Integrated Visualization of Structure and Attribute Similarity of Multivariate Graphs

Philip Berger*
University of Rostock, Germany

Mohammad Chegini†
TU Graz, Austria

Heidrun Schumann‡
University of Rostock, Germany

Christian Tominski§
University of Rostock, Germany

ABSTRACT

An interesting question one may ask about multivariate graphs is if and how structural characteristics are reflected in multivariate attribute similarities and vice versa. In this work, we propose an integrated visualization of structure S and attribute similarity A in a single overview matrix. The idea is to show S and A in the lower and upper triangular matrix, respectively. Dynamic local rearrangement of matrix cells allows the user to create a side-by-side view of S and A for a detailed inspection of both aspects for a selected node. The approach is applied to a multivariate graph of soccer players.

Index Terms: Human-centered computing—Visualization;

1 INTRODUCTION

A multivariate graph consists of nodes, edges, and multivariate data attributes. Multivariate graphs play a role in many application domains, for example, in medicine, communication, social science, or sports.

When analyzing multivariate graphs, the structural characteristics are of primary interest. Many techniques exist to visualize a graph's structure, for example, via node-link diagrams with an appropriate layout or via matrix representations with a suitable ordering. On the other hand, the attributes are important as well. Classic techniques such as scatter plots or parallel coordinates plots can be used to visualize multivariate attributes. Kerren et al. [2] provide an overview of approaches for the combined visualization of the structure and the attributes of multivariate graphs, including multiple coordinated views, attribute driven layouts, and hybrid layouts.

In our work, we are particularly interested in the relationship between the graph structure and the similarity of nodes with respect to their multivariate attributes. To understand this relationship, analysts may ask the following questions. Given a subset of nodes being similar in their attributes, do they exhibit similar structural properties? Or, given a certain substructure of the graph, do the nodes in that substructure exhibit similar attribute values?

Our goal is to support analysts in finding answers to such and similar questions. To this end, we propose (1) an integrated visualization of structure S and attribute similarity A in a single overview matrix and (2) a dynamic local rearrangement of matrix cells to lay out S and A side by side for studying their relationship in detail for selected nodes of the graph. We consider undirected graphs.

As a running example dataset, we will be using a graph of soccer players of 16 clubs participating in the Champions League. Player attributes include the number of matches played, the number of minutes played, pass accuracy, ball recoveries, and many more. Edges exist between any two players who played for the same club at some time during their career. The edge weight corresponds to the number of clubs that two players have in common.

*e-mail: philip.berger@uni-rostock.de
†e-mail: m.chegini@cgv.tugraz.at
‡e-mail: heidrun.schumann@uni-rostock.de
§e-mail: christian.tominski@uni-rostock.de

2 OVERVIEW MATRIX

The goal of the overview matrix is to show S and A in an integrated fashion. For this purpose, the lower triangular matrix and the upper triangular matrix are dedicated to showing S and A, respectively.

The structure of the graph is visualized in a standard way in the lower triangular part of the matrix. A matrix cell $c_{i,j}$ is filled with a color, if there exists an edge between the i-th and j-th node. The darker the color, the higher is the edge weight. This standard matrix encoding can support analysts in various tasks related to the graph connectivity [1].

The multivariate similarity of node attributes is visualized in the upper triangular part of the matrix. More specifically, each cell shows the similarity of a pair of nodes with respect to their associated data values. Therefore, in the upper matrix, a darker cell $c_{i,j}$ means that the i-th and the j-th node are quite similar. A brighter cell indicates dissimilar nodes.

A prerequisite for visualizing the similarity is to compute it in the first place. The computation of similarity, in particular of multivariate similarity, is a topic on its own. Many alternatives exist to calculate it, for example, by various distance metrics, projection methods, or self-organizing maps. As a pragmatic solution, we use the Euclidean distance. Additionally, we allow users to select which of the attributes should be included in the similarity calculation. This way, it is possible to focus the analysis on selected subsets of attributes.

Together, the lower and the upper triangular matrix form a visual representation where structural aspects and attribute similarity are visible in a single coherent overview. Fig. 1 shows matrices for our example soccer data. To identify patterns in the matrix the rows and columns can be sorted. Ordering nodes based on the graph structure reveals club affiliation as in Fig. 1(a). Sorting according to attribute similarity groups similar players as in Fig. 1(b).
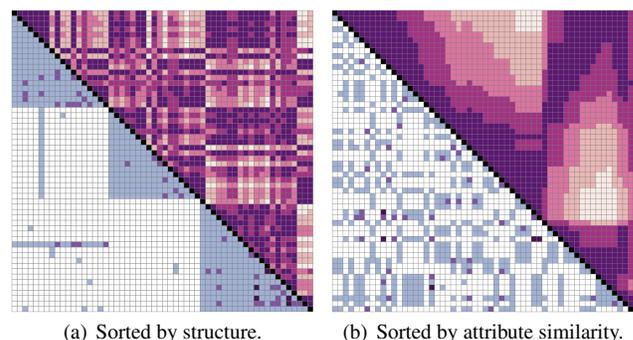

(a) Sorted by structure. (b) Sorted by attribute similarity.

Figure 1: Integrated visualization of structure and attribute similarity.

3 DETAILS THROUGH DYNAMIC REARRANGEMENT

For a detailed analysis of structure and attribute similarity, both aspects would ideally be visible side by side. However, the matrix naturally does not offer such a side-by-side view. Instead, it is necessary to read rows and columns of the matrix along L-shaped

(a) Select node. (b) Insert gaps. (c) Rotate partial columns. (d) Aligned view.
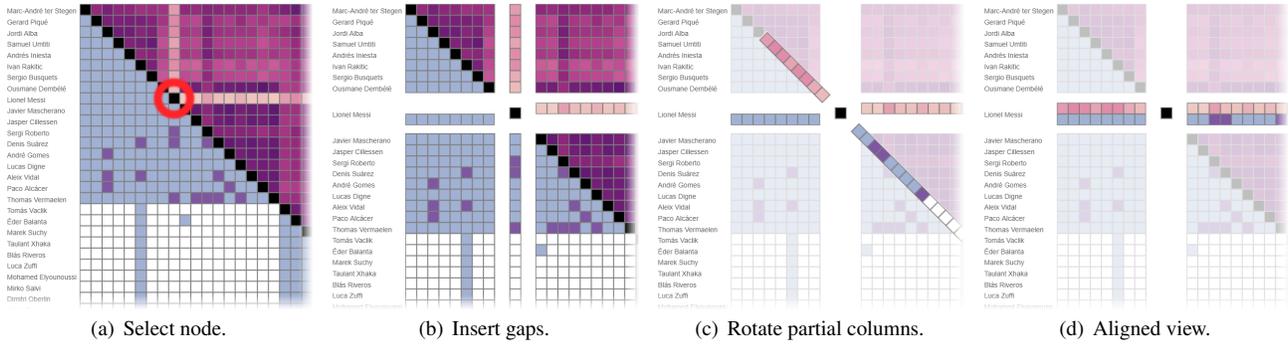
Figure 2: Dynamic rearrangement of matrix cells to create an aligned view for the detailed inspection of structure and attribute similarity.
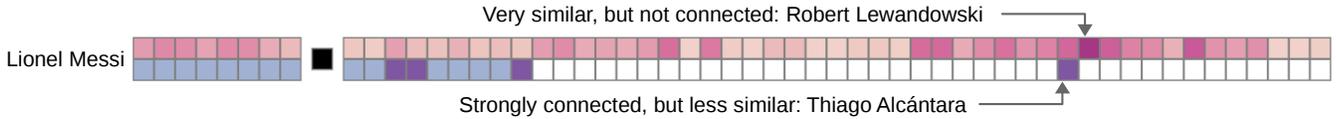

Figure 3: Detail view of player Lionel Messi. Top row: Pairwise similarity to other players. Bottom row: Connection to other players.

paths, where cells to be compared are in different parts of the matrix. So, while the matrix is good for an overview, additional means are needed to support the detailed investigation of structure and attribute similarity.

Therefore, we integrated a mechanism to dynamically rearrange matrix cells for a selected node to create a side-by-side view of its structure and attribute similarity on the fly. The first step is to let the user select the node for which structural aspects and similarity should be studied in detail. Given the selected node, gaps are inserted into the matrix to let the partial rows and columns that contain the node's information better stand out. Then the columns are rotated to align them to create two complete rows, one showing the similarity values and the other showing the structural information. To make the dynamic rearrangement easier to comprehend, it is carried out in a smooth animation. The procedure is illustrated in Fig. 2.

4 EXAMPLE USE

In order to demonstrate our approach, we apply it to an extract of the soccer player data. Fig. 4 shows the overview matrix consisting of only three clubs. The lower triangular matrix shows who shared the same clubs during the career. Sorting the matrix according to this structural aspect reveals triangular cliques of players within the same club. Darker cells show players with similar transfer history. The colored cells outside of the triangular groups represent former affiliations of the players besides their current club. If those cells form horizontal or vertical lines, this indicates that the corresponding player has been a former member of the current club.

The upper triangular matrix in Fig. 4 displays the similarities of players with respect to shot goals, attempts at goal, successful passes, and failed passes. Two players appear to be rather dissimilar to the rest of the players as indicated by brighter rows and columns. The players are Lionel Messi and Robert Lewandowski.

In order to study Lionel Messi in detail, the dynamic rearrangement is applied. In Fig. 3, one can see that Messi is dissimilar to all players who played for the same clubs as he did. There is only one player being similar (darkest cell in the similarity row), which is Robert Lewandowski, but we can see that Messi and he have never played for the same club. Yet, immediately next to Lewandowski is a player with whom Messi already played together (dark cell in the structure row). This player is Thiago Alcántara.
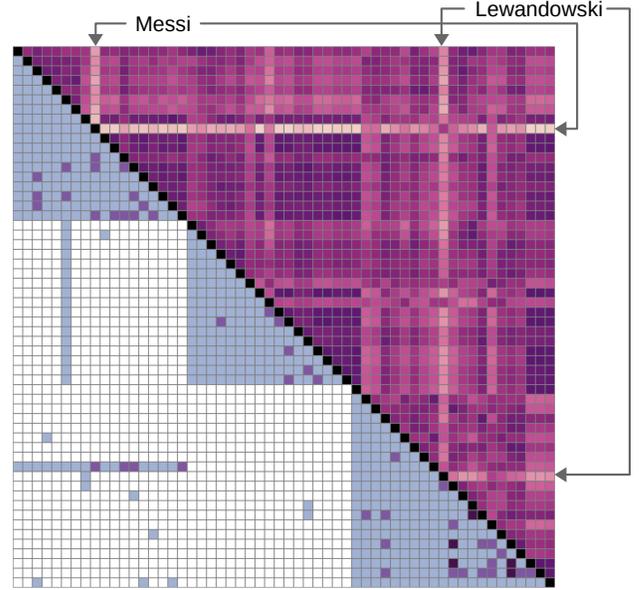

Figure 4: Overview matrix with two players standing out.

5 CONCLUSION

We think that the integrated visualization of structural characteristics and attribute similarity can be an interesting complement to alternative approaches for analyzing multivariate graphs. Our work is still in an early phase. We plan to extend it in the future with more sophisticated similarity calculations and additional interactions and dynamic rearrangements to support additional analysis questions.

REFERENCES

[1] M. Ghoniem, J. Fekete, and P. Castagliola. A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations. In *Proceedings of IEEE Symposium on Information Visualization*, 2004.

[2] A. Kerren, H. C. Purchase, and M. O. Ward, editors. *Multivariate Network Visualization*. Springer, 2014.