EUROVIS 2019
R. S. Laramee, S. Oeltze, and M. Sedlmair
(Guest Editors)

Volume 38 (2019), Number 3
STAR – State of The Art Report

The State of the Art in Visualizing Multivariate Networks

Carolina Nobre[1], Miriah Meyer[1], Marc Streit[2], and Alexander Lex[1]

[1]University of Utah, Utah, USA
[2]Johannes Kepler University Linz, Austria

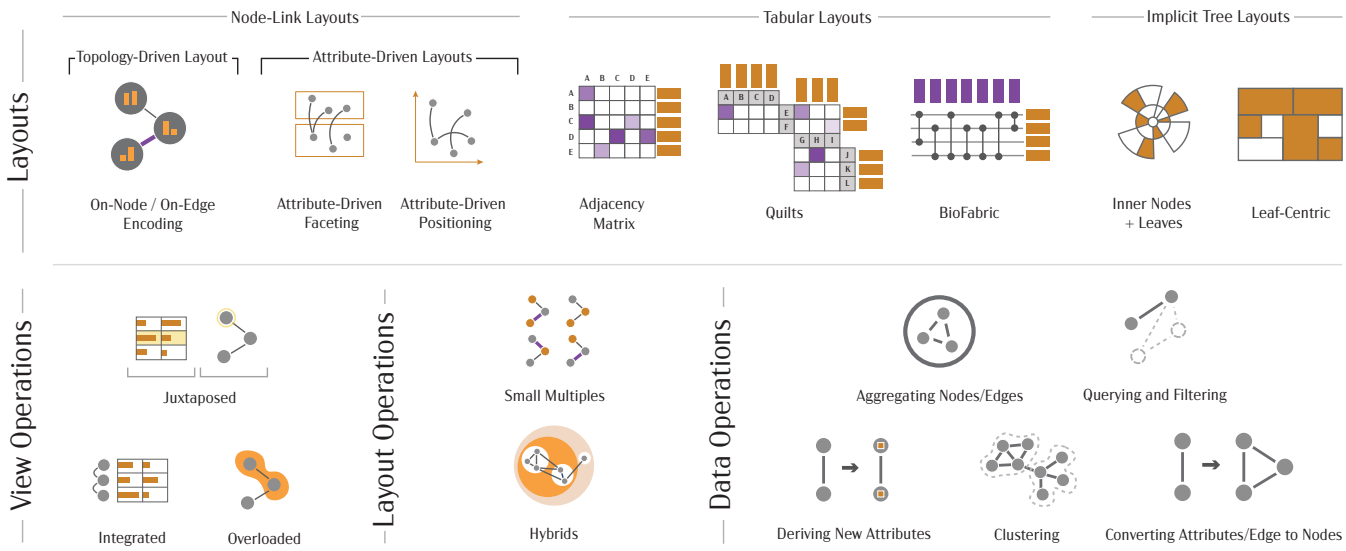

Figure 1: *A typology of operations and layouts used in multivariate network visualization.* ***Layouts*** *describe the fundamental choices for encoding multivariate networks.* ***View Operations*** *capture how topology and attribute focused visualizations can be combined.* ***Layout Operations*** *are applied to basic Layouts to create specific visualization techniques.* ***Data Operations*** *are used to transform a network or derive attributes before visualizations. Icons are colored to reflect node attributes (orange), edge attributes (purple), and topology (grey).*

Abstract

Multivariate networks are made up of nodes and their relationships (links), but also data about those nodes and links as attributes. Most real-world networks are associated with several attributes, and many analysis tasks depend on analyzing both, relationships and attributes. Visualization of multivariate networks, however, is challenging, especially when both the topology of the network and the attributes need to be considered concurrently. In this state-of-the-art report, we analyze current practices and classify techniques along four axes: layouts, view operations, layout operations, and data operations. We also provide an analysis of tasks specific to multivariate networks and give recommendations for which technique to use in which scenario. Finally, we briefly survey application areas and evaluation methodologies.

1. Introduction

Networks are ubiquitous in today's society. A network, also called a graph, consists of nodes (or vertices) and links (or edges) connecting these nodes. Examples range from social networks, to physical networks such as power grids, to networks modeling cellular processes in biology, to trees describing evolutionary relationships between species. Nodes and links can represent people, locations, genes, or ultimately any entity that has a connection with other entities. Most real-world networks consist of nodes and links that are also associated with attributes. Attributes can range from minimal information, such as node labels, to complex attribute vectors, such as the age, skills, and gender of a person, or the activity levels of genes from hundreds of samples. Networks that also contain attributes are commonly referred to as multivariate networks (MVN) [KPW14]. When the structure (the topology) of the network needs to be analyzed together with these node or edge at-

© 2019 The Author(s)
Computer Graphics Forum © 2019 The Eurographics Association and John Wiley & Sons Ltd. Published by John Wiley & Sons Ltd.

tributes, we employ *multivariate network visualization techniques*, which are the topic of this state-of-the-art report.

Visualizing MVNs is particularly challenging, as both the *topology* of the network (also called *structure* of the network), and *attributes* associated with the nodes and links need to be shown. When visualizing topology and attributes in the same view, choosing efficient encodings for one aspect often interferes with the ability to effectively visualize the other. For example, when attempting to encode multiple node attributes directly on the node in a node-link diagram, a visualization designer has to make compromises with regard to the node size, the network size, and the number of attributes shown. Conversely, when visualizing attributes and topology in separate views, an analyst's ability to judge interactions between nodes/links and attributes is limited. Most of the published techniques on multivariate networks attempt to mitigate the trade-off between visualizing topology and visualizing attributes well by introducing novel encodings for certain data types or analysis tasks.

In this paper, we introduce new typologies for multivariate network tasks and multivariate network techniques. Our *task typology* (Section 4) is based on two fundamental MVN tasks: identifying a topological structure based on attributes or characterizing the attributes of a topological structure. We apply these tasks to a series of relevant topological structures, which results in ten MVN tasks.

The *visualization technique typology* (Section 5), shown in Figure 1, consists of four dimensions: layouts, view operations, layout operations, and data operations. The layouts category includes the eight unique multivariate network layouts. View operations describe how multiple views can be used for multivariate network visualization. Layout operations either combine multiple layouts or multiply layouts based on attributes. Data operations describe possible network wrangling operations such as aggregation and computing new attributes. We also provide explicit guidelines on recommended usage of each technique, as well as on how to select an appropriate technique given a network type, attribute characteristics, and a set of tasks the user aims to support. A companion website, available at `https://vdl.sci.utah.edu/mvnv/` provides an overview of the taxonomy and a wizard to select a technique for a specific data and task combination.

Finally, we briefly discuss our methodology (Section 6), common application areas of MVNs (Section 7), and strategies for evaluating multivariate networks (Section 8).

2. Related STAR Reports

The body of literature in the space of visualizing multivariate networks is consistently growing, but a comprehensive survey of the techniques used has not yet been done. STARs exist on related topics such as visualizing large graphs [vLKS 11], trees [Sch11], implicit hierarchies [SHS11], group structures in graphs [VBW15], dynamic graphs [BBDW14], edge bundling techniques [LHT17], among others. However, no survey paper introducing a typology of multivariate network visualization techniques exists.

The survey on multifaceted graph visualization by Hadlak et al. [HSS15] covers the visualization of four facets of graph visualization, two of which are relevant to our work—partitions and attributes. Their work distinguishes three overarching categories of MVN visualizations—visualizations with a topology-driven layout, an attribute-driven layout, and coordinated views—but is not exhaustive in describing MVN techniques, including tabular techniques and hybrid approaches. Most related to the proposed STAR is a survey book edited by Kerren et al. based on a Dagstuhl Seminar [KPW14] that reports on selected topics in the field, but does not introduce a taxonomy of existing techniques.

This review paper aims to fill that gap by describing the design space of techniques used to encode multivariate networks, classifying these techniques into a formal typology, and surveying existing evaluation methods. We only briefly discuss application areas for MVNs, as these are covered in depth in the survey book by Kerren et al. [KPW14].

3. Network and Attribute Characteristics

Here we introduce definitions and describe different types of networks. We consider multivariate networks with nodes and edges where either the nodes, the edges, or both, have attributes associated with them. We refer to the nodes and edges as the network's *topology*. Attributes can be of varying data types, such as textual, quantitative, ordinal, nominal, or set data. Attributes can also be computed based on topological properties. For example, the degree of a node, or its centrality, are common derived attributes. Analogously, topological structures can be derived from attributes. For example, two nodes can be connected if they share an attribute.

Both nodes and edges can be of different types. We consider nodes and edges of a type to be of a semantic unit that has a different set of attributes associated with it. For example, a node type *actor* could have a name, current residence, age, etc. associated with it. Another node type in the same network could be *movies*, with attributes such as movie title, budget, revenue, genres, etc.

In this survey, we discuss the properties of a visualization technique on a number of dimensions, including network size, network type, node and edge attribute size and type, and their suitability for tasks. Here we discuss the data properties, while the tasks are described in Section 4.

For *size*, we distinguish between three categories of networks: small (less than 100 nodes), medium (100-1,000 nodes), and large (more than 1,000 nodes). For our evaluation, we consider the number of nodes simultaneously displayed, not the number of nodes a technique can handle through querying or filtering.

We distinguish between the following network types:

Complex networks, i.e., networks with nontrivial topological features, such as a diverse degree distribution (the distribution of the number of links per node), clusters, hubs, etc. [WC03]. Among the complex networks are scale-free networks, with a degree distribution that follows the power law distribution. These networks are characterized by a small number of nodes with a very high degree that exhibit preferential attachment, i.e., that nodes added to the network are likely to attach to these central hubs. An example is the world wide web, where major websites are linked to much more frequently than others. Another type of

© 2019 The Author(s)
Computer Graphics Forum © 2019 The Eurographics Association and John Wiley & Sons Ltd.

complex network is a **small world network**, with a degree distribution that peaks around the average value. These networks exhibit significant clustering and are characterized by a short average path distance (hence small world). Social networks are a typical example.

Layered and K-Partite networks are networks where the nodes can be partitioned into discrete sets and where links are mostly (in the case of layered networks) or exclusively (in the case of k-partite networks) between the different sets. Layered networks are ordered (e.g., as in genealogies, where there is an order of generations that is represented as layers) and have links only between successive layers, whereas k-partite networks can have links between multiple partitions and are not necessarily ordered.

Trees are acyclic networks, commonly with a defined root, where each node is either a leaf or the root of a subtree. Trees have exactly one path between two nodes and have a low average degree.

For complex networks, we also distinguish between *sparse* networks, with a low average degree, and *dense* networks, with a high average node degree. Layered networks, k-partite networks, and trees are sparse by their nature.

For *node attributes*, we distinguish between *few (less than five)* and *many (more than five)*, but also give more detailed ranges when discussing a specific technique. We considered a category for a high number of attributes (e.g., hundreds), but realized that techniques that support five or more attributes well would also support many attributes. We also discuss data type properties when appropriate. We further distinguish between *homogeneous networks with respect to node types*, i.e., all nodes are of the same type (e.g., only actors), and *heterogeneous networks*, i.e., the network has more than one node type (e.g., actors and movies).

We categorize analogously for *edge attributes*, albeit we consider few to be less than three and many to be three or more attributes. We also distinguish between homogeneous and heterogeneous edges.

4. Multivariate Network Analysis Tasks

Here we introduce a taxonomy of tasks focusing specifically on multivariate network analysis. Our taxonomy is designed to evaluate MVN techniques for their suitability to address a task and is complementary to general network task taxonomies. This task analysis, combined with considerations of network type and network scale described in the previous section, form the basis of our assessment of individual visualization techniques and of our guidelines for MVN visualization (Section 5.1).

There are various data-type independent visualization task taxonomies. The works by Brehmer and Munzner [BM13] and Schulz et al. [SNHS13] are examples of recent, generic visualization task taxonomies. An example of a data-type-specific task taxonomy is the work by Valiati et al. [VPF06] which covers multidimensional visualization techniques.

Specialized task taxonomies for networks also exist. The well-known taxonomy of network tasks by Lee et al. [LPP 06] does not cover MVN tasks in detail. In addition to topology-focused tasks, the taxonomy includes selected attribute-based tasks. For node attributes, these tasks are *finding nodes having a specific attribute value*, and *reviewing the set of nodes*. For edges, the two tasks are about identifying neighbors of a node connected by an edge with a specific property (edge type, specific edge value). Shneiderman and Aris [SA06] also discuss examples of MVN tasks but do not take a systematic approach to the topic.

The most comprehensive analysis of tasks for multivariate networks was done by Pretorius et al. [PPS14]. The authors generalize the tasks proposed by Lee et al. to cater to cases where the network's attributes are of interest to the user. This approach is useful in that it provides a systematic and comprehensive approach to MVN tasks. They list 25 tasks in the categories of structure-based (topology-based), attribute-based, browsing, and estimation tasks. They also discuss how these tasks are composed of lower level analytical tasks proposed by Valiati et al. (identify, determine, relocate, compare) [VPF06].

We believe that the task taxonomy by Pretorious et al. is comprehensive and refer to it for a detailed review. Pretorius et al. also discuss that in the abstract, a task is an analytical activity performed on an entity and a property of an entity, which is called "actions" applied to "targets" by Brehmer and Munzner [BM13]. For the purpose of this STAR, we believe that a slightly simplified approach is useful for both characterizing existing techniques and recommending techniques based on tasks specified by analysts.

In our taxonomy, any task can be expressed as a combination of two fundamental tasks, as applied to different topological structures of a network. These two basic tasks are:

analyzing topology for given attributes (TgA), which are aimed at identifying, characterizing, or comparing topological structures that have certain attributes, and

analyzing attributes for a given topological structure (AgT), which are aimed at identifying, characterizing, or comparing the attributes of a given topological structure.

We do not further distinguish between more detailed tasks, such as identifying, summarizing, or searching, because we believe that in most practical scenarios these go hand in hand, and because most visualization techniques that support one of these also support the others. A notable exception is comparison, which can often be achieved without special considerations, but can benefit from dedicated support by a visualization technique in certain situations.

The targets in the context of network tasks are the *topological structures*. We consider major types of topological structures,

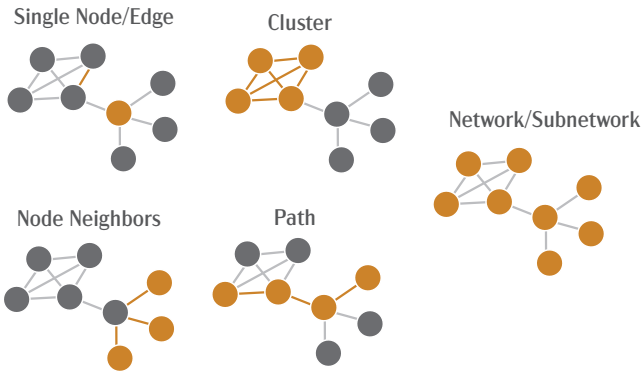

Figure 2: *Topological structures we consider as targets for tasks.*

© 2019 The Author(s)
Computer Graphics Forum © 2019 The Eurographics Association and John Wiley & Sons Ltd.

shown in Figure 2, but recognize that others, such as cliques or spanning trees, could be relevant for speci c analysis questions. The structures include individual nodes and edges, paths of sequential nodes and edges, neighborhoods of a particular node or edge, clusters of nodes and edges, and larger networks or subnetworks that can be considered as complex networks by themselves. We use the term "cluster" to refer to a set of nodes that are well-connected to each other, but less so with other nodes with regards to the topology of the network. A cluster is also often called a community, especially in the context of social networks. Conversely, a "subnetwork" is a subset of a network that is not limited to a speci c structure.

We selected these speci c structures based on our literature analysis of MVN visualization techniques, where we identi ed many published techniques that are optimized for performing MVN tasks on speci c topological structures. For example, approaches such as Path nder [PGS16] and Entourage [LPK13] are explicitly developed for path-based MVN tasks. GraphCharter [TS13], on the other hand, is optimized for analysis on node neighbors.

A comprehensive analysis of multivariate network tasks can thus be described as a combination of general network tasks (e.g., as described by Lee et al. [LPP06]) and our MVN tasks. By reducing MVN tasks to a combination of two basic tasks and a topological structure, we are able to decompose complex tasks into a simple combination of tasks and targets. The task "Estimate the average value of a speci c attribute of all nodes connected to a speci c node" (a concrete example: what's the average income of University of Utah graduates in a social network) can be deconstructed into three basic tasks: (1) nd the node with label "University of Utah" (analyzing topology for given attributes, on target "network"), followed by (2) Lee et al.'s adjacency task (nd the set of nodes adjacent to a node), followed by (3) estimate the average value of an attribute of a node neighborhood (analyzing attributes for a given topolgy on target "neighbors").

Table 1 gives examples of the two basic MVN analysis tasks as applied to the set of topological structures listed in Figure 2. For each task, we also provide an example for a domain-speci c task from the literature.

Although we observed both TgA and AgT tasks in our corpus, we saw a preference for one of the task types in certain application areas. Techniques developed for social network analysis, for example, tend to rst focus on topology-only tasks, followed by attribute tasks on the selected topological structures (AgT). For example, a commonly described task is to identify particular structures, such as communities, followed by an analysis of the attributes in these structures [HF07].

We also observed that identifying a topological structure for given attributes (TgA) is frequently supported by query or lter operations, instead of being a purely visual task. For example, an analyst may want to identify a fraudulent bank account by querying for patterns of transactions, and then expand a network of interactions between that account and others [vHP09]. Similarly, Juniper [NSL19], Path nder [PGS16], and Enroute [PLS12] use queries or lters to isolate a subnetwork or path of interest, followed by inspection of the attributes in that structure.

5. Multivariate Network Visualization Typology

The primary contribution of this paper is a typology of multivariate network visualization techniques, illustrated in Figure 1. The typology is made up of four components: a taxonomy of MVN layouts, and three types of operations that can be applied to these layouts. Every technique and tool discussed in this paper can be described by selecting (at least) one layout, and optionally applying operations to it.

The layout taxonomy encompasses core approaches to encoding the topology and attributes of multivariate networks and trees. The taxonomy of layouts contains three classes: node-link, tabular, and implicit layouts, with various layouts within these classes.

The view operations capture different ways of combining multiple views with each other. For example, view operations include juxtaposition for placing an attribute and a topology visualization side-by-side.

The layout operations allow for different combinations of the layouts to produce hybrids and small multiples.

The data operations category contains methods that can be applied to the data either as a preprocessing step or during the analy-

Toplogical Structure	Task Type	Example	Domain-Speci c Example
Node/Edge	TgA	Which node/edge has the given attribute value?	Who is the oldest person in a social network? [HB05]
	AgT	What is the attribute value of a given node/edge?	What house is Jon Snow in? [NSL19]
Neighbors	TgA	Which neighbors have a given attribute value?	Who is Mary's tallest friend? [LPP06]
	AgT	What are the attribute values of a node's neighbors?	What are the zodiacs of Joe's friends? [TS13]
Paths	TgA	Which path between two nodes has the lowest cost?	Which path between two genes is most frequently cited in pathway databases? [PGS16]
	AgT	What are the attribute values along a path?	How does a metabolite increase along a pathway? [MWS20]
Clusters	TgA	Which cluster has certain attribute characteristics?	Which cluster has the most educated people? [HF06]
	AgT	What are the attribute characteristics within a cluster?	What is the average education level in Mary's cluster? [HF06]
Network	TgA	Find the subnetwork of nodes with a speci c attribute.	Which subnetwork contains only people with a PhD? [HF06]
	AgT	What are the attribute values in this network?	What is the average age in this network? [NSD09]

Table 1: List of generic and domain-speci c examples of MVN analysis tasks composed of our two basic tasks and the described topological structures. Each domain-speci c example references a paper that describes that task.

c 2019 The Author(s)
Computer Graphics Forum c 2019 The Eurographics Association and John Wiley & Sons Ltd.

		Size			Type				Node Attributes				Edge Attributes				Topological Structures				
		Small <100 nodes)	Medium (<1,000)	Large (>1,000 nodes)	Complex (sparse)	Complex (dense)	Layered/K-Partite	Trees	Few (<5)	Several (5)	Homog. (1 type)	Hetero. (>1 type)	Few (<3)	Several (3)	Homog. (1 type)	Hetero. (>1 type)	Single node/edge	Neighbors	Paths	Clusters	Entire/sub network
Node-Link Layouts	On-node/edge encoding																				
	Attr.-driven faceting																				
	Attr.-driven positioning																				
Tabular Layouts	Adjacency matrix																				
	Quilts																				
	BioFabric																				
Implicit	Inner nodes & leaves																				
	Leaves																				
View Operations	Juxtaposed																				
	Integrated																				
	Overloaded																				

Table 2: Scores for how well each technique performs on different network types and different multivariate network tasks. 0 means does not support at all, 1 is supports poorly, 2 encodes supported but with limitations / may require interaction, and 3 means well supported.

sis. Example data operations are aggregating nodes into supernodes or filtering parts of a network.

In describing MVN visualization techniques along these four orthogonal dimensions, we not only cover the techniques observed in the set of papers surveyed, but also provide a design space in which to decompose and analyze any MVN visualization technique.

When describing and discussing these approaches, e.g., with respect to scalability, we assume a static visualization technique. However, we call out opportunities for interaction where appropriate. Certain types of interactions, such as zooming and panning, or visualizing attributes such as labels in tool tips, can be applied to most techniques. Other, more sophisticated interactions, such as linking and brushing, sorting, changing the layout algorithms, or manually adjusting a layout, depend on the visualization technique used. The data operations we describe can frequently be invoked interactively, and especially operations such as filtering or aggregating can drastically change how much and what types of raw data can be visualized with a basic layout technique.

We scored each of the layouts and view operations along 20 dimensions, which include network characteristics such as size and number of attributes, as well as different types of MVN analysis tasks. Table 2 shows the scores for each type. Our methodology for developing these scores is described in Section 6.4. We use the following scores:

- 0 This technique cannot support this data type or task.
- 1 This technique supports this data type or task very poorly.
- 2 This technique can support this data type or task, but is not ideal and may require interaction to achieve it.
- 3 This technique supports this data type or task very well.

© 2019 The Author(s)
Computer Graphics Forum © 2019 The Eurographics Association and John Wiley & Sons Ltd.