GraphLDA: Latent Dirichlet Allocation-based Visual Exploration of Dynamic Graphs

Lijing Lin¹ ∗ Siming Chen¹ Fan Hong¹ Chufan Lai¹ Shuai Chen¹ Xiaoru Yuan¹,²

1) Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University
2) Beijing Engineering Technology Research Center of Virtual Simulation and Visualization, Peking University

Abstract

In dynamic graph visualization and analysis, it is challenging to visualize both the overall evolution of trends and the detailed changes of structures simultaneously. In this work, we propose a latent Dirichlet allocation (LDA)-based visual exploration method for dynamic graphs. With the LDA-based analysis, we can reveal important structures in the dynamic graph based on the extracted semantic topics. To gain a deeper understanding of the derived structures and their evolution, we propose a visual analytics pipeline enabling users to interpret and explore the dynamic graph. To experiment with the proposed method, we provide a visual analytics system to test with real-world data. Our case on the datasets of dynamic collaboration network has demonstrated the effectiveness of the proposed method.

Keywords: Dynamic Graph, Latent Dirichlet Allocation, Graph Structure, Evolution

∗e-mail: {lijing.lin, siming.chen, fan.hong, shuai.chen, xiaoru.yuan}@pku.edu.cn. Xiaoru Yuan is the corresponding author.

Introduction

Dynamic graphs are ubiquitous, such as the social network in Twitter, biochemistry, software engineering, etc. By studying such evolution of dynamic graph, we can get a better understanding of how relationships change over time, how information spreads, etc. Animation and small multiples are two basic approaches directly visualizing the evolution of graph structures [1]. Both of them are difficult for users to relate and compare graphs of different time steps with long periods. van den Elzen et al. [4] propose a method of reducing graphs in time steps to points based on dimension reduction method. Their method could show an overview of the evolution, but could not give reasons why they are similar or different, which is the challenge we would tackle. Henderson et al. [3] apply LDA model to discover the group in dynamic graphs. They treat each source node at each time step corresponds to a document and links from this node as words in the document. Though they provide a good overview of groups, they can not facilitate the temporal structural patterns of the dynamic graph, which is the focus of our work.

In this work, we propose a latent Dirichlet allocation (LDA)-based visual exploration method, to reveal both the overall patterns, such as stable patterns, recurring patterns, and the evolution of important structures in dynamic graphs. Our method regards each link as one word and the graph in each time step as a document. Impor-
tangent structures could be derived as LDA topics, which involve the
temporal patterns in the dynamic graph. Our method can not only
tell what the overall patterns are but also give reasons why such
patterns happen based on the identified structures.

2 GRAPHLDA
We model a dynamic graph $\Gamma$ as a sequence of graphs, $\Gamma = (G_1, G_2, ..., G_T)$, where $T$ is the total number of time steps, and $G_j$ represents the graph in the $j$th time step. We directly define $G_j$ to correspond to the document $d_j$ in the LDA model (Figure 3).

For graph $G_j$, we denote the set of nodes and links as $(V_j, E_j)$. Then we define all links $e_{ij} \in E_j$ in graph $G_j$ as its words. The weight of a link $e_{ij}$, corresponding to the frequencies of words in each document, is defined as, $w(e_{ij}) = a \times val(e_{ij}) + (1 - a) \times imp(e_{ij})$, where $val(e_{ij})$ can directly use quantitative attributes of the link, while the $imp(e_{ij})$ is derived from the topological information of the link in the graph $G_j$. We use $a$ to balance these two factors. We use 0.5 in our case.

With the correspondences defined above, we are able to represent
the graph $G_j$ as a set of links $E_j = \{e_{1j}, ..., e_{N_jj}\}$ with frequencies $w(\cdot)$. Then, the whole dynamic graph is transformed to a matrix, where rows and columns represent graphs and links respectively, and the values of cells donate the frequencies. We feed the graph-link matrix as the input of LDA model. The output contains a set of $K$ topics, probability distributions of topics in each graph, and probability distributions of words in each topic. The extracted topics are the meaningful structures of the input graph.

Figure 3: Illustration of the dynamic graph LDA model.

3 PRELIMINARY RESULT
The visual analytics workflow is shown in Figure 2. It contains three parts, including data preprocessing, topic extraction, and visual exploration. Further, we would provide a case in analyzing dynamic collaboration network.

The data contains papers published in IEEE VIS conferences from 1990 to 2015 [2]. In the coauthor graph, the nodes are authors. A link is added when two related authors collaborated to publish papers. In the preprocessing step, the length of sliding windows is set to three years, and their overlap is two years. We set the number of topics $K$ to 10, and run the LDA algorithm with 2,000 iterations. The results are visualized in Figure 1.

In the projection view (Figure 1a), we could observe that in the early times of the conference, the circles are closed. It means that the coauthor relations are stable since the researchers are still in a small number. The circle, mapping to coauthor graph during 1994 to 1996, is far from others. It shows that coauthor network is different from those in other stages, which could be brought by the development of VIS and the rise of IEEE InfoVis. From about 2000, the circles are closer and finally form a cluster. The coauthor networks are gradually becoming stable since the field tends to be mature.

Further, we could explore the details of extracted topics. We are interested in the dominating topic in 2013-2015 and check the details of the topic (Figure 1c). We find the active researchers, e.g. Hanspeter Pfister, Eduard Gröller, Huamin Qu, Kwan-Liu Ma, etc, and their collaborations as the backbone of this topic. We select a subset of nodes (Figure 1e - nodes with orange stroke) and show the evolution of their collaboration using small multiples (Figure 1e). We can find that these authors are becoming more active and have more collaborations from 2009 to 2015. From the wordles, we could observe that their collaborations mainly happen on InfoVis and VAST, such as visual analytics, information diffusion, etc.

4 CONCLUSION
In this work, we present a novel LDA-based visual exploration method for analyzing the evolution of dynamic graph. Important structures in the dynamic graph can be interactively extracted.

5 ACKNOWLEDGEMENTS
This work is supported by NSFC No. 61672055 and also funded by PKU-Qihoo Joint Data Visual Analytics Research Center.

REFERENCES