Abstract—The visual analysis of multivariate graphs is a challenging problem. We address the particular task of studying relations between the structure of a graph and the multivariate attributes associated with it. To facilitate this task, we propose a novel interactive visualization approach. The core idea is to show structure and calculated attribute similarity in an integrated fashion as a matrix. A table can be attached to the matrix on demand to visualize the underlying attribute values in detail. To support the visual comparison of structure and attributes at different levels, several interaction techniques are provided, including matrix reordering, selection and emphasis of subsets, rearrangement of sub-matrices, and column rotation for detailed comparison. To demonstrate the utility of our techniques, we apply them to explore relations between structure and attributes in a network of soccer players.

Keywords—Graph visualization, Human computer interaction, Multivariate graphs

I. INTRODUCTION

Multivariate graphs comprise two key aspects: the structure as defined by nodes and edges, and the multivariate data attributes associated with them. Multivariate graphs are relevant in various domains. For example, social scientists study groups of friends and their social media behavior. Sports analysts investigate team compositions based on player affiliation, relationships within teams, and individual performances.

In general, the analysis of multivariate graphs aims at understanding the graph’s structure and the graph’s attributes. We are particularly interested in the visual analysis of relationships between structure and attributes. For example, 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? Or, given two similar substructures, are their associated attributes similar as well?

Similarity obviously plays a central role in this context. It is relevant at different levels. At the level of individual data elements, it is of interest whether nodes or edges are similar. At the level of subsets, it is interesting to see whether local parts of the data exhibit similar characteristics. Finally, at the global level, we are interested in the overall similarities between structure and attributes.

This work aims to support the visual exploration of relations in multivariate graphs by a novel visualization approach. To provide a similarity-oriented overview of the data, we show structure and calculated attribute similarity in an integrated fashion as a matrix-based representation. On-demand details are available in a table-based representation showing the underlying attribute values. Flexible interaction techniques are incorporated to dynamically adapt and rearrange the visualization so that selected parts of the data can be compared visually with ease. This includes the reordering of the matrix and the table, the extraction of subsets, and the dynamic alignment of rows and columns for detailed comparison of individual data elements.

As an illustrating example, we analyze a network of soccer players. We will demonstrate how the proposed techniques can help users to identify and investigate relations between the network structure and the characteristics of the players.

Before we describe our approach in detail, we will first look at related work in the context of multivariate graph visualization.

II. RELATED WORK

Visualizing multivariate graphs requires showing the multivariate attributes and the graph structure. The general visualization literature provides a wealth of techniques for multivariate attributes. The most frequently applied techniques are parallel coordinates and scatter plot matrices [1]. Graph structures are typically visualized as node-link diagrams or matrix representations. While node-link diagrams require appropriate layout algorithms [2], matrix representations rely on suitable ordering mechanisms for the matrix [3].

Visualization approaches for multivariate graphs typically combine ideas from multivariate data visualization and graph visualization. Kerren et al. [4] provide an overview, including multi-view and single-view approaches.

Multi-view approaches show the different aspects of multivariate graphs in separate, but linked views. For example, Shannon et al. [5] combine node-link diagrams and parallel
coordinates. Lex et al. [6] use a combination of parallel coordinates, heat map, and a bucket technique to visualize pathways. In related works, Partl et al. [7], [8] visualize the graph structure via node-link diagrams, while attributes are shown in a tabular visualization.

Single-view approaches show structure and attributes in an integrated fashion. That is, aspects of attributes are integrated in the visualization of the structure or the other way around. For example, Cao et al. [9] use node-link diagrams where nodes are represented as glyphs to encode multivariate attributes. Van den Elzen and Van Wijk [10] aggregate selected nodes and show within them the value distributions of node attributes. Attribute-driven approaches arrange the graph structure according to the underlying attribute values. For example, Shneiderman and Aris [11] group nodes based on attributes and show the structure-forming edges only for selected parts of the data. In follow-up work, Rodrigues et al. [12] improve the readability of the graph structure. Wattenberg [13] uses a grid-based layout of aggregated nodes and edges to show relations between attributes and edges. Bezerianos et al. [14] and Eichner et al. [15] use bi-variate scatter plots and superimpose the graph structure. Major and Basole’s Graphicle [16] supports dynamically switching between layouts that facilitate structure-related and attribute-related analysis. The technique by Nobre et al. [17] juxtaposes a filtered graph structure and a table of attribute values.

The existing approaches have different pros and cons. Providing multiple views allows for using the most appropriate visualizations for structure and attributes. However, structure and attributes are shown in different parts of the display, which requires additional mental effort to integrate the two aspects. The simultaneous display of structure and attributes in a single view does not have this drawback. Relations between structure and attributes can be discerned for individual data elements and for local subsets. Yet, the increased information density can lead to increased cognitive load, which can make it more difficult to study relations on a global scale.

In summary, visualizing multivariate graphs remains a challenge. Our work aims to strike a balance between integrated visualization and additional on-demand views. As we will describe in the following sections, our approach uses an overview matrix that integrates major structural and attribute characteristics and offers a linked table view for further detailed analysis on demand.

III. APPROACH OUTLINE

Our focus is to support users in analyzing the relation between structure and attributes in multivariate graphs. To this end, the following requirements need to be fulfilled:

R1 Show similarity Similarities between structure and attributes should be identifiable easily.

R2 Integrate structure and attributes Graph structure and attributes should be displayed in the same view.

R3 Enable drill-down The user should be enabled to concentrate on subsets and individual data elements.

R4 Support exploration Interaction techniques need to be provided for exploring structure and attributes.

Addressing these requirements, we design our approach as follows. The focus on similarity motivates the calculation of attribute-wise node similarity. An integrated view of structure and attribute similarity is provided by a novel form of matrix representation. The similarity values are color-coded in the matrix along with the usual display of the graph structure. From the matrix overview, users can drill down into details as provided by a table representation on demand. The table visualizes the attributes in detail to make clear to users how the underlying data values contribute to the calculated similarity. Further drill-down and exploration is facilitated through dynamic adaptations and rearrangements of the aforementioned visual representations.

In the next section, we describe the integrated matrix visualization and the on-demand table visualization in detail. After that, we explain the developed interaction techniques.

IV. VISUALIZATION OF STRUCTURE AND ATTRIBUTES

As indicated, our approach utilizes matrix and table representations. Both will be explained next.

A. Matrix Representation

Matrix representations have proven to be useful for communicating the structure of graphs [18]. In a regular matrix visualization, rows and columns correspond to graph nodes, and a matrix cell is marked if an edge exists between the row-node and the column-node. Our idea is to extend matrix representations such that the similarity of node attributes can be visualized along with the graph structure.

To this end, we use the upper and lower triangular parts of the matrix for different purposes, as illustrated in Fig. 1. One triangular part encodes the graph structure as usual by marking the edges. Integrated in the marking is a color coding to represent values of edge weights. The darker the color, the higher is the weight. The other triangular part does not show the edges, but instead visualizes attribute similarity of the nodes.

In a first step, the pair-wise similarity of the nodes needs to be calculated. We implement a basic approach that uses the Euclidean distance of the underlying attribute values. Where appropriate, additional derived attributes can be included, for example, graph theoretical measures such as degree, betweenness or centrality. Note, however, that there are many different ways of calculating multivariate similarity and that specific applications might need different, potentially more complex methods, such as projection techniques [19] or self-organizing maps [20].
Once computed, the attribute similarity of the nodes is visualized in the matrix by color-coding the cells. That is, the color of cell $c_{i,j}$ represents the attribute-wise similarity of the $i$-th and $j$-th nodes. Depending on the calculated similarity, different color scales can be used for the visualization. If the similarity is in the interval $[0, 1]$ (0 for not similar and 1 for similar), then we use sequential color scales from ColorBrewer [21]. In this context darker cells indicate similar nodes. In cases where similarity is in the interval $[-1, 1]$ ($-1$ for dis-similar, 0 for not similar, and 1 for similar), diverging color scales would be appropriate.

Fulfilling requirement R1, the matrix visualization shows the data in a way that facilitates the investigation of similarities. Moreover, the matrix integrates the visual representation of structural characteristics and attribute similarity in a single coherent view as demanded by requirement R2. Relations between structure and attributes can be discerned by comparing cells or regions in the upper diagonal matrix with the corresponding cells or regions in the lower diagonal matrix and vice versa.

Finally, we can utilize the matrix cells in the main diagonal. Usually, these cells represent loop edges, that is, edges that connect a node to itself. For graphs without loop edges, the main diagonal can alternatively be used to visualize properties of the nodes, for example, by color-coding a selected node attribute, as illustrated in Fig. 1.

### B. Table Representation

While the matrix representation allows users to see which nodes are similar, it is not possible to understand why nodes are considered similar. In fact, the calculation of similarity values is a kind of data abstraction that condenses down the multivariate attributes of two nodes to a single similarity value.

To enable users to understand why data elements are similar, we provide an on-demand table visualization of the attribute values. As illustrated in Fig. 1, the table is directly attached to the matrix to establish a tight link between attribute similarity and attribute values. Every row of the table corresponds to the node and its attributes in the juxtaposed matrix row. The table cells visualize data values by means of two-tone pseudo coloring. This visual encoding combines the two visual variables of color and length to provide at the same time an overview and the possibility to read data values precisely [22].

With the help of the table representation, users can drill down and study the similarity of nodes in detail. Yet, this
is only a first means to address requirement R3. In the next section, we describe interaction techniques that enable users to further drill-down into details to explore the data for relationships between structure and attributes.

V. Interaction Techniques

The visualization introduced so far provides an overview of graph structure, attribute similarity, and attribute values. As formulated in requirement R4, we also need to provide interaction techniques that facilitate a detailed exploration of the data. In light of our focus on similarity, we need to support the interactive visual comparison of different subsets of the data. This involves the selection of subsets and the dynamic adaptation and rearrangement of the visualization to suit the comparison task [23].

A. Similarity Selection and Matrix Reordering

The first step in any comparison task is to identify interesting subsets to be compared. For our overview matrix, the identification of such interesting subsets depends on two aspects: the attributes in the similarity calculation and how rows and columns are ordered [3].

To focus on specific subsets of attributes, users can select which of them should be included in the similarity calculation. Attributes can be added and removed by choosing the corresponding table headers. The similarity visualization in the upper triangular matrix is updated on the fly.

To reveal interesting subsets in both parts of the overview matrix, it can be sorted by (i) structural characteristics, (ii) attribute similarity, and (iii) attribute values. Depending on the applied strategy, different patterns can be made visible, as shown in Fig. 2.

By ordering the rows and columns based on structural characteristics, such as node degree or betweenness, different aspects of the graph structure can be emphasized. For example, the lower triangular matrix in Fig. 2(a) clearly shows cliques as triangular patterns.

Sorting the matrix based on the attribute similarity can be helpful to reveal groups of similar and dissimilar nodes. We can sort either by the similarity of all pairs of nodes or by the accumulated similarity of nodes. Fig. 2(b) shows an example with two discernible groups after a pair-wise sorting according to the similarity with respect to a selected node.

Another strategy is to sort rows and columns based on the node attribute values. This is facilitated by the headers in the table very much like in common spreadsheet applications. After clicking a header, table and matrix are sorted by the values in the chosen column. This can lead to interesting patterns. In Fig. 2(c), a single node, distinguished by a differently colored horizontal and vertical line, stands out in the first third of the matrix.

B. Subset Selection and Emphasis

As mentioned before, relations between structure and attributes can exist on different levels of the graph. The matrix and the table allow users to discover relations on a global level. To find local patterns, it is necessary to focus the exploration on subsets of the graph. To this end, we integrate an interactive highlight and filter technique.

The first step is to select subsets of nodes that should be investigated in more detail. The selection is carried out by marking cells of the matrix or the table. Fig. 3(a) illustrates the marking of two local patterns in the matrix. Marking cells in either triangular part of the matrix automatically selects the corresponding cells in the other triangular part as well. After the selection, the corresponding nodes are determined, as indicated in Fig. 3(b). By marking rows in the table or cells of the matrix' main diagonal, all associated nodes are selected. The selected subsets are then to be emphasized.

For a brief contemplation about the selected subsets, we provide a rather subtle highlighting based on dynamically inserting gaps into the visual representations. Fig. 3(b) illustrates how these gaps divide the matrix into sub-matrices. For a stronger emphasis of the selected subsets, it is possible to filter out those parts of the visualization that do not contain any selected nodes. This is illustrated in Fig. 3(c).
C. Sub-matrix Rearrangement

Emphasizing selected parts of the data can already support their comparison. Yet, the subsets could still be far apart in the matrix, which unnecessarily complicates the comparison. Therefore, it makes sense to dynamically rearrange the selected sub-matrices for a closer inspection. Side-by-side layouts are well suited for comparison tasks [24].

To facilitate the comparison of subsets, we allow sub-matrices to be detached from the main matrix. This effectively creates a hybrid NodeTrix layout where selected subsets are shown as separate matrices being connected via links [25]. Fig. 4 provides an example with three detached matrices, where two of them also show their associated attribute table. Once the sub-matrices have been studied in detail, they can be integrated back into the main matrix.

D. Column Rotation

The techniques described so far allow users to study the data at a global level and at the level of local subsets. Finally, we also need to facilitate the analysis of individual data elements. In particular, we want to support the detailed comparison of structure and attribute similarity of individual nodes.

The matrix already visualizes the necessary information, and the selection mechanism can be used to emphasize an individual node as illustrated in Fig. 5(a) and 5(b). However, the separated partial rows and columns in Fig. 5(b) form \(\uparrow\) and \(\downarrow\) shapes along which data values must be read from the visual representation. This hampers an effective comparison of structure and attribute similarity. Ideally, both aspects would be arranged side by side like so =.

Therefore, we dynamically rotate the matrix column for a selected node, which creates a side-by-side view of its structure and attribute similarity on the fly. As shown in Fig. 5(c), the partial columns are rotated to align them with the partial rows. As a result, we obtain the two complete rows in Fig. 5(d), 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.

In summary, choosing attributes for the similarity calculation, reordering, selection and emphasis, sub-matrix rearrangement, and column rotation provide the user with the interactive tools needed to explore and compare the relation between structure and attributes in multivariate graphs in detail. Next, we demonstrate our approach with an example soccer player network.

VI. USE CASE

We use a dataset of soccer players of 16 clubs participating in the Champions League season 2017/18. Nodes of the network are players with attributes such as the number of matches played, the number of minutes played, pass accuracy, ball recoveries, and many more defensive and offensive characteristics. Edges exist between two players if they have 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.

The goal of the analysis of this multivariate network is to find out whether patterns occurring in the structure are re-
Figure 5. Dynamic rearrangement of matrix cells to create an aligned view for the detailed inspection of structure and attribute similarity.

Figure 6. Extract of the soccer network ordered according to team affiliation. Cliques in the lower matrix depict different soccer clubs. The upper triangular matrix shows the similarity between players based on their matches and minutes played.

related to the attributes, or vice versa. Of particular interest are the team compositions and individual player performances that stand out from others. To this end, multiple clubs have to be compared to each other before they are studied more closely.

To identify structural patterns, the matrix is sorted based on the current club affiliation. This reveals triangular cliques in the lower part of the matrix containing players in the same club. Fig. 6 shows an extract with three clubs. Darker cells indicate that players have played for the same club before in their career, which indicates similarities in their 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, the corresponding player has been a former member of the club whose triangular clique is parallel to the line.

To identify relations between club affiliation and team composition, we select the attributes **matches played** and **minutes played** in the attribute table. As a result, the upper triangular matrix shows the similarity of players with respect to these two attributes. In Fig. 6, we can see that for the two clubs at the top left and in the center, their triangular cliques in the lower triangular matrix are each mirrored by two smaller triangular patterns of similar players in the upper triangular matrix. Moreover, between the darker triangles of similar players, there are rectangular regions with brighter
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 Messi in detail, the matrix cells for the node are dynamically rotated. In Fig. 7, 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, indicated by a dark cell in the similarity row, which is 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, presented by a dark cell in the structure row. This player is Thiago Alcántara.

Taken together, the findings presented in this section demonstrate how well our techniques support the visual exploration of relations between structure and attributes in multivariate graphs. Starting with sorting the integrated matrix, users can apply different interaction techniques to study selected parts in detail. This allows users to quickly identify key characteristics of interesting data elements and to compare them with other parts of the data.

VII. CONCLUSION AND FUTURE WORK

In this work, we presented a novel approach for visually exploring relations between structure and attributes in multivariate graphs. The core idea is to display structural aspects and multivariate attribute similarity in an integrated fashion as a matrix, and to provide details of the attributes in an on-demand table. Several interactive techniques can be used to adapt and rearrange the visualization to facilitate the comparisons of both structure and attributes at different levels of granularity. In conclusion, we think that the simultaneous representation of structure, attribute similarity, and multivariate attributes can provide new insight into the relation between structure and attributes of multivariate graphs.

In the future, we plan to further improve the utility of our techniques. For example, tracing paths through the graph by jumping from one matrix cell to another is known to be difficult [26]. We plan to better support path-related exploration tasks by further strengthening the interplay of matrix and node-link representations. The NodeTrix layout provides a good starting point.
Another promising research direction would be to improve the dynamic rearrangements to facilitate comparison tasks. So far, the rearrangements are triggered manually by the user. An alternative could be to automatically create suitable arrangements based on the calculated similarity of data elements [27].

This would also require improving the calculation of similarities. We already allow users to choose the attributes to be considered for the similarity calculation. But this is only a first step. Additional methods and control mechanisms would be needed for a flexible task-based calculation of similarities.

REFERENCES


