# Network Visualization

Lothar Krempel

September 14, 2009

In: John Scott & Peter J. Carrington (Eds.): Sage Handbook of Social Network Analysis. London/New Delhi: Sage (forthcoming).

## 1 Introduction

With the spread of the network paradigm to many disciplines from sociology and ethnology to computer science, biology, physics, and economics, many computer programs have become available that allow networks to be represented visually. While graphical representations of network data are easier to produce than ever before, the quick dissemination of these new technologies has not unlocked their potential as yet. The knowledge of how to improve visual representations requires a much more active understanding of these new exploratory tools.

This article seeks to contribute to a more general understanding of visualization technologies. The aim is to set out some of the basic principles of network visualization and to disseminate knowledge about how the efficiency of network visualizations can be enhanced. How and why visualizations have the potential to supplement the numerical analysis of networks with a more exploratory approach needs to be better understood.

Collecting network data was for a long time cumbersome – even for small networks. The Internet has changed all this. Today, vast amounts of information have become available and allow us to analyze the interplay of thousands or even millions of individuals, technical units, or semantic units, linked into large systems.

Progress in network research has been driven by several developments which have eased access to ever larger network data in recent decades. Besides the tremendous increase in computing power and database technologies, the development of efficient algorithms has become a technological driver in the last twenty years. While matrix algebra was the formal language that allowed many concepts of social network analysis to be formalized previously (Wasserman & Faust 1994), it is not well suited to programming computers today, since large networks are typically sparse (Brandes & Erlebacher 2005). All this allows us to analyze huge networks, large amounts of data that can hardly be overseen in numerical form. These days, the analysis of network data is typically accompanied with visualizations, which allow us to view the overall system easily.

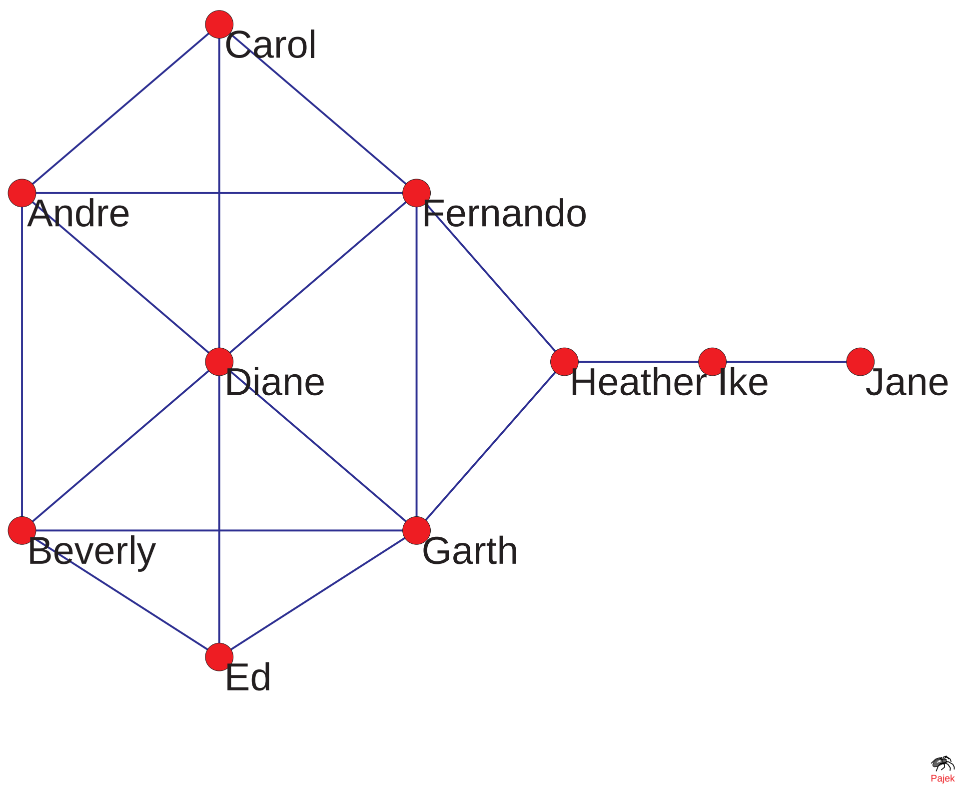
While computers became available after the 1950s, the early computing technology deployed was especially suited to numerical computations. Graphical display devices were expensive, and the computing power was limited. Low-resolution graphical devices for small computers did not appear before the mid eighties, and color output devices not before the early nineties. Today, graphical depictions provide the means to examine networks of some thousands of nodes visually, which is not possible for numerical data.

Advances in the field are driven by different communities: the *social networks community* is often motivated by substantive and methodological questions, the *mathematical graph drawing community* studies all sorts of mappings to 2D and 3D space under various constraints [1], and the *information visualization community* is especially concerned with interfaces that enable us to gain additional insights into network data. A fourth group is the *statistical graphics community* (Chen et al. 2008), which is actively strives to integrate many of the new visual options to enhance more traditional statistical diagrams. Motivated by the fascination of dynamic systems and the possibilities of digital visualization, artists too have been turning their attention to the analysis and depiction of data flows and network topologies (Ars Electronica 1994).

[1] A large number of publications is referenced under http://www.graphdrawing.org

## 2 Networks as Mathematical Graphs

Figure 1: A schematic network drawing




Networks are composed of *nodes* that are *linked* to each other. Nodes are entities of the real world: individuals, organizations, nations, and technical or logical instances that are connected by links.

Links can be of various kinds: contact, friendship, control, command, exchange, investment, trade, or information. Links can also describe co-occurrences, co-authorships, citations, and much more.

The formal definition of a *mathematical graph* describes observations as a set of *nodes* that are linked by *relations*. The links (pairs of nodes) are a subset of all possible pairs, the Cartesian product of the node set.

Links can be *undirected* or *directed.* If links in a network have different strengths, the graph is deemed to be *valued.* When *nodes* are connected by different types of *links*, it is a *multi-graph*.

Graphs connecting two *distinct sets* of nodes are referred to as *bipartite* or *two-mode* graphs (Borgatti and Everett 1997).

Especially rich graphs describe the interrelations of numerous concepts. Network text analysis (NTA), which seeks to represent the content of natural language by formal graph grammars (Diesner & Carley 2004), results in rich relational data sets that describe relations among *n different sets of nodes.* Graphs that represent text through various instances such as *actors, places, resources, institutions*, etc. are *n-mode graphs.*

# 3 Mapping Networks

The most important task in mapping networks is to determine the 2D or 3D locations of the nodes from the links of a graph. Such a *layout* encodes certain features of a network that maintain as much information as possible relating to the embeddedness of the nodes.

While higher dimensional spatial representations have greater degrees of freedom, a factor which allows them to disentangle more of the complexities of a network in the image space, intuitive navigation interfaces are needed to explore such orderings in greater detail. Additional transformations are also needed to explore 3D representations through a 2D window by perspective projections.

Some mappings of network data result in landscapes in which proximity in the image corresponds to the strength of the observed linkages. *Entities* linked up in networks are typically represented as points or rings, and *linkages* are represented as straight lines between the nodes. In contrast to geographical maps, proximity in networks is defined by functional references: who is strongly connected to whom, or who is connected in the same way to whom. Proximity in ordered social networks describes *spheres of influence*, *potential scopes of action*, and *contexts of entities that are mutually significant*. Their significance varies with the type of relationship (i.e. friendship, contact, communication, cooperation, exchange, commerce, transference of information, energy flows, or food chains).

Assigning nodes to coordinates in planar space was for a long time performed by trial and error. Today, we employ algorithms that are capable of moving connected nodes close to each other, while nodes connected by indirect paths or unconnected nodes are mapped at a distance. Various kinds of spring embedders are currently the most used algorithms, whereas multivariate statistical procedures such as factor analysis, correspondence analysis, and SVD (singular values decomposition) are other tools employed to produce spatial embeddings (Freeman 2000, 2005).

## 3.1 Network Maps

While, for centuries, the humanities had a very distant relationship to any form of graphic representation of scientific knowledge, cartography was among the first sciences to represent information in visual form. Cartographic maps inform us about "what is where."

Cartographers in the 16th century knew how to use triangulations that allowed them to represent places in maps. In the plane, the distance from location A to location C can be computed by measuring the distance between AB and the *angles* of the lines connecting A and B to C. Mapping streets, shipping routes, trade, and movements between places with graphical symbols has proved to be a very informative way of proceeding, which has many uses in social life.

Since many social phenomena exhibit a more or less strong dependence on physical distance, geographical space is a natural frame of reference for social networks. Trade has been mapped into geographical maps for more than 200 years (Playfair 1807, see also Friendly 2008); mobility patterns are another form of information that has been extensively studied by geographers, mapping their flows to geographical space (Tobler 1987, 2004).

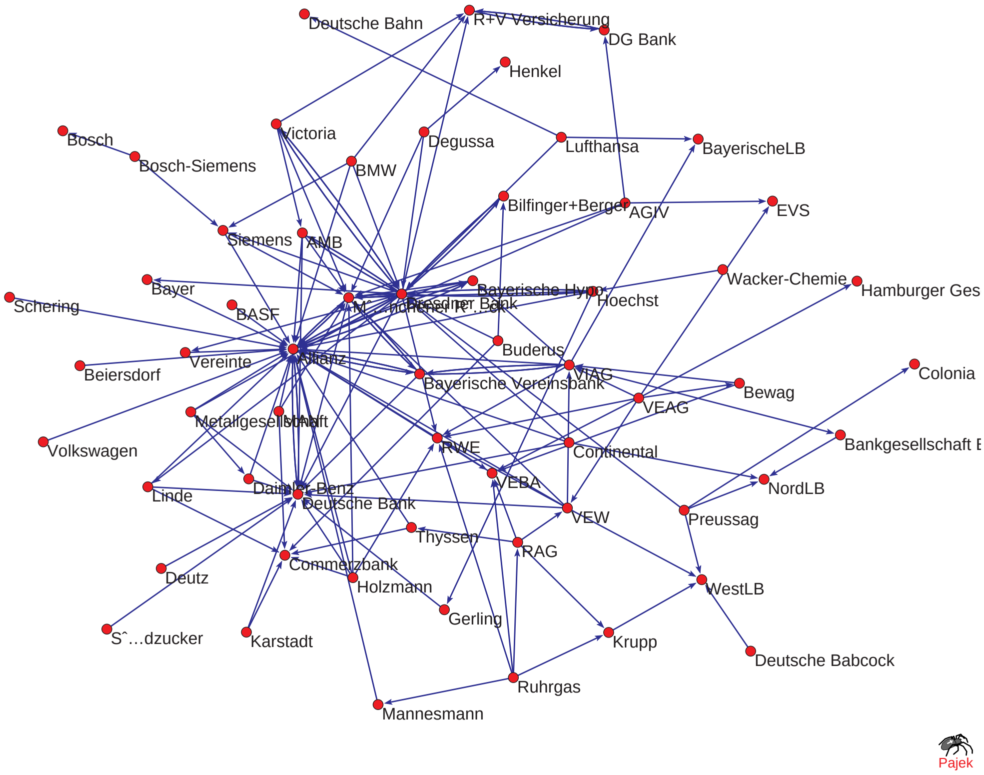
The comparison of geographical and network mappings has much potential for helping us understand how social activities change the world by overcoming geographical distance by means of modern communication and transport technologies. Surprisingly, the field has not been very intensively studied, though it seems to offer a great opportunity for understanding human behavior in a changing world.

## 3.2 Algorithms

Various planar encodings of network data can carry different messages. Figure 2 depicts the capital ties of the largest hundred German Companies in 1996 with the help of a spring embedder (Höpner and Krempel 2004). Visualizing the capital ties in this way produces a layout in which companies are placed close when company A holds shares in B. A second layout of the same data in figure 9b encodes status differences and displays these with a radial layout. Companies with a high status are found at the center of the drawing, while the distance to companies on the periphery corresponds to their status differences, with the result that strongly linked companies are not necessarily placed close to each other. Additionally, the location of companies with a similar status (same distance from the center) is optimized so that companies with ties are placed close.

The invention of statistical procedures to represent similarities and distances in statistical and network data goes back to the pioneering work of Togerson (1958)

Figure 2: A spring-layout of German capital ties for 1996




and Kruskal (1964) at Bell Labs in the 1960s and 1970s. In light of observations on similarities or distances, they designed algorithms that allowed them to embed observations in metric space. Inconsistencies in the distance data were resolved by a type of least squares procedure. This statistical treatment follows many of the ideas on how cartographers map geographical distance into two-dimensional maps. The fit of these mappings can be inspected with the Togerson diagram, which relates the distances in the image to the data distances.

Today the placement problem is typically solved by employing various kinds of spring embedders (Eades 1984, Fruchterman & Reingold 1991, Kamada & Kawai 1984), which arrange the nodes of a graph by translating links into mechanical forces that are counterbalanced by repulsive forces mimicking the repulsion of "electrical fields" to enforce a minimal distance around each of the nodes. The repulsive forces can be scaled, which leads to smaller or larger distances in the image: close neighbors are spread, while large distances are shrunk. The scaling does not affect the readability of a layout as long as the neighborhoods around the nodes are preserved. Central nodes are found in the center of such drawings, while nodes with low or only local connectivity are placed at the periphery.

While the algorithm of Fruchterman & Reingold draws networks using direct links, the Kamada Kawei algorithm requires distance data. Typically the *geodesic distances*, the shortest paths connecting any pair of nodes, are used to compute the layout. Depending on the weights of the links, a single strong link can outperform many weak links attached to a node which moves strongly linked nodes closer to each other.

The relative attractiveness of spring embedders results from their ability to achieve very different representations. Layouts can approach metric space but also achieve grid-like layouts when the repulsing forces are increased. The main characteristic of the embedders' layouts is, however, that spring algorithms provide information about the *local connectivities*, i.e. who is linked to whom and the strength of this linkage. Connected nodes are typically placed close to each other.

## 3.3 Many More Planar Spaces

Since spatial embeddings are the most effective graphical means to map network metrics, many approaches seek to use planar space to convey network properties.

Traditional statistical displays map bivariate relations as scatter plots. The x and y coordinates in the image depict two attributes of an observed unit measured on a sequential (metric) scale. These scales determine the spacing of the x or y-axis and the location of the observed instance in a drawing. As observations in statistical surveys are sampled, there is no information about how any units are linked.

*Relational data*, however, describe connections between observed units. A mapping that preserves the *local environments* of the linked units allows us to trace who is connected to whom: friends are placed closer to each other than friends of friends. *Direct influences* can be read by examining close nodes that are directly linked.

To view the system-wide *potential* of social action, network metrics such as *centralities* (closeness or betweenness) or *status measures* are needed. These measures evaluate not only the direct links but also *the indirect links* to access the potential global impact of social action. This shifts the focus of the analysis from *bilateral action* to a *system perspective*: actors occupying higher ranks in a distribution are considered to be more influential.

The *Graph Drawing Community* explores mathematical graphs with algorithms that produce *planar*, *orthogonal*, *grid-based*, *hierarchical*, or *circular* layouts. Other types are graphs with *curved* or *orthogonal* lines. These algorithmic approaches seek to explore meaningful representations under all sorts of different constraints. Aesthetics that are known to enhance the overall readability of a visual representation such as c*rossing minimizations* are additionally performed at certain stages of the algorithms.

Drawings can be *simplified* if the image space is confined to a number of equidistant points defining a *grid* in 2D or 3D space. Reducing the resolution of an image by means of grids produces *rank orders* of connectivity. Dense centers are spread while large distances are shrunk. Such transformations, however, limit the visibility of links when dense graphs are displayed.

A layout that approaches *hierarchies* in networks is the *layered map*: the y-axis is used to convey information about actor status, while nodes are placed on the x-axis so that connected nodes are positioned close (Brandes et al. 2001).

*Centrality maps* are radial orderings that group nodes around the center of a drawing, where the distance from the center reflects difference in *centrality*, *authority*, or other network metrics. Nodes with lower values are placed on more distant concentric circles. Given these constraints, links between the units can still be optimized so that connected nodes on different circles are positioned close.

# 4 Visual Layers of Network Attributes

Translating numerical information visually does not just provide a researcher with a *more complete multivariate* view, enabling them to observe how actors are embedded in the global structure and facilitating *communication* of structural findings to scientific and non-scientific audiences. As Jacques Bertin (a French cartographer) notes in his "Semiology of Graphics" (1981), the advantage of using the visual system is that it permits several pieces of numerical information to be communicated simultaneously, whereas numbers, mathematical formulae, and written language have to be read sequentially.

From a more general perspective, visualizations translate numerical information into the visual sign system. *Encoding numerical information into visual layers* is best thought of as using independent communication channels, each of which transmits a separate piece of information. Choosing various visual modes to map network properties allows properties of the network to be studied with respect to the ordering principles of a given layout. Efficient data visualization requires knowledge of both computer graphics (how to *encode* information) and properties of human visual perception (how humans *decode* graphical information) and when this *decoding is very fast*.

## 4.1 Encoding

Efficient map-making is described by Bertin (1981) as the encoding of numerical data into *elementary perceptual* tasks. Applying the "natural orders" of human perception makes visual communication almost automatic. If universal codes are used, the graphic language becomes instantaneous and international. If there are strict translation rules that enforce a bijective mapping (so that a visual representation maintains the order and the relations between the observations), visual signs can decoded into human impressions that exactly match the information that is contained in the numerical data.

Human impressions of visual stimuli have been studied in psychology for more than a hundred years. *Psycho-physical scaling* and *psychometric functions* describe how physical dimensions of visual stimuli are related to human impressions (Stevens 1975). To efficiently *encode* information into visual signs it is necessary to know how observers can read (*decode*) given graphical information. Apart from the depiction of lines, most of these functions are nonlinear.

## 4.2 Decoding

Visualizations are more *effective* if they can be interpreted more quickly, enable us to discern more distinctions, and offer fewer errors than alternative presentations. Rules that allow an observer to retrieve the encoded visual information very fast make visualizations *effective*. As has been noticed from the reports of many practitioners, certain visual encodings can be read almost instantaneously (Tufte 1983,1990), while others create visual puzzles.

That certain perceptual tasks can be read extremely fast is explained by the fact that the human brain processes elementary perceptual tasks in parallel through specialized centers. *Pre-attentive perception* needs less time than the movement of the human eye (less than 200 milliseconds). Complex graphical symbols, however, which may combine several pieces of information as icons or use metaphors, are typically much *less efficient*. Their meaning is also limited to specific cultural domains, while elementary perceptual tasks are not.

# 5 A Visual Alphabet for Networks

Nodes and lines can be of *different sizes*, can have different *colors* or *textures*, and can be rendered with *two* or *three-dimensional cues*. Most of these are already found in Bertin's list of elementary perceptual tasks (locations, sizes, textures, color, shapes, directions, angles). A more complete list, as identified by more recent vision research, identifies 2D versus 3D cues, movement and flicker, which are read *pre-attentively*. Combinations of several of these tasks, however, are not pre-attentive but need a longer time to be decoded.

## 5.1 Sizes

How depictions of graphical signs are translated into sense impressions has been the subject of research in psycho-physics for more than a hundred years. Stevens (1975) proposed a general relationship between the magnitude of a *physical stimulus* and its *perceived intensity* or strength, which is described as a power law: , where $I$ is the magnitude of the physical stimulus, $f(I)$ is the psycho-physical function relating to the *subjective magnitude of the sensation* evoked by the stimulus, $a$ is an exponent that depends on the type of stimulation, and $k$ is a proportionality constant that depends on the type of stimulation and the units used. This permits many kinds of physical stimuli and how they are related to corresponding impressions to be explored.

While the human impression of lines is linear (and has an exponent of 1), the "visual area" of a marker is related to the physical area scaled by an exponent of 0.7 – a rule that has been independently discovered in cartographic praxis and is regularly applied when depicting the size of cities on a map.

## 5.2 Shapes and Symbols

*Classes of nodes* can be rendered onto a layout using *shapes, icons*, or *symbols*. *Color codes* are an alternative. Elementary shapes such as *circles, triangles, quadrangles, stars*, or three-dimensional elements such as *cubes or cones* are shapes that can be used to communicate different classes of nodes. Symbols and trademarks are other ways of marking entities in network representations, but their meaning is limited to specific cultural domains.

Pictograms are used today in many public sign systems to communicate information. Such signs originated with the Vienna school of image statistics and the work of Otto Neurath and Gerhard Arntz in the 1920s, who developed simplified symbols ("isotypes"), isigns designed to provide everyday people with insight into complex social phenomena (Neurath 1936).

## 5.3 *Lines*

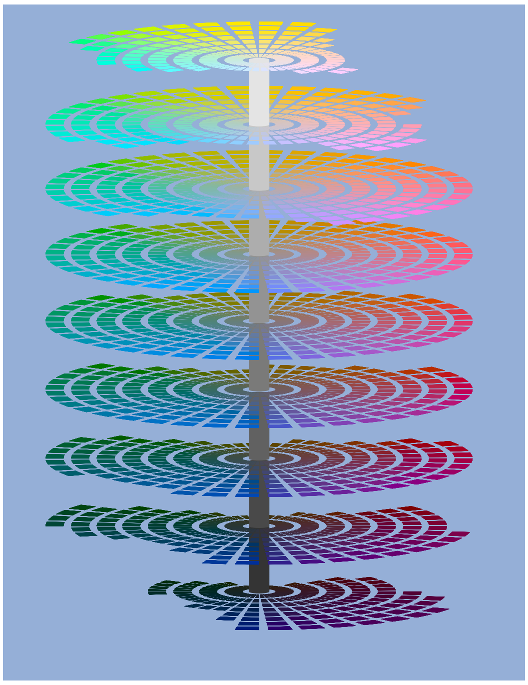
*Lines* can have different sizes (widths), which can easily lead to overlapping if graphs are dense. This can be compensated for to a certain degree if lines are arranged by size, so that strong (short) lines are drawn on top of weak (long) links.

Color-coded lines can be derived from node attributes of the source or target nodes, but can also represent attributes of the links themselves. If links have quantitative attributes, it is possible to use quantitative color schemes that assign equidistant colors to numerical ranges.

## 5.4 Color

Encoding attributes with *color* is very a complex topic, which can hardly be sketched in this article, but it has enormous potential. Even today *color perception* is only understood at very basic levels. Colors can be used to create *distinguishable and ordered impressions*. On higher levels of color perception, colors are also related to aesthetical impressions, cultural meanings, and physiological reactions.

Figure 3: The Munsell System differentiates between nine levels of lightness and ten hues which are organized radially on each of the vertical levels. Colors with identical saturation (chroma) are equidistant to the center. It is a perceptually uniform color system.



*Human color impressions* are organized according to three dimensions: *hue*, *lightness,* and their *saturation*. This provides three *layers* to communicate information. Each of these visual cues can carry its own signal.

Although there are many perception-oriented color systems that differentiate colors according to tone, brightness and saturation, almost all of these systems fail to describe uniformly perceived gradations. The Munsell color system in figure 3 is considered to be perceptually uniform. Today's psychometric colors "unbeknownst to many" have already gained entry into our everyday life. In 1976 they were introduced as international standards (CIE Lab). They are the results of not only decades of quantification by an ambitious group of colorimetricians but also the identification of mathematical functions by means of which the psychometric Munsell System can be applied to the physical model of colors.

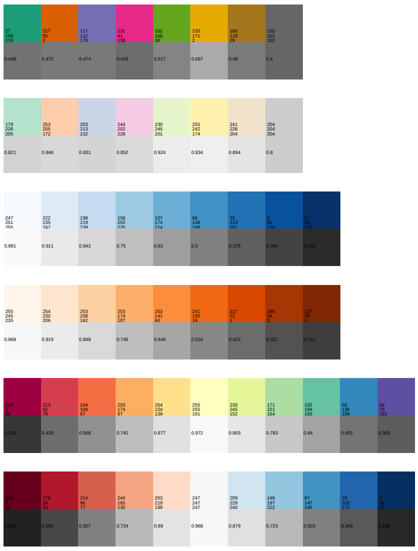
Colorimetricians have been exploring color phenomena since the beginning of the 20th century and have succeeded in mapping the physical properties of light (red, green, blue) into human color sensations (hue, lightness, and saturation). After a century's effort, they have identified complex formulae that describe how combinations of physical light waves are related to (barely) noticeable differences in human color perception (Wyszecki and Stiles 1982).

To make things more complicated, the appearance of single colors is modified by additional variables, most importantly by the *contrast to the background of a drawing*. The impressions of a color scheme vary greatly in their contrast to the surrounding background (Jacobson and Bender 1996). A dark background lets color appear brighter, while white backgrounds dim the appearance of the same color scheme. Communicating information with colors is thus highly dependent on the overall composition of a drawing, the use of *hue*, lightness, and *saturation contrasts* as well as the *contrast to the background* of a drawing.

*Psychometric color systems* describe levels of *color* that are perceived in the human brain as *equidistant*. They are the key to using color to communicate *ordered* and *quantitative* information. Modern perceptually equidistant color systems like CIE Lab have been international standards since 1976 and allow grades of hues, saturation, and lightness to be chosen so that the values encoded appear *equidistant* to human beings. This enables color schemes to be developed that communicate nominal, ordinal, and even metric information. HSB and HSV are related standards that display color in similar dimensions, but do not scale the dimensions in a perceptual way.

If attributes are to be communicated with colors, it is worthwhile taking a look at the work of Cynthia Brewer (Brewer 1999, 1994), a geographer at Penn State who has put much effort into devising color schemes for geographical maps that are very well informed about the potential of modern color systems. Not only has she developed color schemes that allow *qualitative*, *sequential*, or *diverging* dis-

Figure 4: Color schemes for the communication of qualitative, sequential and divergent distributions. (Cynthia Brewer, www.colorbrewer.org)



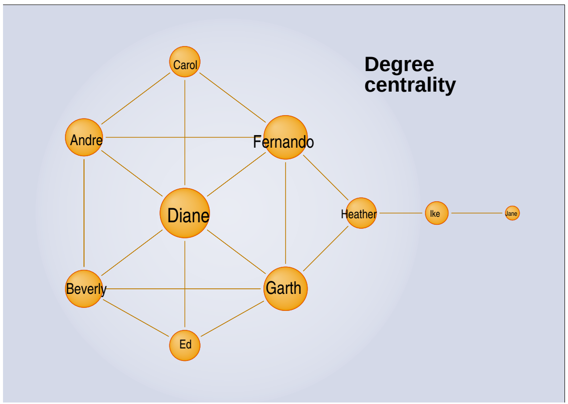
tributions of attributes; she also addresses special topics such as color blindness. Her color schemes are a good starting point when quantities or distributions of exogenous data have to be mapped onto layouts of networks (cf. figure 4).

The requirements for encoding statistical information into color schemes have been spelled out by Rogowitz & Treinish (1996): To encode *nominal* classes, colors should not be too dissimilar. For rank orders, colors should be perceived *as ordered*. To encode quantitative information, *color gradients* are needed where the *different levels appear equidistant*.

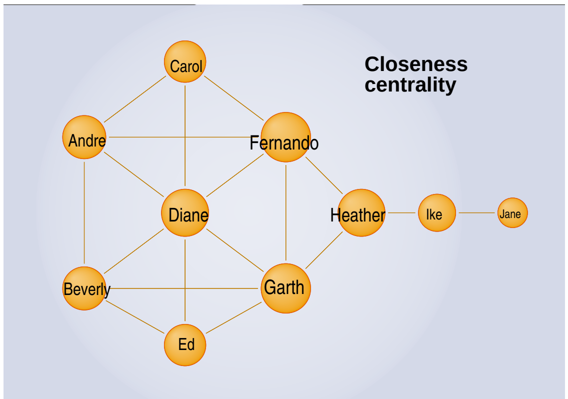
# 6 Mapping Graph Properties

A second type of picture emerges when the effort is made to depict the special qualities of a network, its component entities, or certain subsystems in the form of additional graphic features. This necessitates the use of additional graphic attributes: *sizes, colors*, or *forms* that graphically ascribe these characteristics to the layout of the network. In this way, graphical-theoretical qualities derived from the linkages are integrated into the depiction and can thus be read simultaneously.

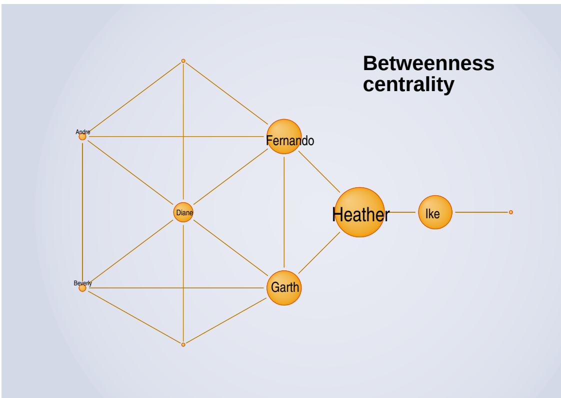
Figure 5: Using size to different centralities makes it easy to compare their distributions. (compare fig. 1)



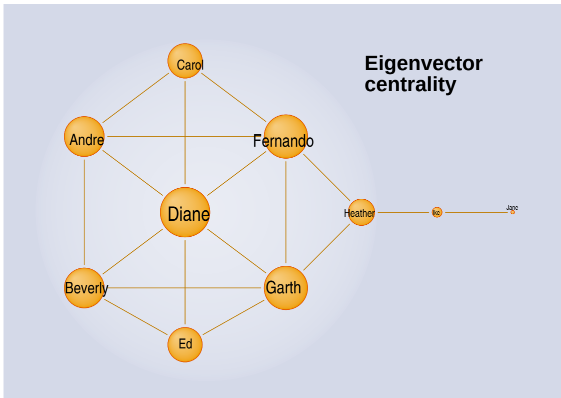

(a) Degree




(b) Closeness




(c) Betweenness




(d) Eigenvector centrality

If the *centrality* of the particular entities is portrayed by means of the *size* of the symbols, then a reading of the graphic representation provides additional information about who is involved in an especially large number of relationships *(degree)*, who can reach many agents via particularly short *paths (closeness)*, and who controls an especially large number of the *shortest linkages* to an adjacent network *(betweenness)*. Coding node properties using the size of the symbols makes for a simple way to study global and local positions. Figures 5a-d map different centralities onto the schematic layout of fig. 1 through the size of the nodes. This permits us to analyze their distributions in detail and to read which actors are considered important by means of a specific type of centrality.

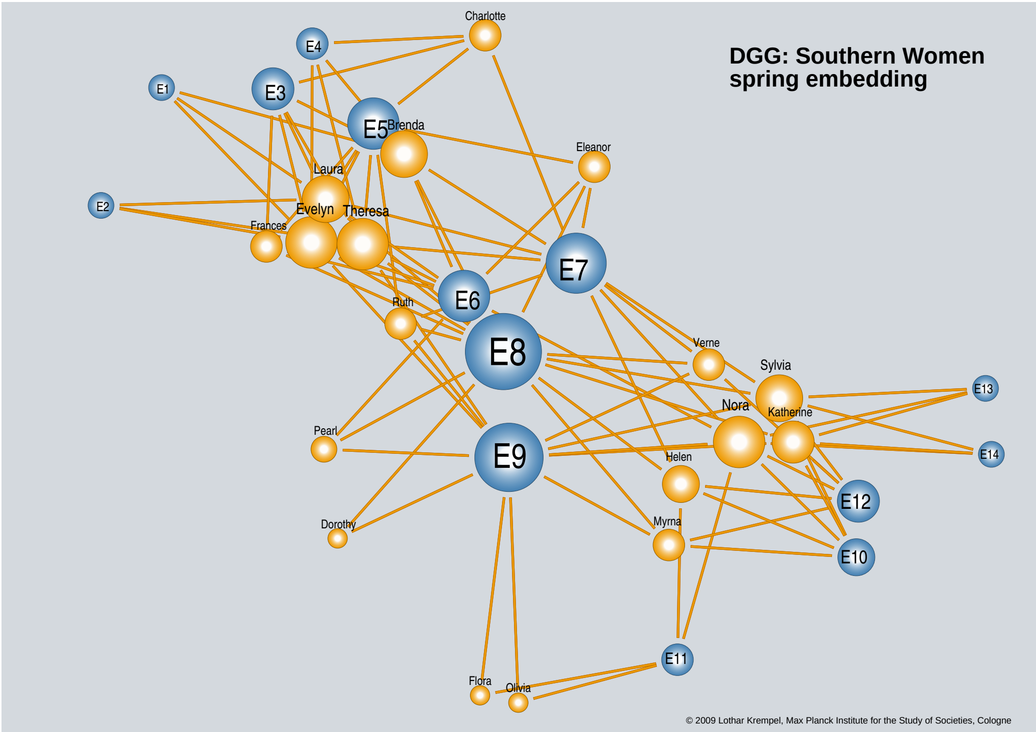
Mapping graph properties as a second layer of information onto the planar layout of a network produces information-rich landscapes, which allow us to explore the layout in greater depth. They provide orientation in a similar way to geographical maps. The fascinating thing about these charts is that they fit together a multitude of observations, like pieces of a jigsaw puzzle, into a picture of the system as a whole. The human eye can discover particular patterns in them with relative ease.

Of special interest are *all intermediate levels* of social structures. *Dense areas* are subsets of nodes that are closely connected. A zoo of concepts exists that can be applied to identify *cohesive subsets* in networks (*components, cores, cliques, n-cliques, clans, clubs*).

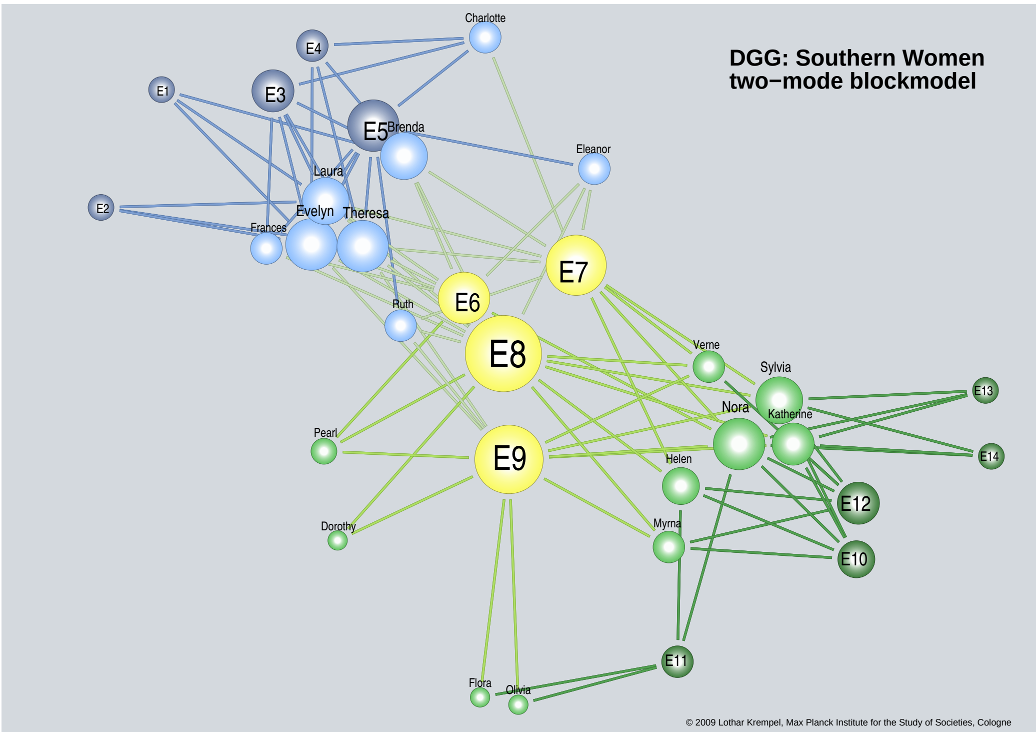
To demonstrate how various network metrics can be mapped simultaneously onto the layout of network, we use the classic "Southern Women" network data set (Davis, Gardener & Gardener 1942). This data set describes the interaction patterns of 18 women participating in 14 informal events over a period of nine months. This bipartite graph has often been used to demonstrate the usefulness of new network algorithms in the literature (Breiger 1974, Freeman 2003). Figures 6 and 7 demonstrate how various properties of this well-studied graph can be mapped with different graphical layers onto the layout which has been produced with the help of a spring embedder.

Figure 6a displays the two type of nodes (women and events) using two colors (yellow and blue) and maps the degrees of the women and events with sizes.

Figure 6: A bipartite graph: Davis Southern Women




(a) Layout, degrees and sets



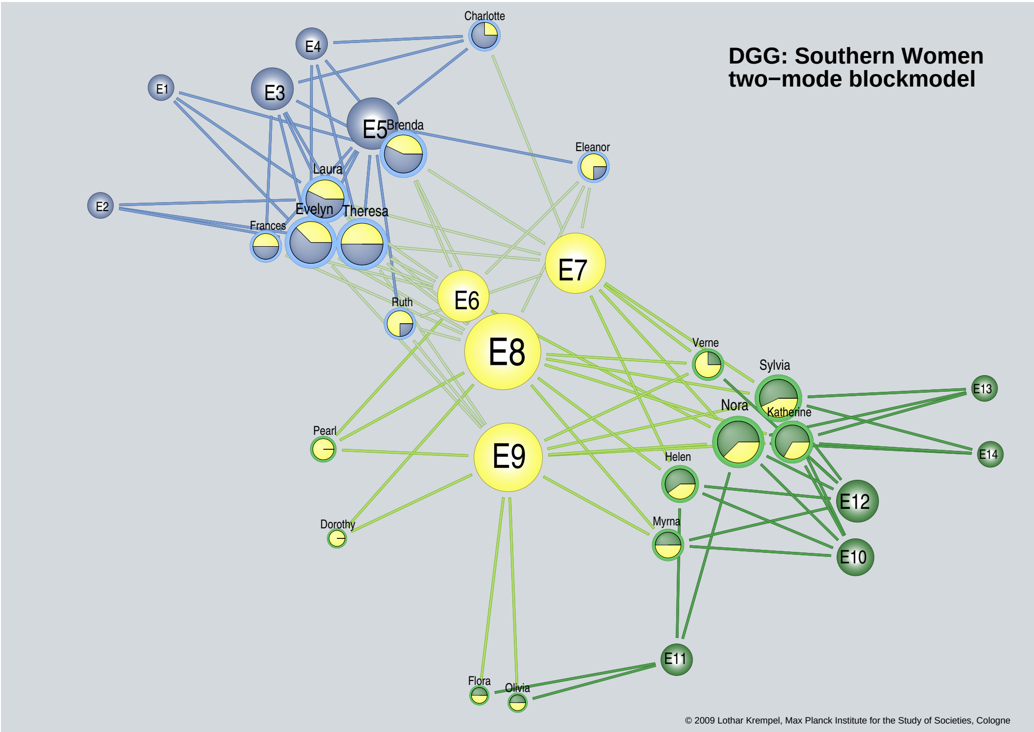

(b) A two-mode blockmodel: 2 women and 3 event blocks

*Positions* and *roles describe* sets of structurally equivalent actors, i.e. actors who have similar profiles of connections and certain comparative advantages: high levels of autonomy or competition (Burt 1992). As long as intermediate structures such as cores or blocks are used to detect subgraphs, the nodes are decomposed into non-overlapping subsets (partitions). To display these classes, it is sufficient to use different hues. A computation of a two-mode block model for the Southern Women dataset yields three structurally equivalent blocks for the events and two for the women. Mapping the block membership with colors (dark green, yellow, and dark blue for the events; light green and light blue for the women) onto the layout of figure 6a results in figure 6b.

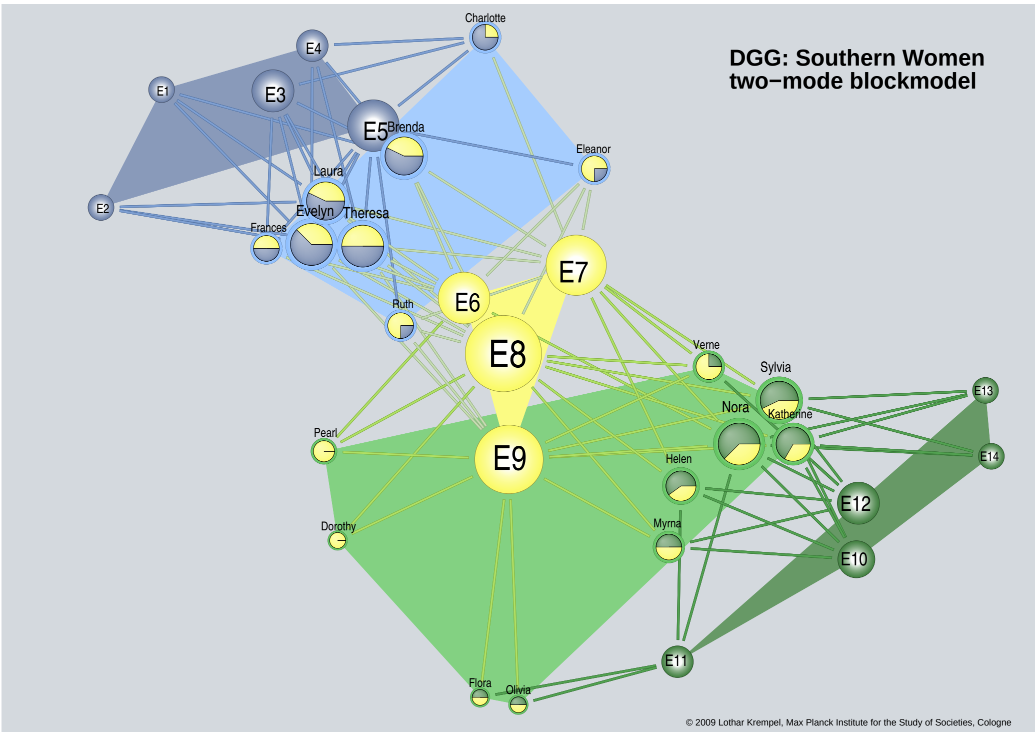
Pie charts allow the connections of the women to the different event blocks to be viewed in greater detail. They depict the number of links to different blocks and reveal which women are connected to a particular event block. Figure 7a shows that all women are only connected to two of the event blocks (yellow/blue or yellow/green). The yellow events are central and are visited by both groups.

Another graphical element is the *convex hull*, which is useful for identifying sets of actors in a given layout. A *convex hull* is a concept from computational geometry and is used here to observe how partitions of the nodes are distributed in a network layout. A *hull* wraps all nodes of a given class by identifying the *area* that is covered by these nodes. If *hulls* are computed for all node sets of a classification, their *intersections* identify areas where members of different sets are placed close to each other, while indicating exclusive areas that only contain members of a specific set (cf. Johnson & Krempel 2003). Mapping the blocks for the Southern Women data set yields figure 7b. In this case, we find that the spring embedder layout has positioned all blocks in separate areas of the drawing. Reading the pie charts of the women also allows us to discover that Dorothy and Pearl are both connected to the central block of events only and therefore hold a distinct position in the network (cf. Doreian, Batagelj and Ferligoj 2005, pp. 257-65).

Figure 7: A block model: Southern Women

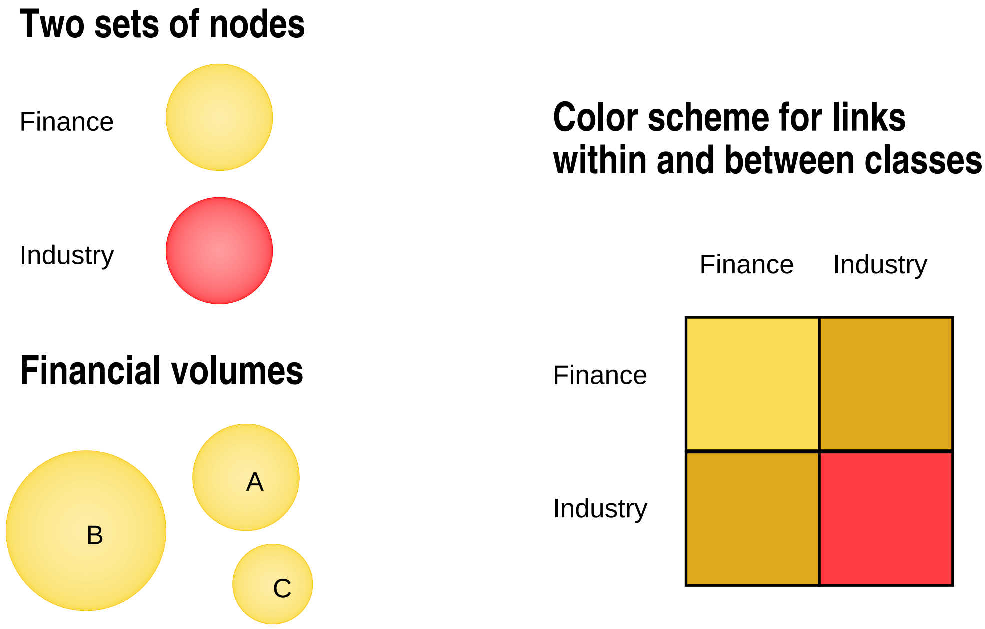



(a) Pie charts as node symbols




(b) using convex hulls

Figure 8: Mapping external attributes




Whereas visual approaches to single networks allow us to examine the complete distributions of the nodes and their properties in a layout, comparisons of different networks need additional normalizations, correcting for the number of nodes or lines. Comparisons of different networks typically make heavy use of traditional statistical methodologies. Exceptions to this are dynamic networks that trace network changes over time.

If the focus is to compare *systems*, a whole range of statistics can be used. Graphs can be compared on the system level by their *density,* their *degree,* their *transitivity and clustering*, and the *number of dense areas or positions.*

# 7 Mapping Explanations

A third class of "analytical graphics" emerges if, in similar fashion, external information about the component units or their interrelationships (e.g. theoretical classifications or independently gathered data) is introduced into the representation. In mapping explanations, typically *color schemes* are used.

Attributes of the links can greatly help to understand how relations between different subsets of external classes of actors ("*catnets*" as defined by White 2008) are organized. Color codes of partitions and color schemes derived from these external classifications can be used to display who is connected to whom and to what extent.

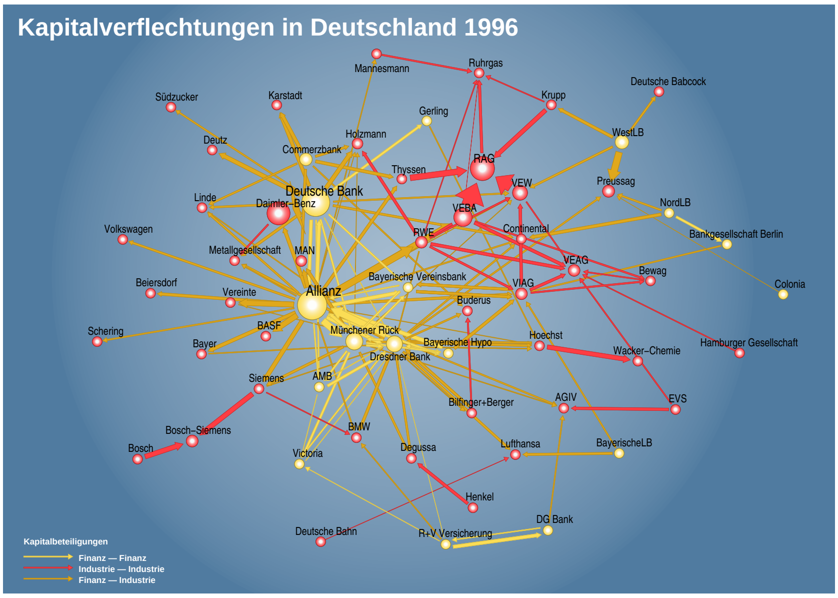
Mapping exogenous attributes onto the layout of a graph makes it possible to explore how actors of a certain kind interact in the network and how a given pattern of external attributes relates to the layout of a graph. Again, correlations will appear as local patterns that may allow us to understand the possible causes of emergent social processes.

In an analysis of equity capital interrelationships, for example, classifying firms as industrial enterprises, banks, and insurance companies, and selecting a different color for each category facilitates recognition of particular concentrations in the network, with any preponderance of units of the same color displayed indicating internal interrelationships.

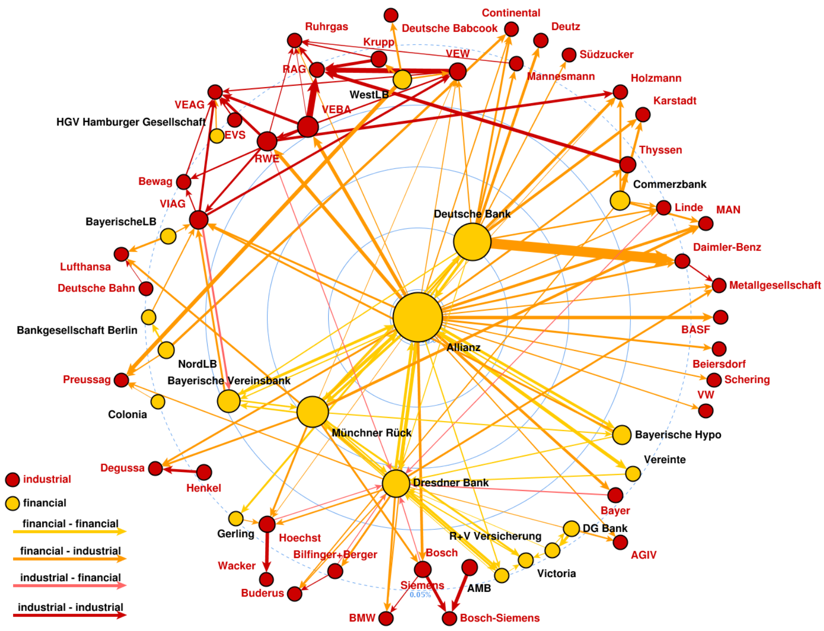
These can be examined more closely by means of lines in a *derivative color scheme*: the extent to which such investments are held exclusively by banks and industrial enterprises or whether the majority of the interrelationships consists of equity capital interpenetration by banks and industrial enterprises. In this case, utilizing different colors projects a theoretically significant classification onto the arrangement of a network.

The depiction makes it possible to ascertain whether the theoretical process of differentiation exhibits systematic patterns in the optimized arrangement of the network. In contrast to a purely statistical treatment, weak local interconnections also emerge in networks. They indicate the structure's potential for development.

Figure 9: Capital ties and attributes




(a) attributes and spring layout




(b) A radial layout of status differences (Baur, Brandes & Wagner 2008)

The spring embedder solution reveals a central core formed of cross-linked banks, which appear as a yellow pattern in the center of the drawing and a second (dark) industrial cluster in the north-east of the drawing (figure 9a).

In the radial drawing (figure 9b), which is based on the same data, status differences are encoded. Such a drawing rests on the assumption that the bilateral control rights that are affiliated with a capital link go beyond direct control. Here the insurer Allianz holds the most dominant position in the center, whereas the other cross-linked banks appear on the semi-periphery, i.e. have lower status. Looking at the color patterns, it is easy to see that the banks hold the most dominant positions, while the industrial cluster can be detected in the north-north-west of the outer circles.

# 8 Complex and Large Drawings

Networks of several thousands of nodes are not easily displayed, even if large output formats are used. The analysis of large networks has to use *formal* or *substantial strategies* that illuminate processes in the overall structure.

A common strategy is to apply some *sort of filtering* which reduces networks to the *most connected nodes* or *most dominant lines*. The idea is to discern some sort of *backbone* to the overall structure. *Node-cuts* or *line-cuts* are two strategies that can help to identify the most important parts of a network. A *node-cut* results from a decision to keep only nodes that surpass a certain *threshold* – for instance nodes with a certain degree or centrality. The subgraph will contain only lines between the selected elements. If, however, we impose a threshold for the lines (*line-cut*), the subgraph is defined by all nodes that are connected by lines that surpass a certain threshold. Only nodes that are linked by at least by one of the selected lines are contained in the subgraph.

Another approach is to select *dense areas* that surpass a certain threshold of connectedness. *Cores* are node sets that are linked by a certain number of ties. While many definitions of dense clusters generate overlapping cliques, cores are nested hierarchically, which results in non-overlapping partitions of a network.

Block modeling is the traditional approach to analyzing social networks and allows graphs to be reduced to block structures (partitions of the nodes), which provide information on who is connected to similar actors in the system. *Structural equivalence* seeks to identify *positions* (sets of the nodes) that are connected to (identical) actors. Social positions need not be densely connected; they describe actors that are engaged in the same contacts. *Aggregating* nodes and ties for blocks yields a reduced graph (simplified structure and image) that describes interrelations between *positions* (Krempel 2005).

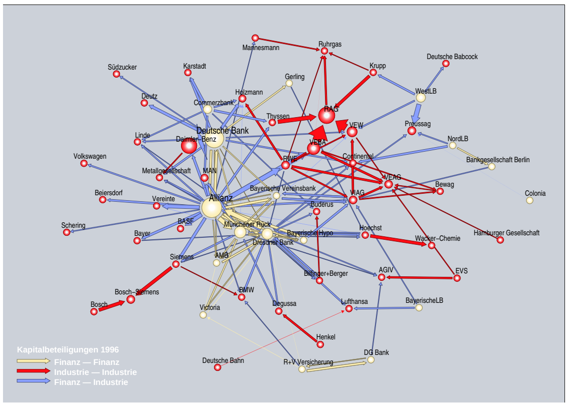
*Multi-graphs* are composed of different kind of links. In such a case, a *peeling approach* is advised. Which type of link achieves the highest levels of connectivity? Which type of links contributes to the overall connectivity once the most dominant ties have been removed?

*Temporal networks* are one of the frontiers of today's network research and visualization.

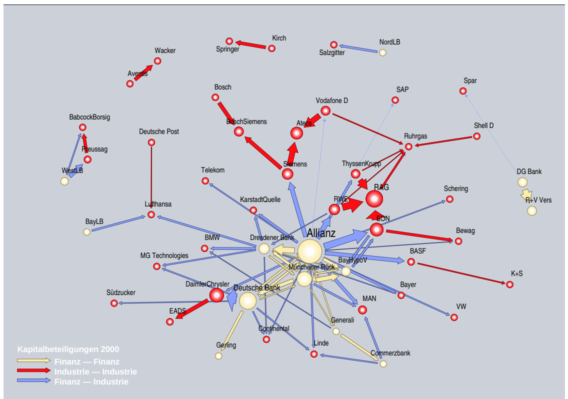
> The combined use of timing and relational information is usually named 'Dynamic Network Analysis.' ... Confusingly, the term 'dynamic network' is often used in the literature to describe various specific subclasses:
>
> - Networks in which the edge and node sets remain fixed, but values of attributes on nodes and edges may vary in time (transmission models)
> - Networks in which edges are added or deleted over time (computer networks, friendship relations)
> - Networks in which the weights of edges change over time (neural networks, exchange networks)
> - Networks in which nodes are added or removed in time (ecological food webs, organizations)
>
> Clearly these categories are not exclusive ... (Bender-deMoll, Morris & Moody 2008).

Many real world problems, however, exhibit more complicated dynamics: if we trace capital ties between companies over time, we find that new firms *emerge*, companies *cease* to exist, pairs of companies *merge* into a *new legal entities*, while others *split* or *spin off* into new companies.

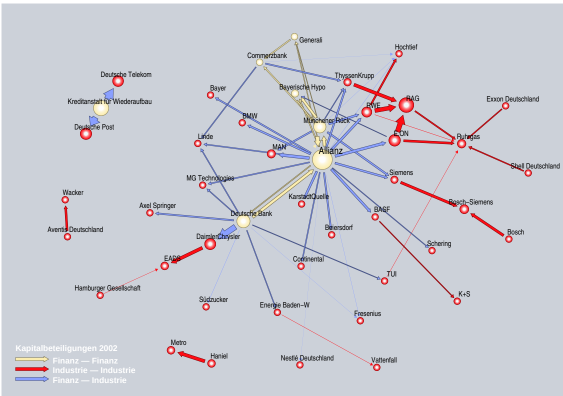
Figure 10: Visualizing the comparative static development of the German company system (1996 - 2006)



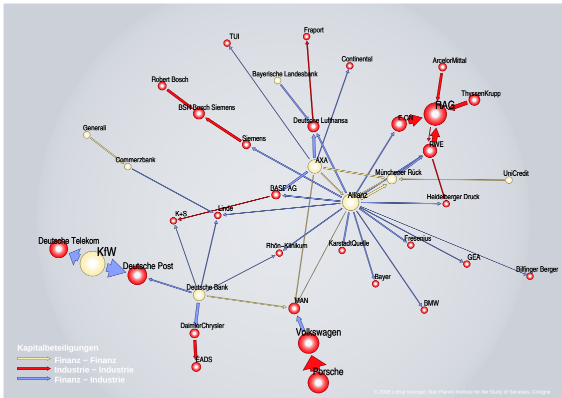
(a) 1966



(b) 2000



(c) 2002



(d) 2006

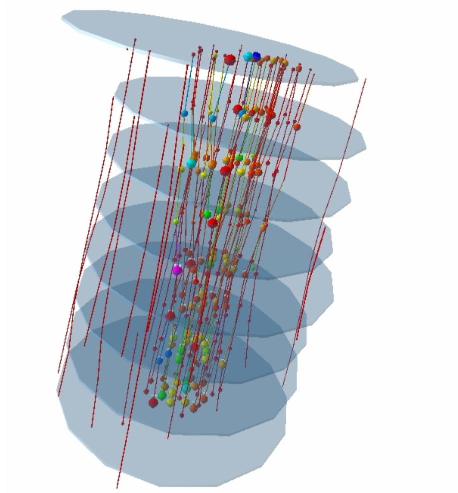
Temporal networks, which describe graph changes diachronically, are attracting growing research interest in their analysis. As *time* adds an additional dimension to the data, there is also a tendency to use 3D drawings. 2D representations can be an alternative if *sequences of drawings* are used, when there is only a small turnover in the population over time.

Structural growth processes represent simple cases , where a growing number of relations reshape a population of nodes over time. Links attach to an early core and produce snow-flake-like growth. New links (that do not decay) connect an increasing number of nodes over time. Looking at the development of collaborations among graph drawing scientists, Graphael (Forrester, Kobourov et al. 2004) present a network that has developed around an initial core.

To trace networks diachronically, *sequences of drawings* can be used. However, if the layout changes too much between single points in time, it becomes too difficult to trace what happens. The Graph Drawing Community deals with this problem by implementing algorithms that preserve *mental maps*, limiting changes to the nodes between different points in time in the overall layout. (Purchase, Hoggan & Görg 2006). One strategy is to compute the layout for the *supergraph* (the union of all graphs in sequence) and to display only active elements at certain points.

Skye Bender-deMoll & McFarland (2006) present various ways of how to visualize network processes as *sequences of images* or as *films*. The authors use *smoothing techniques* such as *moving averages* that aggregate network information in larger *time frames*, and/or the *interpolation* of layouts between discrete points in time, to achieve easier readable results.

Figure 11: Geomi: Emails between scientists over time



A temporal extension of radial drawings can be found in GEOMI (Ahmed, Dwyer et al. 2006).

> The 2.5 D method is one of the solutions to represent temporal network data. In such a method, a graph snapshot at a particular time is placed on a 2D plane, in which a layout algorithm can be applied; a series of such planes are stacked together following time order to show the changes. In order to identify a particular node in different time plane, same nodes in different planes are connected by edges. Combined with navigation tools in GEOMI, users can trace the change of each individual node's relationship to others and also can evaluate the evolution of the whole network in general.

As an example, fig. 11 shows the e-mail connections of a certain research group. Each plane represents one month, while each node is one person. The edges between nodes on the same plate depict the e-mail traffic between persons. In addition, degree centrality is mapped to node size, while node color represents betweenness centrality.

# 9 Summary

The potential of such "visual statistics" is strongly dependent upon the resolution of a series of additional questions. How can quantitative information be communicated? In what cases can depictions of manifold information be interpreted especially easily and quickly? French cartographer Jacques Bertin already provided an important key to understanding such fundamental problems of information processing in his 1974 work "The Semiology of Graphics."

What distinguishes visual symbols from other systems of signs (writing, language, and music) is their ability to simultaneously communicate different types of information. Converting numerical information is a process of translation into elementary graphic signs. With the elementary graphic attributes of size, color, and form, multiple sets of information can be communicated independently of each other and at the same time. If the natural categories of human perception are exploited in doing this, then the translation is especially effective.

If relational observations are ordered according to systematic rules and additional external information is pictorially projected onto these orderings in a way that takes psycho-physiological principles into consideration, then the results are highly optimized, graphical information landscapes – artificial worlds that fit together manifold descriptions of the same objects and reconstruct these objects. This makes it possible to view local, multidimensional patterns and to study the positioning of the elements that have been multiply described in this way within the system as a whole.

The use of colors in particular expands the possibilities of discovering within these structures concentrations of characteristics that identify multivariate linkages. Both the technologies for automatically generating colors and the facility for utilizing different technologies to evoke similar color impressions on the part of different people are based upon an enormously improved understanding of the human perception of color. Although the use of these color technologies has quickly become very widespread in our everyday life, the scientific use of colors in the investigation of complex issues is still pretty much in its formative phase.

The extent to which we are able to better understand and apply these rules will determine how well we can exploit the natural attributes of human perception for scientific purposes. To this end, ergonomically optimized graphics use the particular capabilities of human perception in a systematic way. This makes it possible to combine the potential of automatic procedures with the special capacities of human perception.

The historical concern that visual mappings create artifacts can be overcome if we apply systematic encoding rules to represent numerical information and choose encodings that can be almost automatically decoded. Along with procedures that let us map social space into meaningful planar representations, a new world of scientific images becomes available. If done well, these images are bijective mappings that depict nothing else but the information that is encoded in the numerical data. If this is the case, the researcher can move back and forth between the numerical data and their visualizations, making comparisons that offer many new insights. Combining more traditional statistical exploration with exploratory visual inspection is especially promising when it comes to generating new knowledge.

Visualizations can supplement statistical procedures to capture local events, which are typically undetected in statistical analysis because the observed local regularities only result in medium effects if computed on a system-wide scale. The potential of visualizations is that they can identify local combinations of external attributes which are linked into clusters that are forerunners of emerging social processes. Visualizations of multidimensional network data are more sensitive than traditional statistical approaches: while linear statistics identifies causalities by revealing all the instances of certain combinations of exogenous variables, the visual layers disclose *maximal connected local patterns* that are *homogeneous* for a specific combination of external variables. Such patterns need more attention because they are candidates for emerging social processes – processes which are currently not completely understood because our knowledge and information is too limited. Visualizations can hint at where additional information is needed and help to direct our attention to domains which need further exploration. Mapping network data provides a starting point for all sorts of inquiries, no matter whether they are quantitative or qualitative.

# References

Ahmed, A. T. Dwyer, M. Forster, X. Fu, J. Ho, S.-H. Hong, D. Koschützki, C. Murray, N. S. Nikolov, R. Taib, A. Tarassov, K. Xu P. Healy, P. Eades (Eds.) (2006): GEOMI: GEOmetry for Maximum Insight, Springer Verlag, Proc. Graph Drawing, GD 2005, LNCS 3843, pp. 468-479

Ars Electonica (2004): Language of Networks. Conference and Exhibition on networks. (curated by Gerhard Dirmoser, Lothar Krempel, Ruth Pfosser and Dietmar Offenhuber) http://www.aec.at/en/festival2004/programm/LON_folder_lowres.pdf

Baur, Michael, Ulrik Brandes and Dorothea Wagner (2008): Attribute-Based Visualization in Visone,

Bender-deMoll, Skye, Martina Morris and James Moody (2008): Prototype Packages for Managing and Animating Longitudinal Network Data: dynamic networks and RSonia. Journal of Statistical Software, February 2008, Volume 24, Issue 7 http://www.jstatsoft.org

Bender-de Moll, Skye and Daniel A. McFarland (2006): The Art and Science of Dynamic Network Visualization, Joss Journal of Social Structures, Vol. 7, No. 2

Bertin, Jaques (1981): Graphics and Graphic Information-Processing. Berlin: Walter de Gruyter

Borgatti, Stephen and Martin G. Everett (1997): Network analysis of 2-mode data. Social Networks, 19, 243-269

Brandes, Ulrik and Thomas Erlebacher (ed) (2005): Network Analysis. Methodological Foundations. Heidelberg: Springer.

Brandes Ulrik, Jörg Raab, Dorothea Wagner (2002) Exploratory Network Visualization: Simultaneous Display of Actor Status and Connections. JOSS Journal of Social Structure ,Vol.2, No.4.

Breiger, R. L. (1974) The Duality of Persons and Groups. Social Forces 53, p.181-190

Brewer, Cynthia A. (1999): Color Use Guidelines for Data Representation, Proceedings of the Section on Statistical Graphics, American Statistical Association, Baltimore, pp. 55-60.

Brewer, C.A.(1994): "Guidelines for Use of the Perceptual Dimensions of Color for Mapping and Visualization," Color Hard Copy and Graphic Arts III, edited by J. Bares, Proceedings of the International Society for Optical Engineering (SPIE), San José, February 1994, Vol. 2171, pp. 54-63.

Chen C., W.Härdle, A. Unwin (ed) (2008): Handbook of Data Visualization. Springer-Verlag: Berlin-Heidelberg

Davis, A, B.B. Gardener and M.R. Gardener (1941) Deep South. Chicago: The University of Chicago Press.

Diesner, Jana & Carley, Kathleen. (2004). Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a novel method for Network Text Analysis. Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations., Harrisburg, PA: Idea Group Publishing

Diesner Jana; Kathleen M. Carley (2004): AutoMap 1.2 Extract, analyze, represent, and compare mental models from texts, CASOS Technical Report January 2004 CMU-ISRI-04-100.

Eades, Peter (1984): A heuristic for graph drawing. Congressus Numerantium, 42, 149-160

Forrester D, S. Kobourov, A. Navabi, K. Wampler and G. Yee (2004): Graphael: A System for Generalized Force-Directed Layouts. Department of Computer Science, University of Arizona [http://graphael.cs.arizona.edu ]

Freeman. Lincton C. (2005): Graphic Techniques for Exploring Social Network Data . In: Peter J. Carrington, John Scott and Stanley Wasserman (ed) (2005): Models and Methods in Social Network Analysis. p. 248-269

Freeman, Linton (2004): The Development of Social Network Analysis, A Study in the Sociology of Science. Vancouver: Empirical Press.

Freeman, Linton C. (2003): "Finding Social Groups: A Meta-analysis of the Southern Women Data" In Ronald Breiger, Kathleen Carley and Philippa Pattison (eds.) Dynamic Social Network Modeling and Analysis. Washington, D.C.:The National Academies Press, 2003.

Freeman, Linton (2000): Visualizing Social Networks. JOSS Journal of Social Structure. Vol. 1, No. 1

Friendly, Michael (2008). "Milestones in the history of thematic cartography, statistical graphics, and data visualization". pp 13-14. Retrieved 7 July 2008http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf

Fruchtermann, Thomas M.J and Edward M. Reingold (1991): Graph A large number of publications is referenced under http://www.graphdrawing.orgDrawing by Force directed Placement. Software-Practice and Experience, 21, 11, 1129-1164

Hartmann, Frank , Erwin K, Bauer (ed) (2002): Bildersprache, Otto Neurath : Visualisierungen, WUV Universitätsverlag, Wien

Höpner, Martin and Lothar Krempel (2004): The Politics of the German Company Network. Competition & Change, Vol.8, No.4, 339-356, December 2004

Höpner Martin and Lothar Krempel (2003): The Politics of the German Company Network. Cologne: Max Planck Institute for the Study of Societies. MPIfG Working Paper 03/9, September 2003http://www.mpi-fg-koeln.mpg.de/pu/workpap/wp03-9/wp03-9.html

Jacobson, N. and W. Bender (1996): Color as a determined communication. IBM Systems Journal, 36, NOS 3&4, 526-538.

Johnson, Jeffrey C. and Lothar Krempel (2004): Network Visualization: The "Bush Team" in Reuters News Ticker 9/11-11/15/01. JOSS Jounal of Social Structure. Vol. 5, No. 1.

Kamada, T and S.Kawai (1989): An algorithm for drawing general undirected graphs. Information processing Letters, 31, 1, 7-15

Krempel, Lothar (2005): Visualisierung komplexer Strukturen. Grundlagen der Darstellung mehrdimensionaler Netzwerke. Frankfurt: Campus.

Kruskal, Joseph B. and Myron Wish (1978): Multidimensional Scaling. Beverly Hills: Sage

Kruskal, Joseph B. (1964): Nonmetric Multidimensional Scaling: A Numerical Method. Psychometrika, 29, 2, 115-129

McGrath, Cathleen and Jim Blythe (2004): Do You See What I Want You to See ? The Effects of Motion and Spatial Layout on Viewers' Perceptions of Graph Structure. Journal of Social Structure, Vol. 5, No. 2

Moody, James, Daniel A. McFarland and Skye Bender-DeMoll. (2005): "Dynamic Network Visualization: Methods for Meaning with Longitudinal Network Movies? American Journal of Sociology 110:1206-1241

Neurath, Otto (1937): Basic by Isotype. London, K. Paul, Trench, Trubner & co., ltd.

Neurath, Otto (1936): International picture language; the first rules of Isotype. London : K. Paul, Trench, Trubner & co., ltd.

de Nooy Wouter, Andrej Mrvar and Vladimir Batagelj (2005): Exploratory Social Network Analysis with Pajek (Structural Analysis in the Social Sciences), Cambridge University Press.

de Nooy, Wouter (2008): Signs over Time: Statistical and Visual Analysis of a Longitudinal Signed Network. Journal of Social Structure, Vol. 9, No. 1

Moody, James, Daniel McFarland and Skye Bender-deMoll (2005): Dynamic Network Visualization. AJS Volume 110 Number 4 (January 2005): 1206-41

Playfair William (1807). An Inquiry into the Permanent Causes of the Decline and Fall of Powerful and Wealthy Nations, p. 102.

Purchase H., Eve Hoggan and C. Görg (2006): How Important is the "Mental Map"? – an Empirical Investigation of a Dynamic Graph Layout Algorithm. In Proceedings of 14th International Symposium on Graph Drawing, Karlsruhe, Germany, September 2006.

Rogowitz, Bernice E. and Lloyd A. Treinish (1996): Why Should Engineers and Scientists Be Worried about Color ? http://www.research.ibm.com/people/l/lloydt/color/color.HTM

Scott, John (2000): Social Network Analysis. London,

Spence, Ian and Howard Wainer (1997). "Who Was Playfair? ". In: Chance 10, p. 35-37.13

Stevens, Stanley S. (1975): Psychophysics. Introduction to its Perceptual, Neural, and Social Prospects. New York: John Wiley

Sugiyama, Kozo and Kazuo Misue (1994): A Simple and Unified Method for Drawing Graphs: Magnetic Spring Algorithm. In: Robert Tamassia and Ioannis G. Tollis (eds): Graph Drawing. Springer, Lecture Notes in Computer Science 894, 364-375

Tobler, Waldo (2004): Movement Mapping. http://csiss.ncgia.ucsb.edu/clearinghouse/FlowMapper

Tobler, Waldo (1987): Experiments in Migration Mapping by Computer, American Cartographer, 1987.

Tufte, Edward R. (1990): Envisioning Information. Ceshire: Graphics Press

Tufte, Edward R. (1983): The Visual Display of Quantitative Information. Ceshire: Graphics Press

Torgerson, Warren S. (1958): Theory and Methods of Scaling. New York: John Wiley

Wilkinson. Leland (2008): Graph-theoretic Graphics. In: C. Chen, W. Härdle, A.Unwin (ed): Handbook of Data Visualization. ; Springer-Verlag: Berlin, chapter II.5, p122- 150

Wilkinson, Leland (2005): The Grammar of Graphics. NY, Springer Science and Business Media

Wasserman, Stanley and Katherine Faust (1994): Social Network Analysis: Methods and Application. Cambridge: Cambridge University Press

White, Harrison: (2008): Identity and Control. How Social Formations emerge (second edition). Princeton. Princeton University Press.

Wyszecki, Günter and W. S. Stiles (1982): Color Science. Concepts and Methods, Quantitative Data and Formulae, 2nd Edition. New York: John Wiley & Sons.