PolyViz - a Visualization System for Special Kind of Multipartite Graphs.

Tolga Uslu *　　Alexander Mehler†

Goethe University of Frankfurt

Abstract

In this paper we present PolyViz, a new visualization system that can efficiently display a special kind of *k*-partite graphs with the benefit that the k groups themselves can also have internal links. PolyViz not only allows for the generation of the visualization, but also for the adaptation and the analysis of the underlying data. This was achieved by providing various means of interaction. We illustrate the visualization in the context of two conducted experiments. One of these experiments includes the analysis of the topic distribution of the German Wikipedia and the linkage of these topics. The other experiment is about the visual representation of sentence similarities including their analysis. In any event, PolyViz is not limited to these applications but can be used for visualizing any multipartite data.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Multipartite graphs; Human-centered computing—Visualization—Visualization design and analysis

1 Introduction

Many visualizations serve to refine papers and to display the calculated data in a readable way. However, it often takes a lot of time and adaptation to create these visualizations. In this paper, we present PolyViz, a visualization tool that can display a special kind of multipartite graphs and adapt them according to various requirements. To this end, one has to transfer the data into the requested JSON input structure and can create, adapt, analyze and download the resulting visualization. In a standard multipartite graph, no connections between nodes of the same part are allowed. In our approach, we allow for such links in order to increase the number of use cases. Group internal links are displayed by spanning an arc diagram for each group. The paper demonstrates two use cases in which PolyViz has already been used.

In Section 2 we will describe the visualization and how it was implemented. In Section 3 we describe the system around it, with which the visualization can be created, adapted and analyzed. In section 4 the example applications of the visualization are described and in section 5 we give a short summary and an insight to future work.

2 Visualization

In this chapter we will discuss the implementation of the visualization. It is based on the Javascript library called *D3* [1] for creating dynamic and interactive visualizations in the web browser. In our case we wanted to visualize a special case of a *k*-partite graph. A *k*-partite graph is partitioned into *k* groups each of a certain number of nodes. A standard *k*-partite graph does not allow for edges linking nodes of the same group. However, in order to increase the number of use cases addressed by our visualization method, we allow for such an extension.

*e-mail: uslu@em.uni-frankfurt.de

†e-mail:mehler@em.uni-frankfurt.de

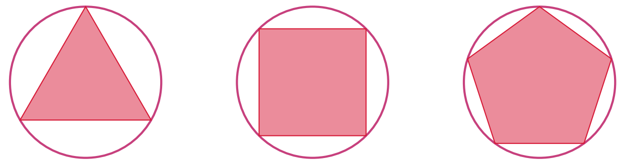

Figure 1: Generation of the paths for k = 3, 4 and 5.

First of all, each group is uniquely mapped onto a path. This is achieved by drawing a *k*-corner in a circle and using the intersections of the corner with the circle to determine the beginning and end of the corresponding path. Figure 1 illustrates the construction of paths for $k = 3..5$. In the next step, the target nodes are mapped onto these paths by respecting a uniform distribution of distances among nodes of the same path. The maximum size of a node equals the smallest distance between two nodes. All other nodes are scaled accordingly. The same procedure applies to the edges. This has the benefit that node and edge sizes do not become too large so that nodes/edges do not overlap. For edges between nodes of different groups, Bézier curves are drawn using the center of the graph as the focus point. For edges between nodes of the same group, a semi-circle is drawn outwards, resulting in an arc diagram for each group – see Figure 3 for a good example.

For determining the colors, we utilize the color palette of D3. Each group has its own color and all nodes are colored according

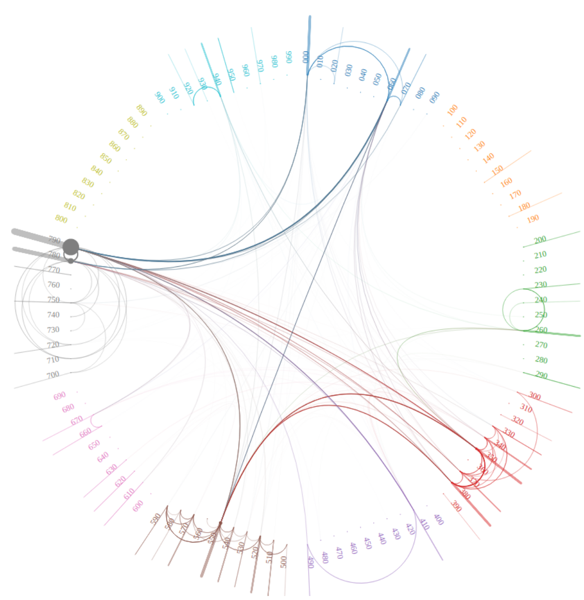

Figure 2: Example of a 10-partite graph visualizing the DDC-related topic distribution of the German Wikipedia: nodes denote classes of the Dewey Decimal Class (DDC) of the second level. Edges denote links between articles subsumed under the corresponding topic nodes.

to the color of the corresponding group. Group-internal edges are assigned the color of the group; edges between different groups are represented by a color gradient between the colors of the groups involved. This makes it easier to find out where outgoing edges end and where incoming edges come from.

3 The System

Our system consists of a web application that allows for uploading a file and for creating the visualization interactively. The format of this file is JSON, a widely used data format in support of browser communication. The input data first declares the existing nodes with all their information such as:

- group
- index
- name
- size

Then the edges are defined by means of the following information:

- sourceGroup & sourceIndex
- targetGroup & targetIndex
- name
- size

This information is sufficient to display the targeted visualization. After creation, it is possible to optimize the visualization according to specific requirements by using various modes of interaction:

- enlarging/reducing the size of nodes and links;
- enlarging/reducing the size of node and link labels;
- filtering links by referring to specific values;
- filtering groups;
- selecting nodes and displaying only adjacent edges.

4 Examples

This chapter exemplifies two use cases using PolyViz.

4.1 Topic distribution and referencing

In [3] we analyzed the topic distribution and linkage of the German Wikipedia. For this purpose we developed a state-of-the-art DDC topic classifier and used it to categorize all articles of the German Wikipedia. The link structure of Wikipedia was used to analyze which topics refer to which other topics (inter-topic links) or are thematically closed (intra-topic links). We visualized this information by means of our visualization technique. Since topics in the DDC are hierarchically organized (10 topics on the 1st level, 100 on the 2nd and 1000 on the 3rd), we referred to the 10 main topics to define the groups (paths) while the topics of the second level are taken to define the nodes on the paths. This results in a 10-partite graph as shown in Figure 2. This visualization states that a few topics dominate the German Wikipedia.

4.2 Sentence similarity

A second example is the usage of PolyViz within the TextImager [2]. TextImager is a tool that performs various *Natural Language Processing* (NLP) procedures and visualizes their results. One of these procedures concerns the computation of (semantic or structural) similarities between sentences. Figure 3 depicts the sentence similarities within a document as well as cross-document sentence similarities. Each document is given its own color and each sentence in a document is represented as a node. The thicker and stronger the links between the nodes, the more similar these sentences are. By this example, one can see which texts contain similar content – without reading the documents. One also sees that sentences of the same text (arc diagram) tend to be more similar among each other than sentences of different texts (internal links). This usage scenario is in support of distant reading.

Figure 3: Example of a 9-partite graph showing similarities of documents on the sentence level.

5 Conclusion

We present PolyViz, a new and interactive visualization system that allows for depicting a special kind of multipartite graphs. While in conventional multipartite graphs it is not allowed to have edges among members of the same group, we allow this case in order to increase the number of use cases. We additionally introduced the web application that can be used to create instances of our visualization technique and to adapt it. Further, we exemplified our PolyViz by means of two uses cases. Since the range of applications is wide, we provide this visualization technique open source via GitHub (in the case of being accepted). Currently, the visualization works for k-partite graphs with k greater than 2. In future work we want to offer the same advantages also for bipartite and for 1-partite graphs, where the latter is equivalent to a single arc diagram.

References

[1] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.

[2] W. Hemati, T. Uslu, and A. Mehler. Textimager: a distributed uima-based system for nlp. In *Proceedings of the COLING 2016 System Demonstrations*. Federated Conference on Computer Science and Information Systems, 2016.

[3] T. Uslu, A. Mehler, A. Niekler, and W. Hemati. Towards a DDC-based topic network model of wikipedia. In *Proceedings of 2nd International Workshop on Modeling, Analysis, and Management of Social Networks and their Applications (SOCNET 2018), February 28, 2018*, 2018. accepted.