

Visualizing Explicit and Implicit Relations of Complex Information Spaces

Marian Dörk, Sheelagh Carpendale, Carey Williamson
University of Calgary, Canada

ABSTRACT

In this work, we describe how EdgeMaps provide a new method for integrating the visualization of explicit and implicit data relations. Explicit relations are specific connections between entities already present in a given data set, while implicit relations are derived from multidimensional data based on similarity measures. Many data sets include both types of relations, which are often difficult to represent together in information visualizations. Node-link diagrams typically focus on explicit data connections, while not incorporating implicit similarities between entities. Multidimensional scaling considers similarities between items, however, explicit links between nodes are not displayed. In contrast, EdgeMaps visualize both explicit and implicit relations by combining graph drawing and spatialization techniques. We have applied this technique to three case studies (philosophers, painters, and musicians) and explored how integrated visualizations of explicit and implicit relations reveal novel patterns and relationships.

Keywords: Information visualization, explicit and implicit relationships, graph drawing, dimensionality reduction, aesthetics, design, WWW.

1. INTRODUCTION

An important goal of information visualization is to reveal different types of relationships within abstract data. Through interaction, the viewer can be enabled to find and understand connections between bits of information. Relations can be explicitly present in a data set as links that specifically connect information items or implicitly by inferring relations based on similarity of attributes. For both types of relationships—explicit and implicit—several visualization techniques have been proposed and refined over the recent years. Two of the most popular techniques for visualizing relations are node-link diagrams (NLD) and multidimensional scaling (MDS). On the one hand, NLD techniques are usually applied to explicit relations or connections that are visualized as edges between nodes representing, for example, online communities, computer networks, or linked web pages. On the other hand, MDS is typically used for implicit relations between documents or other types of multidimensional data. MDS spatializes attribute similarities between items by

placing similar items in close proximity to each other and less similar items further apart.

While NLD and MDS techniques are widely studied and increasingly deployed, they both have significant limitations with regard to readability and interpretability of the resulting visualization. Layout algorithms for NLD visualizations are typically optimized for reducing edge crossings, with the side effect that node positions are not utilized as a meaningful visual variable. Furthermore, as the number of edges increases, it becomes hard to distinguish directionality, if present, and identify high-degree nodes. The resulting visualization of an MDS algorithm, on the other hand, lacks a guiding structure to put elements into context with each other. It is often difficult to understand the meaning of the positional proximity of elements. We argue that the limitations of both techniques could be attributed to the fact that they are constrained to either explicit or implicit relations, yet, many data sets feature both types of relationships.

In this paper, we explore how both explicit and implicit relations can be visualized as *EdgeMaps*, integrated views that combine graph drawing and spatialization. EdgeMaps integrate NLD and MDS techniques utilizing both visual linkage and proximity for the representation of complex—explicit and implicit—relations between items. The intent behind this approach is to make effective use of visual variables that have been underutilized in NLD and MDS techniques.

As case studies for this paper, we have chosen data sets of philosophers, painters, and musicians from the Freebase data community. While there are many biographical records associated with these prominent personalities of philosophy, art, and music, a particular interesting aspect is the existence of influence connections between people, which are a type of explicit relations. On the other hand, birthdates, interests, movements, and genres are attributes that indicate implicit relations between philosophers, painters, and musicians. We chose these dimensions as they provide a compelling use case for the visualization of explicit and implicit relationships and let us explore complex data relationships. Visualizing influences between musicians or philosophers as edges may indicate who had more impact, yet, it is not possible with these links alone to

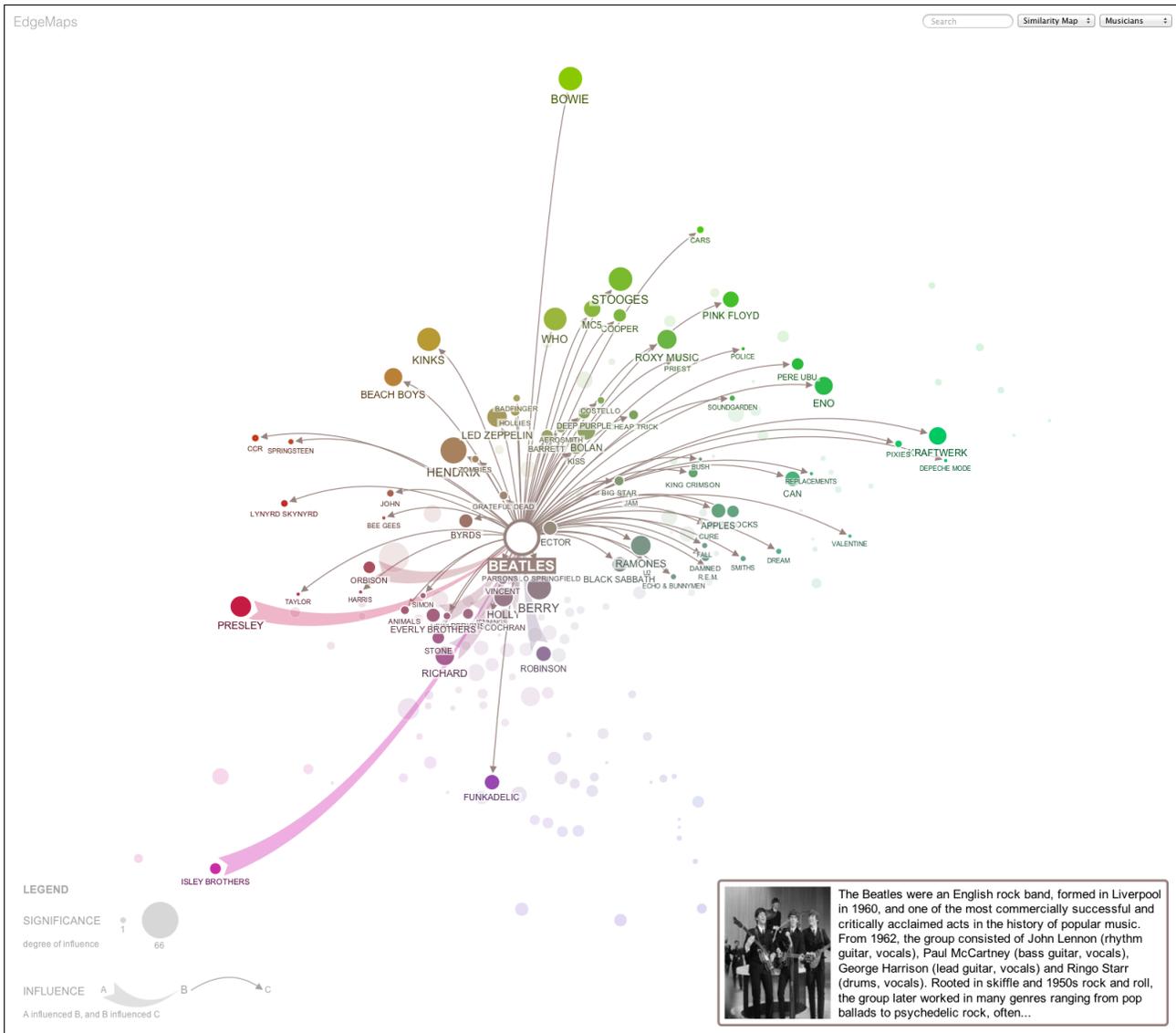


Figure 1. Visualizing relations among musicians. The influence of The Beatles is visualized in the similarity map.

see the extent of the impact. By encoding meaning into both position and edges, it becomes possible, for example, to explore the influence of musicians across genres or of philosophers over time (see Figures 1 and 2).

The remainder of the paper is structured as follows. First, we provide an overview of prior work (Section 2), after which we explain our design goals (Section 3) and the data sets we use as case studies (Section 4). We then introduce the visual representations provided with EdgeMaps (Section 5) and describe the web-based interface design (Section 6). Using the case studies, we illustrate new ways for exploring complex relations (Section 7). We then discuss the limitations and open questions of this work (Section 8) and conclude the paper.

2. RELATED WORK

As visualizing relationships is at the heart of information visualization, our work builds upon many previous contributions in the field, in particular with regard to the use of visual variables, graph drawing methods, and casual visualization.

While not part of his visual information-seeking mantra (“Overview first, zoom and filter, then details-on-demand”), Shneiderman notes the challenge of being able to explore relationships between information items.¹ He stresses the importance of relating data entries, however, equally if not more important are the appropriate visual representations of differ-

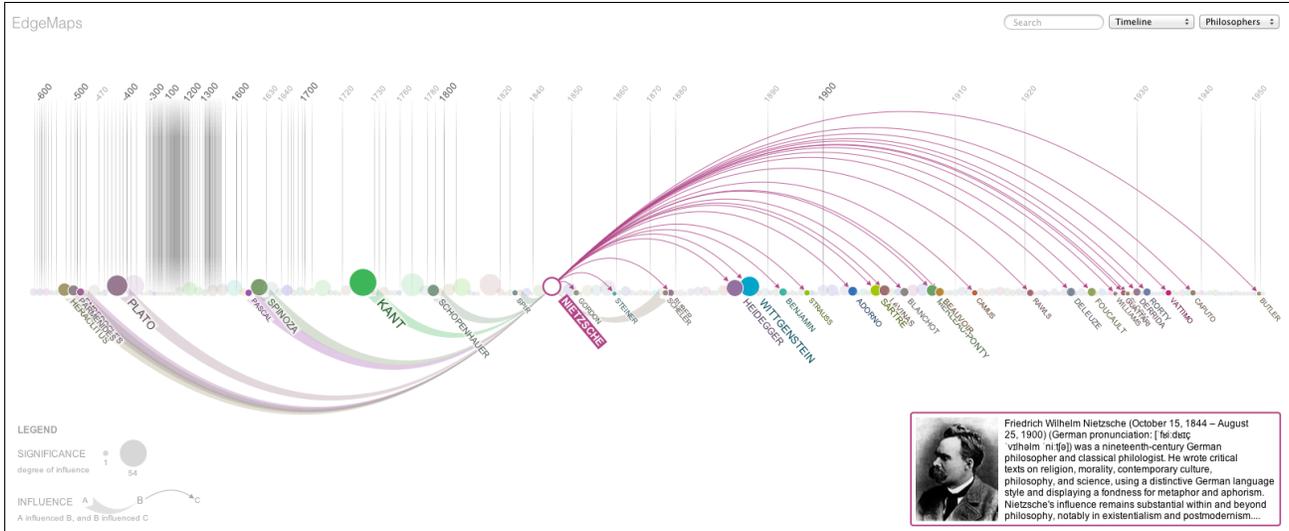


Figure 2. Visualizing influence relations between philosophers; Friedrich Nietzsche is selected in the timeline view.

ent types of relations. To think about representing relationships visually it is worth considering the visual variables that are at our disposal. In *Semiology of Graphics*, Bertin distinguishes between eight visual variables: size, value, texture, colour, orientation, shape, and the two dimensions for the position on the plane.² MDS renderings use planar position as the primary visual variable, while NLDs typically rearrange position in order to minimize edge crossings. Stone makes the case that colour can make visualizations more effective and beautiful when used well.³ She shows how colour can be used for labelling and quantifying data. It would be interesting to explore the use of colour for conveying similarity between items as a degree of association in Bertin’s terms.

There has been extensive research on drawing and interacting with NLDs,⁴ often aiming at reducing edge crossings, which is one of several geometrical and graph-theoretical metrics for graph aesthetics.⁵ Recent additions to this research include *EdgeLens*, a technique for interactively exploring overlapping edges,⁶ and *Edge Bundles*, a method for combining edges with similar paths.⁷ Another problem of large graphs is occlusion, especially when arrowheads of directed edges impair the perception of the actual nodes. A study of directed graphs examined a range of visual cues for directionality and their effect on determining direct and two-step connections.⁸ While these contributions can significantly improve the readability of large NLDs, we argue that contextual attributes of graph elements need to be more acknowledged. This perspective is supported in earlier work on computer network visualization, where edge

and node attributes (e. g., flow, capacity, utilization) of regional and international Internet links were regarded to be more important than the network topology.⁹ As part of a social network visualization it was shown how the visual representation of number of friends, gender, and community structure enriches the NLD and allows for interactive filtering.¹⁰

While conventional NLD techniques focus almost entirely on explicit relations, MDS can be seen as a complementary approach focussing on proximity as a visual representation of implicit relations or similarity. MDS has been used for document visualizations with the goal to visually convey “thematic patterns and relationships” of text collections.¹¹ While the idea of spatializing document collections based on their similarities or differences is promising, the resulting galaxies and themescapes still appear abstract and difficult to interpret. An approach to making MDS more interactive focused on steering the algorithm, but did not look at using interactivity to make the MDS view more meaningful and accessible.¹²

Besides linkage and similarity, another important relation is based on the temporal dimension and references between temporally structured items. Considering that time is generally seen as a linear dimension, the challenge is to visualize relations between items that are mapped onto a linear axis. Arc diagrams show cross-references along a linear axis by adding visual semi-circles to linear views of documents, music pieces, and DNA sequences.¹³ In work that explored the use of arcs to visualize email threads, it was shown how the combination of arcs displayed above and below the main axis

improved readability.¹⁴

Several contributions to information visualization have looked at enriching and combining techniques for visualizing different types of relations. For example, a visual document hierarchy was accompanied with arcs representing cross-references between different sections.¹⁵ Furthermore, NLDs were made more readable by assigning nodes into multiple regions and allowing for interactive edge filtering.¹⁶ A related approach places graph nodes on a multi-variate grid allowing for new forms of exploration.¹⁷ Another technique provides a three-dimensional juxtaposition of different visualizations on panes linking corresponding nodes using edges between the panes.¹⁸ All these techniques underline that there is a need for integrating explicit and implicit relations and enabling their interactive exploration, however, there is a tendency to emphasize one over another.

A related area within information visualization focuses on social, casual, and artistic aspects of visualization. For example, Many Eyes is an online community that allows web users to upload their data sets, create and customize visualizations, and discuss these with other community members.¹⁹ Many Eyes is one example among many web-based visualization projects that make visualizations more readily available within the web browser, which has become a major trend in visualization research and practice.²⁰ The development that visualization is becoming part of everyday activities, moving beyond professional, specialized domains is also described as casual information visualization, implying novel design challenges and audiences.²¹ One example of this development is the increasing use of visualization as a medium for artistic expression,²² which also aims at creating beauty and triggering curiosity besides supporting data analysis and sense-making. Against the backdrop of casual information practices with growing information spaces, considerations of aesthetics and curiosity also enter the domain of information seeking.²³

3. DESIGN GOALS

The motivation behind this work is the multitude of data sets that feature both explicit and implicit relations and preliminary research on complementing the visualization of one type of relation with aspects of the other. For this work, we understand *explicit* relations as data relations that specifically connect data entities and are already present in the data set. As *implicit* relations we see data relations that are not defined in the data set and need to be inferred based on similarities between data items.

Explicit linkage between items has been the mainstay of graph drawing research. On the other hand, there are numerous implicit similarities based on various parameters and dimensions that can be used for data spatialization. With this work we are exploring the space of integrating the visualization of both explicit and implicit data relations in order to reveal previously unseen patterns. In particular, we aim at supporting people in viewing complex data sets and exploring relationships between information entries in a pleasing and engaging way. This translates into the following design goals:

- **Integrate multiple relationships.** Explicit and implicit relations should be visualized as linkage and layout to mutually support each other.
- **Show invisible data patterns.** By integrating explicit and implicit data connections, the visualization should provide novel, interesting shapes and relations that were not visible before.
- **Support serendipitous exploration.** The visualizations should allow viewers to find unexpected insights and easily follow their interests. The interactivity necessary should be readily accessible without requiring training.
- **Display additional information.** The interface should provide detail-on-demand operations allowing the viewer to learn more about particular data entries and to go back to the data source.
- **Provide aesthetic visuals.** The colours, shapes, and transitions used by the visualization should satisfy both utility and visual appeal, making the interaction pleasant and evoking curiosity.

4. DATA SETS AND DIMENSIONS

To illustrate the notion of complex information spaces and explore their visual representation we choose personalities from philosophy, visual art, and popular music as exemplary data sets. We collected the data from the Freebase website using their web-based API.²⁴ Freebase offers structured information about many entities. In the case of philosophers, painters, and musicians, it provides data about birthdates, influence connections, and a range of domain-specific keywords such as philosophical interests, artistic movements, and musical genres. Besides programmatic methods of accessing the data, Freebase also publishes tabular views of individual data entries on their web site (see Figure 3). As the group of musicians has only sparse influence connectivity on Freebase, we complement our data set by influence information from the AllMusic website.²⁵

| Simone de Beauvoir topic | |
|--|-----------------------------|
| Also known as: Simone de Beauvoir edit | |
| Simone de Beauvoir (pronounced [simon da bo'vwair] in French) (January 9, 1908 – April 14, 1986) was a French author and philosopher. She wrote novels, monographs on philosophy, politics, and social issues, essays, biographies, and an autobiography... more | |
| Read article at Wikipedia Write new description for Freebase.com | |
| People | |
| Person | edit |
| more options | edit |
| 10 empty fields | edit |
| Date of birth: | Jan 9, 1908 |
| Place of birth: | Paris |
| Country of nationality: | France |
| Gender: | Female |
| Profession: | Writer |
| | Philosopher |
| Philosopher | edit |
| more options | edit |
| 2 empty fields | edit |
| Era: | 20th-century philosophy |
| Interests: | Ethics |
| | Politics |
| | detail view |
| Influenced By: | Edmund Husserl |
| | Søren Kierkegaard |

Figure 3. One example data set for visualizing implicit and explicit relations is philosopher data from the Freebase community. Here is the entry about the French philosopher Simone de Beauvoir: description, birthdate and profession (top), interests and influences (bottom).

We focus on philosophers, painters, and musicians who influenced at least one other member of their corresponding group. The selection is also based on the availability of keywords corresponding to the selected similarity criteria that are most meaningful for the given group of people. For philosophers, we include their philosophical interests (e.g., ethics, logic, nature) and professions (e.g., writer, physicist, teacher). The keywords of painters include art forms (e.g., painting, mural, fresco) and artistic movements (e.g., renaissance, impressionism, surrealism). For musicians, we included their genres (e.g., jazz, punk, electronic).

Based on the requirements of keyword and influence information the resulting data sets comprised 196 philosophers, 226 painters, and 200 musicians. For each person we keep the name, birthdate (for musical groups, year of formation), description, an image, and the keywords. Furthermore, we store the directed influence links between people, which we consider as explicit relations, and the keywords, which form the basis for computing similarities among people as implicit relations. The dates can also be regarded as another type of implicit relation as it implicitly links people of similar time periods or epochs. For storing these records, we use a MySQL database easily accessible from the server-side component of our system, which was written in PHP.

Using these attributes, several interesting dimensions can be inferred. For example, we can calculate the de-

gree of influence as a measure of significance using the sum of outgoing influence connections. In other words, the more people a particular person has inspired and influenced, the more significant this philosopher, painter, or musician is. Based on birthdates, people can also be grouped into similar epochs. Likewise, using domain-dependent keywords philosophers, painters, and musicians can be grouped by their similarity. Combining time periods and keywords with the influence connections, one can look at the impact of, for example, musicians across epochs and interests. For example, considering all the musicians that were influenced by The Beatles, their impact extended over a wide range of musical genres (see Figure 1). As Friedrich Nietzsche was considered influential by many subsequent philosophers, his impact is one of great temporal extent (see Figure 2). By juxtaposing and integrating both explicit and implicit relations it is possible to reveal these relations, which are not easily accessible by just looking at data tables or reference pages.

5. VISUALIZING EXPLICIT AND IMPLICIT RELATIONS

EdgeMaps provide a visualization that represents both explicit and implicit relations by integrating spatialization and graph-drawing techniques. Explicit relations are encoded as curved edges and implicit relations as node position. Other visual variables are used to double-encode these data relations and introduce additional information such as directionality and distinctness. In the following, we describe our design considerations and decisions concerning visual representation.

5.1 Implicit Relations as Layout

To represent the implicit relations, we designed two general layouts: a similarity map and a timeline, as illustrated in Figure 4. While both layouts represent people as nodes and influences as edges, as we will describe in more detail later, the layouts differ in the way the positions on the plane are utilized. The similarity map represents the topical nearness among people based on their domain-specific keywords such as interests, movements, and genres. The timeline uses birthdates as an ordering criteria to arrange philosophers, painters, and musicians along a temporal axis. While neither of these layouts are new by themselves, they provide new context for graph visualizations.

Similarity map. The node positions for the similarity map were generated using an MDS algorithm²⁶ on the basis of the people’s keywords. To do this, the keyword data is transformed into a vector model that

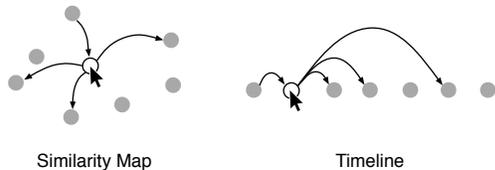


Figure 4. Layouts for visualizing explicit and implicit relations based on similarity (left) and birthdates (right).

the MDS algorithm uses as a basis to map relative distances in the high-dimensional vector space to distance in the two-dimensional plane. The result is a pair of x,y coordinates for each philosopher, with each coordinate value between -1 and 1. Depending on the size of the window, these coordinates are then scaled to the actual display resolution. However, the aspect ratio of the MDS output is not modified, since the proximity of the resulting plane configuration is generated on the basis of similarity. Stretching the layout would confound the representation of similarities, which are mapped to positions on the plane using their Euclidian distances. As there may be people with identical keywords, the MDS algorithm can return items with the same positions, posing an occlusion problem. Considering that MDS is an approximation after all, the similarity map places overlapping nodes slightly apart so that they are still close but not occluding each other. The positions in the similarity map are used to generate unique colours for all circles as discussed in more detail later. The resulting layout is shown in Figure 5.

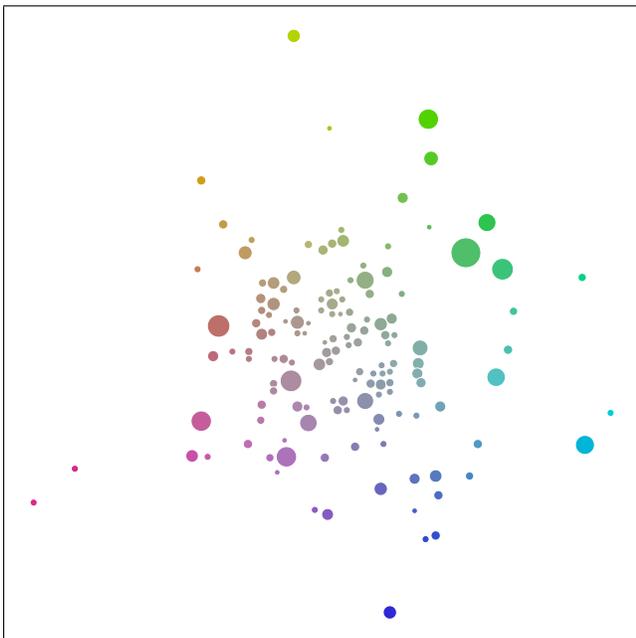


Figure 5. Similarity map of philosophers without selection.

Timeline. The timeline maps birth year to position on the plane. Initial trials with a linear scaling were not satisfying, because of dense clusters of people in some time periods and very sparse or empty periods during other times. We decided to neglect exact temporal intervals and focussed on temporal sequences of people instead. This still allows for relative temporal comparisons of ‘earlier’ and ‘later’ people and, at the same time, accommodate all people’s nodes along the time axis. The result of this ‘stringing’ of nodes along the axis has the effect that certain periods take up much more display space than others. A time legend displayed on top of the time view and a subtle grid representing decades and centuries indicate this temporal folding. As the data for the philosophers was especially sparse between about 300 B.C. and 1200, there is a clustering of grid lines in this particular time area. The resulting timeline is shown in Figure 6.

5.2 Explicit Relations as Curved Edges

The layouts for topical and temporal similarities represent only the implicit relations between data items. In order to represent explicit relations, i. e., influences between philosophers, painters, and musicians, directed edges are drawn between the nodes. However, if all edges were to be shown for all nodes at the same time, there would be far too many edges to be readable, let alone interpretable. Instead, by activating only one node at a time, it is possible to read individual edges and differentiate between two types of influence:

- *Incoming influence.* The person is inspired or affected by other people and builds upon their work.
- *Outgoing influence.* The person has inspired or affected other people who built upon their work.

A recent study on representing directed edges suggests that simple visual cues better help detecting connections between two nodes than curves or arrowheads.⁸ However, our aim is to reveal the extent of many connections of one node with the all the remaining nodes and discern incoming and outgoing influences. Incoming influence can be seen as the basis of a person’s work, the outgoing influence is the impact they had on other people. To visually differentiate between these edge types, we used curvature, directionality, shape, opacity, and colour (see Figure 7).

Curvature. Incoming edges are curved downward and outgoing edges are curved upward. The idea is that incoming influence stands for the foundation upon which a philosopher, painter, or musician builds their

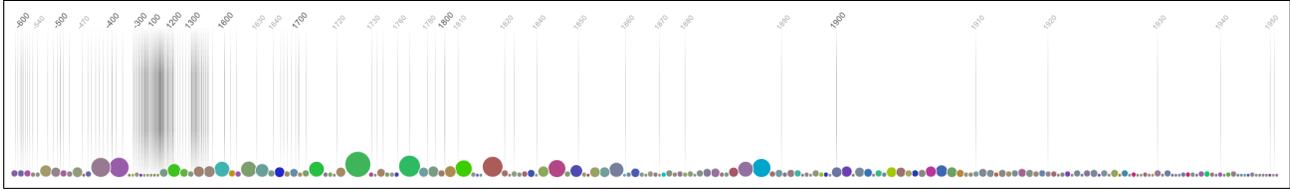


Figure 6. Timeline of philosophers without selection.

work. For the outgoing influence, the edge is curved upward, as a visual depiction of reach beyond previous work. Taken together, both types of edges form a wave-like shape when nodes are arranged in a sequential order. The extent of curvature of an edge depends on its distance between the connected nodes. This way, the wider the reach of an influence, the more salient the edge appears. To reduce the visual footprint of edges in the similarity map, the curvature is slightly smaller than in the timeline view.



Figure 7. By selecting one node at a time, we can distinguish between two types of edges: incoming, B is influenced by A (left), and outgoing, B influences C (right).

Directionality. Incoming and outgoing edges differ also with regard to how directionality is represented. As we assume that only one person is activated at a time, it is most likely that this person’s node will have multiple incoming and outgoing edges, while all other connected nodes will mostly have only one associated edge or in some cases two edges (if the influence is mutual). This indicates that to most clearly represent edge directionality, there is much less clutter around the linked nodes rather than around the activated node. We exploit this fact, by situating the visual representation of directionality at the linked, non-active nodes: incoming edges have arrow-like cuttings at their source nodes, whereas the outgoing edges coming from the active node have arrows at the destination nodes. This ensures that the edge endings at active person’s node are simple and thin allowing for many discernible edges.

Shape. An additional way to differentiate the edge types is by their shapes. Incoming edges are drawn thicker than the outgoing edges. While there is no inherent reason why incoming edges should be thicker than outgoing edges, giving the different types of edges these two distinct shapes allows for easy distinction when multiple edges are displayed at the same time.

Opacity. To balance out the different visual weights resulting from thick and thin edges, the opacity of the incoming edges is decreased. Using opacity instead of adjusting brightness for the incoming edges also reduces the occlusion of other elements due to incoming edges.

Colour. Since the nodes have different colours, the colour of an edge corresponds with the colour of the node, where the influence is coming from. The outgoing influence edges take on the colour of the selected person, and the incoming edges share the colour with the originating node. The mapping of nodes’ position to their colour is discussed in the following section.

5.3 Redundancy through Size and Colour

EdgeMaps visualize data sets primarily using position to represent implicit relations (time and similarity) and edges to represent explicit relations (influence). As secondary visual variables we include node size and colour to accentuate significance and similarity of date items.

We understand significance of a person as the relative degree of their outgoing influence. The rationale is that a philosopher, painter, or musician who had more influence on their peers and successors is arguable more significant than those that had less impact. Outgoing influence can be viewed by selecting the corresponding node and viewing the outgoing edges. We decided to additionally expose the significance of a node without any interaction, by encoding significance of a person as their circle’s area. We experimented with different types of glyphs, however, we decided that simple circles with varying sizes would allow for sufficient data encoding without adding significant perceptual complexity.

Furthermore, we decided to use colour to double-encode similarity based on people’s keywords. Since the output of the MDS algorithm provides a spatialization of this relationship, we used it in combination with the HSV (hue, saturation, value) colour space as the basis for the colour calculation (see Figure 8). We decided to map broad interest regions to hue and the distance to the centre to saturation, while keeping the brightness constant. The idea behind this encoding is that the further out a person is located, the more distinct they are from all other people. Translating this

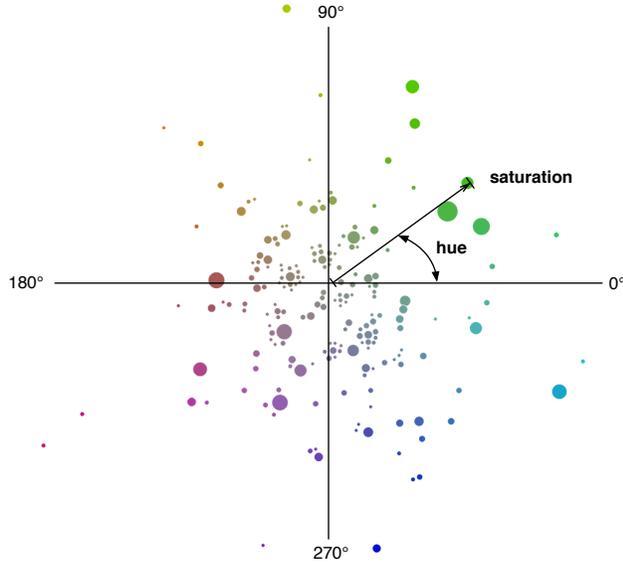


Figure 8. Colours are derived from MDS positions using distance from centre for saturation and angle for hue.

notion of distinctness to a colour space, it would seem intuitive that the more distinct items would be more saturated and the more common items would be less saturated. However, keeping the value constant means that either the nodes in the centre would be too bright or the nodes on the periphery too dark. Therefore, we took the distance to centre into account for the brightness value resulting in well-visible, grey-like nodes in the centre and colourful nodes towards the outside.

Based on this approach, nodes receive distinct colours, which are also used for their outgoing influence edges. As shown, for example, in Figure 9, bottom, the active philosopher's outgoing edges have the colour of this node and the incoming edges have the colours of their respective philosophers. This also helps to distinguish edge types and associate edges with corresponding source nodes. Furthermore, as colours and sizes are consistently used between layouts the viewer may better recognize previously visited nodes.

5.4 Yarn Balls vs. Fireworks and Waves

After having discussed the visual representation of implicit and explicit relations individually, we will now discuss how layout, edges, and additional encodings come together in novel formations.

While exploring several design options, we examined the possibility of activating multiple nodes at the same time. The result is an example of the “yarn ball” effect for complex NLDs, conveying neither overview nor structure (see Figure 9, top). In contrast, displaying

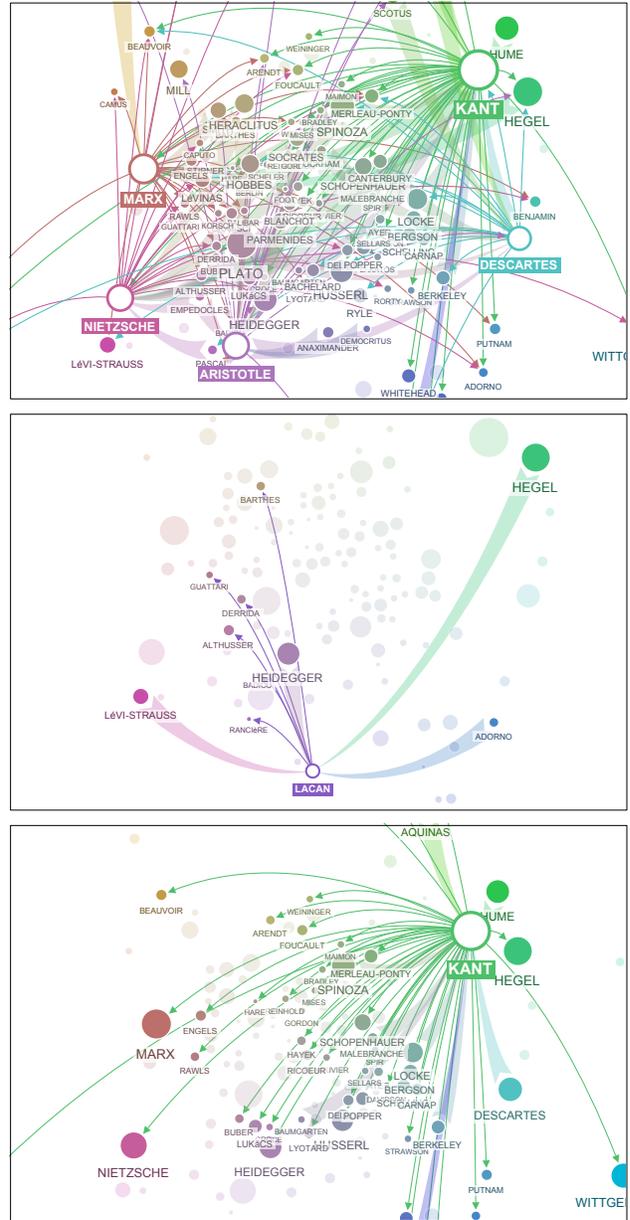


Figure 9. Displaying influence edges in the similarity map for five philosophers (top) and one philosopher at a time (middle and bottom). With a single activated philosopher, influence edges arguably resemble the shape of fireworks. Less influential philosophers have sparser edge patterns (middle) than more influential ones (bottom).

only the edges associated with one individual node allows for interesting visualization and interaction possibilities (see Figure 9, middle and bottom). As discussed before, in this way it becomes possible to better represent edge types and directions. Furthermore, considering that the node size reflects the number of

outgoing edges, differing node sizes may be a more effective way to convey the general overview of a data set. Instead of displaying many edges that make readability more and more difficult with increasing number of nodes and edges, it appears to be more effective to use colour, position, and size for nodes to provide context and overview.

An interesting structure emerges when selecting a node in the similarity map with incoming and outgoing edges rendered in patterns of fireworks (see Figure 9, middle and bottom). Relatively less influential people (e.g., Lacan) result in more modest fireworks than the most influential people, such as Kant. Besides the number of edges, the spatial extent of edges conveys the topical scope of a person’s incoming and outgoing influences. In the temporal layout, the influence edges lead to distinct wave-like patterns (see Figure 10). The resulting edge layout is particularly interesting as the waveform can also be seen as an engaging representation of the dynamic evolution of philosophical ideas, artistic movements, and musical genres.

In the similarity map, the spatial extent of edges stands for topical scope and in the timeline the extent of edges represents temporal scope of influence. It becomes possible to distinguish influences within close temporal or topical proximity from influences that span multiple time periods and similarity regions. There is a qualitative difference in influencing people of the same time with similar characteristics than influencing people distributed over a larger time interval across a range of interests, movements, or genres. For example, the influence between two jazz singers is different from the influence that a blues musician has on a punk musician. In EdgeMaps these differences are revealed by the position, distribution, and length of edges, and the colours and positions of associated nodes.

6. CREATING A WEB-BASED VISUALIZATION INTERFACE

In addition to the representation of multiple types of data relations, it is important to consider how the interaction and visual design of the interface can be realized to support the exploration of complex relationships. In the following, we describe how we designed a web-based interface for EdgeMaps aimed to support established information visualization principles and integrate the tool into the web-browsing experience.*

*Demo online: <http://mariandoerk.de/edgemaps/demo>

6.1 Interacting with the Visualization

The interface provides a range of interaction techniques for manipulating the view, data, selections, and details. The primary way of interacting with the EdgeMaps visualization is selecting a node representing a person and revealing the corresponding influence connections coming towards this node and going out to other nodes. The viewer can then change the selection by either selecting another node (possibly one that is associated with the current one) or by deselecting this node by clicking on the background or clicking again on this node.

A text search allows the viewer to search through names, descriptions, and keywords. When a search term is entered, the visualization shows only the labels of the nodes representing data items that are matching the search query; the opacity of the remaining nodes is dimmed and their labels are not shown. This way it is possible to quickly see where, for example, painters of a certain genre are situated, making the underlying similarity layout easier to interpret (see Figure 11).

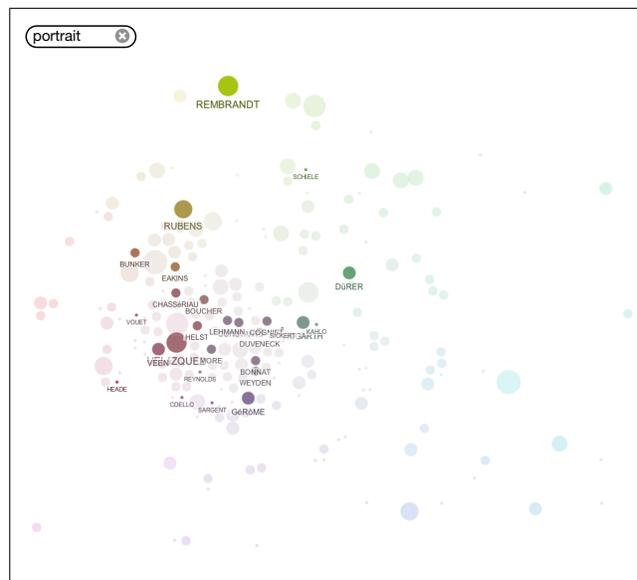


Figure 11. The text search allows the viewer to explore, for example, the distribution of painters of portraits.

Additional controls are provided in the top right of the interface as drop-down menus to change the layout and data set of the visualization. Changing the data set initiates a simple blending transition from one visualization to the next. Switching between layouts triggers a transition between the similarity map and the timeline gradually moving the nodes into their new locations. As the colours are based on similarity it is possible to rediscover nodes from the similarity map in the timeline view. The transition between the views is

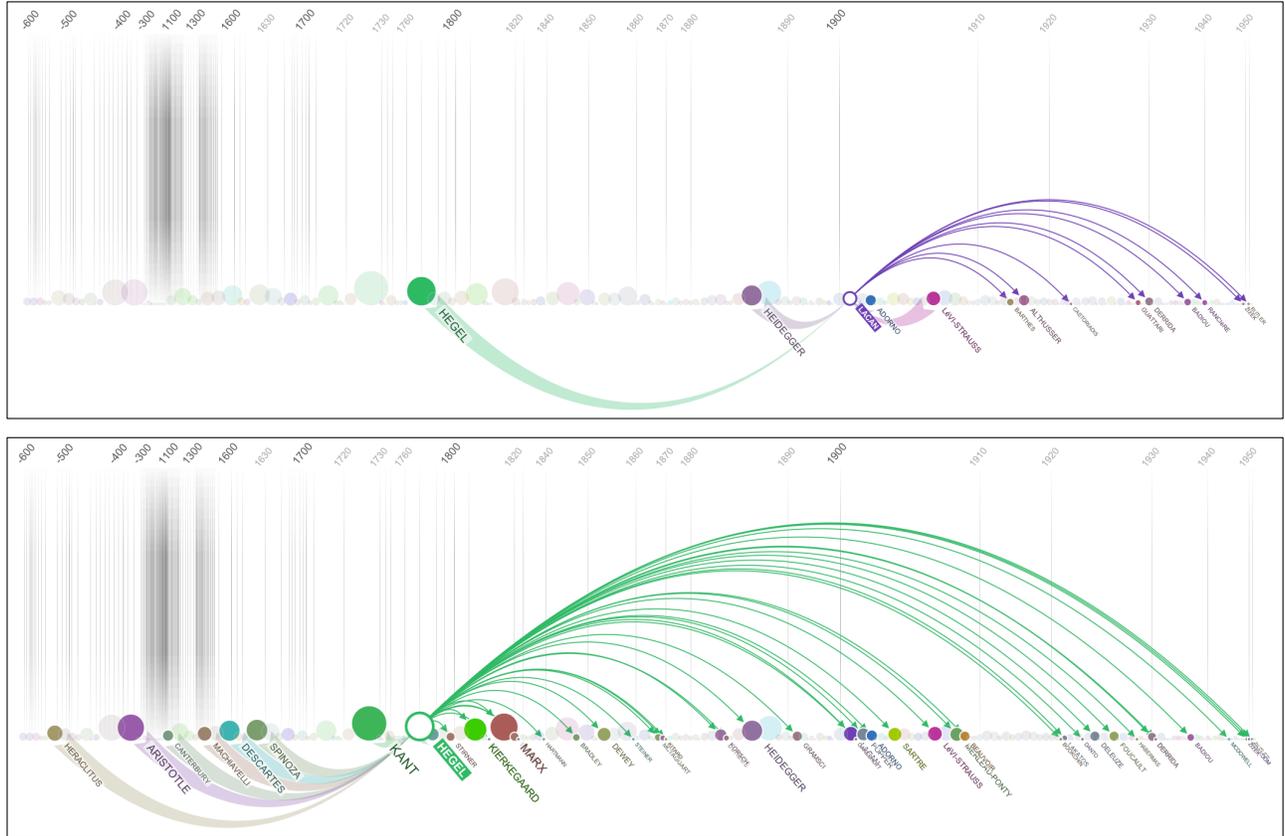


Figure 10. In the timeline view, influence edges result in a wave-like form indicating the propagation of philosophical ideas over many time periods. The more significant a philosopher is the larger is the resulting wave (below).

animated to allow the viewer to follow nodes between the views and establish an understanding of temporal and topical interrelations. For example, by comparing how these transitions unfold it is possible to visually examine whether keyword-based similarities correspond with time periods people were active in.

To learn more about a person, a detailed view with additional information about the selected node is displayed in the lower right portion of the window (see Figure 12). A short description of the person is displayed and accompanied with a visual depiction (e.g., photo, painting, sculpture) of the person. Selecting the thumbnail opens the corresponding entry on Freebase allowing the viewer to go to the data source and learn more about a given person. The border of this detail display uses the colour of the corresponding person’s node in the EdgeMaps visualization.

6.2 Visual Information Design

As an attempt to reduce visual clutter, only the name of the person that is currently selected and the people associated with the selected person are displayed.

The nodes for the remaining people are dimmed and do not have a label. This allows the viewer to focus on the current selection, but it also alleviates a label-occlusion problem that still occurs when many edges of associated nodes are displayed (see Figure 9, right). It is possible to hover with the mouse over any node to make occluded or hidden labels visible. To ensure aesthetic proportions between circles and labels, the font size is set relative to the size of the circle. Furthermore, only the surname of the person is used, which is typically unique and sufficiently known.

Legends are displayed to summarize the different mappings that are used. For example, for the size of circles and types of edges, a legend for both layouts is always displayed in the lower left of the screen (see Figure 12, lower left). The two circle sizes displayed correspond to the smallest and largest nodes in the visualization, giving a sense of the extent of significance between people. In the timeline view, years indicate the temporal distribution along the time axis. The legends are drawn in shades of light grey to avoid distraction from the information visualization.

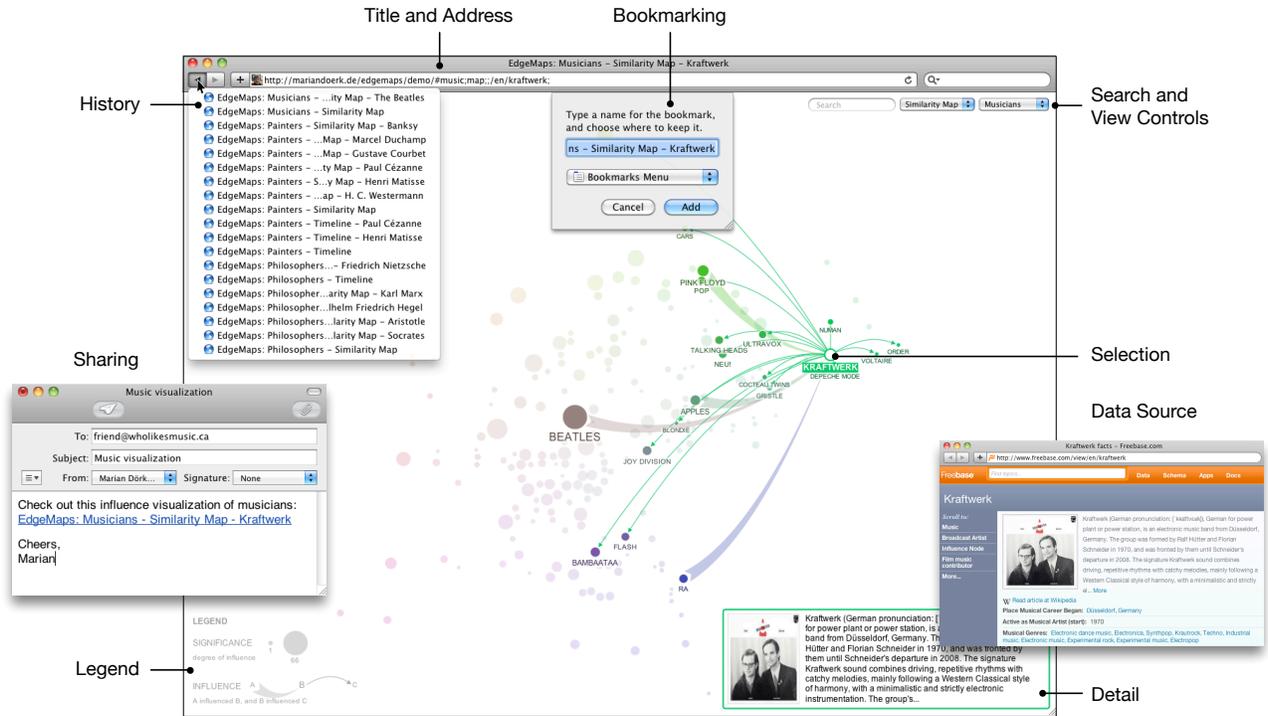


Figure 12. The web-based interface of EdgeMaps ties in with common browser functions and use patterns on the Web.

6.3 Web Integration and Native Graphics

One goal behind realizing EdgeMaps as a web-based system, was to align the interactivity with existing use patterns supported by web browsers. As shown in Figure 12, our prototype supports the browser’s history, bookmarking, and sharing features. The back button in web browsers provides a common way of backtracking on the Web allowing the surfer to return to a previous page; it serves a similar function in the EdgeMaps tool, the viewer can use it to return to a previous state of the visualization. The state of the visualization, i. e., the type of layout, data set, selection, and search query, is encoded into the address as part of the fragment identifier of the URL. In this way the interaction steps can be added to the browser’s history, a given state can be bookmarked or shared as a link via email, instant messaging, or on social networks. We also adjust the title to reflect the current state of the visualization to make the history, bookmarks, and link titles easy to understand. For example, when dragging the address into a rich-text document, modern browsers and word processors automatically turn it into a hyperlink with the corresponding title (see Figure 12, bottom left).

To draw interactive visualizations in the browser, there are several options. On the one hand there are browser plug-ins, such as Flash and Java, that are used

traditionally to provide advanced graphics and interactivity. On the other hand, new web standards—in particular HTML5—are currently emerging that strive for native support of interactive graphics in the browser. We are interested in exploring the browsers’ native, standards-based graphics capabilities and in not relying on proprietary technologies. This approach has also the benefit that the tool would be viewable on mobile devices that do not support browser plugins (mainly Apple’s iOS devices). The two main options for browser-native graphics are the Canvas element for bitmap drawing and SVG for scalable vector graphics. As native browser graphics are currently going through rapid developments, we considered both approaches. To make a quick performance comparison between Canvas and SVG, we have drawn 1000 random circles with differing opacities, positions, and sizes. In this experiment the Canvas element was about twice as fast in terms of rendering than the SVG. While the output of both approaches looks identical, browser-based SVG allows for straightforward event handling, as it retains its own database of drawn elements. In contrast, the Canvas element requires custom picking logic for the objects to be interacted with, which would lead to significant performance overhead. Based on these considerations we chose SVG in combination with the vector graphics JavaScript library Raphaël.²⁷

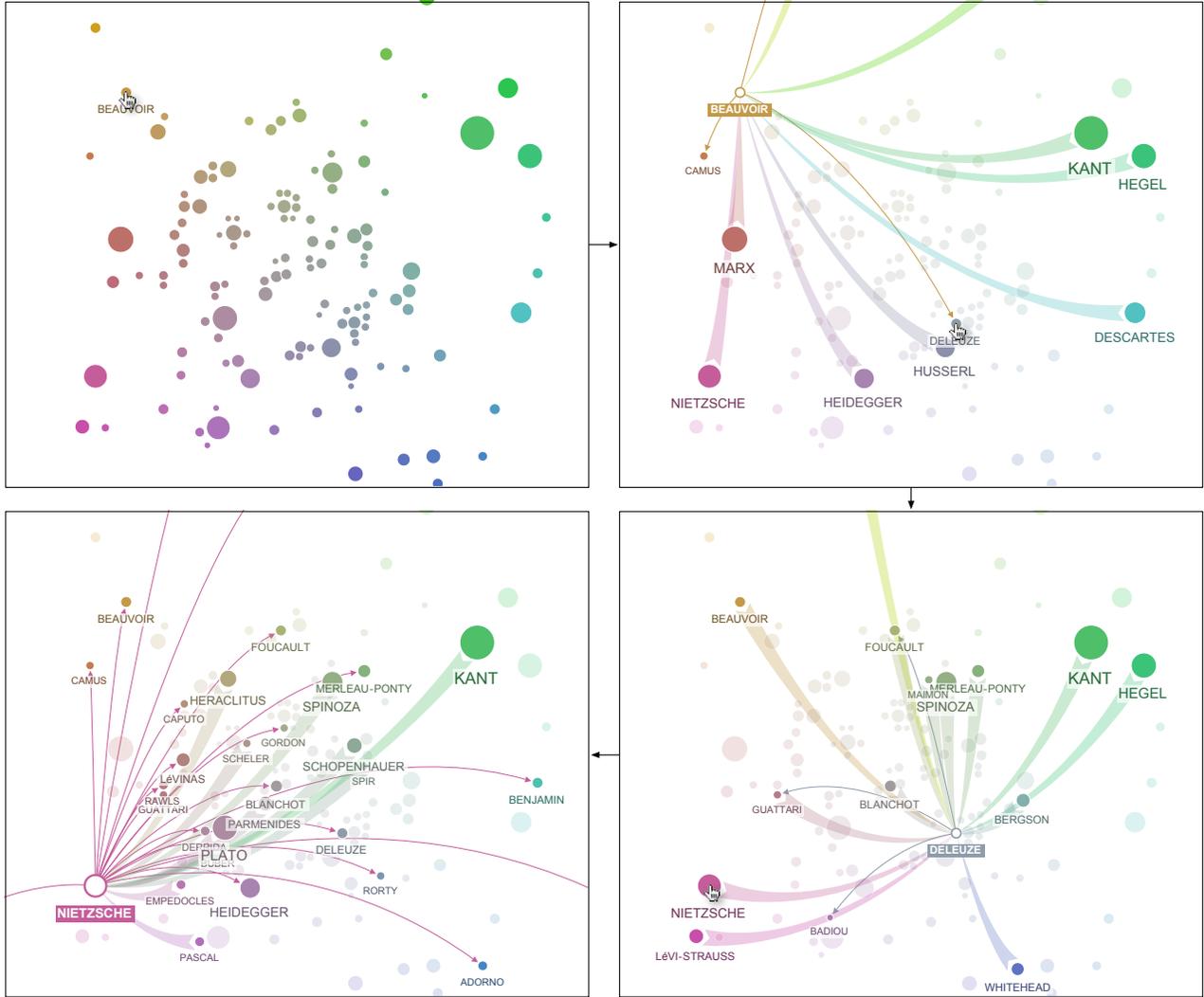


Figure 13. Exploratory pivoting along philosophers' influences from Beauvoir over Deleuze to Nietzsche.

7. REVEALING COMPLEX RELATIONSHIPS

By visualizing explicit and implicit relations, EdgeMaps allow the viewer to explore data patterns that are difficult to see otherwise. In the following, we illustrate some of the new kinds of exploration that can be pursued and the types of insight that can be gained.

Pivotal Data Exploration. Limiting selections to one node at a time has interesting implications on the interaction with the visualization. Activating one person draws the incoming and outgoing edges, highlights the respective nodes, and displays their names. As the linked nodes are emphasized, the viewer is more likely to follow the displayed edges and activate one of the linked people. In a sense, exploring the visualization

along influence edges can be seen to be like pivoting through the data set around one person at a time. Each step triggers a novel visualization sharing the positioning of the former state, yet providing a different influence networks centred around the current node. Consider selecting, for example, the node of the philosopher Beauvoir, which is one of the smaller circles in the periphery of the visualization (see Figure 13, top). After having selected her node it is now possible to follow one of the philosophers that was influenced by her. In this case one could select Deleuze (bottom right) and afterwards a philosopher with a larger and more saturated node, in this case Nietzsche (bottom left). The path of exploration depends somewhat on serendipity, intuition, and interest, all of which are affected by the overall visualization and the viewer's inclinations.



Figure 14. EdgeMaps reveal different patterns of how musicians influence each other. There is a difference between reaching members of similar genres (top left) versus reaching further across the similarity map (top right). Influential and innovative musicians tend to show a spatial split between incoming and outgoing influences (bottom left and right).

Qualitative Differences in Influence. The influence edges reaching across the view generate visual patterns that can help develop a rich understanding of a person’s influence. For example, the similarity map of the musical artists’ data allows us to see the reach of a musician’s influence. Taking a closer look at the influence network of The Beatles shows that their influence spans most regions of the similarity map except those inhabited mostly by jazz, country, and rap singers at the bottom of the visualization (see Figure 1). In contrast, the influence network of the country singer Hank Williams appears much more constrained to his immediate neighbours (see Figure 14, top left). Furthermore, it is possible to compare the relative location

of nodes that are connected by incoming and outgoing influence edges. For instance, John Coltrane appears to be influenced by musicians of similar genres, while his influence reached relatively far and wide across the similarity map (see Figure 14, top right). The influence networks of Jimi Hendrix and The Ramones are in each case located between the musicians they got influenced by and the singers and bands they influenced (see Figure 14, bottom left and right). Taking into account that Jimi Hendrix is considered a pioneer of the electric guitar and The Ramones are seen by many as the first punk-rock band, it makes intuitive sense that their incoming and outgoing influences are disparate groups of musicians on the similarity map.

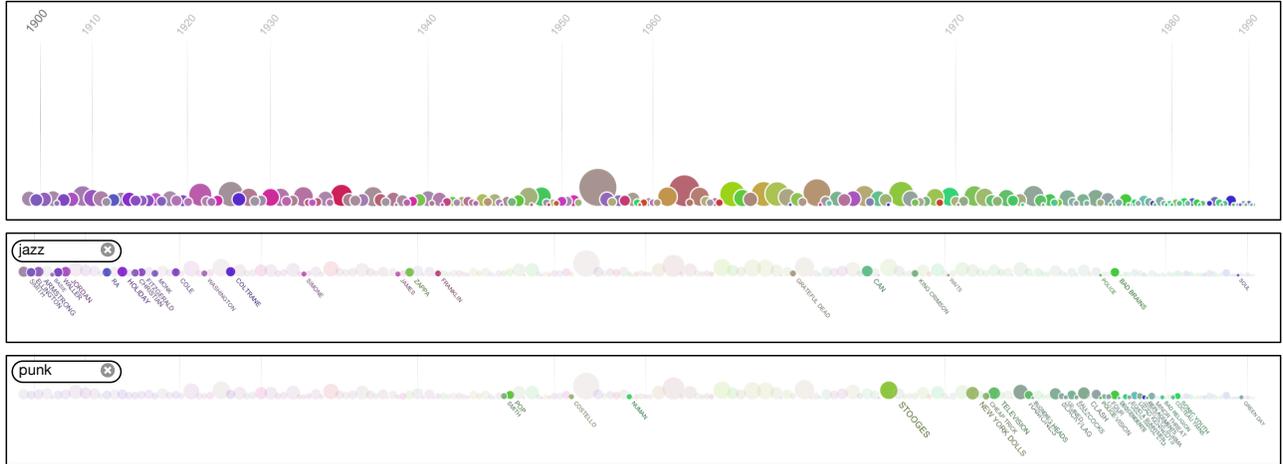


Figure 15. The gradual colour changes in the time layout of musicians suggests a correlation between genres and time periods. This observation can be examined more closely by highlighting nodes matching search terms of genres.

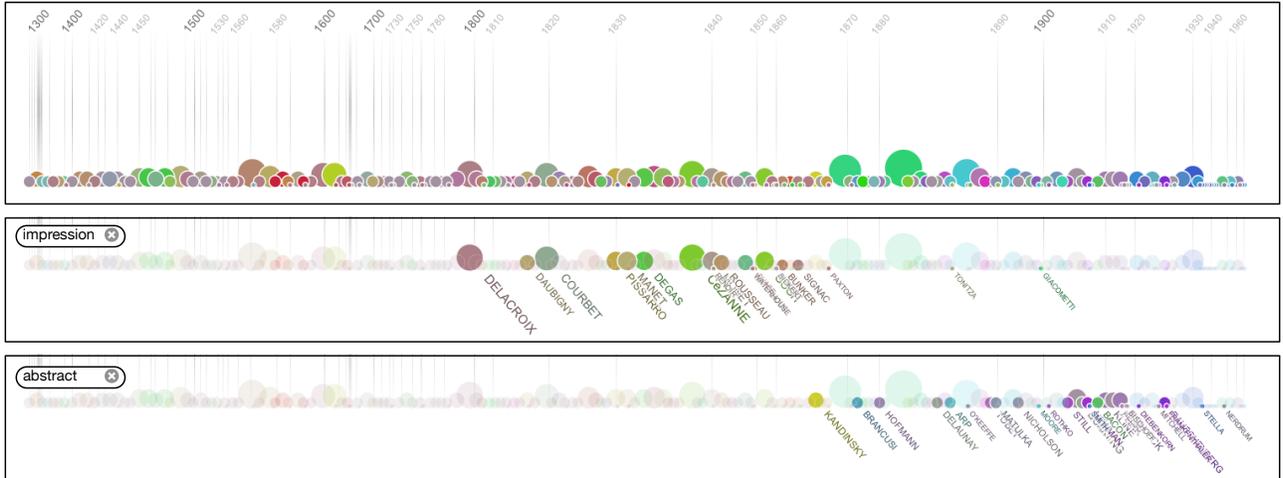


Figure 16. In the painters’ timeline view, there is a slightly higher concentration of blue and purple tones on the right. Searching for terms indicating artistic movements confirms the temporal development of impressionist and abstract art.

Time-Similarity Comparisons. By switching between time and similarity layouts it appears as if time periods and genres of musicians have a relatively high correlation. This hunch is further supported considering the colour distribution of nodes along the time axis (see Figure 15, top). The nodes appearing in the bottom left in the similarity map in shades of purple, are mostly located on the left side of the timeline, i. e., in the first half of the last century. On the flip side, the nodes that appeared in tones of green in the top right of the similarity map are located towards the right of the timeline. Examining the distribution of nodes highlighted for the search terms ‘jazz’ and ‘punk’ further supports this observation (see Figure 15, middle and bottom). The painters’ timeline does not indicate such a strong correlation between similarity map lo-

cations with time periods, however, searching for ‘impression’ and ‘abstract’ indicates relatively discrete intervals along the time axis (see Figure 16). The more heterogenous colour distribution is likely due to the inclusion of art forms in the similarity keywords that are fairly independent from artistic movements.

8. DISCUSSION

The design and implementation of EdgeMaps constitute a first step towards integrating explicit and implicit data relations. While the idea of integrating graph drawing and spatialization techniques is promising, we discuss several limitations and challenges that suggest opportunities for more research on the visualization of complex data sets featuring diverse relations.

Edge congestion. We have argued that displaying edges for only one node, i. e., philosopher, painter, or musician, at a time allows for novel exploration methods and avoids some of the edge congestion problems of larger NLDs. However, once a node has a wide and dense influence network, similar edge congestion challenges occur. It would be interesting to examine how lens and bundle techniques^{6,7} could be combined with visualizations that use a layout representing implicit relations with edges for explicit relations.

Overlapping nodes. With growing information spaces, the challenge is not only edge congestion, but the overall perceptual scalability of visualizations with more and more overlapping nodes. In the timeline view, it would be possible to position the nodes vertically, for example, ordered by significance, i. e., size. The interest map currently places smaller nodes around larger nodes with the same position. In order to support large numbers of overlapping nodes, it is possible to indicate spatial togetherness using bubble sets.²⁸

Aggregating nodes. Another promising approach is to introduce aggregation that can simplify the graph visualization and provide higher-level perspectives.^{17,29} Considering the time and similarity layouts of EdgeMaps, nodes could be aggregated along temporal or topical proximity. The resulting influence networks of clusters would allow for higher level exploration of mutual influences of groups of philosophers, painters, or musicians. For example, the viewer could explore how painters of the Renaissance influenced the impressionists, or how jazz and blues singers affected electronic and new wave musicians.

Data set. The data sets used to illustrate the potential of EdgeMaps are limited both in size and also cultural scope that is centred around North America and Europe. This bias likely reflects the demographics of the Freebase community. It would be interesting to use the visualization as a starting point to add missing connections, scrutinize the keywords, and add data entries that are beyond the cultural horizon. Besides visualizing historic personalities, we are interested in examining other data sets such as citation networks, migration patterns, and international trade.

Spatialization. While we used an existing MDS algorithm,²⁶ it would be beneficial to explore its parameters and, for example, consider planes with dynamic rectangular shapes besides squares as it is realistic to rely on windows of fixed aspect ratios. The difficulty to interpret the meaning of position and proximity is alleviated with the display of influence edges, yet further exploration is necessary to find other techniques

that make the output of MDS algorithms more accessible and meaningful. One of the ideas that arose during this work is to label regions in the MDS plane based on representative keywords that are more common among nodes that are positioned closer to each other. Besides the aim of making the MDS layout more comprehensible, it could be useful to create a flexible MDS algorithm allowing the viewer to change how items are positioned. The great challenge for this would be to make this algorithm interactive.

Folding time. The timeline features an irregular scaling that is due to removing spacing between nodes in favour of larger information density. While the textual time legend indicates this temporal folding, the added grid lines visually expose the scaling. Ideally the viewer would be able to adjust time foldings according to their interests and exploration steps. Currently a person is represented as a single point on the timeline, which makes it difficult to represent their lifetime duration or longevity of active influence. It would be interesting to represent the actual lifespans and active years of philosophers, painters, and musicians.

Layouts. Supplementing the graph visualization with meaningful node positions led to the creation of novel visual representations that expose temporal and topical qualities of data connections. Besides time and keywords, many more dimensions can be utilized as implicit relations for underlying layouts. For example, positions in a hierarchy or geographic space could be used for node placement. It is also feasible to combine two types of similarities for the spatial dimensions of the plane. For example, a one-dimensional MDS projection could be juxtaposed with the temporal distribution of nodes. In addition to enhancing the understanding of influence connections, such a hybrid layout could help the viewer to compare time relations with similarities based on interests, movements, or genres. As artistic movements can be viewed as correlations between time and style, a corresponding layout could be very helpful.

Insight and experience. Having applied the EdgeMaps visualization to three data sets allowed us to closely examine the insights and experiences that can be supported. However, we have only anecdotal evidence for the potential of visualizing explicit and implicit relations this way. It would be beneficial to study how domain-experts would explore the data sets using EdgeMaps. Particularly, we would be interested in the role of pivoting around individual nodes to gain a bottom-up understanding of a complex data set. It may be interesting to compare how domain experts experience the visualization of a given data set in contrast to viewers that are unfamiliar with the data.

9. CONCLUSION

As information spaces grow in size and complexity, there is a growing need for visual representations that help us make better sense of diverse data relations and patterns. With this work we have approached this challenge by making the following contributions:

- A novel information visualization technique that combines implicit relations (topical and temporal similarity) with explicit relations (influences).
- Detailed discussion of the design considerations concerning visual representations and interaction techniques of the EdgeMaps visualization.
- Implementation of a web-based system incorporating basic use patterns on the Web such as browser's history, bookmarks, and link sharing.
- Application to three case studies exploring data about personalities from philosophy, art, and music examining the types of novel insights and uses.

The EdgeMaps technique represents implicit relations as the underlying layouts for node-link diagrams in which nodes stand for people, and edges for explicit influence connections between them. By constraining the selection to one node at a time, it is possible to visually distinguish between incoming and outgoing edges. Novel visual patterns resembling the aesthetics of fireworks and waves may not only look more appealing, they also allow for a closer examination of influence differences. In contrast to the notorious yarn-ball effect of dense graph visualizations, we have suggested that an interactive pivotal exploration along edges between nodes may better allow viewers to grasp network structure than complicated overviews. While much more work is necessary to help us better explore and understand complex information spaces, we see EdgeMaps as a promising step towards this effort.

ACKNOWLEDGEMENTS

Funding was provided by SMART Technologies, NSERC, iCORE, and NECTAR.

REFERENCES

- [1] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, 1996.
- [2] Jacques Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press, 1983.
- [3] Maureen C Stone. *A Field Guide to Digital Color*, chapter 12 Color in Information Display, pages 277–300. AK Peters, 2003.
- [4] Ivan Herman, Guy Melançon, and M. Scott Marshall. Graph visualization and navigation in information visualization: A survey. *TVCG: Transactions on Visualization and Computer Graphics*, 6(1):24–43, 2000.
- [5] Helen C. Purchase. Metrics for graph drawing aesthetics. *Journal of Visual Languages & Computing*, 13(5):501–516, October 2002.
- [6] Nelson Wong, Sheelagh Carpendale, and Saul Greenberg. EdgeLens: An interactive method for managing edge congestion in graphs. In *InfoVis 2003: Symposium on Information Visualization*, pages 51–58. IEEE Computer Society, 2003.
- [7] Danny Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *TVCG: Transactions on Visualization and Computer Graphics*, 12(5):741–748, 2006.
- [8] Danny Holten and Jarke J. van Wijk. A user study on visualizing directed edges in graphs. In *CHI '09: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2299–2308. ACM, 2009.
- [9] Richard A. Becker, Stephen G. Eick, and Allan R. Wilks. Visualizing network data. *TVCG: Transactions on Visualization and Computer Graphics*, 1(1):16–21, 1995.
- [10] Jeffrey Heer and danah boyd. Vizster: Visualizing online social networks. In *InfoVis 2005: Symposium on Information Visualization*, pages 32–39. IEEE Computer Society, 2005.
- [11] James A. Wise, James J. Thomas, Kelly Pennock, David Lantrip, Mark Pottier, Anne Schur, and Vern Crow. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *InfoVis 1995: Symposium on Information Visualization*, pages 51–58. IEEE Computer Society, 1995.
- [12] M. Williams and T. Munzner. Steerable, progressive multidimensional scaling. In *InfoVis 2004: Symposium on Information Visualization*, pages 57–64. IEEE Computer Society, 2004.
- [13] Martin Wattenberg. Arc diagrams: visualizing structure in strings. In *InfoVis 2002: Symposium on Information Visualization*, pages 110–116. IEEE Computer Society, 2002.

- [14] Bernard Kerr. Thread Arcs: An email thread visualization. In *InfoVis 2003: Symposium on Information Visualization*, pages 211–218. IEEE Computer Society, 2003.
- [15] Petra Neumann, Stefan Schlechtweg, and Sheelagh Carpendale. ArcTrees: visualizing relations in hierarchical data. In *Eurographics-IEEE VGTC Symposium on Visualization*, pages 53–60, 2005.
- [16] Ben Shneiderman and Aleks Aris. Network visualization by semantic substrates. *TVCG: Transactions on Visualization and Computer Graphics*, 12(5):733–740, Sep/Oct 2006.
- [17] Martin Wattenberg. Visual exploration of multivariate graphs. In *CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 811–819. ACM, 2006.
- [18] Christopher Collins and Sheelagh Carpendale. VisLink: Revealing relationships amongst visualizations. *TVCG: Transactions on Visualization and Computer Graphics*, 13(6):1192–1199, Nov/Dec 2007.
- [19] Fernanda B. Viégas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matt McKeon. Many Eyes: A site for visualization at internet scale. *TVCG: Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, Nov/Dec 2007.
- [20] Jeffrey Heer, Fernanda B. Viégas, Martin Wattenberg, and Maneesh Agrawala. Point, talk, publish: Visualization and the web. In Robert Liere, Tony Adriaansen, and Elena Zudilova-Seinstra, editors, *Trends in Interactive Visualization*, pages 269–283. Springer, 2009.
- [21] Zachary Pousman, John T. Stasko, and Michael Mateas. Casual information visualization: Depictions of data in everyday life. *TVCG: Transactions on Visualization and Computer Graphics*, 13(6):1145–1152, Nov/Dec 2007.
- [22] Fernanda B. Viégas and Martin Wattenberg. Artistic data visualization: Beyond visual analytics. In *Online Communities and Social Computing*, volume 4564 of *Lecture Notes in Computer Science*, pages 182–191. Springer, 2007.
- [23] Marian Dörk, Sheelagh Carpendale, and Carey Williamson. The information flaneur: A fresh look at information seeking. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011.
- [24] Freebase. <http://www.freebase.com> (Retrieved 2011-04-29).
- [25] Allmusic. <http://www.allmusic.com> (Retrieved 2011-04-29).
- [26] Jan de Leeuw and Patrick Mair. Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, 31(3):1–30, 2009.
- [27] Raphaël. <http://raphaeljs.com> (Retrieved 2011-04-29).
- [28] Christopher Collins, Gerald Penn, and Sheelagh Carpendale. Bubble sets: Revealing set relations with isocontours over existing visualizations. *TVCG: Transactions on Visualization and Computer Graphics*, 15(6):1009–1016, 2009.
- [29] Frank van Ham and Jarke J. van Wijk. Interactive visualization of small world graphs. In *IEEE Symposium on Information Visualization 2004*, pages 199–206, 2004.