1

Introduction to Multivariate Network Visualization

Andreas Kerren, Helen C. Purchase, and Matthew O. Ward

Information Visualization (InfoVis) research focuses on the use of techniques to help people understand and analyze data. In particular, it considers how abstract data (i.e., without correspondence to the physical world) can best be visually represented. A variety of different abstract data types are addressed in InfoVis research (e.g., numerical, ordinal, categorical [29]), all of which can be arranged in different ways: for example, in linear, tabular, or network form. Common representations of statistical data (e.g., pie charts, bar charts, or scatter plots) are all visualizations of abstract numerical data.

A "multivariate network" (MVN) is an abstract data type that provides particular challenges for the information visualization research community. It permits the representation of complex relational data (stored in the form of a network) as well as the association of attributes with that data. The attributes themselves may use a range of different abstract data types.

A MVN therefore consists of a set of objects, each of which has information associated with it. In addition, the objects are connected to each other in a network that represents the relationship between the objects. Further complexity is added when information is also associated with the inter-object relationships themselves. For example, a social network representation may consist of people (the objects), each of which has information associated with them (their age and post code). Friendship relationships between the people form the network, with additional information about each friendship between two people being stored (e.g., the last time they communicated with each other or how long they have known each other).

In InfoVis terms, the objects in the network are called *nodes* or *vertices*, the relationships between the objects are called *links* or *edges*, the network is mostly called a *graph*, and the information associated with the objects and relationships are called *attributes*, *features*, *dimensions*, or *properties*.

MVNs prove particularly challenging for InfoVis researchers because of the wealth, richness and variety of the information that can be stored in them. Any number of attributes can be associated with nodes and edges, and nodes can be associated with any number of other nodes. Depicting all

In A. Kerren et al. (Eds.), Multivariate Network Visualization, LNCS 8380, Chapter 1, pages 1-9, Springer, 2014.

this information in visual form so as to help people understand it is a clear challenge. In many cases, given the limits of human perceptual and cognitive abilities, it is impossible to clearly show all the information at once in a useful form. Such problems may be addressed by prior knowledge of the user tasks to be performed, by providing facilities for interacting with the data, or by a clever choice of representation. These problems are of course made even more complex if the information changes over time, if there is more than one network, or if the network is particularly large.

Despite the challenges of depicting MVNs, their existence is common, and they are (and have been) used to represent abstract domain knowledge for many years (for example in software engineering, biology, evolution, environmental sciences, meteorology, and sociology). The popular educational tool of "concept mapping" demonstrates the wide range of applications to which MVNs can be applied [23]. Studying ways in which computational techniques may be used to help with the effective visualization of MVNs is therefore an area of study with wide applicability as well as the potential to be highly useful in other research areas.

In this chapter, we define MVNs formally, introduce and classify existing related InfoVis research, and provide a brief summary of the rest of the book.

1.1 Multivariate Networks: Definitions and Terminology

A *(simple) graph* $G = (V, E)$ consists of a finite set of vertices (or nodes) V and a set of edges $E \subseteq \{(u, v)|u, v \in V, u \neq v\}$. Based on this, a variety of general graph properties and characteristics can be found in the literature; the most important ones are introduced in the following list (cf. also the book chapter [18] or the books [7, 14]).

- An edge $e = (u, v)$ with $u = v$ is called a *self-loop*.
- If an edge e exists several times in E then it is called a *multiple edge*.
- A *simple graph* has no self-loops and no multiple edges.
- The *neighbors* of a node v are its adjacent nodes.
- The *degree* of a node v is the number of its neighbors.
- A *directed graph* (or digraph) is a graph with directed edges, i.e., (u, v) are ordered pairs of nodes.
- A directed graph is called *acyclic* if it has no directed cycles, i.e., there is no directed path where the same node is visited twice.
- A graph is *connected* if there is a path between u and v for each pair (u, v) of nodes.
- A graph is *planar* if it can be drawn in the 2D plane without intersections of edges (*edge crossings*).

In contrast to the above definition of a simple graph, a *multivariate network* N consists of an underlying graph G plus n additional attributes $A = \{A_1, \ldots, A_n\}$ that are attached to the nodes (and/or edges). For node attributes, A_i represents a column in a table of attributes $A = (a_{ji})\,(j = 1 \ldots |V|\,; i = 1 \ldots n)$ and contains one attribute value per node (similar definition for edges). Thus, $a^u = (a_{u1}, \ldots, a_{un})$ describes all attribute values for node u given that there is no missing data.

Also in network theory, researchers have developed a set of useful measurements and metrics that can be used to get an impression about the most important characteristics of the graph topology such as central actors in a social network [22]. Those measures can also be applied to multivariate networks. *Community analysis* based on specific clustering techniques is one such approach. Another example are so-called *network centralities*, i.e., measures that quantify how important a node or edge in the network is. More formal, a network centrality C is a function that assigns a value $C(u)$ to a node $u \in V$ of a given graph $G = (V, E)$. This function supports centrality comparisons according to their importance, i.e., u is more important than v iff $C(u) > C(v)$ [8, 20]. A simple example of a network centrality is the degree of a node in an undirected graph.

1.2 Existing Visualizations

In the following, we briefly highlight the most important visualization techniques for arbitrary multivariate graphs/networks, i.e., we do not consider special cases such as the visualization of hierarchies (trees) or directed acyclic graphs (DAGs). The interested reader is referred to the vast amount of literature on these topics as for instance [11, 30].

Before we continue our discussion on multivariate network visualizations, we turn the reader's attention to standard techniques for the visualization of multivariate data itself. Multivariate (or multidimensional) data sets can mostly be described as data tables with n data objects and m attributes/features, i.e., for each object exists an attribute vector with m dimensions. The attribute values can be classified—for instance—into nominal, ordinal, or quantitative. In practice, we often have a large amount of data objects and many attributes with different types. Finding a suitable visual representation is thus challenging, and the right choice might depend on further parameters like application domain, integration into a larger visualization environment, or support of specific interaction techniques. In general, visual mappings for multivariate data can roughly be categorized into *point-based approaches* (e.g., scatterplot matrices [6], projection methods like MDS [21, 34], etc.), *axis-based approaches* (parallel coordinate plots [10], Kiviat diagrams [3], etc.), *icon-based approaches* (Chernoff faces [5], stick figures [24], etc.), and *pixel-based approaches* (e.g., recursive patters [15], pixel bar charts [16], etc.). There are many good textbooks that provide a good overview of those methods; we recommend the books of Spence [29], Kerren et al. [19], and Ward et al. [31].

A View on Graph Drawing

Traditional graph drawing (GD) methods compute a 2D/3D layout of the nodes and the edges, mainly based on *node-link diagrams* [32]. They play a fundamental role in network visualization. Particular graph layout algorithms can give an insight into the topological structure of a network if properly chosen and implemented. The graph readability is affected by quantitative measurements called *aesthetic criteria* [7]. Thus, graph drawing generally deals with the ways of drawing graphs according to the set of predefined aesthetic criteria [4]. These criteria are often contradictory, and problems which aim to optimize the criteria are often NP-hard. Therefore, many GD algorithms are heuristics. For the sake of completeness, we want to note that there are also so-called space-filling methods that try to solve some conceptual problems of node-link diagrams, such as the high space consumption or edge crossings. *Matrix displays* fall into this category (like the approach proposed in [1]). These visualizations represent a graph directly via its adjacency matrix, where a matrix element (i, j) represents the existence of an edge between the two nodes i and j. A disadvantage of matrices is that the perception of the graph topology is depending of the node order in the matrix.

In both visual representations (i.e., node-link and matrix displays), multivariate data can be integrated in various ways. For instance in node-link diagrams, multivariate glyphs that replace the node representations (usually a dot or circle) can be used to show data attached to nodes; edge attributes can be represented by different link colors, thicknesses, labels, or edge shapes. In matrix representations, the cells can be color-coded or be replaced by small icons to show edge attributes; node attributes might be shown as colored node labels, for instance. Usually, all mentioned efforts to integrate multivariate attributes into network representations do not scale well and get easily cluttered. The next subsection provides a more detailed classification of techniques which go beyond the traditional graph drawing approaches.

Classification of Approaches

Good drawing algorithms as previously described cannot solely solve the problem of MVNs. There are several reasons for this statement. First, the most traditional graph drawings do not scale well, i.e., they are not able to represent huge data sets with many thousands of nodes and/or edges. Second, additional multivariate data cannot be intuitively embedded into a standard drawing. The InfoVis community has tried to address those issues by visualization approaches that provide filtering and interaction possibilities in order to reduce the number of graph elements under consideration as well as by methods to visually analyze attributes in context of the underlying graph topology. According to Jusufi [12], several approaches can be found in the literature that offer solutions for the problem of visualizing multivariate networks: *multiple and coordinated views*, *integrated approaches*, *semantic substrates*, *attribute-driven layouts*, and *hybrid approaches*.

Multiple and coordinated views: Solutions in this category combine several views and present them together. This strategy allows the user to choose the most powerful visualization techniques for each specific view and data set [9, 25]. As an application example, we highlight the work of Shannon et al. [27]. Their approach consists of two distinct views: one view shows a parallel coordinate approach for the visual representation of the network attributes, and the other view displays a traditional node-link drawing of a graph. The tool is equipped with a variety of visualization and interaction techniques; both views are coordinated by linking and brushing [29] techniques. The drawback of multiple views is that they split the displayed data because of the spatial separation of the visual elements.

Integrated approaches: To provide a combined picture, attributes and the underlying graph can be displayed in one single view. "Integrated views can save space on a display and may decrease the time a user needs to find out relations; all data is displayed in one place." [9]. In Borisjuk et al. [2], small diagrams (e.g., bar charts) are employed instead of representing the nodes as simple circles, dots, or rectangles. Each diagram shows experimental data that is related to the regarded node. This approach provides a view of all available information, but the embedding of the visualizations into the nodes consumes a lot of space. This issue may affect the readability of the network due to the visual clutter that may appear when the number of nodes and the attributes is high [17]. However, the problem of space usage and additional clutter can be alleviated by interaction techniques.

Semantic substrates: In order to further avoid clutter in multivariate network visualizations, some researchers realized the idea of so-called semantic substrates that "are non-overlapping regions in which node placement is based on node attributes": Shneiderman and Aris [28] introduced this idea and combined it with sliders to control the edge visibility and thus to ensure comprehensibility of the edges' end nodes. Their tool efficiently improves the situation of visual clutter that happens with large MVNs. However, one conceptual drawback of such approaches is that the underlying graph topology is not (completely) visible.

Attribute-driven layouts: Those layouts use the display of the network elements to present insight about the attached multivariate data instead of visualizing the graph topology itself. In contrast to semantic substrates, this technique does not necessarily place the nodes into specific regions. Instead, it controls the placement of a node in the graph layout by considering the node's attributes. An example is *PivotGraph* [33] which shows the relationships between (node) attributes and links within a 2D grid-layout. This concrete approach scales well for some situations because of the inherent node aggregation (nodes on the same grid position share the same attribute values) but is restricted to discrete attribute values and only two attribute dimensions.

Hybrid approaches: They combine at least two of the previously discussed techniques. The most common combinations are multiple coordinated views with any of the integrated approaches. For instance, Rohrschneider et al. [26] integrate additional attributes of a biological network inside the nodes and edges. The authors also use other visual metaphors for creating multiple coordinated views to show time-related data of the network. Another hybrid approach is the JauntyNets tool [13] which combines multiple coordinated views with an attribute-driven layout.

1.3 Outline of this Book

The book is divided into two parts. The first three chapters (Chaps. 2-4) present three application domains in which multivariate networks are commonly used: software engineering, social networks and the life sciences. Written by experts in the three respective fields, these chapters describe how multivariate networks play a crucial role in the study of the comprehension of programs for the purposes of maintenance and evolution (Chap. 2), the analysis of personal and social networks defined by a wide variety of relationships (Chap. 3), and the exploration and analysis of biological data at several levels of detail (Chap. 4). Not only do these chapters describe the use of multivariate networks in these domains, they also consider how these networks can be effectively and appropriately visualized so as to support domain-specific tasks, and discuss the challenges facing these three rapidly evolving fields.

The second part of the book covers a range of topics associated with the visualization and use of multivariate networks, focussing first on fundamental visualization aspects (tasks, interaction, and representation), and then addressing broader issues (time, multiple networks, and large networks). Chapter 5 presents a new framework of tasks specifically associated with multivariate networks, based on existing taxonomies of general visualization tasks and simple graph-reading tasks. These multivariate network tasks are shown to be composed of lower-level visualization tasks, and are then illustrated with domain-specific examples. Chapter 6 highlights the fact that effective completion of user tasks when using a visual representation requires interaction, allowing the information landscape to be navigated, and more of the information to be perceived. It describes the range of different methods of interacting with multivariate networks, as well as guidelines for novel interaction techniques. Chapter 7 focuses on the means by which multivariate networks can be visually represented—beyond the traditional node-link method—by proposing and discussing a range of alternative (and novel) visual metaphors inspired by nature, geography or manufactured objects. It concludes with a gallery of potential new metaphors.

The important issue of time is covered in Chap. 8. This chapter provides essential definitions for temporal multivariate networks, and shows how two applications (biology and social networks) relate to a structure-behaviour-evolution model originally proposed for characterizing temporal networks in software engineering. A survey of existing visualization methods for temporal networks is presented. The heterogeneous networks chapter (Chap. 9) is primarily concerned with multivariate networks that are associated with each other at different levels and at different scales, and demonstrates the concepts with examples from the three application domains of biology, social sciences, and software engineering. The challenges of visualizing such linked networks are discussed. The final chapter (Chap. 10) considers the ever-present visualization challenge of scalability—what to do when the networks are so large that they cannot be displayed effectively. Based on considerations of cognitive and architectural limitations, suitable visualization approaches for large networked data sets are explored, and their effectiveness discussed.

References

1. Abello, J., van Ham, F.: Matrix zoom: A visual interface to semi-external graphs. In: Proceedings of the IEEE Symposium on Information Visualization. pp. 183–190. IEEE Computer Society, Los Alamitos, CA, USA (2004)
2. Borisjuk, L., Hajirezaei, M.R., Klukas, C., Rolletschek, H., Schreiber, F.: Integrating data from biological experiments into metabolic networks with the dbe information system. In Silico Biol 5(2), 93–102 (2005)
3. Chambers, J.M., Cleveland, W.S., Kleiner, B., Tukey, P.A.: Graphical Methods for Data Analysis. Belmont, CA: Wadsworth (1983)
4. Chaomei, C.: Information Visualization. Beyond the Horizon. Springer-Verlag, London Berlin Heidelberg, 2nd edn. (2004)
5. Chernoff, H.: The use of faces to represent points in k-dimensional space graphically. Journal of the American Statistical Association 68, 361–368 (1973)
6. Cleveland, W.C., McGill, M.E.: Dynamic Graphics for Statistics. CRC Press, Inc., Boca Raton, FL, USA (1988)
7. Di Battista, G., Eades, P., Tamassia, R., Tollis, I.G.: Graph Drawing: Algorithms for the Visualization of Graphs. Prentice Hall (1999)
8. Dwyer, T., Hong, S.H., Koschützki, D., Schreiber, F., Xu, K.: Visual analysis of network centralities. In: Misue, K., Sugiyama, K., Tanaka, J. (eds.) Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation (APVis'06). pp. 189–198. Australian Computer Society, ACM International Conference Proceeding Series, vol. 164, Darlinghurst, Australia (2006)
9. Görg, C., Pohl, M., Qeli, E., Xu, K.: Visual Representations. In: Kerren et al. [17], pp. 163–230
10. Inselberg, A., Dimsdale, B.: Parallel coordinates: A tool for visualizing multidimensional geometry. In: IEEE Visualization. pp. 361–378 (1990)
11. Johnson, B., Shneiderman, B.: Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In: Proceedings of the 2nd conference on Visualization '91. pp. 284–291. VIS '91, IEEE Computer Society Press, Los Alamitos, CA, USA (1991)
12. Jusufi, I.: Multivariate Networks: Visualization and Interaction Techniques. Ph.D. Thesis, Linnaeus University, Växjö, Sweden (2013)
13. Jusufi, I., Kerren, A., Zimmer, B.: Multivariate Network Exploration with JauntyNets. In: Proceedings of the 17th International Conference on Information Visualisation (IV '13). pp. 19–27. IEEE Computer Society Press, Los Alamitos, CA, USA (2013)

14. Kaufmann, M., Wagner, D.: Drawing Graphs: Methods and Models, Lecture Notes in Computer Science Tutorial, vol. 2025. Springer (1999)
15. Keim, D.A.: Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics 7(1), 1–8 (2002)
16. Keim, D.A., Hao, M.C., Dayal, U., Hsu, M.: Pixel bar charts: A visualization technique for very large multi-attribute data sets? Information Visualization 1(1), 20–34 (2002)
17. Kerren, A., Ebert, A., Meyer, J. (eds.): Human-Centered Visualization Environments. LNCS Tutorial 4417, Springer, Heidelberg (2007)
18. Kerren, A., Schreiber, F.: Network visualization for integrative bioinformatics. In: Chen, M., Hofestädt, R. (eds.) Approaches in Integrative Bioinformatics, pp. 173–202. Springer Berlin Heidelberg (2014)
19. Kerren, A., Stasko, J.T., Fekete, J.D., North, C. (eds.): Information Visualization: Human-Centered Issues and Perspectives, Lecture Notes in Computer Science, vol. 4950. Springer, Berlin, Heidelberg (2008)
20. Koschützki, D., Schreiber, F.: Comparison of centralities for biological networks. In: R. Giegerich, J.S. (ed.) Proc. German Conf. Bioinformatics (GCB?04). pp. 199–206 (2004)
21. Mardia, K.V.: Multivariate Analysis. Academic Press (1979)
22. Newman, M.E.J.: Networks: An Introduction. Oxford University Press (2010)
23. Novak, J.D.: Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations. Routledge, 2nd edn. (2010)
24. Pickett, R.M., Grinstein, G.G.: Iconographic displays for visualizing multidimensional data. In: Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics. vol. 1, pp. 514–519 (1988)
25. Roberts, J.C.: Exploratory visualization with multiple linked views. In: MacEachren, A., Kraak, M.J., Dykes, J. (eds.) Exploring Geovisualization, chap. 8, pp. 159–180. Elseviers (2004)
26. Rohrschneider, M., Ullrich, A., Kerren, A., Stadler, P.F., Scheuermann, G.: Visual network analysis of dynamic metabolic pathways. In: Proceedings of the 6th International Conference on Advances in Visual Computing – Volume Part I (ISVC '10). pp. 316–327. Springer, Berlin, Heidelberg (2010)
27. Shannon, R., Holland, T., Quigley, A.: Multivariate graph drawing using parallel coordinate visualisations. Tech. Rep. 2008-6, University College Dublin, School of Computer Science and Informatics (2008), `http://www.csi.ucd.ie/files/ucd-csi-2008-6.pdf`
28. Shneiderman, B., Aris, A.: Network visualization by semantic substrates. IEEE Transactions on Visualization and Computer Graphics 12, 733–740 (2006)
29. Spence, R.: Information Visualization: Design for Interaction. Prentice Hall, 2nd edn. (2007)
30. Stasko, J.T., Catrambone, R., Guzdial, M., Mcdonald, K.: An evaluation of space-filling information visualizations for depicting hierarchical structures. International Journal of Human-Computer Studies 53, 663–694 (2000)
31. Ward, M., Grinstein, G., Keim, D.A.: Interactive Data Visualization: Foundations, Techniques, and Application. A.K. Peters, Ltd. (2010)
32. Ware, C.: Information Visualization: Perception for Design. Morgan Kaufmann, 2nd edn. (2004)
33. Wattenberg, M.: Visual exploration of multivariate graphs. In: CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems. pp. 811–819. ACM, New York, NY, USA (2006)

34. Williams, M., Munzner, T.: Steerable, progressive multidimensional scaling. In: Proceedings of the IEEE Symposium on Information Visualization (InfoVis '04). pp. 57–64. IEEE Computer Society Press (2004)