Supporting Collaborative Sensemaking in Genealogical Research

A master thesis
submitted in partial fulfillment of the
requirements for the degree of
Master of Science
in
Computer Science
of
Stanford University

Nam Wook Kim
June 2010

Abstract

Genealogy is an attractive domain for studying collective information aggregation and interpretation. Genealogical researchers engage in a social process of collecting, organizing, and disseminating family records. Visualization constitutes an important part of this process, as visual depictions of family structuretypically in the form of traditional family tree diagramsare an important means of aggregating, verifying, and presenting genealogical data. Along with the rise of the web, online genealogy web sites have begun to support genealogical research. However, they have two major issues. First, family trees used in the existing sites are often limited to prioritizing the display of generational relations, accordingly missing other important dimensions in particular time. Second, the storytelling component, usually a timeline of events, lacks a family context.

To respond to these issues, this thesis contributes TimeNets, a novel visualization technique for genealogical data that simultaneously conveys family structure and temporal patterns. Unlike traditional family tree diagrams, TimeNets can depict additional patterns such as divorce, remarriage, plural marriage, and out-of-wedlock births. We conducted a controlled, comparative evaluation of TimeNets and standard family tree diagrams, finding that TimeNets accelerate analysis tasks involving temporal data. Using TimeNets as our base representation, we built Akinu, a new website for the collaborative curation of genealogical data. The site allows people to collaboratively build rich family stories that weave both family and time together.

Acknowledgements

This thesis would never have been possible without my mentors, Prof. Jeffrey Heer and Dr. Stuart Card. They made my life so happy at Stanford. I have learned many things from them. I owe a huge debt of gratitude to both of them. Thank you very much, Prof. Heer, for giving this life-changing opportunity. You have paved my way!

I would also like to give special thanks to Prof. Pat Hanrahan for reviewing this thesis. He was actually one of the reasons that I decided to come to Stanford.

My friend Juho Kim also helped me keep motivated by being my role model as a masters student at Stanford. Thanks Juho and good luck on your PhD at MIT!

Finally, I dedicate this thesis to my parents, Seong Ho Kim and Seong Ja Jeong, and my brother, Nam Gyun Kim, who always believe in me and my dreams.

Contents

Chapter 1

Introduction

The combination of networking, database technology, visualization, and content analysis algorithms is creating new possibilities for the collective aggregation and interpretation of information. In this thesis, we take a specific domain of collective information aggregation-genealogy-and develop an improved visualization that will eventually anchor a social sensemaking system. Genealogy, or the study of families, is an attractive domain for studying collaborative sensemaking practices. It is a popular activity pursued by millions of people, ranging from hobbyists to professional researchers. It engages millions in a social process of foraging for data, evaluating uncertain data sources, analyzing the data, and then disseminating the resulting products.

The genealogical research process involves determining when and where people lived, as well as their biographies and kinship [8]. It begins with collecting family records from relatives or local archives. Family trees are then used to keep track of collected records and organize them for future reference. By sharing the results with family members or other genealogists, it become quicker and easier to gather more information. The research process is often collaborative, and leads to diverse knowledge of religious histories, migration trends, and historical social conditions; tracing ancestry gives us an understanding of our history.

Both family trees and collaboration constitute core parts of genealogical research. Many genealogy web sites [18, 1] have begun to support genealogical research by providing a social space where users can collaborate on building their family histories.

They typically employ family trees as a data collection tool; people use them to enter family data. In particular, the family tree visualization plays a role in generating visual depictions of family structure that are an important means of aggregating, verifying, and presenting genealogical data. These sites facilitate genealogical research more efficiently than traditional offline research. However, they have limitations as well. First, family trees are limited to prioritizing the display of generational relations, thereby missing other important dimensions, in particular time. Second, the storytelling component, usually a timeline of events, lacks a family context.

To address such limitations, we first present TimeNets, a timeline-based visualization technique for representing genealogical data. To enable the analysis of families over time, TimeNets prioritize temporal relationships in addition to family structure. Individuals are represented using timelines that converge and diverge to indicate marriage and divorce; directional edges connect parents and children. This representation both facilitates perception of temporal trends and provides a substrate for communicating non-hierarchical patterns such as divorce, remarriage, and plural marriage. In addition, historical or family events can be incoporated TimeNets, allowing the user to tell a family story. We conducted a controlled, comparative evaluation of TimeNets and standard family tree diagrams, and found that TimeNets accelerate analytical tasks involving temporal data. We then introduced Akinu, a prototype social genealogy web site where people collaboratively build rich family stories that weave both time and family together by using TimeNets as a data entry interface.

Chapter 2

TimeNets: A Timeline-Based Visualization for Genealogical Data

2.1 Introduction

The most common task confronting genealogists is to correctly identify individuals and their familial and temporal relations. To keep track of their findings, people typically use genealogical diagrams, or "family trees," such as ancestor (pedigree) charts and descendant charts (Figures 2.1a-b). By aligning people by generation, the charts prioritize the display of kinship relations, facilitating the identification of marriages, parent-child relations, siblings, and cousins.

However, such representations often omit other aspects of genealogical data, particularly time. For instance, genealogists must frequently cope with temporally ambiguous evidence in order to establish kinship [26]. To infer genealogical relations, a researcher may compare estimates of an individual's birth date with the marriage dates of potential parents; misapprehension may lead to an incorrect reconstruction of the family. Most existing genealogical diagrams (e.g., [12, 17, 19, 27]) share a common set of limitations:

1. They do not show family networks well. Families are networks of relationships, not trees. The most popular visualizations are ancestor charts (trees of generations of parents), descendant charts (trees of generations of children) and

hourglass charts (combined ancestor and descendant charts for an individual). This approach assumes a hierarchical structure that does not fit real-world families [27].

2. They do not show complex relationships well. Traditional diagrams are unsuited for communicating complex patterns such as divorce, remarriage, out-of-wedlock births, and polygamy. These are part of real family histories and may have different meaning in world cultures.

3. They do not show temporal attributes well. Temporal attributes such as birth, death, marriage and divorce dates are either omitted or depicted only by text labels.

4. They do not scale well. One of the major advances in genealogy in recent years has been the online availability of family data, making it easier to construct larger family relationship networks. Yet, unless heavily edited by hand, automatically generated diagrams are not suited to depict these larger networks. They tend to show perhaps eight generations, sacrificing depth or breadth of relationships.

5. They do not show the relationship between nodes at a distance. It is hard to see the relationship to a famous person or between two people co-mentioned in an historical record if they are not close together in the family network.

To address the limitations of traditional genealogical diagrams we contribute TimeNets, a visualization technique for genealogical data. TimeNets encode both family kinship and timelines of individual life events; interactive degree-of-interest filtering is used to scale to large data sets. TimeNets address complex relationships by laying them out on individual lifespan timelines (Figure 1). These timelines also express temporal attributes, such as birth or marriage date. Scale is handled using focus+context techniques: a degree-of-interest function filters the display based on a user's indicated interest in some nodes and their relationship to other nodes. The same mechanism also allows for the display of nodes at a distance and the contextual nodes that relate them.

2.2 Related Work

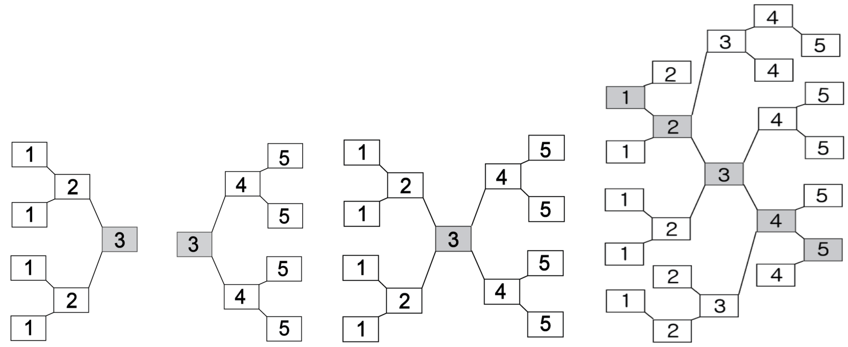
To place TimeNets in context, we review existing techniques for both timeline and genealogy visualization.

2.2.1 Genealogy Visualization

In a broad sense, there exist two types of genealogical relations. Parent-child relationships (*consanguine* relations) define a hierarchy in genealogical data. Relationships through marriage (*conjugal* relations) are non-hierarchical and merge family trees. Together these form a network of relationships—complex but simpler than a general graph. The most common genealogical research is ancestral research—tracing ancestry of self—and descendant research—finding descendants of an ancestral couple. They correspond to constructing a tree of ancestors and a tree of descendants. This observation verifies why ancestor (pedigree) and descendant charts (Figure 2.1) are canonical charting methods for genealogical data.

Other depictions have also been applied. An hourglass chart combines both a pedigree and a descendent chart centered on a specific individual (Figure 2.1c). Fan charts are ways of drawing these trees without connecting lines and with more space available to the leaves of the chart [12]. These charts make it easy to understand the basic hierarchical relationships of direct dependency at the cost of suppressing other relationships. Specialized charts, such as a Table of Consanguinity or a Canon Law Relationship Chart [7] are used to determine the degree of relationship between people who share a common ancestor, such as great aunt or third cousin twice removed. Sometimes these basic genealogical charts are combined with pictorial artwork.

McGuffin and Balakrishnan [27] introduced Dual Trees (Figure 2.1d). Dual Trees generalize the hourglass chart by offsetting the roots of the trees with respect to each other; multiple roots are connected along the hierarchy and each root has its own hourglass chart. As a result, more information can be shown at a time without introducing edge crossings. To maintain readability, however, only a limited number of nodes are shown on a computer screen. An interaction technique for expanding or collapsing a node is used to explore large data and transition between different



(a) (b) (c) (d)

Figure 2.1: Genealogy diagrams. (a) Ancestor chart. (b) Descendant chart. (c) Hourglass chart. (d) Dual tree.

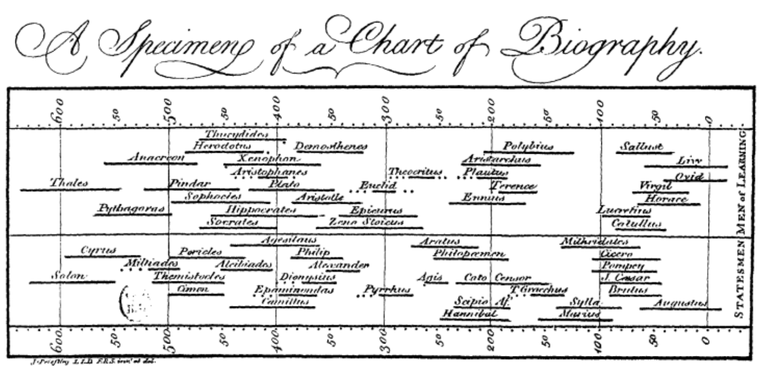


Figure 2.2: Biographical lifelines by Priestley, 1765 [34].

dual-tree subsets.

The genealogical techniques described are widely used in published genealogies and software. They are successful in showing a limited number of hierarchical relationships, but have the five limitations we have previously described.

2.2.2 Timeline Visualization

An inspiration for making genealogical diagrams more expressive has an impressive pedigree itself. In 1765, Joseph Priestley used timelines to depict the lifespans of two thousand famous people from 1200 B.C to 1750 A.D (Figure 2.2). He also invented using dots to indicate uncertainty in birth and death dates. The horizontal axis is

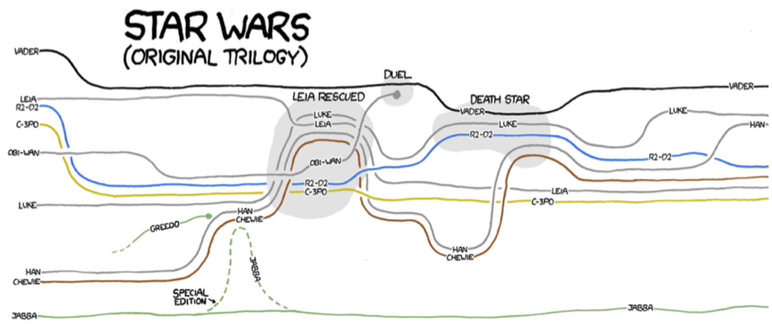


Figure 2.3: XKCD Movie Narratives: Star Wars.

time and people's vertical position is ordered by "importance." Kinship is not shown, but Priestley's diagram makes clear who was a contemporary of whom and who was living during world events.

Timelines have been used to visualize life events in a number of domains, such as medical records and criminal justice. One of the best known of these is Lifelines [33] and its successors. Lifelines uses timelines to visualize personal histories based on medical records. Each timeline shows different sections of the record such as diagnosis and medications. Users can drill down into timelines for details-on-demand. Temporal and causal relationships among different sections also can be inferred, but require significant cognitive effort. The Pattern Finder [13] is a descendant of the Lifelines work that visualizes mined temporal patterns in multivariate data.

Randall Munroe of XKCD [29] hand-crafted timelines of interactions among movie characters (Figure 2.3). Each character is represented using a lifeline differentiated by color. Lines converge and are grouped using a gray background to indicate which characters are together at a given time. A hierarchy is not defined on the data and accordingly not shown.

Timelines have also been applied to the visualization of family networks. Geno grams [19] are like family trees, but lines depicting a marriage represent ordinal time. Genograms can depict more complex relationships like divorce and remarriage, but depend on special symbols. Genograms are most useful when the number of people depicted is moderate and they are easiest to use when most relationships are hierarchical. Genelines [17] depict people as timelines (Figure 2.4) and are good

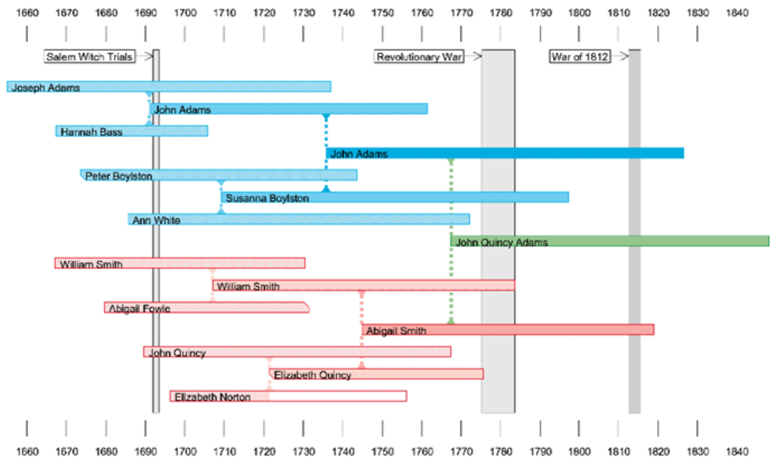


Figure 2.4: Genelines pedigree chart [17].

at showing temporal attributes. How they show non-hierarchical patterns such as divorce and remarriage and how they scale to large data are unclear, however.

2.2.3 Degree-of-Interest Techniques

As family networks become larger, they no longer fit on the screen using any of the techniques discussed so far. Degree-of-Interest techniques, introduced by Furnas [15], compute a score for each node in the network based on which nodes are presumed to be of most interest to the user. Nodes below a threshold score are suppressed. Using versions of this technique, Heer and Card [22] were able to display large DOI Trees on the order of a million nodes. Card et al [5] combined DOI Trees with time-varying

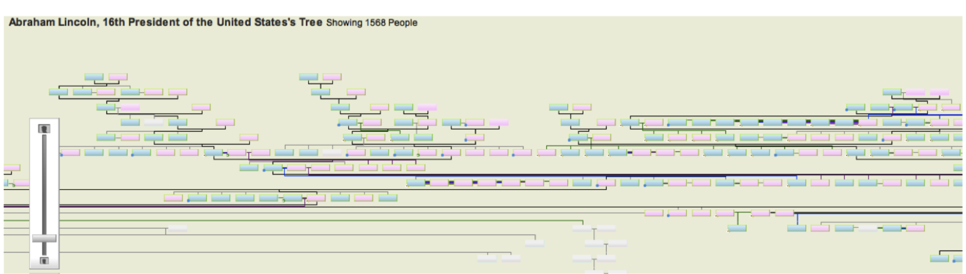


Figure 2.5: Geni.com [18]: Family tree of Abraham Lincoln. It is hard to find where he is.

organizational data to display changes over 50 years of leadership of a medium sized country. van Ham and Perer [39] recently extended DOI techniques to general graphs. DOI techniques might also aid the scaling and filtering of family network diagrams (Figure 2.5). In this project, we apply it to a neighbor of trees—genealogical lattices. Interest might be assigned based on the relatives of a focal person, the relatives two people might have in common, or search results, such as every relative named "Christopher."

2.3 TimeNet Design

As the related work suggests, three visualization paradigms have promise for genealogies: hierarchies, timelines, and degree-of-interest techniques. Our challenge is to bring all three of these techniques into correspondence through unified visual encoding and layout algorithms. In this section, we describe the series of visual encoding decisions and associated trade-offs involved in crafting TimeNets. We focus on high-level design goals and defer discussion of implementation and interaction details to the next section.

In designing TimeNets, our goal was to support simultaneous graphical representation of ancestor and descendant relations, complex conjugal relationships, temporal attributes, and data uncertainty—all in a scalable fashion. In addition to generational structure, non-hierarchical relationships can get complicated. Divorce and remarriages are frequent in modern family settings. Furthermore, in non-traditional family arrangements, one might have more than one spouse at a time. A timeline is a natural way to visualize these relationships as well as other important temporal attributes such as marriage dates. Taken together, both hierarchical and temporal information will enable effective understanding of relationship dynamics and story telling of family history.

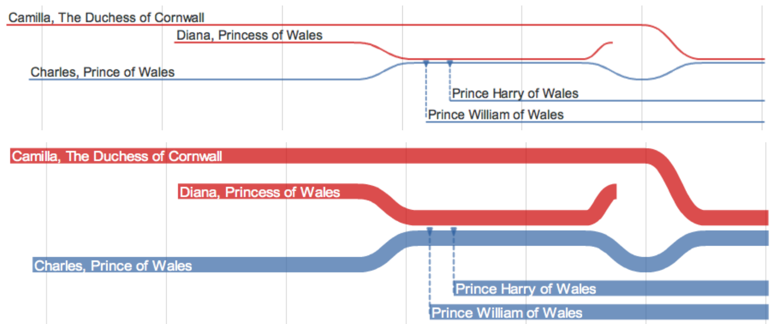


Figure 2.6: TimeNets with different styles. (a) Thin lines with external labels. (b) Thick lines with internal labels.

2.3.1 People as Individual Lifelines

To make temporal attributes salient, we started with the common convention of a linear timeline. A TimeNet's horizontal axis represents time progressing from left to right. The examples in this paper use metric timelines; ordinal timelines are possible with minor modifications. Similar to prior genealogical timelines [17, 29, 34], we represent a person as an individual lifeline (Figure 2.6). The left end of the lifeline represents a person's birth and the right end represents their death; thus the horizontal extent of the line depicts a person's lifespan. By default we use line color to depict sex (blue for male, red for female). Lifelines include text labels consisting of a person's name and potentially other data. Line width is left as an aesthetic design parameter; if the lifeline is thick enough, we place the label within it (Figure 2.6b).

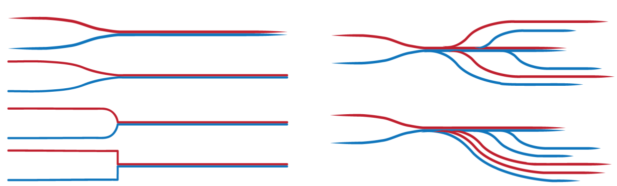


Figure 2.7: Early design prototypes. (a) Lifeline interpolation techniques. (b) Different children layouts.

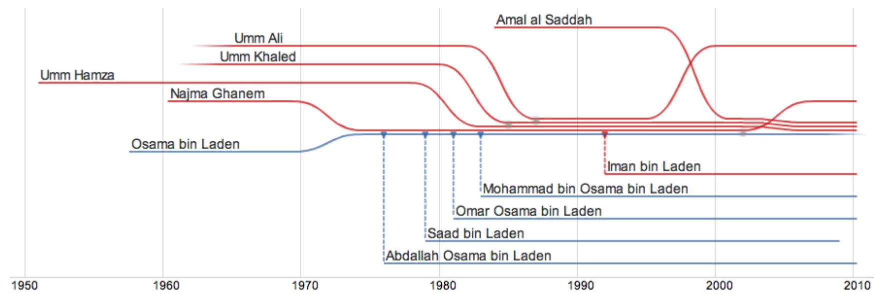


Figure 2.8: The marriages of Osama bin Laden. Gradients indicate uncertainty of birth or marriage dates.

2.3.2 Marriage/Divorce as Converging/Diverging Lifelines

With the horizontal axis devoted to time, the vertical axis is free to represent relationships. We use vertical proximity to encode conjugal relations: two or more lifelines converge into a bundle of adjacent lines to denote a marriage (Figures 2.6–2.8). The point at which the lines meet represents the marriage date. Conversely, lines diverge to indicate divorce. This representation naturally encodes a variety of marriage patterns, as sequences of marriage, divorce, and remarriage are depicted by the convergence and divergence of lifelines. Plural marriage (polygamy) is represented by more than two lines converging into a shared marriage bundle (Figure 2.8).

In the case of multiple marriages, the question arises of how to order spouses. Our default approach is to vertically order spouses by their first marriage date; hence a vertical scan visits spouses in chronological order. A different approach is to alternatively place spouses above and below a focal person (Figure 2.10). An alternating placement reduces line crossings, but makes it more difficult to determine spouse ordering. In either case, when a person divorces we return their lifeline to its original position, facilitating consistent placement and enabling a horizontal scan to determine if a divorce occurs.

We have also explored a variety of lifeline interpolation strategies (Figure 2.7a). Orthogonal lines and circular arcs clearly depict the dates of marriage and divorce, however, splines with continuous curvature are easier to follow (particularly for line crossings) and elicit higher user preference ratings. We default to using cubic Bézier curves, but users can change the interpolation settings if desired.

Unfortunately, line crossings due to multiple marriages are sometimes unavoidable. To alleviate this problem, we use the aforementioned spline interpolation and can apply alpha blending to facilitate line-following. In some cases, a divorce and subsequent remarriage may be in close temporal proximity, resulting in nearly vertical line crossings. In such cases, we slightly exaggerate the time period to enable better perception of the crossing (Figure 2.9).

2.3.3 Parent-Child Relationships as Drop Lines

To represent consanguine relations, we depict children as lifelines emanating from their parents. Our first design iteration initiated a child's lifeline directly on the parents' marriage line; the child line then diverged into its own space (Figures 2.7b, 2.11c). Informal user testing revealed that this representation is confusing, as it is often ambiguous which line corresponds to a child and which to a divorced spouse. Furthermore, this representation can result in lifelines with very long vertical stretches that both add visual noise and complicate perception of temporal patterns (Figure 2.11c).

Instead, we adopt a strategy similar to Genelines [17]: we depict parent-child relations using a directional edge (or "drop line") that connects the parents to the start of the child's lifeline. To make lines perceptible but not distracting we render parent-child edges using faded dashed lines. Parent lines are annotated with a visual marker indicating the directionality of the edge. One disadvantage of this approach is that tracing from parent to child requires more complex line-following. However, there are a number of compensating advantages: drop lines enable more accurate perception

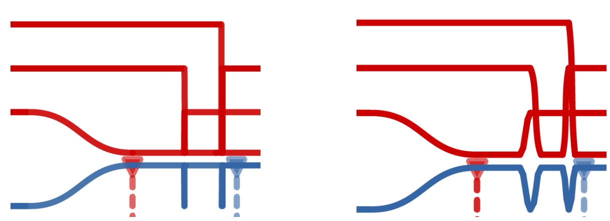


Figure 2.9: Divorce and remarriage in close proximity. (a) No perturbation. (b) Perturbed event points.

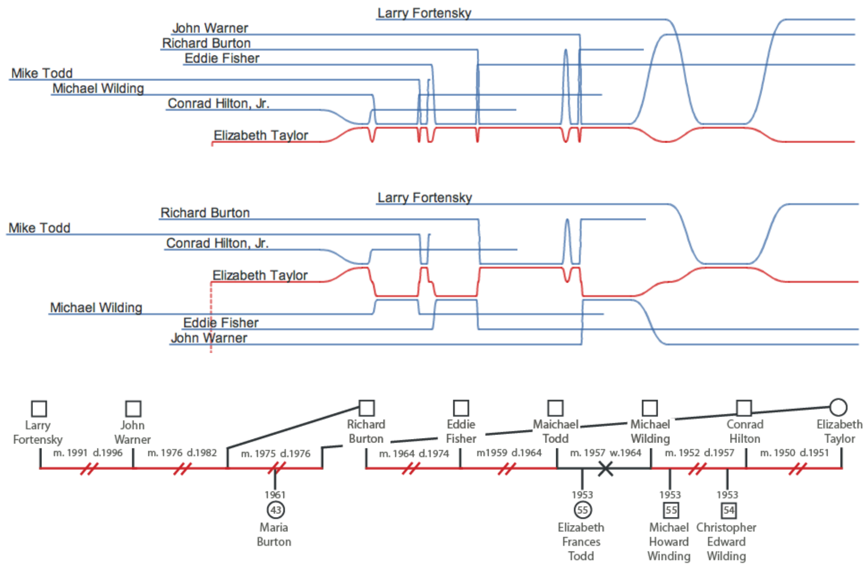


Figure 2.10: Marriages of Elizabeth Taylor. (a) Spouses ordered chronologically. (b) Alternating spouse layout. (c) Genogram representation [19].

of temporal attributes (e.g., birth date and lifespan) and reduce the saliency of edge crossings when child lines are positioned far from their parent lines (compare Figures 2.11b and 2.11c). Moreover, drop lines easily accommodate children born out of wedlock: we simply place markers on each parent's lifeline and connect them with the drop line (Figure 2.11a).

By default, we vertically sort children by birth date. We place younger children closer to their parents, as this arrangement helps minimize line crossings. We can also alternate child placement above and below parents (Figure 2.7b), but such alternation

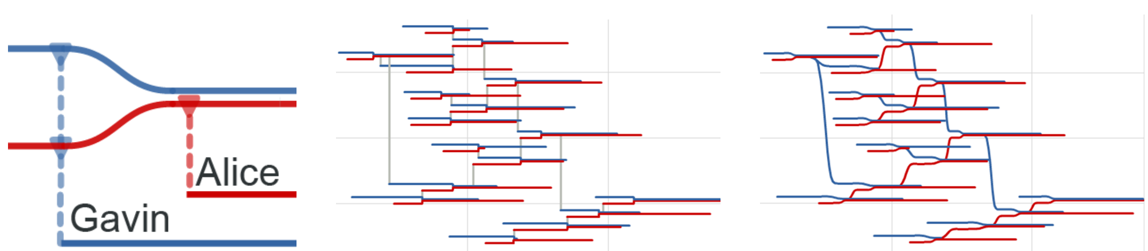


Figure 2.11: Child layouts. (a) Children born out of and in wedlock. (b) With drop lines. (c) Without drop lines.

impedes quick apprehension of birth order.

2.3.4 Uncertainty

Genealogical data regularly suffer from missing or approximate values. Without indications of uncertainty, visual analysis may lead to inaccurate conclusions. Missing temporal attributes such as birth and death dates can be particularly problematic for our time-based layout. As described in the next section, we first estimate missing attributes to derive potential birth, death, and marriage dates. We then include visual markers to convey missing and uncertain values to viewers (Figure 2.8). By making uncertain data values more apparent, we hope to assist users as they clean and curate their data. For uncertain birth and death dates, we fade lifelines using a gradient; the lifeline takes on full saturation at the estimated date of birth or death. For uncertain marriage and divorce dates, we draw an underlying marriage marker and again use a gradient to indicate uncertainty. By clicking an uncertain value a user can then enter a revised date.

2.3.5 Other Patterns and Attributes

While TimeNets directly show marriage, ancestry, and temporal patterns, they can also be used as a substrate for conveying additional data. For example, the color encoding of lifelines can be changed to communicate attributes other than gender. A variable color encoding scheme may show changes in geographic location (e.g., continent or country) over time, or the occurrence of different diseases. TimeNets can also highlight structural patterns: one might highlight an ancestral path or view the output of a graph analysis routine. We can also add annotations for historical or personal events of interest (Figure 1), allowing one to tell a family story.

2.3.6 Focus + Context Techniques

To navigate large genealogies, we use degree-of-interest (DOI) estimation to determine the most salient aspects of the data and then filter the elements deemed less

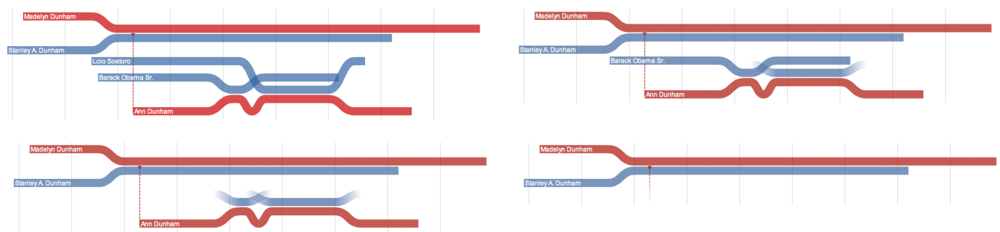


Figure 2.12: Progressive elision by DOI (left-to-right).

interesting. TimeNets visually communicate the existence of elided elements in two ways. First, when a person has a DOI value beneath the visibility threshold but is married to someone with above-threshold DOI, a segment of their lifeline is shown to indicate the duration of their marriage (Figure 2.12). Second, to handle low-interest descendants, drop lines are still used, but are faded out (c.f., [39]). These marks provide an indication of the elided context, and thus serve as "information scent" [31] for further exploration.

2.4 Implementation

TimeNets are constructed in a two-stage process: data processing and visual encoding. In the data processing stage, we ingest genealogical data and apply a series of data transformations, including estimation of missing temporal attributes. In the visual encoding stage, we calculate degree-of-interest values and use them to layout the graph and label visible elements. In this section, we detail each of these steps.

2.4.1 Data Model

Although a variety of genealogical data formats exist, the de facto standard within the genealogical community is GEDCOM [16]. Accordingly, we parse GEDCOM files as one data source for TimeNets. Unfortunately, the GEDCOM specification can not represent many types of interpersonal relationships, including same-sex marriage, polygamy, and incest. In response, we developed our own data model for genealogical data. The first step in our pipeline is thus to ingest data from an external source—such as a GEDCOM file or web repository such as Freebase [14]—and map it to our data model.

We use a basic relational data model. At its simplest, the model consists of two relational tables: a list of individual people and a table of relationships. For individuals, we assume the presence of at least five attributes:

<*id, name, sex, date_of_birth, date_of_death* >

We encode relationships using foreign keys for two people and require relationship type and temporal attributes:

<*person1_id, person2_id, relationship_type, relationship_start_date, relationship_end_date* >

Here *person1_id* and *person2_id* refer to individual records in the person table. Relationships involving multiple people are represented by multiple entries (rows). The primary *relationship_type* values are *Child-of* and *Spouse-of*, though these types are extensible. The data model can be extended by introducing additional columns (e.g., geographic data) or by introducing additional tables (e.g., historical events).

2.4.2 Missing Data Estimation

TimeNets rely on temporal attributes such as birth date and death date in order to compute a layout. However, it is common for genealogical data to have missing or incomplete temporal values, e.g., a data set may have birth and death dates but lack marriage dates. To address this issue, we estimate missing data values as part of our data processing stage.

We use a rule-based method to estimate missing dates. The basic idea is to take advantage of the regularities among temporal attributes. We define an ordered chain of rules for each attribute, and use the first applicable rule in the chain:

· **birth** ← parents' marriage; mean sibling birth; mean spouse birth; ...
· **death** ← mean sibling death; mean spouse death; ...
· **marriage** ← oldest child's birth; ...
· **divorce** ← assume no divorce

We use default estimates if no applicable rule exists. For instance, we offset a person's birth date (e.g., by 20 years) to estimate a missing marriage date and assume a the person is alive if their lifespan is under a threshold (e.g., 85 years).

The main goal of our estimation rules is to ensure that we have at least reasonable values for missing attributes for subsequent visualization. However, our current solution is only a stopgap method. While we have attempted to select suitable defaults, analysts can modify the estimation rules or add new ones; in the future we plan to improve the estimation process using machine learning techniques. As discussed previously, TimeNets also visualize the uncertainty of estimated dates so that analysts can identify and repair missing values if desired.

2.4.3 Degree-of-Interest Calculation

Once the data has been suitably transformed, we calculate degree-of-interest (DOI) estimates. These DOI values provide a rank-ordering of the "interestingness" of people within the genealogical graph based on a current set of focal nodes (e.g., clicked elements or search result hits). These values are in turn used to subsequently filter and layout the graph; after the DOI values are computed, only the nodes whose DOI values are above a chosen threshold are visualized. Our approach is based on previous models [4, 5, 22], with modifications to support non-hierarchical marriage relationships.

Our default DOI function is as follows. Starting with a set of maximally interesting focus nodes, we traverse the genealogical graph and assign lower DOI values with increasing distance. If a root element (e.g., central matriarch) is defined, maximal DOI values are assigned both to focus nodes and their relatives along the path to the root. Otherwise, DOI values decrease linearly across consanguine relations. Across conjugal relations, DOI values decrease more slowly using fractional DOI increments. Thus for a given focal node, spouses will be given higher interest than either parents or children. For both spouses and children, additional fractional DOI increments are assigned based on date order; for example, first spouses have slightly higher DOI than later spouses.

Of course, other DOI functions are possible. For instance, one might be interested in exploring cousin relationships and thus assign cousins higher interest values. Our system is modular and can be extended to incorporate alternative schemes. However, we leave the specification of new interest functions by genealogical analysts to future work.

2.4.4 Layout

Once DOI values are calculated, we compute the layout. The layout algorithm works by grouping genealogical elements into a three-level scenegraph consisting of nodes, local blocks, and global space (Figure 2.13). Nodes represent the bounding region for a specific lifeline. People either directly or transitively connected by marriage are grouped together to form a local block. Our algorithm first segments the graph into local blocks and performs a local layout for each, determining node bounds in the process. Blocks are then positioned by a global layout pass.

Local block segmentation

A directed acyclic graph of blocks is constructed by traversing the genealogical structure in depth-first fashion and grouping conjugally-related people. Blocks may have more than one parent due to intermarriage. Our current approach has one limitation: it assumes that cross-generational incest (e.g., mothers marrying sons or sons-in-law) does not occur. We believe this to be a reasonable assumption for most real-world data sets.

Local layout and lifeline generation

To perform local layout, we first arrange visible nodes along the time dimension relative to the origin of the block and determine node lengths. We then generate lifelines and set their vertical ordering. We also compute the position and style attributes of marks representing elements beneath the current DOI threshold (e.g., the partially elided elements in Figure 2.12).

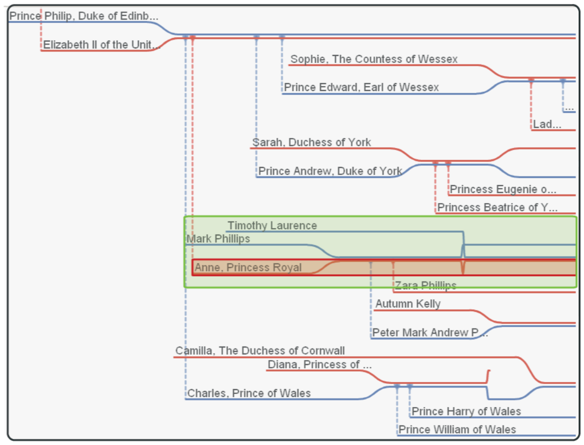


Figure 2.13: A three-level scenegraph groups nodes into local blocks within a global coordinate space.

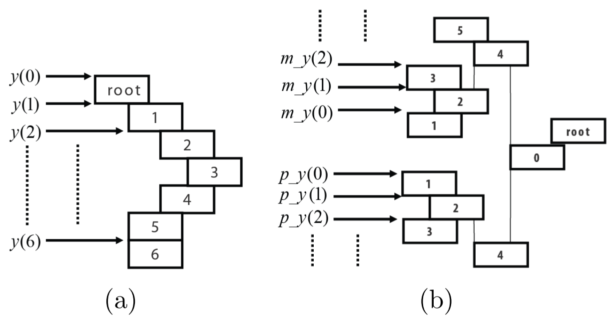


Figure 2.14: Global layout. (a) Descendants placed by pre-order traversal, (b) Ancestors by in-order traversal.

Lifelines are generated according to the design principles in the previous section. We maintain event points for temporal attributes of each person, including dummy event points to aid spline routing (e.g., between birth and marriage points). If divorce and re-marriage events occur in close spatial proximity, we perturb the event points along the horizontal dimension to ensure better perception of line crossings. We then place a label along (or in) the lifeline. If necessary, we truncate the label to fit the horizontal bounds of the lifeline.

The vertical placement of event points depends on a person's computed DOI. We start by finding the person with maximal DOI in the block. We vertically oscillate this focal lifeline between a married and non-married position. The focal lifeline then serves as a reference line for spouses, whose lifelines converge to and diverge from the reference line (Figure 2.10). Different orderings are possible (Figure 2.6); the default is to order spouses vertically above the reference line.

Global layout

Once the block hierarchy is built, global layout is performed by positioning each block. First, we arrange blocks along the horizontal axis according to the minimum birth date in each block. Second, we perform the vertical layout, ensuring that the bounding boxes for local blocks do not intersect. We use different placement schemes for ancestors and descendants. For descendants, we traverse descendant blocks in pre-order, ensuring that the visit order is from the youngest child to the oldest child within each generation (Figure 2.14a). Each block's position is then assigned according to the visit order. As a result, the first-visited block is positioned below the root and the second-visited block is positioned below the first block, and so on. For ancestors, we visit blocks using in-order traversals (Figure 2.14b). Once layout is performed, we check if the vertical size fits the screen space. If not, we iteratively cull low-interest nodes and update the layout until it fits.

2.4.5 Interaction and Animation

Interactive navigation of TimeNets is similar to previous DOI-based visualizations [22]. Clicking a node makes it the current focus and updates the layout; control-clicking multiple elements defines multiple foci. In this way, one can navigate the graph and build up views of interest. Alternatively, one can type a search query; the result set is used as focal nodes. We use staged animations to communicate changes between interface states (c.f., [32]): the first stage fades out elements whose DOI has dropped beneath threshold, the second stage animates previously visible elements to their new positions, and the third stage fades in newly visible elements.

2.5 Evaluation

To inform the iterative design of TimeNets, we conducted a formative evaluation comparing the effectiveness of generational "family tree" diagrams with TimeNets. Subjects were shown a genealogical diagram and asked comprehension questions. We hypothesized that (**H1**) traditional tree diagrams support faster and more accurate perception of structural family relations but that (**H2**) TimeNets better facilitate the apprehension of patterns with a temporal component.

2.5.1 Method

We asked subjects to complete tasks with two different visualizations: a modified descendant chart (Figure 2.15a) and a TimeNet chart (Figure 2.15b). We augmented the descendant chart design to support multiple marriages: spouses are listed in chronological order and each marriage is indicated by a curved edge annotated with marriage and divorce dates. Edges to children originate from these marriage markers. We used 600×600 pixel images depicting a fictitious family of 36 people. Each person was labeled with a common first name with either 5 or 6 letters. To avoid ambiguity all names of the same gender have a unique first letter. Names were varied between diagram conditions.

For each diagram, subjects were asked to answer comprehension questions (Table 2.1) grouped into three categories:

· *Structural* questions involving only kinship,
· *Temporal* questions involving only timing, and
· *Structural* $\times$ *Temporal* questions involving both.

There were 36 unique tasks in all, 18 for each diagram. Subjects were instructed to accurately answer questions as quickly as they could. A total of N=22 subjects participated via Amazon's Mechanical Turk [21] and were paid $0.10 USD per task. Before participating, subjects had to successfully complete a suite of qualification practice tasks. To combat known reliability issues with timing on MTurk [21], we

Structural	How many daughters does Irina have? How many half-siblings does Peter have? Who is Isaac and Holly's closest male ancestor? Which person has had the most marriages? Which mother of two is still married to her first husband? Which woman has step-children but not biological children?
Temporal	How many people were alive in 1950? Which person was born during the 1920s? Were Marcus and Carmen alive at the same time? Who was born most recently? Who died in infancy? Who has the longest lifespan?
Temp × Struct	How many couples got married in the 1970s? Which of Leslie's sons was the last to get married? Who did Brenda marry after divorcing Roger? Who was half the age of their spouse when they married? Which uncle is younger than some of his nephews? Who is at least 10 years younger than all their siblings?

Table 2.1: Representative User Study Tasks.

used a "ready-set-go" interaction with each task and timed the tasks ourselves using JavaScript.

2.5.2 Results

We analyzed both task accuracy and response time. To analyze accuracy, we first scored each subject response as either correct or incorrect; the overall accuracy rate was 90%. We found no significant differences between tree diagrams and TimeNets for structural (χ^2(1,211)=1.030, p=0.310), temporal (χ^2(1,210)=0.072, p=0.789), or structural × temporal (χ^2(1,206)=1.603, p=0.205) tasks.

Next, we examined response times. As the data are not normally distributed, we used a non-parametric test (Mann-Whitney U) to compare conditions. For structural tasks, the median response time using TimeNets is 2.8s slower (19.9s vs 17.1s, 14%) than tree diagrams. This difference, however, is not significant (U(108,103) = 5353, p = 0.637). For other tasks, TimeNets exhibit a statistically significant advantage. The median response time using TimeNets is 4.3s faster (14.6s vs 18.9s, 23%) for temporal tasks (U(104,106) = 4408, p = 0.012) and 6.0s faster (18.1s vs 24.1s, 25%)

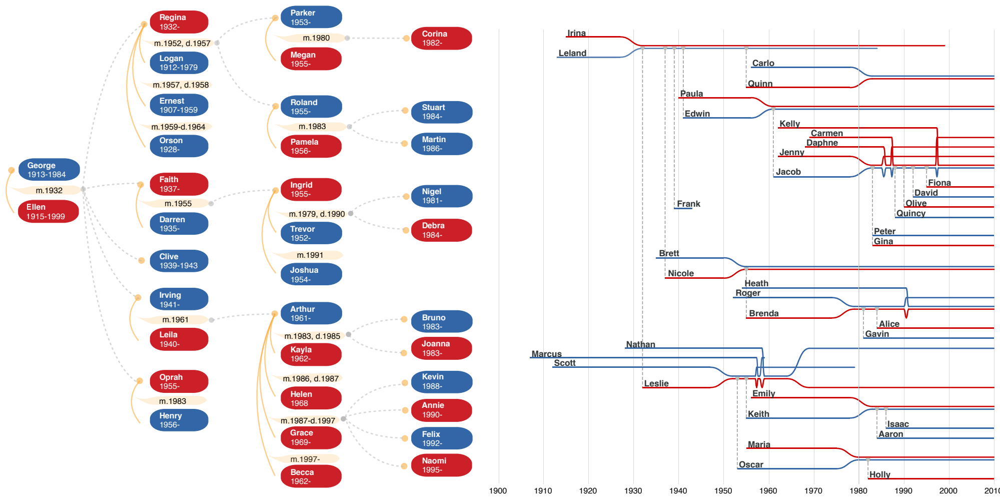


Figure 2.15: Genealogical diagrams used as experiment stimuli. (a) Descendant chart. (b) TimeNet.

for structural × temporal tasks ($U(103{,}103) = 4430$, $p = 0.041$).

2.5.3 Discussion

Our results provide scant evidence for **H1**: we found no significant differences in accuracy across chart types, and while the descendant charts were slightly faster for structural tasks, the difference was not significant. On the other hand, we did find evidence for **H2**: tasks requiring the use of temporal attributes were completed significantly faster using TimeNets, resulting in a ∼25% time savings. Our results suggest that (a) TimeNets can be learned quickly by a lay audience and (b) TimeNets facilitate the perception of temporal trends in genealogical data better than tree diagrams.

In addition to establishing concrete benefits for TimeNets, our study also provided qualitative insights for improving future designs. From subjects' comments and our own test runs we learned that visual search for a person's name often dominates task time regardless of diagram type. This observation suggests that search and highlighting mechanisms for finding individuals could facilitate interactive use of either diagram type. Also, more sensitive studies (e.g., using eye tracking) might be able to separate the effects of diagram type on visual search versus decoding and inference.

These results provide promising formative evidence for the use of TimeNets in genealogical research: TimeNets appear to be well-suited for conveying structural and temporal data in an integrated fashion, and may prove a useful tool for analyses involving temporal attributes and/or complex marriage relations. Still, further evaluation is needed to more deeply understand the strengths and weaknesses of genealogical visualization techniques. New studies might examine depictions of data uncertainty, and case studies with practicing genealogists are necessary to assess the effectiveness of these techniques in real-world contexts.

2.6 Conclusion

In this paper we presented TimeNets, a time-based representation of genealogical data. By depicting individuals as timelines which converge and diverge to depict marriage, TimeNets represent a number of real-world phenomena—including divorce, remarriage, plural marriage, and out-of-wedlock births—that are either difficult or impossible to represent using standard genealogical diagrams. By using degree-of-interest techniques, TimeNets also support scalable, interactive exploration. In a controlled experiment we found that TimeNets exhibited significant advantages over family tree diagrams for tasks involving temporal data: TimeNets accelerated task times $\sim$25% without diminishing accuracy. These results suggest that TimeNets could serve as a useful tool for genealogical researchers and hobbyists.

Though we have focused on human genealogical data, we believe our techniques can be applied to other domains concerned with time-varying branching and merging phenomena. Examples include academic genealogy, biological evolution, artistic movements, computer systems (e.g., multi-threading), and organizational structures (e.g., firms and subsidiaries [25]). Exploring such domains may also suggest new variations of TimeNets. For example, the use of ordinal time, alternative degree-of-interest functions, and additional means of communicating structural units (e.g., a nuclear family) are all potentially useful extensions of our technique.

Chapter 3

Akinu: A Site for Collaborative Genealogical Research

Visualizations are used not only to explore and analyze data, but to communicate findings. - Heer 2008 [24]

3.1 Introduction

Well-designed visual encoding leverages the human visual system to amplify cognition by replacing computation with perception. Most visualization research to date concentrates on finding optimal encoding to help people analyze data. As a result, it is often limited to individual sensemaking, overlooking collaboration around visualizations. Recently, some research has recognized the role of visualization as a communication medium and studied the use of visualizations in a social context. Wattenberg observed, although not by design, the explosive social use of his NameVoyager visualization, including an intense level of conversation across the web [41]. A primary reason for its broad popularity was the encouragement of data analysis as a social activity. This has been followed by a number of research projects [24, 40] that attempt to deeply understand the patterns of social data analysis.

Genealogy also has a potential value as a social sensemaking domain. A typical genealogical research process begins with collecting family documents and oral

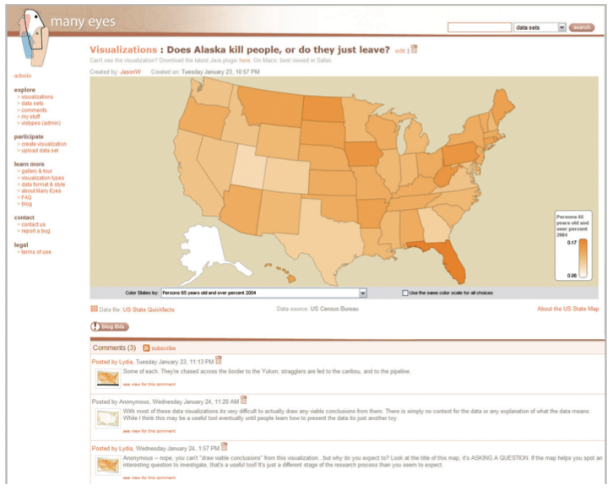


Figure 3.1: The user-created visualization and commenting interface below on Many Eyes. The image is taken from Wattenberg et al [40]

stories from relatives. It is also often necessary to contact local archives in search of additional information such as vital and census records. During this data collection period, people may also collaborate with family members, distant relatives, or professional genealogists. To keep track of findings, standard family trees such as pedigree charts and family group sheets are employed. People often share the collected information to facilitate the research process. Finally, they share their family stories among family members or friends. From this process, we extract three main themes of social sensemaking in genealogical research: collaboration, visualization, and storytelling. In order to support the sensemaking process, we describe the design and implementation of Akinu, where users upload their family data, create TimeNets, and share their family stories.

3.2 Related Work

Recently, the social use of visualizations has gained increasing attention. In this context, visualizations are considered not just analytical tools for individuals, but also social tools supporting communication and storytelling. There are two major research projects that attempt to better understand social dynamics surrounding the collective use of visualizations. The first such system is sense.us [24], a prototype

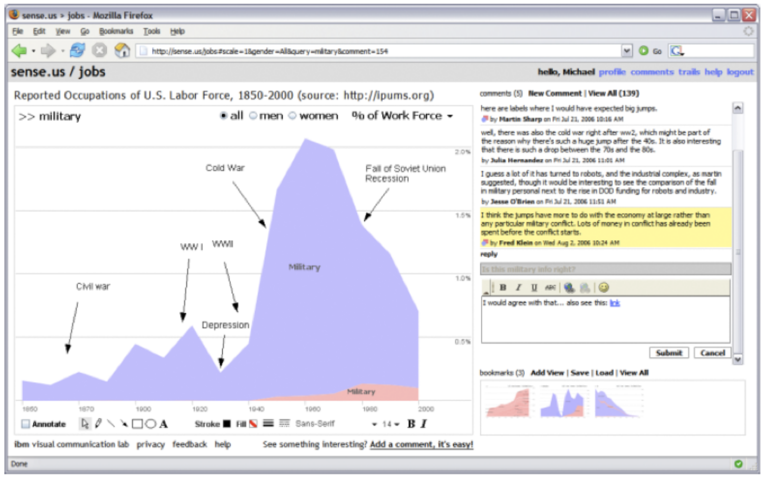


Figure 3.2: sense.us: asynchronous collaborative visualization system. The image is taken from Heer et al [23]

web application for social data analysis (Figure 3.2). The site supports asynchronous collaboration by providing view sharing, discussion across views, graphical annotation, and social navigation of activity histories. It is preset with a suite of interactive visualizations of U.S. census data; users cannot contribute new datasets or visualizations. While this limits the potential social space of the users, it encourages users to easily build common ground.

While sense.us was deployed on a corporate intranet, Many Eyes [40] is a public web site where users upload data, create visualizations, and engage in discussion (Figure 3.1). The site shares similar collaborative features with sense.us. The main difference is that users can contribute their own datasets as well as visualizations. As a result, Many Eyes occupies a broader social space than sense.us, which is limited to pre-populated data and visualizations. Interestingly, despite many community features in Many Eyes, most social interactions occur off the site [10]. This is partly because users come with their own datasets whose audience is not on the site, resulting in a fragmented community.

A number of commercial systems have also explored asynchronous collaboration around visualizations. Swivel [38], Chartall [6], DabbleDB [9], Spotfire [36]and Data 360 [11] are all collaborative visualization systems that allow users to upload data, create charts, and share the results with others. Not all, but most of them, provide

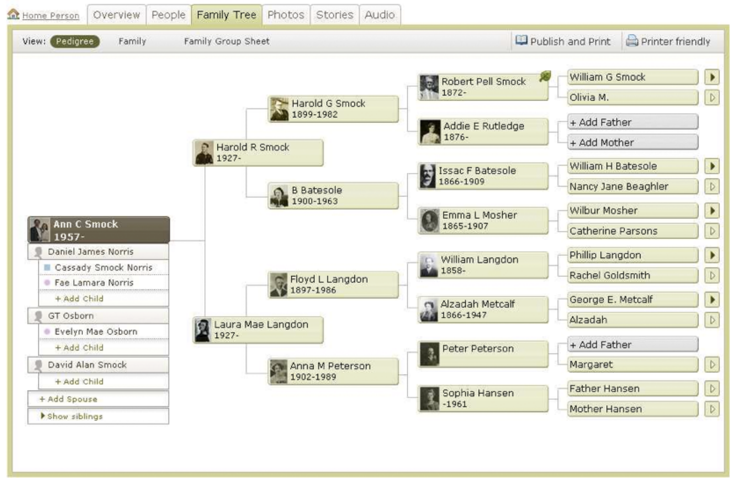


Figure 3.3: Ancestry.com: pedigree chart

discussion models and awareness cues for others activities. Although these systems have proven to be popular, useful, and scalable to a large audience, they all rely on static business graphics. A slightly different system is Wikimapia [?], which provides an online editable map where users can collaboratively and interactively annotate geographic regions.

Online genealogy services have also begun to support asynchronous collaborative genealogical research. Two popular sites are Geni.com and Ancestry.com. They share many similar features, including merging family trees, data sharing among different families, and automatic search for possible family members. Ancestry.com provides a large genealogical database of immigrant and emigration records, census records, and military records. Although these sites are armed with useful features, there are issues when it comes to visualization and storytelling. While Ancestry.com provides standard family trees (Figure 3.3), Geni.com provides an unusual visual interface (Figure 3.4). In any case, they both suffer from the problems mentioned in the TimeNets chapter, including that the temporal dimension is omitted. For storytelling, they provide a timeline component that lists events involved with the given person. However, the timeline is disconnected from the family trees, meaning that there is no family context around events. In addition, no historical reference exists. As a result, it is quite difficult to convey compelling family stories.

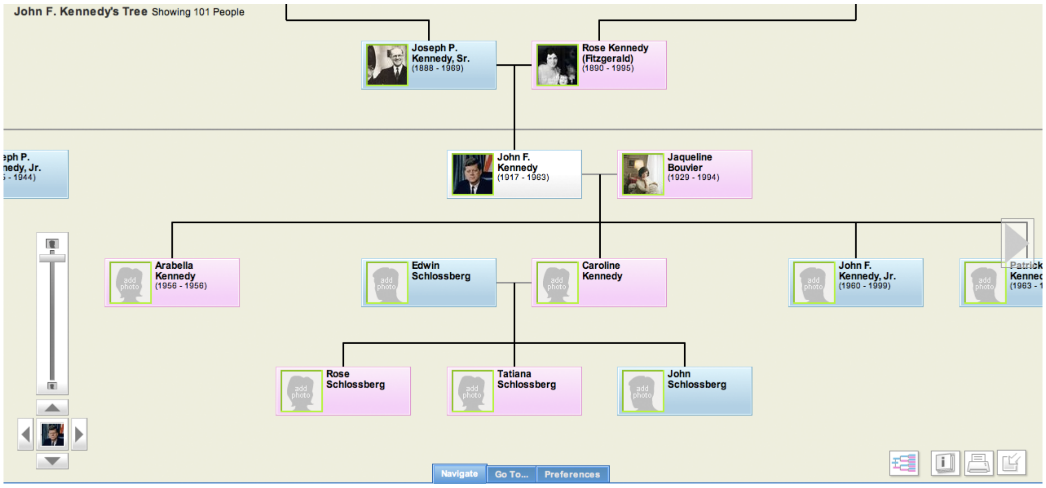


Figure 3.4: Geni.com:graph visualization with vertical depth signifying one's generation.

3.3 Design and Implementation

Akinus primary design goal is to provide a better visual interface using TimeNets in the context of collaborative genealogical research. TimeNets addresses the limitations of existing genealogy visualizations (refer to the third chapter for details) and also enables storytelling of family histories by embedding events into the visualization (Figure **??**).

Akinu is roughly modeled on well-known genealogy sites such as Geni.com and Ancestry.com. A user can collaborate with other family members by inviting them to an existing genealogy project. Recent activities of collaborators are also shown. The user can also create more than one genealogy project. Thus, the user has more freedom to create genealogies of their interest (e.g., the families of public figures). To facilitate transferring data from other genealogy software or web sites, GEDCOM is also supported.

The central activities on the site are creating TimeNets and adding events. Historical and family events are coupled with TimeNets to eventually convey rich family stories that are not possible on other genealogy sites. Users can share the TimeNets of their families off the site using a URL. In addition, TimeNets supports zooming and panning to enable interactive exploration.

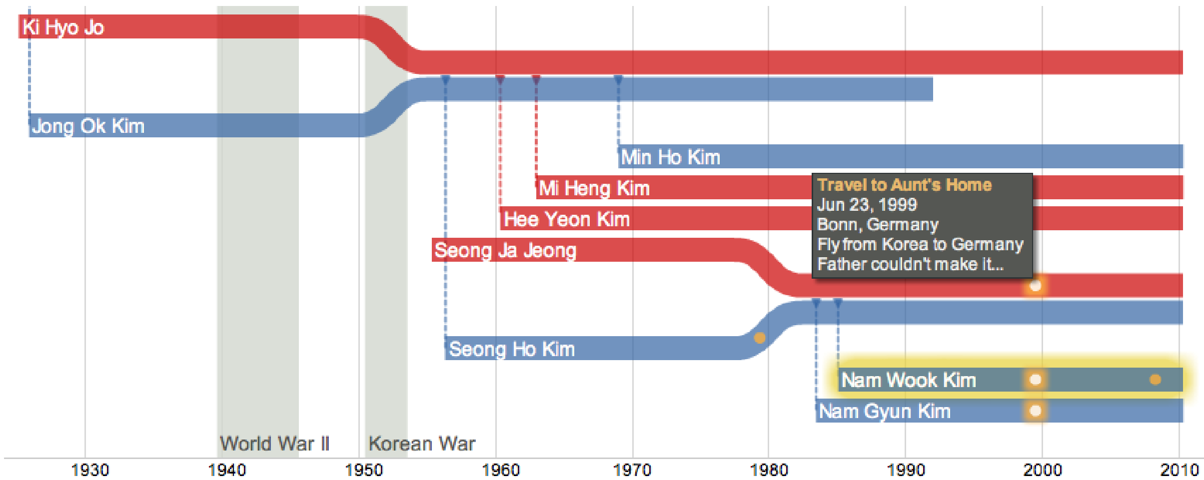


Figure 3.5: TimeNet visualization of the author's family.

Akinu is implemented by integrating TimeNets with web technologies. TimeNets was originally developed on the web. It previously used the Freebase web service and our own data format as a data source. Akinu builds the backend database so that users can upload data directly to the site. Three data types are supported: person record, event, and document. A document can have different forms, including audio, photo, and video. The document can then be referred to from more than one event. Similarly, each person record can incorporate more than one event. All of the genealogical records belong to a genealogy project.

The previous, static version of TimeNets has been modified to accept data on the fly; that is, Akinu supports direct data manipulation using TimeNets. A user can enter new data by clicking a persons lifeline and adding new relationships to the person or editing the persons information on the data input panel (Figure 3.6). Behind the scenes, the Cairngorm framework [2] is used to help data communication between Flex, in which TimeNets is implemented, and a Rails server. In addition, rubyAMF [35] is used to map a Rails ActiveRecord object to an ActionScript class. All other features, such as user account, invitation, and project page, are implemented using XHTML and Javascript that also communicate with the Rails server. The Rails server saves all the genealogical data into the MySQL database.

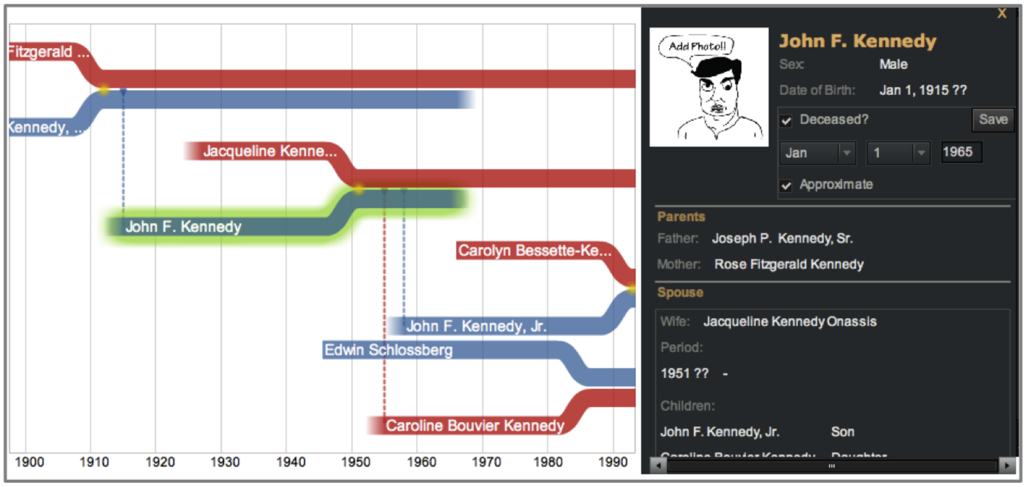


Figure 3.6: Akinu visual interface: to enter new data, first click a person's lifeline and add new relationships or edit the person' info on the data authoring panel on the right side.

3.4 User Study

We conducted two informal user studies to test the usability of the Akinu site and identify further improvements that should be made. All the studies were conducted on Amazons Mechanical Turk [28]. In the first study, we asked participants to create their family trees spanning at least three generations. Forty-seven subjects participated and contributed 376 person records. Most subjects did not submit precise dates (e.g., month and day are not specified), indicating that the subjects did not remember all of the exact birth or death dates of their family members. Even some

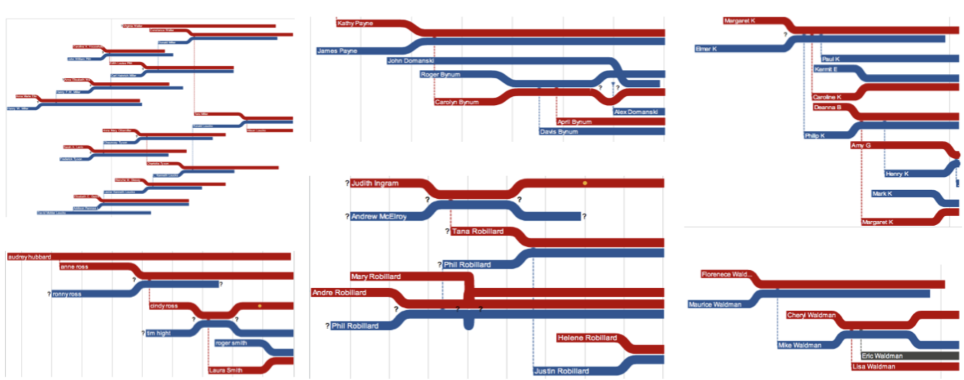


Figure 3.7: Example user-created TimeNets from the first study.

Negative	It was a bit confusing at the start
	Basically worthless. Really, of what use is a visual tree?
	The visual graph is kind of confusing. It is not user-friendly in that sense.
	The data entry was easy but it was kind of hard to figure out at first.
	The family tree is a bit confusing to understand, the layout is a bit unorthodox.
	It is rather hard to understand at first glance.
Positive	I love the look and ease of this program.
	It was a very pleasant and definitely a learning experience.
	Thanks a lot requestor for having such a task. Really appreciate it.
	I love the idea of this tool.
	I like the visual interface, it was quick to load which is a great plus.
	I love how it shows everything simply

Table 3.1: Representative User Study Tasks.

years were uncertain, as indicated by the uncertainty representation(in this study, question marks were used to indicate uncertainties). In the second study, we provided three generations of the Kennedy family tree as base data and asked participants to add a list of four people. Thirty-five subjects participated and contributed 169 person records. Unlike in the first study, people tended to enter precise dates, partly because they referred to online resources to complete the given task. However, not many people added more than requested, indicating less enthusiasm.

Even if the web site is still in an early stage of development, we were pleased that almost all the participants quickly became used to the visual interface. First of all, they enjoyed using it. An enthusiastic user of the first study contributed seven generations of her family by uploading 47 person records (see the top-left side on Figure 3.7). We also share selected user comments in Table 3.1, which have been categorized into positive and negative ones. Most users quickly understood the visual encoding of TimeNets, even with a brief introduction. However, some users still had trouble making sense of it. Nevertheless, most of them seemed to get used to the visualization once they started using it. From the studies, we also found a number of usability issues, including the absence of search and inefficiency when it came to adding relationships. These, along with other necessary improvements, are mentioned in the next section.

3.5 Conclusion

We have described the Akinu web site, which supports collaboration, visualization, and storytelling. To make this web site usable by millions, we have to address a number of usability issues. First, there is no support for fruitful interaction among family members. The site could benefit from a stronger set of community tools such as family discussion or family-friendly activity notification (e.g., your mom added a new family event). Second, we observed that a search interface is necessary to create a relationship from existing people records. A semantic visual search would be even more beneficial, for example, searching for a relative whose name is Christopher or highlighting first cousin once removed. A third natural extension is to support authoring step-by-step narratives. Additionally, borrowing from sense.us and Many Eyes, features such as graphical annotation, embedded discussion, and bookmarking views would be advantageous.

Chapter 4

Future Work

This thesis introduces a social genealogy platform that may benefit future research. A potentially interesting future direction would be studying the effects of visualizations on data collection behaviors. Previous visualization research has been limited to understanding how visualization affects people in analyzing data. By considering the process of sensemaking in which data is collected, organized and analyzed [3], the value of visualization has mostly been recognized as an output device (i.e. analytic tool). Although many existing systems integrate data collection with visualization [many eyes, swivel, wikimapia, geni.com], the question of how visualization influences data collection and refinement remains unanswered.

Heer [24] observed in sense.us that people tend to discuss unclear meanings or anomalies in data collection. This gives rise to the question of what would happen if people can directly modify the underlying data. In the discussion with Heer and Viegas [20], Wattenberg also envisioned that the future of social data analysis tools will consist not only of people engaging in data analysis, but in contributing data as well. In the same sense, there remains the interesting question of how visualization play a role in data collection activity. The question can range from as specific as how visualization affects the quantity or quality of entered data, to as general as what the relationship between visualization and data collection is in the context of collaborative sensemaking.

Relating to TimeNets and Akinu, an initial idea for future directions would be

Figure 4.1: uncertainty representations for a birth date (a) no marker (b) subtle gradient (c) salient question mark

Figure 4.2: uncertainty representations for a marriage date (a) no marker (b) subtle gradient (c) salient question mark

to study how the saliency of visual representations affects people in evaluating and correcting uncertain data (Figure 4.1 4.2). We hypothesize that people tend to fix uncertain attributes with salient question marks. A highly salient representation is expected to provide people with better awareness cues for uncertainties. On the other hand, high saliency might also annoy people or hamper their understanding the visualization. These things would encourage people to react against the uncertainties. For the same reason, the subtle gradient representation will work better than the absence of representation in terms of evaluating uncertainties. We believe that a reasonable compromise would be the subtle gradient because it not only provides awareness cues but also is aesthetically pleasing.

A similar idea can be extended to other collaborative systems. An example domain is a wiki system. A popular wiki system, Wikipedia, is a collaborative encyclopedia that can be edited by anyone. A number of visual analytic tools [37, 30]have been designed to support the analysis of social dynamics in Wikipedia such as editing activities (e.g., the number of contributors or revisions per article). Although the tools have proven to improve the interpretation or credibility judgment of articles, it is still an unexplored idea how such visualization affects the resulting quality or quantity of articles. Other systems that require quality control might also be interesting domains to study. In addition, to study the effects of visualizations on social data collection, any system that uses visualizations as a data-entry interface would be a promising platform (e.g., Wikimapia [42]).

Chapter 5

Conclusion

This thesis identifies shortcomings in the current paradigm of social genealogy web sites. Most genealogy visualization has focused on showing generational relations that are depicted as a hierarchical structure. While undoubtedly valuable, these omit other useful aspects of the data in particular time. In addition, the storytelling component lacks a family context, inhibiting rich storytelling. In this thesis, we have first described TimeNets, a timeline-based visualization for genealogical data. TimeNets prioritize temporal relationships and are able to show complex relationships such as divorce, remarriage, plural marriage, and out-of-wedlock births. In the controlled experiment comparing TimeNets against family tree diagrams, the results demonstrated that TimeNets facilitate temporal analysis tasks while also providing comparable performance for structural tasks. Second, we have introduced Akinu, a public genealogy web site for the collective curation of genealogical data. Using TimeNets as base representation, the site allows users to collaboratively build rich family stories that weave both family and time together.

Bibliography

[1] Ancestry.com.

[2] Cairngorm, June 2010. http://opensource.adobe.com/wiki/display/cairngorm/.

[3] Mackinlay J.D. Shneiderman B. Card, S.K. Readings in information visualization: Using vision to think. Morgan-Kaufmann, 1999.

[4] S. K. Card and D. Nation. Degree-of-Interest Trees: A component of an attention-reactive user interface. In *Advanced Visual Interfaces*, pages 231–245, 2002.

[5] S. K. Card, B. Suh, B. A. Pendleton, J. Heer, and J. W. Bodnar. Time tree: Exploring time changing hierarchies. In *IEEE VAST*, pages 3–10, 2006.

[6] Chartall, June 2010. http://www.chartall.com.

[7] Cousin, Dec 2009. http://en.wikipedia.org/wiki/Cousin.

[8] Cyndi's list of genealogy sites on the internet.

[9] Dabbledb, June 2010. http://www.dabbledb.com/.

[10] Catalina M. Danis, Fernanda B. Viegas, Martin Wattenberg, and Jesse Kriss. Your place or mine?: visualization as a community component. In *CHI '08*, pages 275–284, New York, NY, USA, 2008. ACM.

[11] Data360, June 2010. http://www.data360.org/.

[12] G. M. Draper and R. F. Riesenfeld. Interactive fan charts: A space-saving technique for genealogical graph exploration. In *8th Workshop on Technology for Family History and Genealogical Research*, 2009.

[13] J. Fails, A. Karlson, L. Shahamat, and B. Shneiderman. A visual interface for multivariate temporal data: Finding patterns of events over time. In *IEEE VAST*, pages 167–174, 2006.

[14] Freebase, Dec 2009. http://www.freebase.com/.

[15] G. W. Furnas. Generalized fisheye views. In *ACM CHI*, pages 16–23, 1986.

[16] The GEDCOM Standard Release 5.5, Jan 1996. The Church of Jesus Christ of Latter-day Saints.

[17] Genelines, Dec 2009. http://progenygenealogy.com/.

[18] Geni.com.

[19] Genopro, Dec 2009. http://www.genopro.com/.

[20] J. Heer. A conversation with jeff heer, martin watten- berg, and fernanda vigas. *ACM Queue*, 2010.

[21] J. Heer and M. Bostock. Crowdsourcing graphical perception: Using Mechanical Turk to assess visualization design. In *ACM CHI*, 2010.

[22] J. Heer and S. K. Card. DOITrees revisited: Scalable, space-constrained visualization of hierarchical data. In *Advanced Visual Interfaces*, pages 421–424, 2004.

[23] Jeffrey Heer and Maneesh Agrawala. Design considerations for collaborative visual analytics. In *VAST '07: Proceedings of the 2007 IEEE Symposium on Visual Analytics Science and Technology*, pages 171–178, Washington, DC, USA, 2007. IEEE Computer Society.

[24] Jeffrey Heer, Fernanda B. Viégas, and Martin Wattenberg. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In *CHI '07*, pages 1029–1038, New York, NY, USA, 2007. ACM.

[25] K. Ito and E. L. Rose. The genealogical structure of japanese firms: Parent-subsidiary relationships. *Strategic Management Journal*, 15, 1994.

[26] R. McClure. *The Complete Idiot's Guide to Online Genealogy*. 1997.

[27] M. J. McGuffin and R. Balakrishnan. Interactive visualization of genealogical graphs. In *IEEE InfoVis*, 2005.

[28] Amazon's mechanical turk, June 2010. http://www.mturk.com/.

[29] R. Munroe. XKCD #657, Dec 2009. http://xkcd.com/657/.

[30] Bongshin Lee Nathalie Henry Riche and Fanny Chevalier. ichase: Supporting exploration and awareness of editing activities on wikipedia. In *AVI '10*. ACM, 2010.

[31] P. Pirolli and S. K. Card. Information foraging. *Psychological Review*, 106:643–675, 1999.

[32] C. Plaisant, J. Grosjean, and B. B. Bederson. SpaceTree: Supporting exploration in large node link tree, design evolution and empirical evaluation. In *IEEE InfoVis*, pages 57–64, 2002.

[33] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: Visualizing personal histories. In *ACM CHI*, pages 221–227, 1996.

[34] J. Priestley. *A Chart of Biography*. J. Johnson, St. Paul's Church Yard, 1765.

[35] rubyamf, June 2010. http://code.google.com/p/rubyamf/.

[36] Spotfire, June 2010. http://www.spotfire.com/.

[37] Bongwon Suh, Ed H. Chi, Aniket Kittur, and Bryan A. Pendleton. Lifting the veil: improving accountability and social transparency in wikipedia with wikidashboard. In *CHI '08*, pages 1037–1040, New York, NY, USA, 2008. ACM.

[38] Swivel month = June, year = 2010, note = www.swivel.com/ .

[39] F. van Ham and A. Perer. Search, show context, expand on demand: Supporting large graph exploration with degree-of-interest. *IEEE TVCG*, 15(6):953–960, 2009.

[40] Fernanda B. Viegas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matt McKeon. Manyeyes: a site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, 2007.

[41] Martin Wattenberg and Jesse Kriss. Designing for social data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):549–557, 2006.

[42] Wikimapia, June 2010. http://wikimapia.org/.