SIGRAD 2015
L.Kjelldahl and C. Peters (Editors)

Exploring Time Relations in Semantic Graphs

Ralph Wozelka[1,*], Mark Kröll[1,*] and Vedran Sabol[1,2]

[1]Know-Center GmbH, Inffeldgasse 13, Graz, Austria
[2]Graz University of Technology, Inffeldgasse 13, Graz, Austria
{rwozelka|mkroell|vsabol}@know-center.at

ABSTRACT

The analysis of temporal relationships in large amounts of graph data has gained significance in recent years. Information providers such as journalists seek to bring order into their daily work when dealing with temporally distributed events and the network of entities, such as persons, organisations or locations, which are related to these events. In this paper we introduce a time-oriented graph visualisation approach which maps temporal information to visual properties such as size, transparency and position and, combined with advanced graph navigation features, facilitates the identification and exploration of temporal relationships. To evaluate our visualisation, we compiled a dataset of ~120.000 news articles from international press agencies including Reuters, CNN, Spiegel and Aljazeera. Results from an early pilot study show the potentials of our visualisation approach and its usefulness for analysing temporal relationships in large data sets.

Categories and Subject Descriptors (according to ACM CCS): H.3.3 [Information Search and Retrieval]: Information filtering; H.5.2 [User Interfaces]: Graphical user interfaces (GUI).

1. Motivation

The processing and analysis of information is becoming more and more complex partly due to the vast amount of available information, but also due to our capabilities of automatically explicating relations within the data. This relationship analysis has already found its way into some profession's everyday work including journalists or historians which seek to get a chronological overview of a certain topic. To adequately represent relational data, node-link-based graph representation is widely used. Its simplicity and established formalism facilitate the analysis of relationships between data nodes. Adding a temporal component to graph visualisation, and with it an additional source of complexity, is not easy to achieve since complex graph representations already have a tendency towards clutter due to link crossing and overdraw.

In this paper, we extend the Semantic Blossom graph [RWV*14], an approach supporting focused and context-sensitive exploration of complex graphs along semantic facets, by mapping temporal information to visual properties such as (i) node size, (ii) transparency and (iii) position. Interactive combination of these properties facilitates the identification and thus the exploitation of temporal relationships in the data.

2. Related Work

Network data such as semantic knowledge graphs are commonly visualised using node link diagrams and graph visualisation methods (cf. [Cui07]). The representation of complex graph structures exhibits a tendency towards clutter due to link crossing and overdraw; adding a temporal component to it is thus not an easy to achieve goal. A commonly taken approach of addressing graph evolution is by showing changes through an animated sequence (cf. [BPF*14], [Raf13]). While animation may appear a natural way of conveying change, it suffers from the phenomenon of change blindness and from the human inability to remember previous states. Another approach involves visualising the graph at different time points through multiple snapshots or temporal slices [CKN*03]. While this group of methods enables the viewer to compare multiple graph states within a single static visualisation, and to view development of specific graph parts over time [Dwa05], they add a large amount of complexity to the representation as the graph is drawn multiple times. Additionally, slices are often shown in 3D [FHN*07] which complicates navigation and may lead to occlusion problems. Some approaches combine temporal slicing with animation [EHK*05]. A related approach emphasises the temporal dimension over the graph structure by positioning the nodes along a timeline [BPO*06]. As the connectivity of the graph does not contribute to the layout, such a representation is often cluttered by links connecting nodes placed far from each other, making such a layout suitable only for smaller graphs. A very commonly taken approach is to code the temporal information using visual features of nodes and links, such as colour or transparency [ATS*11]. Nevertheless, the advantage of representing the history of the whole graph at once within a single, stable representation may be outweighed by the complexity of the representation for large graphs. This is due to the fact that all nodes and links, i.e. those within and those outside of the user's time interval of interest, are always displayed.

* both authors contributed equally to this work

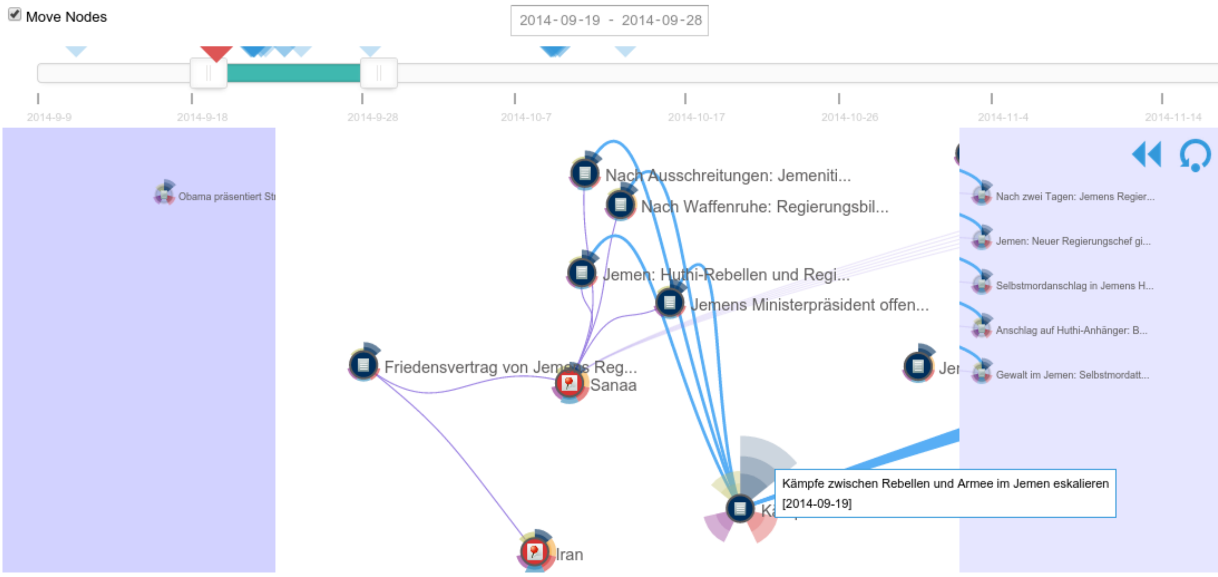

Figure 1 *gives an overview of the visualisation interface. The main part of the interface consists of three views – a past view (left), a main view (middle) and a future view (right). The main view contains only nodes whose time stamp fall into the time span selected with the time slider at the top; the selection corresponds to the date information – in this case 2014-09-19 – 2014-09-28. Nodes before the selection are displayed in the past view; nodes after the time span in the future view. All displayed nodes are assigned a corresponding triangle icon on the time slider representing their time stamp, with selected node's time shown in red. A "mouse over" highlights a node, provides more comprehensive information in a tool tip, in this case the article's full title as well as the publication date, and shows the node's semantic blossom which is used for focused and context-sensitive exploration of the graph.*

Our work combines the advantages of coding temporal information using visual features, in particular, size and transparency, with positioning of nodes which are outside of the user selected time interval at the edges of the visualisation. In this way the user can focus on the part of the graph within the selected time interval which is shown as a classical node-link representation with full exploration capabilities. At the same time the context of the full graph and the temporal information remain preserved.

3. Mapping Temporal Information

The used data set contains ~120.000 news articles from international press agencies including Reuters, CNN, Spiegel and Aljazeera. The articles were in English and German with matching on syntactic but not on semantic level. The articles were crawled from March, 2014 to January, 2015. Each article's time stamp represents its publication date. These articles are then processed using NLP methods to extract named entities (such as persons, locations and organisations) and stored into an index. Associations between extracted named entities are computed based on their co-occurrence and stored into an association index (cf. [Ker12]), which is implemented as a double inverted index (concept – documents, document – concepts). Queries to the association index form the basis for visual exploration of relationships.

We base our visualisation on the Semantic Blossom graph [RWV*14]), a graph visualisation approach which facilitates the exploration of large densely connected graphs in a smart and efficient way by initially showing not the entire graph but only a subset of relevant nodes. Users can expand nodes along semantic facets to explore the graph. We extend this approach by a temporal component which maps a time information attached to nodes onto visual properties such as (i) node size, (ii) transparency and (iii) position. We use multiple visual properties to emphasise the difference between nodes in and outside the selected time interval. Using only a single visual property, e.g. size, would either make the symbol too small to be readable or would not distinguish the two groups adequately.

Figure 1 gives an overview of our interface to visualise the temporal information within the data. Relevant time spans can be selected by handles on the time slider at the top (the minimal time span is 48 hours). This temporal selection triggers the assignment of nodes to three views – a Past view, a Main (Present) view and a Future view. Nodes with a time stamp in the selected time span are shown in the Main view using a node-link representation. All displayed nodes are assigned a triangular icon above the time slider according to their time stamp; a selected node is marked by a red triangle. A "mouse over" highlights a node and provides more comprehensive information; in this case the article's full title as well as the publication date. The highlighted node illustrates the concept of the semantic blossom that shows which and how many relations exist to distinct types of nodes using a "blossom" representation. The colour (and position) of each petal defines the node type to expand, its size indicates the number of connected nodes of that particular type. Context-sensitive expansion, i.e. only of those nodes which are connected to specified nodes already shown in the visualisation, is possible via simple interactions with petals and nodes.

3.1 Node Positioning

Nodes outside of the selected time span are temporally sorted and displayed in their respective views. This means that nodes closest to the temporal selection are positioned at the top of their views, whereby their size as well as their opacity also increase (see Figure 2).

Figure 2 *shows the temporal ordering of nodes in the Future view. The closer to the Main view in terms of time, the larger and the more opaque they are. Temporally more distant nodes are smaller and more transparent. Edges are preserved between nodes in different views; to prevent clutter, they are indicated using highly transparent links.*

By ticking off the "Move Nodes" check box, all nodes are positioned in the Main view. In this case the temporal information is only conveyed by node size and node transparency as illustrated in Figure 3.

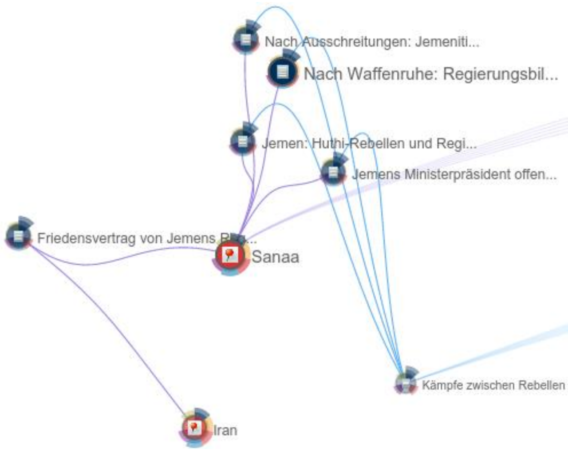

Figure 3 *shows the visualisation in a "Move Nodes" ticked off mode. All displayed nodes moved to the main view; temporal information is conveyed only by node size and node transparency. In this case, large and opaque nodes exhibit time stamps within the temporal selection. The smaller and more transparent a node is, the further it is from the selected time interval.*

It should be noted that temporal information is different for article nodes and facet nodes. Article nodes have a single time stamp. Facet nodes can either represent named entities, e.g. persons, locations or organisations, which are contained in an article or metadata such as source information. Since facet nodes are not unique, i.e. a person can occur in more than one news article, they can have more than one time stamp as illustrated in Figure 4. A "mouse over" in the graph view reveals time stamps in the timeline of all other displayed articles the person "Mohammed Basindawa" occurs in. The time stamps are furthermore highlighted at the time slider via red triangles facilitating to identify temporal relationships amongst the nodes.

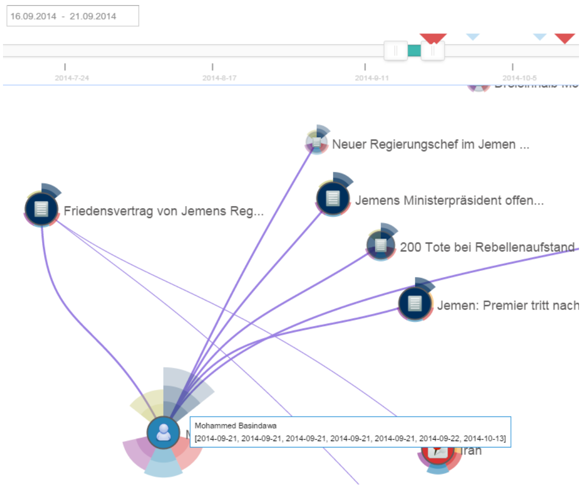

Figure 4 *shows the "mouse over" information of a person facet node ("Mohammed Basindawa") as well as the time stamps of all other displayed articles which mention this person. These time stamps are also marked via red triangles on the time slider. For a possible positioning in the views the average of the time stamps is taken.*

4. Pilot Study

4.1 Task Description

We conducted an early pilot study with two participants in form of a thinking aloud test to collect initial feedback on usability problems and on whether one of the modes would be preferred by the users. The pilot study also serves as a preparation for a subsequent full-scale user study, which will compare the two modes of temporal representation – with and without moving nodes – in detail. The two participants had experience in gathering information; both search daily to once a week for information on people and organisations on the web using a standard search engine.

The evaluation task required the participants to look into the conflict going on in the past months in the Republic of Yemen. Starting point was the search query "Yemen" resulting in the most relevant documents being displayed in our visualisation interface. The two participants were steered through a prepared scenario by posing questions such as "Find out the name of the related Yemen politician" or "When was the cease-fire between rebels and government reported?". At the end, a big picture of the Yemen conflict should have been formed taking into account chronological ordering of the events: "fighting", "peace treaty", "elections" and "prime minister resigns".

4.2 Feedback and Observations

In the following we report selective findings from the thinking aloud test as well as the filled out questionnaires at the end of the pilot study:

- Having the distribution of timestamps visualised (triangles) above the time slider was helpful.
- Highlighting the triangle icons above the slider on node "mouse over" was perceived as a very useful inspection helping find timestamps and establish a temporal context to other nodes.
- The "Move Nodes" representation was easier to read due to a better focus on the selected time interval and less clutter, which is caused by many out-of-context nodes and links in the other mode.
- The "Move Nodes" mode made it easy to overlook new nodes as they were added and moved out of the way.
- If nodes move, context is lost more easily. The movement itself needs to be slow and smooth to avoid confusion.
- The currently missing synchronised highlighting over all views in the "Move Nodes" mode would be helpful.
- Visualisation design is generally suitable for discovering most important relationships in the data.
- The participants agreed that the visualisation, in both modes, supports discovery of temporal information, for instance, when, or how far apart events happened.

To summarise: moving nodes helps avoid clutter, which is welcome, but new nodes added through expansion tend to be overlooked easily, which will be addressed in a future version of the tool. It is also an argument for providing both modes to the user.

5. Conclusions

In this work we presented an approach for visualising time-oriented data by mapping temporal information to visual properties, in particular the node position, however taking a middle approach where only nodes out of the user-defined temporal focus are moved, while the within-focus nodes remain unaffected. Our visualisation enables and facilitates explorative and explanative analyses of time-oriented data graph structures, for instance, for information providers to get an overview of a situation based on (i) respective events in the past or (ii) discovered relations between data entities such as persons or organisations.

To gather evidence on the effectiveness of our approach, we plan to conduct a comprehensive user study with up to 30 professional with information analysis background. Future work also includes an improvement with respect to the user interface, for instance, implementing a proper edge rendering between nodes from different views.

Acknowledgement. This work is funded by the FFG KIRAS program (project nr. 840824). Know-Center is funded within the Austrian COMET Program under the auspices of two Austrian Federal Ministries - BMVIT and BMWFW - and State of Styria. COMET program is managed by the FFG.

References

[ATS*11] AHN J., TAIEB-MAIMON M., SOPAN A., PLAISANT C., SHNEIDERMAN B.: Temporal visualization of social network dynamics: Prototypes for nation of neighbors. In *Proc. of the 4th Internat. Conference on Social Computing, Behavioral-Cultural Modeling and Prediction*, (2011).

[BPF*14] BACH B., PIETRIGA E., FEKETE J.: GraphDiaries: Animated transitions and temporal navigation for dynamic networks. *IEEE Transactions on Visualization and Computer Graphics*, (2014).

[BPO*06] BLYTHE J., PATWARDHAN M., OATES T., DES JARDINS M., RHEINGANS P.: Visualization support for fusing relational, spatio-temporal Data: Building Career Histories. In *Proc. of the 9th International Conference on Information Fusion*, (2006).

[CKN*03] COLLBERG C., KOBOUROV S., NAGRA J., PITTS J., WAMPLER K.: A system for graph-based visualization of the evolution of software. In *Proc. of the ACM Symposium on Software Visualization*, (2003).

[Cui07] CUI W.: A survey on graph visualization. PhD Qualifying Exam (PQE) Report, Computer Science Department, Hong Kong University of Science and Technology (2007).

[Dwa05] DWAYER T.: Extending the wilmascope 3D graph visualisation system. In *Proc. of the Asia-Pacific Symposium on Information Visualisation*, (2005).

[EHK*05] ERTEN C., HARDING P., KOBOUROV S., WAMPLER K., YEE G.: Exploring the computing literature using temporal graph visualization. In *Proc. of the Conference on Visualization and Data Analysis*, (2005).

[FHN*07] FU X., HONG S., NIKOLOV N., SHEN X, WU Y., XU K.: Visualization and analysis of email networks. In *Proc. of Asia-Pacific Symposium on Visualisation*, (2007).

[Ker12] KERN R.: A feature association framework for knowledge discovery applications. PhD Thesis. (2012)

[Raf13] RAFELSBERGER W.: Interactive visualization of evolving force-directed graphs. Lecture Notes in Computer Science, (2013).

[RWV*14] RAUCH M., WOZELKA R. VEAS E., SABOL V.: Semantic Blossom Graph: A New Approach for Visual Graph Exploration. In Proc. of the 18th Int. Conf. on Information Visualisation (IV2014).