Introduction and Overview

This study is about the visualization of social structures. It investigates how the visual representation of graphs permits us to gain insights into complex networks of social relations.

In order to do this, we examine three questions: (1) How can we order the elements of a network to display the essential properties of its graph? Which kinds of information do we need to do that? (2) Which structural properties of graphs and their elements can be integrated into visual representations in such a way that their characteristics are easily and clearly identifiable? (3) How can we translate additional, external, information into the representation, so that its embedding in the structure and the local concentration of its characteristics are visualized?

To address those three questions, we have to translate numeric information into visual markers. Sorting the relations in a graph, we search for orderings of the units that allow us to optimally display the relations between actors. The solutions to this problem determine the location of the elements in a drawing.

To translate the properties of a graph into the resulting layout, we can vary the size, shape, or color of symbols. We explore external characteristics and their structural location by marking nodes and/or edges of the graph, usually with color. This enables us to investigate the distribution of external characteristics in more detail.

The overarching question, therefore, is how we can use visual means to develop insights into complex relationships. In order to address this question, we first present a systematic review of the current knowledge on the visual representation of information, which we base on a wide range of different literatures. Based on this review, we then introduce the procedures and means of representation that we have developed through an extensive investigation of graphs. As a result, we can identify the principles and elements of network visualization, and determine the core and basis of a visual methodology that can be used to generate and explore multidimensional representations of networks.

What we show Graph theory and social network theory have been using numeric algebraic methods to identify the characteristics of graphs and of the networks they represent. By investigating the number of direct relations between individual actors, it is possible to describe the integration of individual actors into a surrounding network and to characterize their position in the system with respect to that of other actors. Extending the investigation to indirect relations, via third actors, it is possible to distinguish between central and peripheral actors in the network. Central actors, then, are able to reach all other actors over very short paths.

Networks may consist of disconnected subgraphs, or they may, in some areas, be weakly connected. In both cases, the system consists *components*. Areas of high *local densities* in networks develop when many actors are mutually connected. Depending on the density of mutual relations, one can identify cliques, clubs, clans, k-cores, k-plexes, or lambda sets.

A third perspective on the analysis of networks identifies classes of actors with similar relations to third actors. Thus, a network can be characterized as a set of *positions* (or *blocs*) whose actors display similar relations.

Finally, it is possible to determine roles in networks by identifying actors who have relations to different persons that occupy similar positions in a network. As a result, roles identify actors with similar relations to third actors that are not necessarily identical.

Arranging This study is about the analysis of social structures. Its central question is how we can graphically represent relational linkages between units using automatized routines.

Graphs describe relations between units of a system. Sorting and comprehensive representation of relations often constitute such complex problems that it is necessary to employ systematic, automatized procedures. Therefore, the first problem is to construct algorithms that can produce representations that take account of a maximum number of characteristics of the underlying graph. The second problem is to determine the information that the algorithm has to utilize.

The layouts that are the result of various ordering methods determine the location of units; strongly connected units are positioned close to each other, while units that are unrelated, or only indirectly related, are positioned apart from each other. “The decisive task in visualizing a graph consists in matching the spatial arrangement within the diagram with the structural characteristics of the graph. This step has been called the layout of the visualization. For an algorithmic solution to the layout problem that results in meaningful visualizations based on objectiated criteria, it is necessary that the layout process take account of sufficiently complex dependencies.” (Brandes 1999, p. III)

The usefulness of layouts therefore depends on the question whether they can identify properties that have been determined by graph analytic methods, such as the centrality of nodes and connected subgraphs. Furthermore, their usefulness depends on their ability to distinguish particular structural configurations, such as positions and roles.

The methodological focus of this study is the single (two-dimensional) image, which aims to offer the most comprehensive insight into networks (cp. Bertin 1983). From this perspective, it is of particular importance to utilize visual elements effectively, so that the solution of representational conflict by means of sequences of images, or higher-dimensional representations, can be avoided. Even with only three dimensions, higher-dimensional representations require a multiplicity of perspectives, as a single perspective cannot avoid occlusions and layerings; as a rule, interactive user interfaces therefore cannot be avoided.

Marking A further question involves the integration of further information into evolving representations, and the visual coding of the properties of graphs and their elements, that is, the use of size, shape, and color to clearly mark structural particularities. Such markings render certain selected properties of the nodes, their local environment, or of particular sub-structures, onto the arrangement and thereby emphasize particular aspects of the graph.

We can improve the readability of a representation to an amazing degree by using size to mark structural properties of the elements of ordered networks. Even lay persons can easily read the resulting images of networks. We use topically widely diverging examples to investigate the integration of graph properties into visual displays, and the corresponding choice of visual signs for this purpose. The examples also differ with respect to the richness of information that describes the relations and attributes of actors. They cover problems ranging from binary, symmetric graphs to valued, directional networks with additional information on the attributes of nodes and edges.

Exploration In a third step, it is finally possible to investigate additional information – external attributes of nodes or edges within a network – with respect to its location within the structure of the network. For this purpose, node or edge symbols are marked according to the available information about them, based on appropriate color schemes. As a result, the concentration of a particular attribute in a particular area of the network appears as a special region and points to patterns of external information and of the embeddedness of units within the structure.

Due to the combined translation of structural properties and external attributes into the arrangement of ordered networks, the resulting informational geographies are often intuitively readable, and even local peculiarities can be quickly identified.

Data Visualization It is common to view the translation of data into graphic representations as a problem of the visualization of information in general. We use such an abstract perspective to put the general problems of data, and data structure, representation by visual arrangement on a systematic foundation. That is, we try to understand how visual elements (the position of markings, lengths, shapes, colors, directions, and patterns) can be used to communicate the information encoded by numeric data. As a result, we can use a unified perspective to investigate comparatively simple diagrams used in mathematical statistics, such as histograms or scatter plots, that make use of only a limited range of visual markers[1], as well as much more complex forms of visual representation.

Under this general perspective, therefore, metric representations of networks compete with schematic arrangements and conventions of display, as long as the latter provide similar information or communicate specific aspects of networks more efficiently. With respect to the location of the units of a graph, Euclidean spaces are only one of several possibilities to generate arrangements of elements. If the goal is to investigate and to group sets and subsets of systems, it might be sufficient to use topologies that identify similarities of elements through neighborhood relations.

It is possible to distinguish at least two steps in the process of data visualization: the identification of structural properties of the raw data, and the translation of the data structures into visual markers[2]. We consider the resulting visual representation to be expressive if their visual structure represents all, and only, information contained in the underlying data structure. We consider displays to be

---

[1] Such diagrams use primarily the location of points and the lengths of lines as signifying markers. Such visual elements have been comparatively well understood for a long time.

[2] “The reference model of information visualization developed in this chapter approximates the basic steps for visualizing information: The first step is to translate raw data to a data table, which can then be mapped fairly directly to a visual structure. View transformation are used to increase the amount of information that can be visualized. . . .

“In visualization data tables are mapped to visual structures, which augment a spatial substrate with marks and graphical properties to encode information. To be a good visual structure, it is important that this mapping preserve the data (MacKinlay 1986). Data tables can be often mapped into visual representations in different ways." MacKinlay 1999, p. 33.

more effective if they can be interpreted more quickly, enable us to discern more distinctions, and are interpreted with fewer errors than alternative presentations[3].

Methodology From an abstract perspective, it is clear that the visualization of data does not constitute an alternative technology to common statistical methods; rather, it can be seen as their extension. The data information, that is, the results of the statistical analysis of the data, is translated further into the medium of graphical signs. This medium possesses a larger bandwidth and facilitates the simultaneous communication of several pieces of information.

From a methodological perspective, it should be noted that the knowledge necessary to produce efficient visualizations cannot be derived endogenously to statistics. The factors that determine the effectiveness of a visualization are properties of the visual representations used, their resolution, and the transformational rules used in the visualization. Effective transformations depend on the properties of human perception. They require detailed knowledge of the information and degree of order that humans can gather from visual patterns, and the conditions under which they can do this.

Once we know the functional relationship between the variation of visual markers and the intensity of the human sensations they cause, we can translate numeric data into visual variables that trigger corresponding sensations. In contrast, if we mapped the data into visual markers but disregarded these functions, it would be possible that the orderings contained in the numeric data would be lost. This would mean that the observer had to sequentially interpret the visual markers and thus reconstruct the ordering of the data.

Today, the way human perception distinguishes the lengths of lines, the areas of circles and squares, but also different colors, can be described by *psychometric functions* (see chapter 2). If one uses psychometric functions to translate numeric information, then the resulting visual representation triggers patterns of human sensation that correspond to patterns of the metric data. Since the ordering relations within the data are thus preserved, the resulting perceptual representations are *homomorphic*[4]. If one uses the natural orderings of human perception, then the visual patterns can be used to determine the same relations and dependencies that can be determined by numeric measurement.

Therefore, sociological measurement[5] tries to represent observations, that is, characteristics of actors and objects, by numbers, so that it is possible to investigate their properties and relationships

---

[3] "A mapping is said to be expressive if all and only the data in the data table are also represented in the visual structure. . . . A mapping is said to be more effective if it is faster to interpret, can convey more distinctions, or leads to fewer errors than some other mapping." MacKinlay 1999, p. 23

[4] "The term “homomorphic representation," in contrast to “representation," implies that empirical objects are not just mapped onto numbers, but are mapped in such a way that the numeric relations preserve the empirical relations as well . . . An *into* homomorphic representation is also called isomorphism, or structural equivalence. Isomorphic representations can always be generated from homomorphic representations, through the appropriate pooling of empirical objects in equivalence classes (f, then, is not a point function, but a set function)". Gigerenzer (1981), p. 45f

[5] “Definition: A measurement is a homomorphic transformation (representation) of an empirical system into (by) a numeric system.

This definition suggests an important use of measurement: The relations between numbers can be used as substitutes for the relations between empirical objects, so that empirical statements (for example, about the strength of correlations between psychological properties) can be derived. . . .," Gigerenzer (1981), p. 46

by means of numeric operations. In addition, the visualization of information subjects the numeric measurements to a second mapping, into sets of visual characteristics. As a result, numeric information, which in itself is already a representation of observations, is being translated into corresponding human sensations. Graphic markers not only communicate information very quickly, they can also communicate several pieces of information simultaneously.

From the perspective of the information visualization, it is therefore important to identify the universe of visual elements that can be used to codify information in a simple way. In this context, we are looking for such visual elements that communicate information independently of each other. This makes the representation of multidimensional information possible.

On the visual dimension, a further step is the identification of those functions that describe how visual characteristics are distinguished in human perception. With the knowledge of those functions, it is possible to translate numeric data into visual signs that preserve the information within the data and its orderings. Visual markers can be designed in such a way that the information contained in the data is communicated quickly, easily, and with precision.

Chapter 1 The first chapter of the study summarizes the role of graphic methods to process large amounts of information. Considering its long tradition to codify information visually, the experiences and conventions of cartography constitute an important starting point. Cartographers have developed representational conventions that are based on a long experience with the use of visual signs and markers to communicate information.

However, the basic question is why visual representations foster the abilities of discovery, orientation, and identification within human thinking and problem solving. For example, there has not yet been a sufficient theoretical answer to the question why the new forms of visual communication within the digital media have been so successful communicating complex contexts.

Considering the complexity of human perception and the multitude of cognitive processes, it seems almost impossible to answer this question. However, even a brief overview of applicable research suggests that human perception and human thinking are able to process, with a high degree of flexibility, the patterns that result from the visualization of numeric information, and that humans can extract a variety of information even from simplified representations.

Obviously, human perception is able to extract information even from non-Euklidean representations. Quantitative information can be even read off topologies; a few points of reference are sufficient to provide orientation within complex sets of information. This also implies that the basic informational content of a representation can be preserved, even if the transformation of numeric results leads to schematic simplifications. In addition, this indicates the usefulness of procedures that flexibly produce simplified patterns of orderings, or that transform metric embeddings into topologies or topological orderings.

Chapter 2 In the second chapter, we identify efficient transformation rules for visual structures, combining results from statistical graphics, the study of search and orientation processes, cognitive psychology, and colormetrics.

In statistical graphics, there has been extensive research on the question which markers are particularly useful to communicate qualitative and quantitative variables. Essentially, this line of research focusses on the visual variables specified by Bertin, which are rated according to their ability to communicate nominal, ordinal, and metric information.

Bertin's variables have also proved to be particularly effective for the study of search and orientation processes. They can be decoded in very short time and permit the quick solution, quasi-automatically and without the conscious direction of attention, of search tasks involving simple templates.

For the communication of additional network information, size and color, besides the position of visual markers, are of particular interest. The perceptual distinction of the strength of physical stimuli, such as the length of lines, has been extensively studied and functionally specified. This is also the case with respect to the areas of simple geometric signs, such as circles and squares.

Although the properties of colors are disproportionately more complex, the metrics of their perception are well known. Psychometric color models describe equidistant perceptual levels that humans distinguish with respect to color hue, brightness and saturation. Thus, we can use color to translate data isomorphically into human perception.

The economics of information provide a useful perspective on the use of effective transformation rules to maximize the efficiency of representation: Visual representations are more efficient if they permit the processing of larger amounts of information within the same period of time. Thus, we also obtain a simple criterion to evaluate visualizations.

Chapter 3 The third chapter deals with the properties of social networks, their visual representations, and simple explanations of social network structures.

We start with a brief summary of the properties of networks, as they are typically identified in network analysis. This implies the identification of structural properties, of connected substructures with particular properties, and the specific positions of actors that are similarly integrated into the surrounding structure.

To explain the structure of social systems, we often use actor typologies. Formally, such simple explanations result in mutually exclusive subsets of the elements of a system. To the extent to which those subsets identify special locations of actors, they can be considered explanations of structures.

Some visual languages that can be used to investigate the relations between subsets are very old. Venn diagrams, for example, permit the representation of relations between subsets, and of the logical implications of statements propositions. Such diagrams make use simple geometric shapes, such as circles, whose relationships they depict with overlapping areas. They also permit the identification of neighborhood relations, orderings, and sequences of subsets if particular sets contain other sets. If systems of subsets can be represented by concentric circles, it is possible to use these (the distances between circles) to determine distances between neighborhoods. Such arrangements have been used for a long time for the visual representation of networks. They are particularly useful as a frame of reference for the positioning of elements; for example, they clarify the position of network elements with respect to the center of gravity of the entire system.

Chapter 4 The fourth chapter is about procedures to order networks. The linkages between the units of a graph, which are described as a network, constitute a highly complex problem. The purpose of visualization is, then, to specify visual designs that reproduce the essential properties of the underlying graphs.

In a first section we discuss how we can use automatic procedures to produce simple diagrams, similar to the traditional Venn diagrams, that reproduce the structure of a graph. For this purpose, we use an algorithm that optimizes the transformation of nodes to fixed positions, such that the aggregate length of all edges is minimized. We show that such an algorithm can represent simple structures, and that it can produce representations that reproduce the results of well-known data sets that can be easily analyzed statistically. Finally, we demonstrate how this method can be used to investigate networks of subsets that have been defined *a priori*. We can now impose geometric restrictions on the solution space for those subsets, for example by associating the subsets with different circles, as in a Venn diagram. Then, the algorithm orders the entire system such that central units are placed towards the center of the system, or proximate to the main actors with which they interact. Peripheral actors, or actors whose relations are within the same subset, in contrast, are located within the system's periphery.

A second class of highly flexible procedures that have been useful for the ordering of graphs are spring embedders. Such procedures order graphs through the introduction of repulsive forces between all units, and of attractive forces between units that are directly related. The resulting equilibria often result in easily readable layouts that can be optimized according to additional criteria. We specify how to modify these procedures for two special classes of graphs (that is, bi-partite and valued graphs), so that we can also flexibly order those problematic classes.

The particular advantage of these procedures is that we can use them to flexibly transform the embedding of graphs, within certain boundaries, preserving the neighborhood relations of the ordering. Such transformations of solution topologies are particularly useful if we want to communicate additional information about networks by means of size markers.

Chapter 5 The fifth chapter investigates how pieces of additional information can be integrated into the layouts to facilitate faster orientation within the representations of complex networks. For this purpose, we integrate visual markers into the layout that allow us to identify particular configurations of elements within the system.

In order to describe the embedding of the elements of graphs into their surrounding structures, we use *zone symbols*. We use *convex hulls* to identify the location of subsystems within the structure. Finally, we demonstrate how we can use the length (shortness) of the paths to all other reachable nodes to illustrate the centrality of particular positions within a network.

Digressing from the main argument, we investigate whether the arrangements produced by spring embedders are able to visualize more complex network analytic results. The goal is to develop types of representation and visual markers that can display blocks of actors (sets of equivalent actors that are similarly embedded into the surrounding structure). For this purpose, we use hulls and aggregations to communicate visually the results of block model analyses.

Chapter 6 As a further digression, we demonstrate in the sixth chapter how transformations of different layouts of the same data set can be used to gain further insights into the underlying network. We arrange a system of exchange relations, subject to geographic constraints, and then transform it into a social system that reveals the core of the underlying exchange system. This facilitates the identification of the social core of inter-regional exchange structures. In order to determine to which extent the system is based on ritual exchange, kinship, or pure economic exchange relations, we inspect the relationships between the core and additional pieces of information.

Chapter 7 In Chapter 7, we address the question how to translate further, external, information into visual layouts, with a particular emphasis on color schemes. The use of color permits the visual exploration of highly complex contexts. In numeric format, such complexities could not be inspected with comparable flexibility, and an investigation of their various aspects would require considerably more effort. Depending on the informational content of additional attributes, we represent their distributions by nominal or sequential color schemes, in form of pie charts located proximate to the respective nodes. By transforming the sample distribution of external attributes into color patterns, we can easily represent the relative frequencies of different attributes. In cases of ordinal or metric attribute distributions, sequential color schemes can be used to represent cumulative distributions, too.

In this chapter, we use a valued graph of world trade data. We order these data with a metric spring embedder. In this context, we investigate how nominal and continuous attributes of nodes (region of a country, GNP) can be rendered, by means of color schemes, onto layouts. Furthermore, we show how to use node attributes to derive topical markers for edges. If we transfer these derivative edge attributes, by means of color markings, into the layouts, connected subsystems with similar actor attributes appear visually as local characteristics. Areas marked by similar colors then point to local concentrations of external attributes within the structure, that is, to correlations between attributes and their embedding in the structure.

A third case deals with external attributes of edges, which can be translated into color markings of the edges of the graph. These are particularly complex pieces of information (over-time growth rates of trade flows, concentration of items of trade, estimation error in statistical models) that we can search for specific patterns. Using zone symbols, we can describe the distribution of edge attributes within the primary environment of the nodes. On the one hand, the visualization of such attributes supplies information on the concentration of particular attributes in a node's local environment. On the other hand, it is easy to visually compare the attribute distribution in a node's local environment with the attribute distributions of other nodal environments, as similar local distributions exhibit similar color patterns. To the extent that neighboring positions or subsystems in ordered networks exhibit similar local distributions, their embeddedness within the structure is characterized not only by patterns of connectedness, but also by locally similar attribute densities. Neighboring locations and densities in the arrangement of a graph therefore are characterized by multidimensionally similar embeddedness. The identification of several multidimensionally homogeneous subsystems opens the possibility of new theoretical perspectives to better understand orderings and processes within the structure.

Finally, we demonstrate the use of selection and aggregation for the further exploration of multidimensional representations. By selecting trade flows that exceed certain volumes, we can reduce the complexity of the representation and investigate more closely the embeddedness of the elements. The aggregation of units and their attributes are a second tool for the targeted investigation of representations under a particular aspect, without having to arrange the network in a different way.

Chapter 8 The last chapter summarizes and appraises the insights won in this study on the ordering of graphs, the visual marking of structural properties, the representation of external quantitative information, and the exploration of visual patterns derived by these methods. Finally, perspectives for further development, and open questions related to the visualization of structures, will be identified.