history [3].

In other words, where individuals and families form identities and cultures based on their genealogy, society also forms an identity based on the narratives and stories that we tell ourselves.

This leads to Arnon Herskovitz's suggestion that genealogy as an academic discipline is focused on the following: people, families, communities, representations, and data [4].

"We are all storytellers, and we are the stories we tell. … adolescents and young adults in modern societies are challenged to formulate meaningful answers to the twin identity questions: Who am I? How do I fit into the adult world?" [8]. In other words, "finding oneself" is nothing more than creating a narrative.

In summary, ***genealogy research is performed in order to understand the identity (e.g. Who am I? Where did I come from?) of the person, family, and/or community involved.***

WHAT IS THE PURPOSE OF VISUALIZING GENEALOGICAL DATA?

One thing that genealogical research papers indicate is that family trees are NOT just graphs of interconnected people. Indeed, it appears that showing the pedigree (e.g. the ancestry) of an individual is just one of the many facets of genealogical research.

So, instead of answering what the purpose of visualizing genealogical data is directly, we will address it by asking instead, when do we ***not*** need to visualize genealogy data?

Since one of the main reasons for doing genealogical research in the first place is to find a self identity – answering the questions of "Who am I?" and "Where did I come from?" – someone that already knows who they are does not need to further investigate the issue. Someone who already understands the data so well that visualizing it does not add any additional insight has no need to see it. They understand the data already; they already have an identity.

However, if that person would like to tell the story or show proof that they are correct in their assertions about their family tree to other people then a visualization is important. It becomes important because the visualization becomes a conduit of communication.

In addition, if one does not yet know the answers to the questions of self-identity, then seeing visual output helps one comprehend self-identity faster than examining raw data.

To summarize Colin Ware's *Information Visualization* book [12], the following are reasons why people use visualizations:

1. Visual displays provide the highest bandwidth channel from computer to the human.
2. Visualization provides an ability to comprehend huge amounts data.
3. Visualization allows the perception of emergent properties and patterns that were not anticipated.
4. Visualization often enables problems with the data itself to become immediately apparent.

To paraphrase, visualizations help people receive and understand data faster than any other medium such as reading text, sound, or tactile input.

GENEALOGICAL VISUALIZATION PARADIGMS

Enumerating the many different types of genealogical visualizations is beyond the scope of this paper. However, to be brief, the reader will most likely be familiar with the most common types such as pedigree charts and fan charts.

Pedigree Charts

Figure I shows historically the most common visualization, a pedigree chart. In essence, it is a binary tree that shows the ancestry of a particular person.

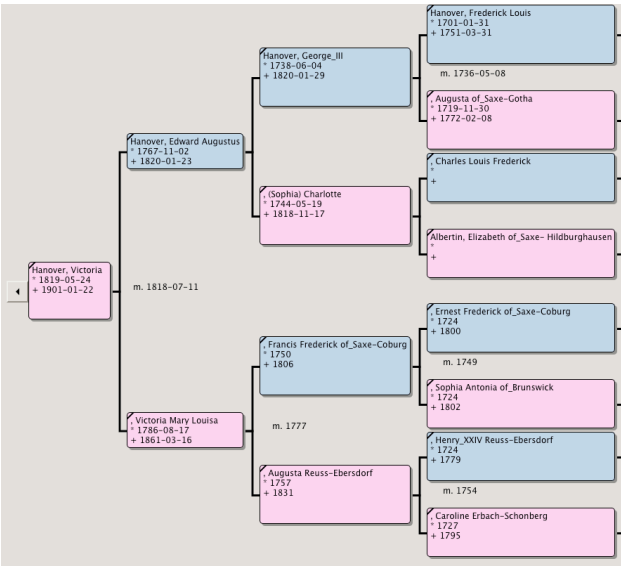

Figure I Pedigree chart visualization of Queen Victoria's ancestral line. A pedigree chart is a traditional visualization tool that shows a little bit of information (e.g. birth, death, and marriage dates) along with showing the structure of ancestors. This particular visualization also has a dual encoding of color to indicate gender.

A pedigree chart is a form of node-and-link visualization that borrows from graph theory. It adequately shows the direct ancestors of the focus person (the person to the farthest left in Figure I).

However it has a number of shortcomings. In terms of this paper, the largest problem is that it does not tell the whole story. In addition, with only 4 generations shown, as in Figure I, the visualization is limited to showing only 15 individuals and 7 marriages. Even with pedigrees that show 5 generations, only 31 individuals can be shown with only 15 marriages shown.

In addition, pedigree charts lack additional information. For instance, there is nothing to show how many marriages an individual had, how many other children besides those shown, etc.

In Figure I the focus person is Queen Victoria. How many children did Queen Victoria have? How many marriages? Who were her contemporaries? How many grandchildren did she have? How many times were her children remarried?

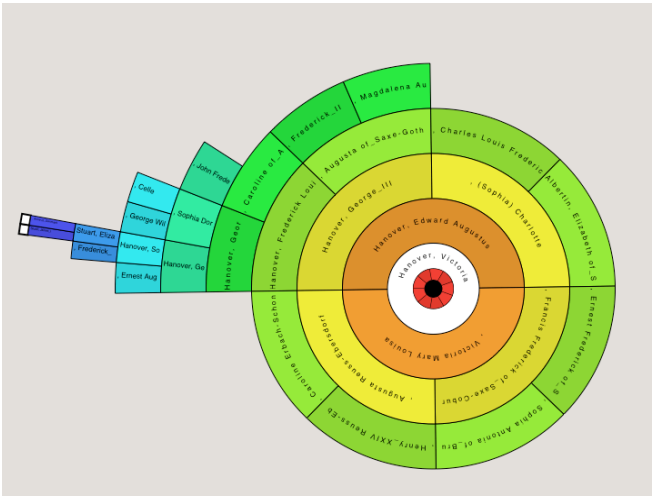

Figure II Fan chart visualization of Queen Victoria's ancestors. Note that it is easy to see where "holes" are in the data. For example, it is easy to see that along the royal line (the line that extends to the left) is where the most information lies and that most of the other family lines have been ignored.

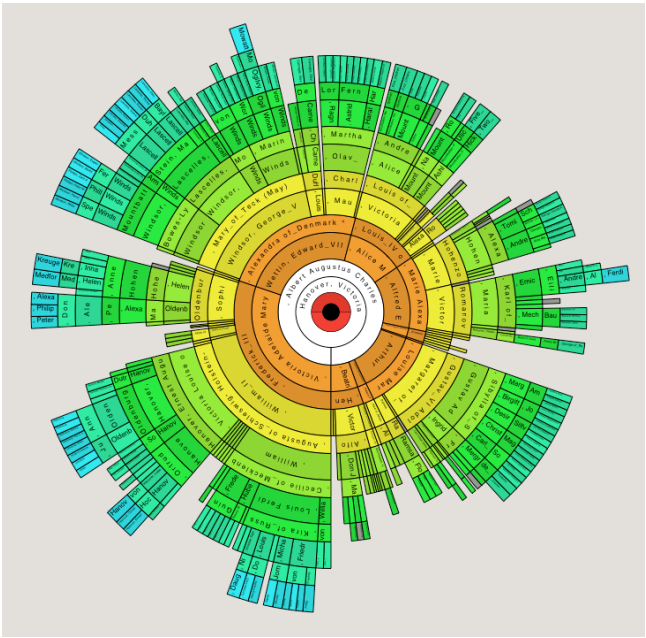

Figure III Fan chart visualization of Queen Victoria's ancestors.

Fan Charts

Another common type of genealogical visualization is the fan chart. Figure II and Figure III both show Queen Victoria as the focus. Figure II shows all of Queen Victoria's direct ancestors in the data file. Figure III shows all of Queen Victoria's direct descendants in the data file.

Fan charts allow a much larger number of ancestors/descendants to be shown than pedigree charts. They are excellent in showing holes in the data – where data is missing from the data file – and can show large

amounts of data. However, they lack in that there is no sense of time reference or much detail at all about any particular individual.

Other common visualizations include the following: radial, node and link, treemaps, hierarchy (e.g. organization charts), H-tree, hyperbolic trees, balloon view, bipartite graphs, quilts, event, and temporal visualizations.

Generational vs. Temporal Perspective

One of the biggest drawbacks of fan charts and pedigree charts is that they are both based on the idea of generations. A generation is a word that designates a parent to child relationship. For example, Figure III shows up to twelve generations of Queen Victoria's ancestry where every generation has a parent relationship to the previous generation.

However, not all the people in a given generation in the pedigree and fan charts are contemporaries. Many of the people of the same generation were much older/younger than others of the same generation. In fact, some of the people in her ancestry that are the same "generation" away from Queen Victoria were old enough to be the grandparents of people in the same generation.

For instance, suppose an individual has twin daughters. The first twin daughter has her own child at the age of fifteen and the second twin daughter has her own child at the age of forty. Although both grandchildren born will be two generations away from their grandmother, there would be twenty-five years difference in the age!

The two grandchildren would not grow up in the same era. It is doubtful that they would listen to the same music, know all the same people, etc. Someone born in 1960 will have a different perspective on life than someone that was born in 1985.

On the other hand, a temporal visualization is one that shows correctly where people fit in terms of time. It does not show generations well, but it does help understand more about individuals' lives by understanding when they were born and grew up.

TEMPORAL FAMILY-CENTRIC VISUALIZATION PARADIGM

We introduce a new way of looking at genealogy visualizations: instead of using the generational individual ancestor/descendant perspective, we introduce a temporal family-centric perspective. Our visualization paradigm focuses on family units instead of individuals and shows them in the correct time frame when they lived.

For example, to fully understand King David's life, one needs to fully understand the people and the places of importance that existed in his life.

- Who was related to whom?
- Were they alive at the same time?
- If they were, would they have known each other? (For example, a person might have been born two years before another person died, but that does not mean that they knew each other.)
- What kind of relationship might have existed?

In essence, our visualization paradigm answers the following question: If one were to go to a family reunion, how are all the living relatives related?

Prototype

We created a genealogical visualization prototype that uses JOGL (Java Bindings for the OpenGL API). JOGL is a set of Java classes that act as a wrapper around the C OpenGL libraries allowing full use of OpenGL and hardware acceleration with Java [6].

The prototype is based on two things:

1. Temporal order of events.
2. Family perspective instead of an individual perspective.

Figure IV shows an overview of the different families that Henry VIII (King of England during the 1500's) was involved in. The different boxes show the families and are ordered along the Y-axis to depict time. The left-most box shows the family that Henry VIII was born into. (Henry VIII is the focus individual in Figure IV and is depicted as yellow). The other six boxes depict the six marriages that he was part of.

Figure IV. View of the prototype depicting Henry VIII, King of England.

Figure IV shows the prototype with Henry VIII as the focus point. Having a focus point is standard in genealogical visualizations as it gives the user someone to concentrate their attention on. For example, Figure I, a pedigree chart, and Figure II, a fan chart, both rely heavily on a focus point (Queen Victoria in those figures).

Following cultural precedent and tradition, males are shown as blue and females shown as pink with the exception that the focus person is shown in yellow.

On the right side of the prototype in Figure IV are various filtering and statistical information about what is shown. They are standard to most genealogical visualizations and do not deserve further explanation.

Family Units

Figure V and Figure VI explain the general meaning of the family units. Each horizontal line represents an event – with one exception. (The exception is that parents are depicted on top of a family unit, right above the marriage line.)

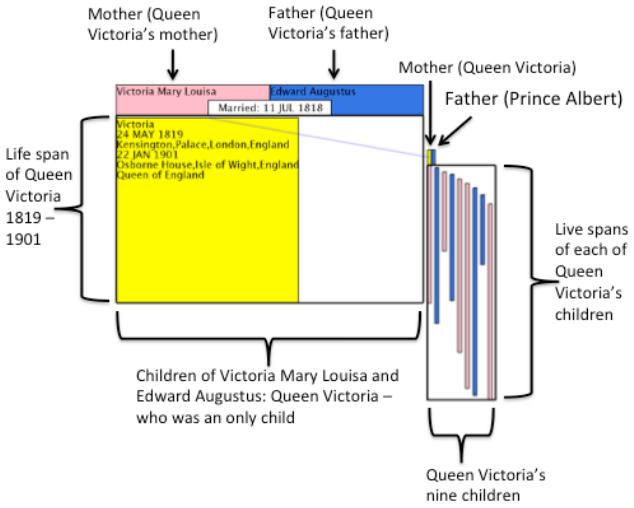

Figure V. Explanation of family units. This screen capture shows the two families that Queen Victoria was associated with.

For example, Figure V and Figure VI show the family unit into which Queen Victoria was born, with Victoria Mary Louisa as mother and Edward Augustus as father (most royalty did not have surnames). Inside the box shows their only child – Queen Victoria – and accurately shows her life span from 1819 – 1901.

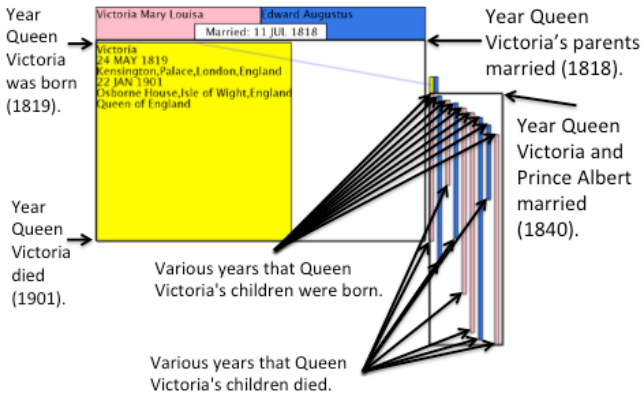

Figure VI. Each horizontal line depicts an event – a marriage, birth, or death event.

The family unit to the right shows Queen Victoria's only marriage. It accurately depicts that she got married in 1840. It also shows the life spans of her nine children – when they were born and died – and places the events at the correct corresponding locations temporally along the Y-axis.

Overview and Detail

For each family unit there are three levels of detail:

- Overview of family
- Greater family detail
- Greater individual detail

Each family unit starts out in overview mode, as depicted in Figure VII.a.

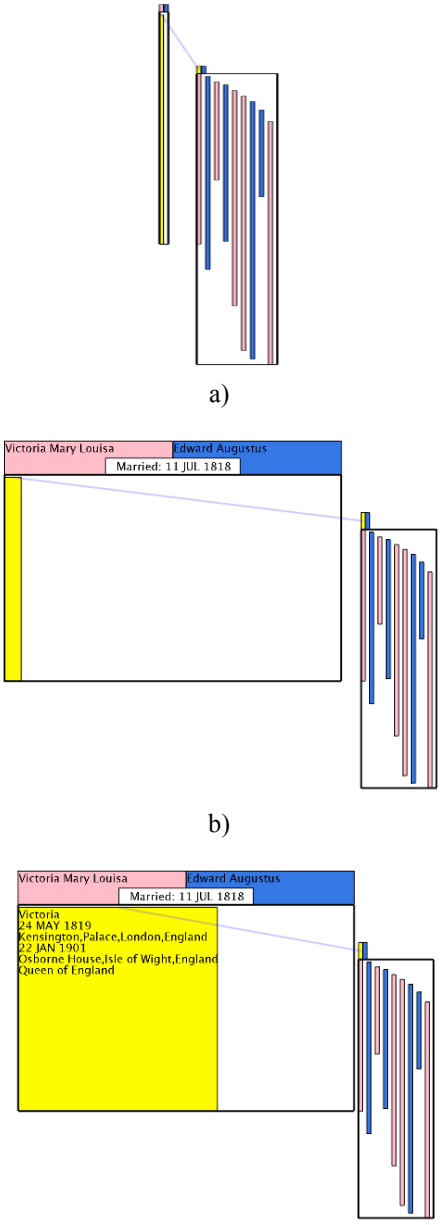

Figure VII. Three levels of detail for each family unit: overview, family detail, and individual detail.

In all three cases the y-axis, which depicts time, is constant. For example, Figure VII.a shows an overview of Queen Victoria's two family units. Figure VII.a shows two families – the family that Queen Victoria grew up in as an only child (left-most box) and the family where she was the wife (right-most box) with nine children.

By double-clicking on the individual/family unit more detail is shown. Figure VII.b shows the left family unit with greater detail and Figure VII.c shows the left family unit with Queen Victoria's individual detail. With family units with more than one child, each child can either be in overview or detail mode.

User Interaction

The prototype is interactive in at least two ways:

- Every individual and family unit has tooltips that show all available data about the individual/family unit.
- Each family unit can be moved by the mouse along the X-axis. This allows the user to move family units around for better comparisons and understanding.

Figure VIII and Figure IX show screenshots of tooltips. Figure VIII shows the tooltip for Henry VIII as an individual and Figure IX shows the tooltip for the entire family into which Henry VIII was born.

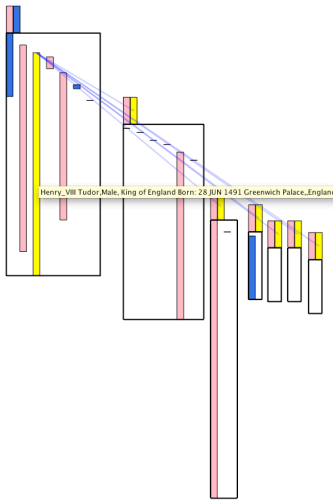

Figure VIII Screenshot of a tooltip showing Henry VIII's individual information.

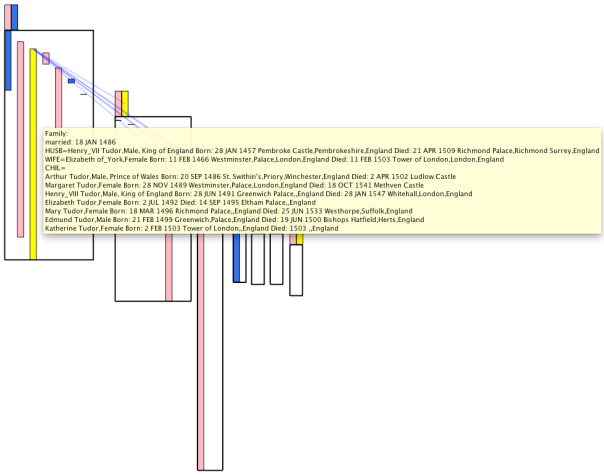

Figure IX Screenshot of a tooltip showing Henry VIII's family into which he was born.

Multiple Marriages

One of the things that pedigree charts and fan charts lack is the ability to see multiple marriages simultaneously. Whether by reasons of divorce or plural marriage, showing all the marriages of an individual helps to understand the story of the person.

This is especially important in understanding King Henry VIII's life. To be brief, King Henry VIII was not happy with his first wife because she kept having stillborn babies.

The stillborn babies can be seen in Figure VIII by looking at King Henry VIII's first and second marriage – the second and third family units – and seeing what appear to be lines instead of long rectangles. The lines show that the babies lived for a short time (usually one day). In addition, the only baby to survive was a girl (who would later by known as Queen Mary or Bloody Mary).

As a result King Henry VIII divorced his wife and remarried. His second marriage resulted in a single girl to reach adulthood: Queen Elizabeth I.

King Henry VIII finally got his only boy in his third marriage: King Edward VI. Unfortunately, Edward only lived to be fifteen. Edward named his cousin as successor, but his half sister – Mary, from Henry's first marriage – had the successor killed and claimed the throne herself. After Mary died, Queen Elizabeth became the next in line, never married, being known as the Virgin Queen, and ended the Tudor line to the throne.

This story of King Henry VIII and his children comes alive with our prototype, but is very difficult to visualize or see in a pedigree or fan chart.

Events

In order to better understand the narrative of people in the past it is important to show corresponding events that occurred in their lives. For example, knowing that an ancestor was born in Germany and died in the United States might be considered mildly interesting. Knowing that the ancestor was born in 1926 and died in 1994 might not add any understanding. Further knowing that the ancestor's religion was Jewish might still be mundane.

However, what if two additional events were shown about the ancestor? If the first event were that he/she migrated to the United States in 1940 and the second event showed when World War II started and ended, one might be drawn to a fascinating conclusion that the ancestor, being Jewish, most likely fled from Germany to avoid the holocaust.

Figure X and Figure XI show the coronation events of monarchs during Queen Victoria's life. One can see that shortly after Victoria was born, George IV was crowned in 1821, was followed by William IV ten years later, then Victoria herself was crowned seven years later in 1838, being only 19 years old.

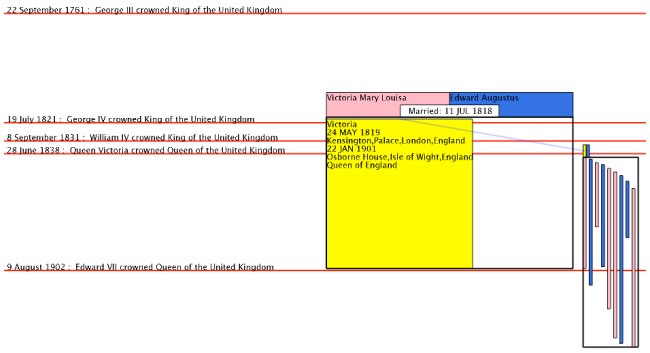

Figure X Screenshot of Queen Victoria's life with events.

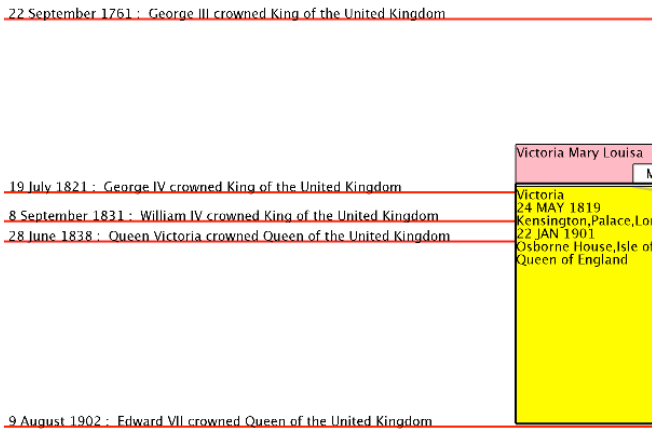

Figure XI. Magnification of Figure X to better show event details.

Showing temporally co-existing events increases the narrative and understanding of the stories behind each of our ancestors and helps understand how they lived.

Estimations and Unknowns

Due to the temporal nature of the visualization, if all dates are not known then estimates are needed. All estimated data is depicted in purple. Through a number of heuristics based on all known data available in the data file, estimates are automatically made to better show in time when people lived.

For example, Figure XII shows the royal family line from Queen Victoria to the current line of succession. It shows Queen Victoria's family into which she was born, her family, King Edward VII's family, King George V's family, King George VI's family, Elizabeth II's (the current queen) family, and Prince Charles's family (the next in line to be king).

Figure XII shows eight people ending in purple. These people do not have death dates and so their deaths are estimated. In this case they are still alive.

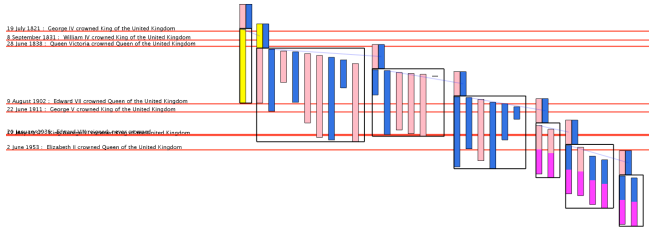

Figure XII The royal families from Queen Victoria (born 1819) to the present. The purple bottom of individuals for the last three family units shows that the death information is missing. In this case, these individuals are still alive.

Comparison of non-related families / Community Perspective

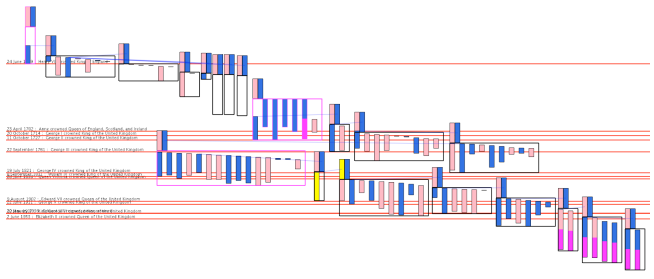

Figure XIII Showing the Tudor family from Henry VII to Elizabeth I in comparison to the Windsor family from Ernest August of Brunswich (King George I's father) to Prince Charles.

In addition to showing related events the prototype also can show non-related families. Figure XIII shows the Tudor family from Henry VII to Elizabeth I while also showing

the Windsor family (the family that has maintained the throne since the coronation of George I in 1714).

By allowing non-related people to be visualized simultaneously allows what Arnon Herskovitz explained as another purpose of genealogy: to understand communities [4].

For example, given a specific year in a village in the Middle Ages in Europe all the families in the village could be visualized. This would allow a user the opportunity to see which villagers might have known each other as contemporaries. It would also show who married whom, when, and just how many of the villagers were related to each other and how.

During plagues and famine the user could understand with a glance that many families had many children, but few of them survived to adulthood.

For example, in Figure XIII shows that Henry VIII married six times, but only a few of his children reached adulthood. One can easily compare his children to the children of the Windsor families and find that the Windsor families generally had most of their children reach adulthood with a few exceptions.

Node & Link Weakness

Probably the greatest weakness of our visualization paradigm is that it a node and link visualization. In other words, it has lines that connect different families together. Although each family unit is separate, an individual might exist in a number of different families. This causes line crossovers and can add to confusion.

One area in graph theory leads to the possibility of reduction of line crossings on a planar surface. Unfortunately, given *n* nodes (family units) and *m* edges (relationships between family units), finding the best way to present the genealogy on a 2D planar surface is a nondeterministic polynomial time (NP-complete) problem [11]. The field that this refers to is called *computational complexity theory.* In general, solutions in computer science that can be solved in polynomial time are considered "reasonable" problems, in terms of hardness. Some problems, however, cannot be solved in polynomial time. However, it might be possible to verify a solution for correctness in polynomial time. These specific types of problems are referred to as NP-complete

Based on computational complexity, it is fairly easy to see that the problem is NP: for every edge *m* in the genealogy one can guess all crossings involved with *m* and their order. Unfortunately, similar to the well-known (to computer scientists, at least) *Traveling Salesman Problem,* it is necessary to go through all possibilities before one is completely sure of finding the optimal order of nodes (family units) to reduce the edges (the relationships of the family units). See [11] for a more thorough proof on the topic.

In other words, presenting our visualization with an automatic layout such that there exist the guaranteed fewest amounts of lines crossing each other in a given layout is very, very difficult to do quickly. It is relatively easy to see if our chosen solution is "good" – but it is very difficult to come up with the "good" solution prior to checking it.

However, in practice, we have found that when using a genealogical visualization such as ours, that people tend to be focused on a few select individuals at a time and that seeing thousands of families visualized at once is used more for the "Wow! I am related to a lot of people!" moments rather than trying to understand the stories of people.

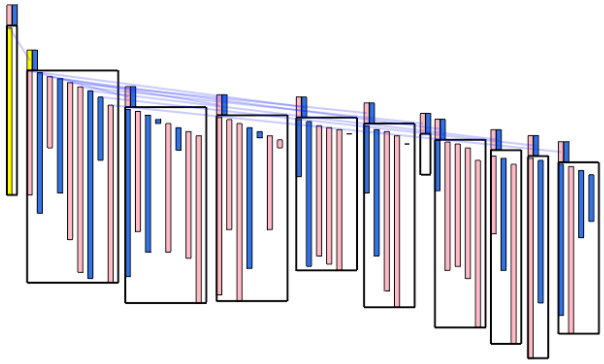

Figure XIV Screenshot showing Queen Victoria's families and all the marriages of all of Queen Victoria's children, which shows an example of how the many lines can be confusing.

Figure XIV shows a screenshot of all the marriages of all of Queen Victoria's children. The lines can be confusing without any interaction. However, in practice, the user can quickly see more information about each family unit and individual with the aid of tooltips.

Scalability Weakness

For any visualization, scalability is an issue. Although our visualization scales fairly well in terms of direct descendants or comparing people of different times, viewing hundreds or thousands of families that were alive at the same time becomes problematic. With the scalability feature of OpenGL thousands of different family units become dots with lines between them and have little general meaning.

However, one of things that make scalability less of an issue is that our visualization paradigm is user driven: the user only sees lots of families because she went through the effort to expand the family relationships she was interested in.

For viewing large families (one of our driving research questions) the visualization scales quite well. For example, Figure XIV shows Queen Victoria's families and all of her children's families. In other words, it shows Queen Victoria as an only child, her one marriage, her children's marriages, all of her grandchildren, and how long everyone lived.

As another example, Figure XIII shows 20 families and 113 individuals all with correct or estimated marriage dates, birth and death dates as well as lines that show how the different people were related.

SUMMARY

Genealogy is as concerned about people's birth dates, death dates, and marriages dates as computers are concerned with zeroes and ones. Raw dates and facts are the backbone of genealogy, but they are only recorded for the sake of understanding people.

We present a novel family-centric visualization paradigm that shows when people lived during the lives of other people. Its primary purpose is to show the family relationships to better understand a complete story of people's lives.

In this paper we presented a new family-centric visualization paradigm. It encodes more information about individuals and their families than traditional pedigree and fan charts. Our family-centric visualization paradigm can be summarized by the following:

- Is temporal based,
- Shows the life span of individuals,
- Shows all the families they were involved in (the families into which they were born and all of their own marriages),
- Shows external events (e.g. wars, economic collapses, or other pertinent information), and
- Can compare non-related families

REFERENCES

1. Basu, Paul. (2004) My own island home. *Journal of Material Culture.* Volume 9, issue 1, pp. 27 – 42.
2. Bishop, R. (2008). In The Grand Scheme Of Things: An Exploration Of The Meaning Of Genealogical Research. *The Journal of Popular Culture*, *41*(3), 393-412.
3. Hackstaff, Karla. (2010) Family Genealogy: A Sociological Imagination Reveals Intersectional Relations. *Sociology Compass*, Volume 4, issues 8, pp. 658 – 672.
4. Herskovitz, A. (2012). A Suggested Taxonomy of Genealogy as a Multidisciplinary Academic Research Field. *Journal of Multidisciplinary Research*, *4*(3), 5-21.
5. Huffington Post, story published November 11, 2013: http://www.huffingtonpost.com/2013/11/11/craig-cobb-white-supremacist-black_n_4256360.html
6. JOGL. http://jogamp.org/jogl/www/
7. Kramer, Anne-Marie (2011). Telling multi-dimensional family histories: Making the past meaningful. *Methodological Innovations Online*, 6(3), 17-26.
8. McAdams, D.P. (1999). Personal narratives and the life story. In L. A. Pervin & O. P. John (Eds), *Handbook of personality: Theory and research* (3rd ed., pp. 242-262). New York: Guilford Press.
9. Merriam-Webster's dictionary. http://www.merriam-webster.com/dictionary/genealogy
10. Saar, Martin. (2002) Genealogy and Subjectivity. *European Journal of Philosophy*, Volume 10, Issue 2, pp. 231 – 245.
11. Tamassia, Roberto, editor. (2013) Handbook of Graph Drawing and Visualization. CRC Press.
12. Ware, Colin. (2004) Information Visualization: Perception for Design. Morgan Kaufmann.