

# **Spatiotemporal Knowledge Visualization and Discovery in Dynamic Social Networks**

**Ralf Klamma, Yiwei Cao, Marc Spaniol, Yan Leng**  
(Lehrstuhl Informatik 5, RWTH Aachen University, Germany  
{klamma|cao|spaniol|leng}@i5.informatik.rwth-aachen.de)

**Abstract:** In this paper, we introduce a so-called DyVT tool (**D**ynamic social network **V**isualization **T**ool) to support spatiotemporal knowledge visualization and discovery in dynamic social networks. The dynamic aspects of social networks refer to contextualized information such as spatial, temporal as well as users' personalized information. We also define an XML-based target language incorporating emerging formats like DyNetML, KML, and GraphML. It also provides means to express, store and interchange the dynamic aspects of complex dynamic social network data. Based on this language, users can animate and personalize spatiotemporal knowledge extracted from social network data like email threads or blogs. In addition, a Java based graphical user interface is also available to enable non-experienced users to customize knowledge visualization easily. A mashup with Google maps for spatiotemporal visualization is provided. With this tool spatiotemporal knowledge on an IBM DB2 Mailing list database containing 69 mailing lists and 56389 mails altogether is well explored.

**Keywords:** Dynamic social network, information visualization, spatiotemporal knowledge, XML, GraphML, DyNetML, KML, interchange format, Google Map, animation

**Categories:** E.1, E.2, H.2.3, H.2.8, H.3.1, H.3.2, H.3.3, H.3.5, J.4

## **1 Introduction**

Social network data generated from computer mediated communication over the Internet has created an enormous interest in the analysis of social interaction using digital networks. More and more research efforts have been put into Social Network Analysis (SNA) [DeFo03]. Social network indicates a special kind of network where social relationships are viewed as nodes and tiers with values in a social context. One of the key characteristics of a social relationship is its dynamic nature. However, most past analyses of social networks are essentially static where the dynamic nature of social relationships indicating when and where social relations take place is discarded. Hence, more recently there is an upsurge of interest in studying dynamic social network in which social data sets change over space and time [BrCa03]. Understanding networks from a dynamic perspective is essential to understand the interaction in virtual worlds. It facilitates reasoning about real groups as complex dynamic systems that evolve over space and time. Spatiotemporal information visualization has been a key research issue in the area of spatiotemporal knowledge representation and reasoning.

Meanwhile, visualization has played a major role in SNA. Instead of creating intriguing pictures, visualization is essential for generating learning situations

[ViDo04]. A standing critique of the research in social network visualization focuses on a “structural bias” that implicitly denies much of the dynamic nature of social relations [Milg67]. Much recent interest in this area revolves around understanding how networks develop and change. So far, the research community has developed a number of powerful tools for collecting, analyzing and visualizing social network data. However, these existing tools are developed for different purposes. Some concentrate on temporal visualization using movies or mashup images, while some aim to deal with spatial data using different map formats. Unfortunately, there is no such a system existing to visualize complex dynamic social network data in a user-friendly tool. From this point of view, a meaningful way to capture this dynamic contextualized information in a highly collaborative way is needed.

This paper presents spatiotemporal knowledge discovery via a so-called DyVT tool supporting dynamic social network visualization. Furthermore, an XML-based language is specified to define the data model integrating the dynamic nature of complex social network data such as temporal data, and geospatial data into a uniform file. All the later visualization tasks are performed based on this uniform XML file. It uses animations showing how relationships emerge over time to discover temporal knowledge. Moreover, a graph with a map as background is effective to represent spatial knowledge. In addition, user-chosen personalized graph appearance can also be visualized dynamically.

The rest of the paper is organized as follows. After reviewing the related work in Section 2, Section 3 is devoted to system architecture and system implementation. In Section 4 we make a summary and give the perspective of future work.

## 2 Related Work

In order to get meaningful knowledge visualization, the first step is to design an interchange format for complex social network data with regard to spatiotemporal knowledge. Secondly, visualization methods can be applied. Thus both interchange formats for social network data and existing packages are discussed in this section.

Interchange formats for social network data can be classified into two categories: unstructured free text based and XML based formats. The typical examples of the former format are DL [UCINET07] and Pajek .net [NMBa05]. These formats are easy to implement by software designers. But they are ambiguously understandable by both users and machines. The interoperability is poor, too. The latter XML based formats can overcome these limitations. They are easy to read and to parse by providing powerful facilities for building expressive and stable data formats with DTDs and XML Schemas. Three languages among them are presented as follows to show how they represent social network data.

**GraphML** (Graph Markup Language) [BrEi02] is an XML language to deal with network data. Its characteristic feature enables users to add modules that implement special extensions or additional data, which affords GraphML a large degree of flexibility. Additional advantage of GraphML is being supported by a large number of graph analysis and drawing software such as JUNG and yFiles etc.

**DyNetML** is provided by CASOS at Carnegie Mellon University [TRCa04]. It represents dynamic network data as sets of time slices. A dynamic network element is

defined as a sequence of *MetaMatrix* elements representing a snapshot of the organization for one time period.

**KML** is used to display geographic data in some special browsers such as Google Earth, Google Maps, and Google Maps for mobile. KML has a tag based structure with names and attributes used for specific display purposes. Thus, Google Earth and Maps act as browsers for KML files [KML07].

Next, the state-of-the-art overview of the available packages to visualize dynamic social network are given. Animations and “movies” are often used to visualize temporal data. DyNet [BrCa03] at CMU, MOVIE MOL [MOVIE MOL07] provided by Linköping University and Uppsala University, as well as SoNIA [SoNIA07] at Stanford University all use this method. In addition, Mapnet [Mapnet07] and London Connections [FaOw07] have been developed to visualize spatial data on maps. However, these packages serve different purposes. Some focus on temporal data visualization, while some concentrate on spatial data visualization.

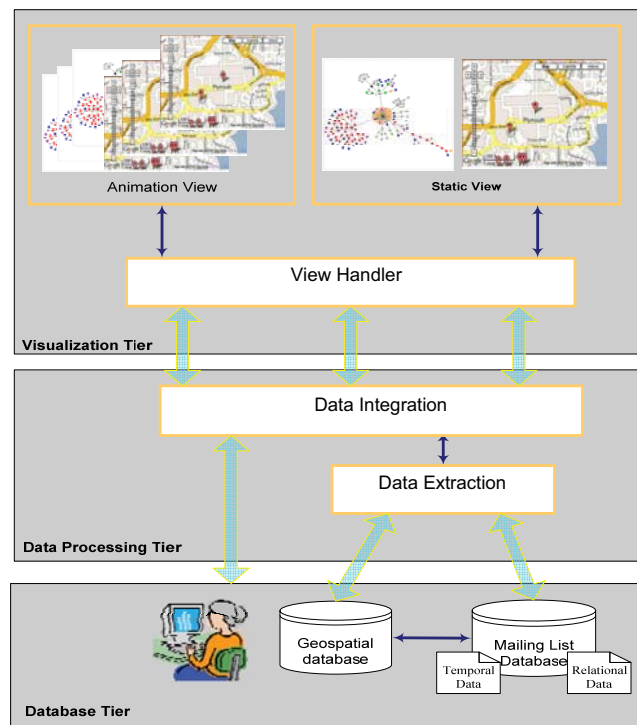


Figure 1: System architecture of DyVT

### 3 DyVT: Dynamic Social Network Visualization Tool

DyVT visualizes spatiotemporal knowledge within one common tool using the XML-based DyVTML. It employs the three-tier architecture based upon the information visualization reference model. The reference model divides the visualization process

into a series of steps ranging from data acquisition and modeling, the visual encoding of data, to the presentation of interactive displays [Chi00]. Figure 1 illustrates the system architecture of DyVT.

### 3.1 Database Tier

The database tier contains diverse social network data sources. In the prototype we use three data sources. The *IBM DB2 mailing list database* contains information of mailing lists, each of which can describe a social network. More detail about the definition of social relations can be found in [KSDe06]. If A sends an email to the mailing list, B and C reply it, then there exist directed relations between B and A as well as C and A. Besides the relational data for social networks, the temporal data can be obtained from it as well. *GeoLiteCity database* [MaxMind07] determines the Internet visitors' city information extracted from the IP address. Currently, it is a database with the size of 19 MB which generally contains IP address range and geographical data for all publicly assigned IPv4 addresses. *User-Chosen network appearance data* allows users to configure the graph appearance they prefer. After the attributes for nodes and edges are chosen, an XML-based ADML (*A*ppearance *D*ata *M*arkup *L*anguage) file including all appearance data can be generated (cf. Figure 2).

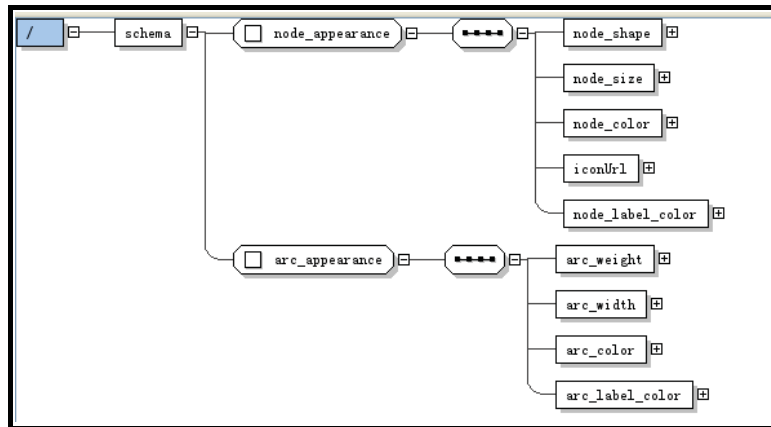


Figure 2: XML Schema for ADML

### 3.2 Data Processing Tier

The data processing tier bridges between the database tier and the visualization tier. It aims to collect all backing data needed in visualization tier. This process is developed through two steps: *Data Extraction* and *Data Integration*.

The step of data extraction is responsible for extracting relational social network data and dynamic nature of relations including spatiotemporal data from the mailing lists and the GeoLiteCity databases. Then the data is transformed into the XML-based DyVTML (*D*ynamic *V*isualization *T*ool *M*arkup *L*anguage). It allows setting two constraints to extract the relational data with time stamps of social networks: *Choose a mailing list* and *Define time interval*. In addition, spatial data is extracted from the

email addresses. To obtain the relevant spatial data, the domain parts of the email addresses should be firstly extracted. Then the host names are mapped to the corresponding IP addresses. Finally, the IP addresses are converted into numeric representation which is used to look up and retrieve spatial data in GeoLiteCity database. The transformation process keeps the visualization tier independent from the varying formats of the data sources. Therefore, it reduces the execution time and is convenient for data migration.

In addition, GraphML, DyNetML and KML are suitable for being used as interchange formats. However, each of them serves different purposes. The starting idea of defining DyVTML is a logical integration of these three formats. The graph topology of GraphML is applied to represent relational data including the definition of nodes and edges. Since the spatial knowledge visualization is developed with the help of Google maps API, the supported KML is used. Temporal knowledge representation derives from DyNetML using sets of time slices. However, we do not use the expression of DyNetML directly, because in DyVT the time slices are able to be defined by users. Furthermore, we use the concept of time slices to list the time stamps for both nodes and edges.

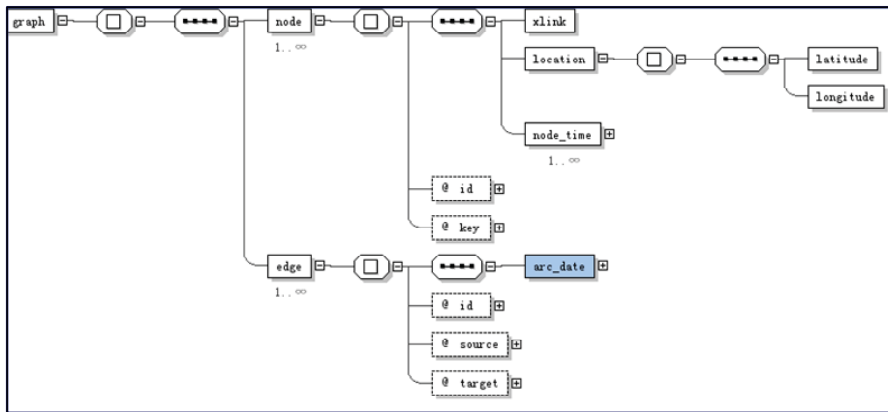


Figure 3: XML Schema for DyVTML

So the step of data integration contributes to parse two separate XML-based files. Appearance data ADML constructed in the database tier and the underlying social network data DyVTML. Both files instantiate series of nodes and edges which are data models including visual features for the visualization tier. The separation of appearance data from social network data allows different users to visualize the same underlying social network using different personalized appearance settings quickly.

### 3.3 Visualization Tier

The actual rendering of the data onto views is done in the visualization tier. To provide varying perspectives onto the dynamic nature of social network data, two handlers are built: *temporal view handler* and *map view handler*.

The temporal view handler is based on the SoNIA package. Images rendered within successive time slices are based on circle layout and the modified Kamada

Kawai layout [MMBe05] where the initial positions are fixed by nodes' position in the previous slice. These images are gathered for an animation to visualize social network within a certain time interval. DyVT enables users to specify the size of time slices which affects the number of nodes and edges shown in each frame of animation. Users can also choose the speed of moving time slices which affects how smoothly the network changes (cf. Figure 4).

In addition, the visualization results can be exported to multimedia formats such as JPEG, GIF, and SVG etc. The map view handler builds the mashup with Google maps [GoogleMap07]. Moreover, animated effects are applied to arrange the order of displaying nodes and edges based on their time stamps. Popup windows and name lists are developed for node information retrieval in order to improve the graph readability (cf. Figure 5).

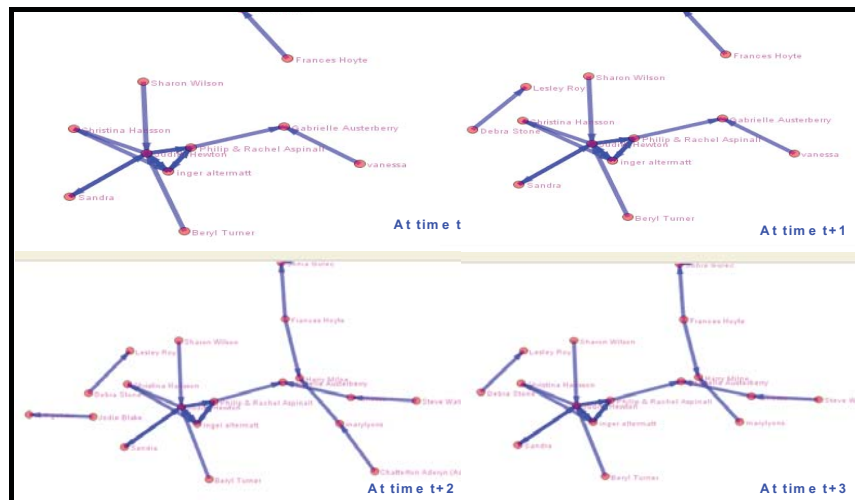


Figure 4: Screenshot of temporal knowledge representation in 4 successive time slices

#### 4 Conclusions and Future Work

In summary, DyVT is a unique tool to capture and handle dynamic social networks. By defining two new XML based languages. DyVTML combines GraphML, DyVTML and KML for the integrated representation of spatiotemporal knowledge of dynamic networks. And ADML allows personalized views on networks visualization. DyVT allows straightforward animations of dynamic networks, mashups with Google maps and exportation of visualization results in available multimedia formats.

DyVT is still a prototype which provides a way to visualize multiple dynamic natures in a uniform platform. The future work will focus on pattern-based spatiotemporal knowledge exploration [AnAn06]. Furthermore, the interchange formats should support multi-relational social network in DyVTML by allowing different personalized appearances to each cluster within a network. Above all, DyVT is expected to be used in various Web 2.0 environments such as Wikis, Blogs, and

social networking software etc, in order to explore spatiotemporal knowledge of large dynamic social networks.

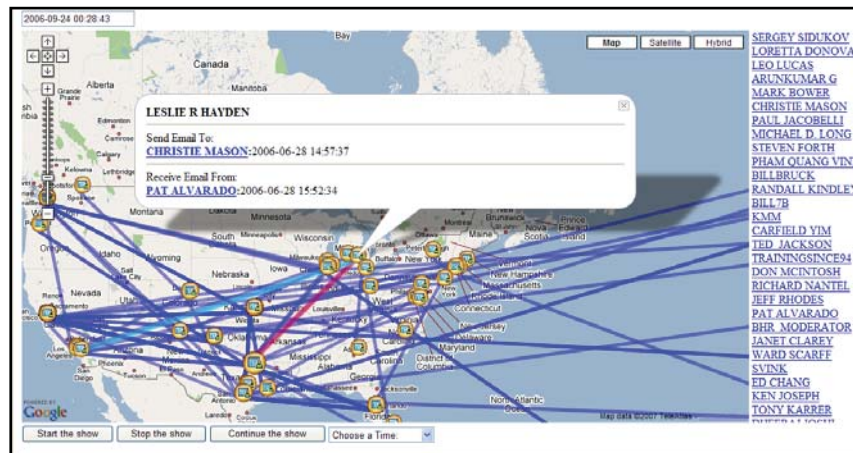


Figure 5: Screenshot of map view with a Popup Window and Name List

### Acknowledgements

This work was supported by the excellence initiative of the German government within the research cluster “Ultra High-Speed Mobile Information and Communication (UMIC)”, by the German National Science Foundation (DFG) within the collaborative research center SFB/FK 427 “Media and Cultural Communication”, and by the 6th Framework IST programme of the EC through the NoE on Professional Learning (PROLEARN) IST-2003-507310.

### References

- [AnAn06] N. Andrienko and G. Andrienko: Exploratory Analysis of Spatial and Temporal Data, Springer-Verlag Berlin Heidelberg, 2006
- [BrEi02] U. Brandes, M. Eiglsperger, I. Herman, M. Himsolt, and M. S. Marshall: GraphML progress report: Structural layer proposal. P. Mutzel, M. Junger, and S. Leipert (eds.) Proceedings of 9th International Symposium on Graph Drawing (GD '01), Springer Lecture Notes in Computer Science 2265, 2002, 501-512
- [BrCa03] R.L. Breiger and K.M. Carley: Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, National Academies Press, 2003, ISBN 0309089522
- [Chi00] E. H. Chi, A taxonomy of visualization techniques using the data state reference mode, Proceedings of IEEE Symposium on Visualization (InfoVis' 2000), October 9-10, 2000, Salt Lake City, USA, 69-76
- [DeFo03] A. Degenne and M. Forse: Introducing Social Networks, Sage Publications, 2003
- [GoogleMap07] Google Map API, <http://www.google.com/apis/maps/> {8.5.2007}

- [KML07] KML homepage: <http://earth.google.com/kml/index.html> {14.3.2007}
- [KSDe06] R. Klamma, M. Spaniol, and D. Denev: PALADIN: A Pattern Based Approach to Knowledge Discovery in Digital Social Networks, K. Tochtermann, H. Maurer (Eds.): Proceedings of I-KNOW '06, 6th International Conference on Knowledge Management, Graz, Austria, September 6-8, 2006, J.UCS (Journal of Universal Computer Science) Proceedings, Springer, 457-464
- [Mapnet07] Mapnet homepage: <http://www.caida.org/tools/visualization/mapnet/> {8.5.2007}
- [MaxMind07] MaxMind LCC, <http://www.maxmind.com/app/city> {14.3. 2007}
- [Milg67] S. Milgram, The small world problem, *Psychology Today* 1, May 1967, 61-67
- [MMBe05] J. Moody, D. McFarland and S. Bender-deMoll: Dynamic network visualization. *American Journal of Sociology* vol. 110, no. 1, January 2005, 1206-1241
- [MOVIEMOL07] MOVIEMOL: "An easy-to-use molecular display an animation program", <http://www.fos.su.se/moviemol.html> {12.03.2007}
- [NMBa05] W. De Nooy, A. Mrvar, V. Batagelj, *Exploratory Social Network Analysis with Pajek*, Cambridge: Cambridge University Press, 2005
- [SoNIA07] SoNIA - Social Network Image Animator homepage: <http://www.stanford.edu/group/sonia/index.html> {14.3. 2007}
- [FaOw02] R. Fawcett-Tang and W. Owen: *Mapping: An Illustrated Guide to Graphic Navigational Systems*, Rockport Publishers, 2002, ISBN 2880467071
- [TRCa04] M. Tsvetovat, J. Reminga, K. M. Carley: DyNetML: interchange format for rich social network data, Technical report, Institute for Software Research International School of Computer Science, Carnegie Mellon University, January 2004, CMU-ISRI-04-105
- [UCINET07] UCINET DL Format: <http://www.analytictech.com/downloadnd.htm> [Last access: 8.5.2007]
- [ViDo04] F. B. Viegas, J. Donath: *Social Network Visualization: Can We Go Beyond the Graph?* Workshop on Social Networks for Design and Analysis: Using Network Information in CSCW, 2004