Graph Exploration by Multiple Linked Metric Views

A. Panagiotidis¹, M. Burch¹, O. Deussen², D. Weiskopf¹, T. Ertl¹
¹Visualization Research Center, University of Stuttgart, Germany
{panagiotidis,burch,weiskopf,ertl}@visus.uni-stuttgart.de
²Dept. of Computer and Information Science, University of Konstanz, Germany
oliver.deussen@uni-konstanz.de

Abstract—The visualization of relational data by node-link diagrams quickly leads to a degradation of performance at some exploration tasks when the diagrams show visual clutter and overdraw. To address this challenge of large-data graph visualization, we introduce Graph Metric Views, a technique that enriches the visualization of traditional layout strategies for node-link diagrams by additionally allowing an analyst to interactively explore graph-specific metrics such as number of nodes, number of link crossings, link coverage, or degree of orthogonality. To this end, we support an analyst with additional histogram-like representations at the axes of the display space for graph-specific metrics. In this way, a cluttered and densely packed node-link diagram becomes more explorable even for dense graph regions: The user can use the distribution of metric values as an overview and then select regions of interest for further investigation and filtering.

Keywords—Graphs, node-link diagrams, metrics.

I. INTRODUCTION

Graphs allow modeling and illustrating relationships among a group of objects. Euler [1] was one of the pioneers who modeled such relationships in a graphical way by using nodes for the objects and links expressing the pairwise relations when he found an abstraction for the problem of the “Seven Bridges of Königsberg”. Nowadays, even sophisticated methods for graph drawing produce cluttered visualizations for large and dense networks. This is even more problematic considering that there are further attributes in the data that cannot be integrated into a node-link diagram.

In this paper, we introduce Graph Metric Views: Histograms attached to the sides of a node-link diagram (see Figure 1), similar to marginal histograms aligned with 2D scatter plots. These views depict information about the layout of the graph, its graph-theoretic properties, and the underlying data. While users are exploring the graph and interacting with it, the metric views are updated accordingly. Furthermore, interesting regions of the metric views can be visually linked to the node-link diagram by coordination strategies like brushing and linking or direct visual linking.

We argue that this approach is advantageous when the graph becomes cluttered [2] or all visual channels are exhausted in the node-link diagram and thus no further information can be shown. Furthermore, an analyst can explore regions that are not clearly visible in the layout due to occlusion and overdraw. For example, cliques may not be perceived due to the high link density in these regions, whereas we can clearly show where and how big they are.

Our interactive technique is designed around the Visual Information Seeking Mantra [3]: overview first, zoom and filter, then details-on-demand. We employ multiple coordinated views [4], [5] and use brushing and linking [6]. To preserve the user’s mental map we introduce the concept of a graph minimap to provide context information. This minimap always shows the whole graph in a smaller view while a user focuses on parts of it in the node-link diagram. Our approach is extensible since we rely on common techniques for the graph layout and its rendering, while further metrics can be added to support new data sets.

II. RELATED WORK

Graph drawing is a well-established field, with many ways to visualize relational data and networks directly [7], [8] or derived from multivariate, temporal data [9]. Typically, a node-link diagram is used that is laid out in a force-directed approach [10] or by stress majorization [11], often motivated by cognitive and aesthetic reasons [12], [13].

Despite sophisticated algorithms, graph layouts still result in local and global hairball-like structures for complex and large input data. This visual clutter is undesirable because it leads to degradation of performance at some exploration and analysis tasks [2], for example due to many link crossings. In general, there are many ways to deal with and reduce visual clutter [14].

Regarding node-link diagrams specifically, there are several ways to improve the presentation to reduce visual clutter and to reveal interesting patterns. Edge bundling is a prominent method to combine links according to their hierarchy [15] or in a force-directed manner [16]. Similarly, Cui et al. [17] bundle links by routing them through points of a virtual mesh that is modified interactively or automatically. Several splatting approaches deviate from the typical link drawing and instead use densities to depict the relations in the data [18], [19], [20]. Level-of-detail approaches combine link cumulation with density-based node aggregation [21]. Besides the discussion and analysis in the respective publications, there are many evaluations on the topic of modifying links for graph visualization [22], [23].
In addition to adapting the visualization, there are many interaction techniques to cope with clutter and dense graphs in general. The MoireGraphs approach [24] arranges nodes in a radial layout and employs a focus+context metaphor to show interesting nodes enlarged with their neighborhood. The EdgeLens method [25] distorts links in the region of a lens to resolve ambiguities. Similarly, Tominski et al. [26] allow for local exploration of dense regions in hierarchies and graph visualization with a fish-eye metaphor. EdgeAnalyzer [27] is an interaction technique to locally group links and render them differently so that their paths can be traced more easily and ambiguities are resolved.

These methods already allow for great flexibility and freedom when dealing with graphs. Thus, our approach is complementary to them and tackles the problem of showing additional information that cannot be included in a node-link diagram. We achieve this by attaching additional views on different aspects of the data directly to the node-link diagram. Our visual design adopts the strategy of multiple coordinated views [4] that we apply and adapt to the specific needs of graph visualization. Network visualization frameworks such as Cytoscape [28] or Gephi [29] are also designed for graph-related tasks and provide additional perspectives on the graph data. However, none of them provides views like ours on the graph data in a similar way. In short, our Graph Metric Views can be combined with any variant of node-link diagram and layout method to facilitate the exploration of complex graphs and additional information attached to those graphs.

Our visualization technique is loosely related to NetLens [30], a tool for visual query refinement. It allows the user to explore relational data by showing different metrics as bar charts. Through selection and filtering, it is possible to drill down into many aspects of a data set. In addition to showing data-related metrics, we also visualize the relational structure of the data set. Furthermore, we utilize layout-related as well as graph-theoretic metrics besides data-inherent information.

Our technique is closely related to VINCENT [31]. There, graphs are laid out radially and have bar charts of network centralities attached. These centralities are node metrics based on graph-theoretic properties of a graph, like the degree or closeness. VINCENT uses histograms of these centralities but shows them detached from the graph visualization. Our approach shares the same motivation, namely that often more information is needed in a graph visualization than can be meaningfully presented in that same visualization. Yet, we do not enforce any layout for the graph and allow for any kind of metric related to the data set. Furthermore, we attach the histograms directly to the graph view to allow for visual linking between the metrics and the node-link diagram.

III. GRAPH METRIC VIEWS

Relational data can be modeled as a (directed) graph $G = (V,E)$ consisting of $n$ vertices $V := \{v_1, \ldots, v_n\}$. The set of edges $E \subseteq V \times V$ denotes the relations between vertices, while their weight $w : E \rightarrow \mathbb{R}$ describes another attribute of the data set. Additional attributes have to be shown through visualizations of the graph, for example by laying out the vertices as node-link diagram or by using different shapes for nodes and edges. In the following, we denote elements of the abstract graph by vertices and edges,
while nodes and links are used when we refer to the layout or visual representation of the graph.

Typically, node-link diagrams suffer from visual clutter and overdraw that are the result of high spatial densities of elements within regions of the graph. In Section II, we mentioned techniques to cope with this problem. Yet, it is not easy to convey further information in node-link diagrams. Thus our goal was to design a technique that amends existing node-link diagrams and allows for enriched interaction methods.

Figure 1 depicts an overview of our visualization technique. A graph view shows the users’ focus, an interesting region of a node-link diagram, while the minimap provides its context. We attach additional linked views to the sides of the node-link diagram that show arbitrary metrics of the data and its representation (see Figure 2). These metric views depict a one-dimensional projection of the display space of the node-link diagram to one of its axes. Section III-A describes the different kinds of metrics, while Section III-B details how these can be visualized.

A. Metric Types

For our metrics, we distinguish between node characteristics and link characteristics. For example, the vertex degree is a property of nodes, while the edge weight and the link length are link properties. In both cases, we can identify three types of metrics that differ in the way their values are determined.

Layout-related metrics are based on the visual representation of a graph, i.e., the rendering of the nodes and links as graphical shapes. As such, these provide straightforward information, like the number of nodes and links or the sum of the edge weights along the abscissa or ordinate. They can be further refined, e.g., by considering or counting only incoming or outgoing links per node, link intersections, or angles between links.

Graph-theoretic metrics provide deeper insight into the relational and hierarchical aspects of graph data sets. While cliques and communities can often be seen in certain layouts easily, they may be hard to spot or discern due to clutter and overdraw. Additional metrics like centralities (e.g., eccentricity, closeness) are often interesting but cannot be embedded in the node-link diagram itself if there are no visual channels available [31]. We show these in our metric views as supplemental information attached to the original graph layout.

Data-inherent metrics can be used to correlate arbitrary values from the data to the layout. As an example, let us consider a software project where typical call graphs contain many attributes besides the call relation: file sizes, lines of code, hierarchy levels, various types of complexity, and so on. There is also information from the development and test environment (e.g., execution durations, code coverage) as well as the version control (e.g., revision count, last committer). These metrics apply to nodes (e.g., lines of code of a method) as well as to links (e.g., the distance of two methods in the package hierarchy).

Since metric values are accumulated layout-related, graph-theoretic, or data-inherent properties of a graph and its visual representation, it is possible to further aggregate them by using the minimum, maximum, or average values instead, where applicable. For example, the average length of links can be used as an indicator for the aesthetics of layout methods.

B. Metric Views

For visualization, the metric values from Section III-A are accumulated into bins with regard to the display space of a graph layout, similar to histograms. For this accumulation, we project nodes and links onto a one-dimensional axis. Thus, the metric views to the left/right relate to the ordinate of the node-link diagram, while the views at the top/bottom relate to the abscissa.

The computation of smooth histograms is a classical problem of statistical graphics. Typically, kernel density estimation techniques are employed [32]. For histograms of node characteristics, we have the traditional problem of building histograms from point samples. For edge-oriented metrics, we build the histograms for intervals that correspond to the projection of the edge to the histogram axis. In either case, we use a box-filter as kernel, and use pixel-sized bins for the representation of histograms. The width of the box filter can be chosen by the user; the default value is given by the width of the node’s visual shape in image space.

The metric views can be stacked in arbitrary order to allow for investigation of multiple attributes of the data set and graph. While we do not limit the number of active views, it is foreseeable that too many views quickly clutter the display. On the one hand, users need to decide how many views they use and how much screen space they allocate. On the other hand, computing and rendering the metrics can be expensive, thus slowing down the user interface as a whole. This issue is alleviated through caching and manual pausing of views, i.e., users can disable data and visual updates to specific views if they are too sluggish.

C. Interaction

All views are linked with each other through brushing and linking. By separate filtering in each metric view, it is possible to drill down into interesting regions of the graph that are occluded in the node-link diagram. Furthermore, marking regions in the metric views highlights the corresponding parts in the graph view so that users can see which parts of the graph contribute to the metric value. This is important to mentally link the elements of the node-link diagram to peaks or other visual features in the metric views. Additionally, visual elements of the graph that have been highlighted through selection in a metric view can be used to
filter the graph itself. For example, nodes that are not marked can be hidden, while marked nodes (and their neighborhood) will remain visible.

The users’ focus is the graph and its visual representation, respectively, which typically employs most of the screen-space (see Figure 1). Exploration is possible through typical navigation like scrolling, zooming, picking, and filtering. Interactions with the graph view automatically update all metric views so that these always correspond to the visible area of the graph. Furthermore, interactions with the metric views affect the graph view, for example scrolling in the metric view or highlighting regions.

To preserve the users’ mental map, we provide a minimap of the whole graph (see bottom left of Figure 1). This shows also the context of the graph view in which the visible region is highlighted. Users can navigate the graph through the minimap in the same way as through the graph view (i.e., zooming, panning), which in turn updates the graph view as well as the metric views.

Information like node labels and edge weights are available as tooltips on-demand. Metadata and settings—for example vertex and edge count, or the layout used—can be shown as overlay anytime. Through saving the graph fully or partially along with the configuration of the metric views, it is possible to collaborate or continue an analysis at a later point in time.

D. Implementation Details

We implemented our technique using C++11 and Qt 5.1. The igraph [33] library is used for all graph-theoretic operations and computing the layouts. While igraph is highly flexible, it is a single-threaded CPU-only library. Thus, computing layouts of large graphs or finding cliques might take a long time. To improve the responsiveness of the user interface we moved all igraph operations to a dedicated thread. A progress bar coupled to igraph’s built-in progress reporting informs the user about the status of these operations. For example, while cliques are sought, we can still navigate the graph or interact with the metric views.

IV. Case Study

We demonstrate how our visualization approach can be used for a typical real-world problem of visual graph analysis in the field of software engineering. In software systems, it is important to analyze properties like coupling and dependencies. These systems are often modeled as directed call graphs where vertices represent full-qualified method names. Edges between two vertices represent the source method calling the target method.
Figure 3 shows such a call graph of Cobertura\(^1\), a code coverage tool for Java. At this stage, the analyst does not know details about the data set, except for very basic high-level information: It is a directed graph with 5,000 vertices and 11,319 edges, each vertex contains the name of a method and its package, and edge weights represent the number of times a method called another one. Now, the analyst wants to obtain insights into that data set, following the Information Seeking Mantra [3]. We use a Fruchterman-Reingold [10] layout to find methods that call many other methods; such dense areas indicate high coupling, which is unfavorable in software engineering. Yet, from the initial layout it is unclear from which packages these methods originate, i.e., from the Java SDK, the Cobertura developers, or some third party dependencies. Though this information is important, it cannot be embedded into the graph layout easily.

We use a topographic color coding, mapping low edge weights to blue, medium weights to green, and high weights to red. This way, we can immediately spot an area with many red links (see Figure 3), which indicate the highest method call frequency and should be investigated first. To explore this area we attach a data-specific metric that shows the distance in the package hierarchy between two nodes as property of the links. We use both the average as well as maximum package distance to reveal outliers that might be concealed in the average (see the outer metrics in Figure 3). This reveals some interesting patterns in the metrics, for example two peaks in the average package distance near the region with the red links. To find out whether these peaks are due to the high density of links in that area, we attach a layout-related metric for the number of edges (see the innermost metric in Figure 3). We see that the peaks are in fact within regions with few links and thus few method calls. We assume that these must be very specific methods that gather information from many places, since they have high package distance and call each other frequently, but not as often as other methods.

To further investigate this, we highlight the peaks in the metric views to correlate them to the node-link diagram. Then we zoom into this region, which updates the metric views, revealing new peaks. We repeat this process until we reach a view of the graph where the second metric, the maximum package distance, is constant for the visible part of the node-link diagram (see Figure 4(a)). Subsequently we remove that metric view, since it serves no purpose at this point. The metric values are very similar in this region, so we filter on the metric views to only show regions where

\(^1\)http://cobertura.github.io/cobertura/
the average package distance is above a threshold. To further improve the view on the graph we hide all nodes and links that are not highlighted by any metric view (see Figure 4(b)).

Our focus is now on a region with high average package distance as well as few links (with respect to the rest of the node-link diagram). We now replace the edge count metric with the node metric that represents the maximum hierarchy level of methods (see Figure 4(c)). We spot an outlier (a singular method with the highest hierarchy level in this part of the node-link diagram) and can easily identify the method as CopyFiles.copy in the package net.sourceforge.cobertura.reporting.html.files, which is not called often. Finally, we increase the thresholds gradually on the metric views and mark the new peaks. Now only few methods are left for consideration (see Figure 4(d)), which can be inspected manually through their tooltips. At this point the analyst can look up these methods in the source code to understand their behavior.

In conclusion, our approach allows for top-down interaction to match structures of the node-link diagram to the metric views as well as vice versa. This interaction between metric views and node-link diagram allows for a drill-down into large data sets. Visual signatures and prevalent structures can be found in the node-link diagram (cluster-like structures) and correlated to the metric views (intervals with high values). Furthermore, outliers are visible as peaks in the metric views as well as in the node-link diagram as links with substantially different colors than the others. Most importantly, our metric views show attributes of the data that were not visible before.

V. LIMITATIONS

A problem of our approach is rooted in the position of our metric views. Let us assume a node-link diagram with two dense regions in the same vertical area. While a vertical metric view will accumulate the values correctly, it is impossible to discern the contributions of the clusters, since they are in the same region of the metric. This ambiguity can already be resolved, for example by moving one cluster out of the graph view, so that their contribution to the metric can be seen separately. Alternatively one can add the same metric horizontally and highlight the affected regions; note though that this fails if there are also nodes or links in that
region that are not part of the cluster. In essence, this is a drawback we need to investigate further, possibly employing a user study with different scenarios.

The metric views are drawn at the sides of the graph view by design, but this is also problematic in some ways. For one, metric views become harder to read with each additional metric view for a given side of the graph view. Then, resizing those metric views to improve the readability reduces the available screen space for the graph. This is a trade-off between available space on the screen and the usability of our approach. We need to investigate this phenomenon in a user study to determine when the usability falls off and the metric views become impractical.

As we detailed in Section III-B, some metrics are expensive, for example calculating link intersections (quadratic complexity) and finding shortest paths for each node. This disrupts the user experience and slows the graph exploration down due to an unresponsive user interface. On the one hand, users can disable offending metrics and enable them later on. On the other hand, it is not obvious which metrics to disable, since their slowness might depend on the graph, its underlying data, and the metric itself. Additionally, the overall user experience is influenced by the compute-intensity and number of active metrics. A reasonable approach could be to monitor the runtime for each metric calculation and disable those above a threshold when many updates or interactions are queued.

VI. SUMMARY AND FUTURE WORK

In this paper, we have introduced Graph Metric Views: supplemental views to a node-link diagram that show various metrics related to network data. The main advantage of our approach is that we show additional information of such data sets that cannot be integrated into the node-link diagram due to visual clutter and exhaustion of visual channels. Thus we can reveal patterns that would otherwise be hidden in hairball-like structures or are not visible at all. We complement existing graph drawing techniques by basing our metrics on layout-related attributes of the node-link diagram, graph-theoretic properties, and data-inherent information. Using the metric views, we can explore aesthetics criteria for graph visualization as well as drill into unknown data sets easily. Our coordinated views are connected through brushing and linking, resulting in an interactive technique for analysis of local and global properties of graphs. We have illustrated the usefulness of this graph analysis approach by exploring a software call graph.

In future work, we will evaluate our technique in a controlled user study with typical tasks that are required when reading and understanding graph structures. We also want to extend our approach to dynamic graphs, i.e., those that change over time, and investigate how it applies to further layouts and non-traditional graph visualizations. Finally, we plan to look into further, more complex metrics as well as alternative visualizations of the metrics themselves.

ACKNOWLEDGEMENTS

This work was partially funded by the Federal Ministry of Education and Research of Germany (BMBF) as part of the FeToL project.

REFERENCES


