Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges
Tatiana Von Landesberger, Arjan Kuijper, Tobias Schreck, Jörn Kohlhammer, Jarke Van Wijk, Jean-Daniel Fekete, Dieter Fellner

To cite this version:

HAL Id: hal-00712779
https://hal.inria.fr/hal-00712779
Submitted on 28 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Visual Analysis of Large Graphs:
State-of-the-Art and Future Research Challenges


1. Introduction

The analysis of graphs is important in many application areas including finance, biology, sociology, transportation, and software engineering; it includes a variety of different tasks. The main aspects relate to the understanding of global and local structure of the graph, the connections between entities, the clusters of highly connected entities, etc. Such high level tasks often consist of a series of low level tasks, in particular when dealing with large and complex graphs.

The analysis of graphs is often supported by their visual presentations. In this respect, graph visualization research concentrates on the development of effective graph layouts and visual mappings. The visualization of large graphs is accompanied by effective interaction techniques, in particular, in cases when the whole graph is too complex or large to be visualized in one static view. The interaction alone may not be sufficient to accomplish certain analytical tasks. Therefore, algorithmic support — such as machine learning, or graph analysis algorithms — needs to be supported in interactive visualization systems. Such integrated visual analysis of large data sets is the main focus of the research field called Visual Analytics, which evolved from Information Visualization and Scientific Visualization. It has effectively started to grow after the publication of the seminal
In this state-of-the-art report, we provide a systematic overview of the main approaches in each of the three aspects of visual graph analysis of graphs applicable across various domains. Owing to the broad scope of the article, we present the main features (strengths and weaknesses) of the techniques as far as they were discussed by the authors of the papers or were mentioned in evaluations. In particular cases, we point out the readers to evaluation papers for more details.

2. Basic Graph Definition and Preprocessing Techniques

In this section, we recall fundamental graph definitions as well as approaches for graph preprocessing useful for subsequent graph visualization.

2.1. Definitions

Graphs are a prominent data structure within Visual Analytics and related research fields. Often, graphs are applied for describing relationships between entities. A graph refers to a set of vertices (nodes) and a set of edges (i.e., links) that connect pairs of vertices. It is a pair $G = (V,E)$, where elements of $V$ are vertices and elements of $E$ are edges [Die05]. Furthermore, attributes can be attached to vertices and edges, e.g., to denote their type, size, or some other application related information.
Graphs are often classified into undirected and directed [HMM00]. For a directed graph (resp. undirected), the edge vertices \( e = (v_1, v_2) \) are ordered (resp. unordered). A graph containing both directed and undirected edges is called mixed.

A path of length \( s \) in \( G \) is a sequence of connected vertices \( \text{path}_G(a_1, a_2, \ldots, a_s) \) where \( a_i \in V \) and \( (a_i, a_{i+1}) \in E \). A cycle is a closed path with \( a_1 = a_s \). A tree is a connected undirected graph without cycles [Die05]. A Tree \( T \) is called rooted when one vertex \( r \) is distinguished as a so-called root node: \( T = (V, E, r) \). Such trees are often treated as hierarchies, where the length of the path to the root denotes the level of a vertex in the hierarchy. However, formally, a hierarchy is a directed acyclic graph so, in a formal hierarchy, a node can have several paths to the root node. In this survey, we use the term hierarchy as synonym to “rooted tree”. Note that a connected graph can be transformed to a tree by removing edges causing cycles while keeping the graph connected.

In graph theory literature, a directed graph with weighted edges is also called a network. In information visualization, the term network is often used in a broader sense denoting a graph with attributes associated with vertices and edges.

An additional graph category are so-called compound graphs. A compound graph \( C = (G, T) \) is defined as a graph \( G = (V, E_G) \) and a rooted tree \( T = (V, E_T, r) \) that share the same set of vertices, such as:

\[
\forall e = (v_1, v_2) \in E_G, v_1 \notin \text{path}_T(r, v_2) \text{ and } v_2 \notin \text{path}_T(r, v_1)
\]

Relationships between vertices are expressed by \( T \): vertices sharing a common parent in \( T \) belong to the same “group”. When two vertices sharing a common parent are connected in \( G \), they share a generic relationship. Many other kinds of relationships can be expressed including hierarchic and cross-group.

Compound graphs can be created by successive aggregation (or clustering) of graph vertices in a bottom-up approach. This operation usually involves creating new nodes as group/cluster parents. In this case, vertices (and implicitly, also edges) of the original graph are aggregated (i.e., added as children of the group parent), thereby creating constructed meta-nodes or super-nodes. The attributes of the meta-nodes can be calculated from the attributes of the merged nodes. Similarly, edges between meta-nodes are aggregated into meta-edges and their attributes can be calculated from the original edges. Compound graphs which are constructed in this way are also referred to as aggregated graphs. The list of operations that can be performed on such graphs is dependent on the particular application and graph type.

Graphs may also evolve over time, thereby forming dynamic graphs (i.e., time-dependent graphs) in contrast to static graphs. Time-dependent changes may affect the attributes of nodes and edges, the graph structure, or both. Figure 3 summarizes the graph classification presented above.

Furthermore, graphs may be distinguished according to their topological properties. There exists a variety of literature on graph theory (e.g., [Die05]) which focuses on graph terminology, classification, and algorithmic graph analysis. In the following, we mention only the most relevant terminology used later in this report. Basic graph properties include the number of nodes, graph density, and connectivity. Properties are often taken into account (or are a prerequisite) for certain visualization techniques. These properties often heavily influence which visualization methods can be used or fall short, with respect to readability and performance. For example, the increasing number of nodes, higher graph density, or both pose a scalability problem in visualization owing to limited display space and human perception capabilities.

The number of nodes (i.e., graph order) is often referred to as graph size \(|V|\). Graph density is the number of edges relative to the maximum potential number of edges \( D = \frac{2|E|}{|V|^2} \). Sparse graphs have around \( O(|V|) \) < \( |E| \) << \( O(|V|^2) \) edges, while dense graphs show density values close to one. Graphs with the maximum number of edges are called complete graphs. A clique is a subset of a graph that is fully connected.

According to the graph size, graphs are often referred to, e.g., as small or large. The definition of large graphs is however not standardized. Often graphs with thousands, hundreds of thousands or millions of nodes are called large. However, not only the number of nodes determines the notion of a “large” graph. Graph density and connectivity also play an important role for the notion of a “large” graph. From the visualization point of view, “large” graphs usually lead to cluttered displays. In algorithmic analysis, “large” graphs refer to long computational times or memory footprint larger than the available RAM size. A discussion about the influence of graph size and density on visualization and construction of graphs for testing visualizations according to these parameters is provided in [Mel06].

Several special graph structures appear often in real-
world cases, and dedicated visualization methods have been developed for these [ACJM03, vHW08, JHGH08, MJW*09]. For example, social networks usually exhibit a structure called small world network: the typical distance between two nodes grows proportionally to \( \log |V| \). Scale-free networks, e.g., protein networks or certain types of social networks have a degree distribution that follow approximately the power law. Bipartite graphs are graphs whose nodes form two disjoint sets: \( V_1, V_2 \) with \( V_1 \cup V_2 = V \) and \( V_1 \cap V_2 = \emptyset \), such that: \( \forall e = (v_1, v_2) \in E, v_1 \in V_1 \) and \( v_2 \in V_2 \).

### 2.2. Algorithmic Graph Preprocessing

In graph visualization, algorithmic graph preprocessing often includes graph simplification to reduce the size, while maintaining the main graph structure. Also pre-processing of graph properties can be used for graph visualization (in algorithms for positioning of nodes and edges) or highlighting of interesting parts of the graph. The modified graph is used then for an easier visual inspection as large and complex graphs are difficult to understand even using advanced node and edge positioning algorithms (layouts). Such preprocessing steps can usually be performed automatically without user interaction. There are two main approaches to graph reduction: graph filtering and graph aggregation.

**Graph filtering** There are two types of filtering: stochastic and deterministic. Stochastic filtering is mainly based on random selection of nodes and edges from the original graph. These methods are compared in [LF06]. Deterministic filtering uses, as its name suggests, a deterministic algorithm for the selection of the nodes/edges to be removed. This filtering can be based on node/edge attributes, on topologic values such as betweenness centrality, or other graph properties. For example, filtering based on edge-betweenness centrality can be used for removal of less important edges while keeping the underlying structure (connectedness and other features such as cliques) of the graph [JHGH08] (see Figure 4).

**Graph aggregation** In this approach, nodes and edges are merged to single nodes and edges, thereby reducing the size of the graph and revealing relationships between groups of nodes. Graph aggregation can be repeated multiple times, creating a hierarchical graph, which is a special kind of compound graph. There are various ways of aggregating a graph, including using predefined node hierarchies, or aggregation according to node attributes, or according to the node clusters [EDG*08, BDL*10], to name a few. Figure 5 (top) shows an example aggregation schema with several aggregation levels. The highlighted rectangle shows the corresponding data in each aggregation level. Figure 5 (bottom) shows the original and aggregated data in a matrix visualization.

![Graph Aggregation](image)

Figure 5: Graph aggregation for multi-scale graph visualization [EDG*08]. ©2008 IEEE. Top: Graph aggregation schema showing several levels of aggregation. Darker rectangles show the corresponding data areas in the aggregation. Bottom: Example of graph aggregation using a matrix visualization.

### 3. Visual Representations of Graphs

Visualization is one of the main means of exploratory graph analysis. It includes the development of appropriate types of visual representations (e.g., matrix or node-link diagrams), efficient placement of graph elements on the screen and ef-
cient visual attribute mappings (design of graph elements for improved readability of the drawing).

In computer-created graph visualization, several so-called aesthetic criteria are taken into consideration. They are usually implemented as objective functions to optimize in layout algorithms. The standard criteria include minimizing the number of crossings, minimizing the total drawing area, maximizing symmetries and many more related to particular types of graphs and edge drawing styles [Pur97, DBETT99, BBD09]. Recently, Beck et al. [BRSG07] extended previous works to focus on both static and dynamic graphs irrespective of their graphic representations (including also matrix representations in addition to node-link diagrams). They consider three groups of criteria: general, dynamic and aesthetic scalability.

- The general criteria include reduction of visual clutter, reduction of spatial misunderstanding resulting from spatial closeness, maximization of spatial matching of items for following paths and maximization of space efficiency.
- For dynamic graphs, the following criteria are desired: maximization of display stability between time points, reduction of cognitive load when analyzing time dynamics, minimization of temporal aliases mainly owing to positioning of different nodes in the same place in two time periods.
- Aesthetic scalability criteria refer to graph readability for larger graphs, i.e., scalability in number of vertices (i.e., increasing graph order), scalability in number of edges (i.e., increasing graph density), and scalability in number of graphs, in particular with increasing number of time steps for which graph data is given.

All these criteria are important but they cannot be simultaneously optimized and are not sufficient to design a good layout which is usually data and task dependent. Therefore, exploratory graph visualization requires more than one layout algorithm to reveal the several perspectives on relationships between nodes.

In this section, we describe the main graph visualization techniques following the graph classification from Section 2. We introduce techniques for static and time-varying graphs. In each part, techniques for hierarchies, generic directed and undirected graphs, and compound graphs are presented. We discuss different ways of visual graph representations and designs of graph drawings.

3.1. Visual Representations of Static Graphs

The visualization of static graphs has received much attention in the Information Visualization community. The section starts with trees that are simpler than general graphs.

3.1.1. Trees

Techniques for displaying trees can be divided into three main groups: space-filling, node-link based, and hybrid (see Figure 6). There have been several studies comparing the different ways of visualizing trees [SCGM00, BN01, vHvW02, Kob04, AK07]. A very useful visual overview of tree visualization has been provided in the poster [JS10]. It is difficult to unify these results as they differ significantly. Recently, Ziemkiewicz and Kosara have shown that the effectiveness of the visualization technique depends not only on the task to be solved, but also on the formulation of the task assignment, i.e., if it reflects a containment or a level metaphor [ZK08].

Node-link techniques These approaches use links between items to depict their relationship. Layout algorithms controlled by optimization criteria or the node positions. Many layout algorithms have been proposed to date in the Graph Drawing community. They include layered, radial or balloon layouts in 2D [HMM00]. Cone trees [RM91] in 3D, point based trees [SSH09], nature inspired Phyllotrees [NCA06], or hyperbolic layouts [LRP95, Mun97, AH98] (see Figure 7). Most of these classic tree layout algorithms have a linear complexity in time and memory so the layout computation is scalable. However, the node-link representation by design leaves significant background space empty and thereby may encounter scalability problems when applied to larger graphs. For the visualization of node attributes, specialized techniques for multi-dimensional data visualization such as glyphs, radial or parallel plots have been used.

Space filling techniques These techniques try to use the full area of the display to present the hierarchy. Instead of employing links for representing node relationships, the spatial positions of nodes are employed, using either closeness
or enclosure. They are mainly applied to visualization of hierarchic partitions of sets of data items, for instance files in a file system. Area size can be used to encode quantitative attributes of nodes, such as file size. Additionally, color and height can represent additional data attributes. In case more complex additional information needs to be displayed, specialized data presentations can be placed in the child nodes such as icons, parallel coordinate diagrams, etc.

Space-filling techniques can be categorized by the placement strategy employed into enclosure, adjacency and crossing (see Figure 8).

- **Enclosures** These techniques recursively layout child nodes within the area of their parent nodes. The most prominent examples are Treemaps – rectangular shapes recursively subdividing rectangular display space according to the underlying hierarchy, introduced by Shneiderman [Shn92] (so called slice-and-dice algorithm). Variants include Voronoi tessellations [BDL05] or bubble layouts [Bed01]. Other types, such as elliptic [OCNF09] or circular shapes have been proposed, but they do not lead to fully space filling visualizations.

  The main advantage of enclosures is the very good usage of the available space, as the child nodes do not need extra space owing to the overlap with the parent nodes. The disadvantage is that the overlapping of the parent nodes may also lead to a more difficult distinction of the hierarchy structure by the user, as it is rather implicitly encoded. For Treemaps, several advanced layout techniques have been developed including ordered (i.e., pivot-based) [BSW02], squarified [BHvW99], and spiral [TS07] Treemap layouts. For example, squarified Treemaps aim at generating subrectangles of square-like aspect ratios, supporting easier comparison of sizes and presentation of additional diagrams or other elements within the rectangles. According to Tu and Shen [TS07], the slice-and-dice algorithm leads to high aspect ratios with good readability. Strip, pivot-based and spiral techniques have medium aspect ratios with medium readability. Squarified Treemaps have very good (low) aspect ratios but low readability. In order to better distinguish the hierarchical structure, cushion Treemaps [vWvdW99] apply shading of the shapes. Treemaps that reflect the geographic distribution of the hierarchical data were presented in [WD08].

- **Adjacency** In contrast to Treemaps, adjacency-based techniques do not overlap the parent nodes by child nodes, but represent the node relationships by placing the child nodes next to their parent nodes. The placement can be in circular layers, such as in the Sunburst method [SZ00], or on linear layers, yielding so-called icicle plots. The advantage of this visualization is that the parent nodes are not overlapped by their child nodes and therefore, their attributes can be more easily displayed and analyzed. However, this visualization is not as dense as squarified Treemaps.

- **Crossings** The crossing method places child nodes across the parent node, thereby only partially overlapping the parent. The Beamtree method [vHvW02] improves over the classic Treemaps when the hierarchical structure may be difficult to visually assess, while still being more space efficient than the adjacency techniques. The main drawback of this technique is that users are unfamiliar with this approach and that it is often less readable than other methods.

Hybrid approaches These approaches combine node-link diagrams with Treemaps: a part of the hierarchy is displayed in a Treemap and the rest as a node-link diagram (see Figure 6c). They present the data in a flexible space-efficient
3.1.2. Directed and Undirected Graphs

Techniques for displaying general graphs can be divided into three main groups: node-link based, matrix-based, and hybrid (see Figure 9). We discuss these in more detail below. In addition, there are specialized graph drawing techniques, which use new graph visualization techniques. Two main examples are: graph splatting and graph maps. The first one forms graphs as two-dimensional scalar fields [vLdL03]. The second one visualizes graphs as maps [GHK10], where the relationships between nodes are represented as adjacency between neighboring areas (nodes). Both approaches create an approximate representation of a graph.

A comparison of node-link and matrix techniques is presented by Ghoniem et al. [GFC04]. According to the study, the advantages of node-link diagrams are their intuitiveness, compactness, and better suitability for path following tasks. They are more effective for smaller and sparse graphs. The Matrix representation inherently does not have edge crossings and node overlapping problems, and is thereby suitable also for dense graphs. When using appropriate node ordering, they can easily reveal dense substructures in the graph. However, they also suffer from scalability in limited display spaces, especially for very large graphs. In visual graph analysis, graph layout and matrix ordering influence the effectiveness of these representations. These issues are therefore in the core of graph visualization research.

![Figure 9: Three types of general graph visualization techniques: a) Node-link diagram, b) adjacency matrix, c) hybrid. From [HFM07], ©2007 IEEE.](image)

Node-link representations. The main challenge is the layout (i.e., the placement of the nodes) so that graph readability and certain notions of graph aesthetics are supported (see Figure 10). Typical requirements include that the nodes do not overlap, the number of edge crossings is minimized, edge length is homogeneous, and in general, graph substructures are easily recognizable. This problem is intensively studied in the graph drawing community. Given these aesthetic goals and constraints, the aim is to find algorithms that efficiently provide good solutions.

Note that a specific group of graphs are graphs with geographic reference, such as transportation graphs. In this case, the nodes and possibly also edges of the graph have an inherent geographic location, which needs to be taken into consideration in their graph presentation. Therefore, a specific graph layout algorithm is not needed for determining the position of each node on the screen. However, the fixed node position may exacerbate graph readability problems, such as crossings and long edges. Visualization of geographic data is a special research field, which we do not address here in detail.

When no position is inherently associated with vertices, a graph layout algorithm is required. The graph layout research field is very large, and an extensive survey of proposed techniques is beyond the scope of this report. The latest survey from Herman et al. dates from 2000 [HMM00] and several new algorithms have appeared since then. The related work part in [AAM07, MM08] as well as the comparison in [HJ07] nicely summarize many current techniques. In our report, we classify the techniques according to the type of node placement.

- **Force-based layouts** These techniques are based on a simulation of mechanical laws by assigning repulsive forces between nodes and attraction forces between endpoints of edges. Several forces have been described in the literature to achieve different properties of the layout. The seminal work of Eades uses an electric force between charged particles to model node repulsion and spring forces between the link endpoints to model edge attraction [Ead84]. Fruchterman and Reingold [FR91] have then improved the distribution of nodes by adaptation of the force models and Noack has further improved it with a more flexible set of force functions to achieve either a good space density or a good clustering of nodes. Kamada and Kawai [KK89] try to layout nodes such as the Euclidean distance between the nodes is proportional to the graph-theoretical distance. This family of layouts, however, does not scale well to graphs of thousands of nodes or more, due to their complexity. Therefore, improvements have been proposed. For instance, faster calculation of forces using an efficient GPU implementation [GHGH09], or using heuristics [FLM95].

- **Constraint-based layouts** This family of layouts extends the force-directed approach with constraints on node position. These constraints include horizontal and vertical alignment of nodes, non-overlapping nodes, edge direction or closeness of grouped nodes [DMW09a]. An example are orthogonal layouts, where the edges are only composed of straight vertical and horizontal lines. These
Multi-scale approaches These techniques rely on a hierarchical decomposition of a graph into simpler nested sub-graphs. They first layout the coarser graph and then include more nodes level by level. Exemplary works include [GK01, KCHO2, HJ05, FT07, MM08] (see Figure 10). These methods are typically much faster than traditional force-directed methods. They can be differentiated according to the technique used for creating the node hierarchy, and the layout of the resulting layers. For example, [MM08] employs node clustering and subsequent positioning of the nodes along space filling curves.

Layered layouts These approaches, also called “hierarchical layouts”, place nodes of the graph on parallel horizontal layers [GKNV93]. They are mainly used for directed graphs and are based on the Sugiyama approach [STT81]. It works in four phases: (1) cycle removal, (2) assignment of nodes to layers, (3) reduction of edge crossings and (4) assignment of coordinates to nodes. Improvements to these layouts, specifically for cyclic graphs, position all nodes of a cycle within one level; examples include the Dig-Cola layout [DK05] and Cyclic Leveling [BBBL09] (see Figure 10b). This algorithm and its variants are quite fast in practice and standard implementations such as [GKNV93] can easily layout several thousands of nodes in seconds.

Non-Standard Layouts Other approaches exist that combine the previous techniques or use completely alternative approaches to graph layouts. Projection of a node layout from high-dimensional to two-dimensional space has been proposed in [HK02]; although it is very fast in practice, the quality of the layout is very sensitive to the structure of the graph. For example, it is very effective for meshes and not effective at all for trees. LGL [ADWM04] first simplifies the graph by computing a spanning-tree; it then computes the layout iteratively in depth order using a force-directed layout. LGL is able to scale to very large graphs (billions of vertices) thanks to the initial decomposition. It is very effective for quasi-trees but has not been thoroughly studied for other kinds of graphs; its results are very sensitive to the spanning-tree computation: choosing different spanning trees will result to quite different layouts for the same graph. The ISOM method [Mey98] applies the Self-Organizing Map algorithm [Koh01] for finding a suitable graph layout. As an alternative to costly layout computation, a graph layout visualization based on the semantics of the graph (on node labels) was presented in [SA06]. Semantically identical nodes (e.g., with the same labels) can be placed in boxes using standard layout algorithms (e.g., force-directed) (see Figure 11) or in layers using their importance for assigning the position within layers [GOB*10]. Furthermore, attributes associated with graph vertices can be used directly to specify the position of these vertices, as with scatter-plots [SA06, BCD*10]: the layout computation is then straightforward and very fast.

Comparison of graph layouts A recent comparison of the readability of graph layouts using eye-tracking [Hua07, PSD09] has shown that force directed layouts outperform orthogonal and layered layouts on various user tasks. Another comparison of advantages and disadvantages of numerous current layouts was published by Hachul and Jünger [HJ07]. They compare the graph drawing outputs according to various criteria finding that the HDE layout [HK02] is very fast but frequently produces layouts with many overlapping edges. In contrast, FM³ [HJ05] creates pleasing layouts in reasonable time. Both algorithms together with

![Multi-level graph layouts](image1)

![Dig-Cola layout](image2)

Figure 10: Graph layout examples. a) A comparison of multi-level graph layouts GRIP, FM³ and Topolayout [AAM07]. ©2007 IEEE. b) Layered layout of cyclic directed graph [DK05]. ©2005 IEEE.

Figure 11: Graph visualization using data semantics [SA06]. ©2006 IEEE.

submitted to COMPUTER GRAPHICS Forum (1/2011).
GRIP [GK01] scale well with graph size. A comparison of user-produced vs. automatically generated layouts [vHR08, DLF’09] found also that the results of physics-based algorithms, such as force-directed layouts, were preferred by the users.

Design of graph drawing The above mentioned techniques cover graph layout. In addition to specific layouts, occlusion and readability of the display can be improved by edge-bundling [Hol06, CZQ’08, TE10, LBA10] (see Figure 12) and the removal of node overlap [GH09, IAG’09]. Drawing of node-link diagrams also includes a suitable design of edge and node drawing primitives. For directed graphs, the representation of edge directions is of importance. There are multiple design possibilities including usage of arrows, color transitions (from color A to color B), thickness transitions (from thick to narrow), curves, and animated textures [TK08, HvW09, BBG’09]. These options may also be combined. A comparison of graph drawing different ways to represent edges was presented in [HvW09]. It shows that arrows, although popular and widely used, do not perform as well as color and thickness transitions. Graph nodes and edges often have associated attributes that are included in the analysis. This study did not concentrate on attributed edges. For such edge attributes, in particular edge weight, coloring of edges or edge thickness can be employed. For the visualization of node attributes, a visualization of multivariate data items (e.g., glyphs or radial plots) is employed. Various possibilities of graph designs can be found in [Kre09].

![Figure 12: The use of edge bundling for improving graph readability. a) original graph b) graph with edge bundling. [Hol06]]. ©2006 IEEE.

Visualization of multiple graph connected components For the visualization of multiple components, first a layout for each individual connected component is calculated and then a specific placement of these components on the screen is performed. The most widely used placement method is called packing. It lays out the components so that they do not overlap and are space efficient. Dogrusoz [Dog02] compares several two-dimensional packing algorithms for graphs which use representation of graphs by their bounding rectangles. They include strip packing, tiling and alternate-bisection. The polyomino algorithm of Freivalds et al. [FDK02] uses a special representation of the graph objects, which substantially reduces the unused display space in comparison to rectangular shapes. Goehlendorf et al. [GKS07] introduce new quality measures to evaluate a two-dimensional placement which yields more compact layouts than the previously mentioned approaches.

Matrix Representation These techniques visualize the adjacency matrix of a given graph, where edge attributes are encoded in the matrix cells. They can display both directed and undirected graphs, where the latter leads to a symmetric matrix. The advantage of this representation with respect to the node-link representation is the non-overlapping display of graph edges, and the readability of the graph especially for larger and denser graphs. The disadvantage is an increased difficulty for users to follow paths, and a possible unfamiliarity of matrices to the users. In a matrix visualization, the ordering of rows/columns plays an important role: similar to layout for the node-link representation. Different strategies to order the matrix can be employed (see Figure 13). Prescientous reordering can reveal clusters in the graph and other patterns. For a discussion of these, we refer to [MML07, DPS02, HF06, EDG’08]. Although matrices are suitable for larger graphs, they also suffer from scalability issues as they use linear order of nodes along the matrix rows/columns. Therefore, interaction techniques and aggregated displays have been proposed [vH03, AvH04, HF06, EDG’08, vHSD09] (see also Sections 4 and 5).

![Figure 13: Examples of matrix reordering on graph presentation. a) Using HDE algorithm. b) Using NNTSP reordering. From [EDG’08]]. ©2008 IEEE.

Combination of matrix and node-link approach Techniques using a combination of the two previous approaches aim at overcoming their limitations by focusing on their strengths. Three main approaches exist (see Figure 14).

- Multiple synchronized views These techniques link the matrix and node-link representation [HF06]. Both views show the same data and are synchronized during exploration. Thereby, the user can concentrate on whatever view is more suitable for the current task.
• **Matrix with link overlay** The Matlink [HF07] approach enhances matrix visualization with links at the border of the matrix (connecting the nodes). Using link highlighting, the paths can be easily spotted in the Matlink view and at the same time, the advantages of the matrix representation are retained.

• **Partial matrix and node-link representation** There are two main approaches. Firstly, Nodetrix [HFM07] combines both representations in one view, where node-link diagrams display the overall graph structure of the network, and adjacency matrices show communities. The work also discusses three ways of link display for this setting: aggregated links, underlying links, and underlying links with full size (see Figure 15). These forms can be also used for attributed links. Secondly, layered graphs (directed acyclic graphs) can be represented by so-called “quilts”. They arrange nodes in a matrix-like form and connect them with orthogonal edges. In this way, a clear view of the graph is created [WBS”08, BDF”10].

### 3.1.3. Compound Graphs

Literature on visualization of graphs with hierarchic structure is relatively rare. We identify three main approaches.

**Node-link graph visualization techniques** These use node-link diagrams for the lowest hierarchy level and then use “bubbles” (enclosures) for various hierarchy levels. Examples include TugGraph [AMA09] and Grouse-Flocks [AMA08]. The advantage of this method is its intuitiveness. However, for large graphs with many links, this view gets easily overcrowded (see Figure 16 a). The edge overplotting problem can be partially solved by edge bundling [Hol06] (see Figure 12). Alternatively, only links between merged nodes can be drawn (see Figure 16 c).

**Treemap-based** A treemap visualization of the node hierarchy uses overlaid links between nodes [FWD”03] (see Figure 16b). This approach may suffer from strong overplotting in case of many links between nodes of the hierarchy. Therefore, edge bundling is advised to improve the readability of the display [Hol06] (see Figure 12). Similarly, also one-dimensional treemaps with links between nodes, so called ArcTrees [BDJ05] can be employed (see Figure 16d), but these do not scale well for large hierarchies.

**Matrix view with links** These visualizations combine the generic node relationship visualization with a tree-based visualization of the hierarchic node relationships. This is an analogy to MatLink [HF07]. This view is very clear, however, it may be difficult to understand the compound relationships between nodes (see Figure 16e).

### 3.2. Visual Representation of Dynamic Graphs

In this section, we discuss two categories of visual display of the time changes on graph elements: Using animation and
using static displays. Animated displays usually employ or enhance static visualization techniques such as presented in Section 3.1. Animation is a natural way of conveying the change of the data over time. However, its effectiveness is limited by human perception capabilities. Usually, users are only able to recognize and remember larger changes in the data. Therefore, highlighting of graph changes is used. It allows for more effective spotting of differences between two successive time points [APP10]. The static view is preferred for more detailed analysis of data changes. Static views that also incorporate the time-dimension of the data are more complex. In the following, we categorize the visualization techniques according to the type of data changes captured into those that affect only data attributes, and those that affect also data relationships. Please note that visual analysis of changes in dynamic graphs is related to comparing graphs. Graph comparison is discussed in Section 5.2.

3.2.1. Trees

![Figure 17: Visualization of time-dependent trees. (a) Time line tree [BBD08], ©2008 ACM. (b) Time series in the treemap nodes [DHKS05], ©2005 IEEE. (c) Animated hierarchic circular icicle plots [TS08a], ©2008 IEEE.](image)

For the visualization of dynamic trees with only data attribute changes, either treemaps with time series in the leaf nodes [DHKS05, SKM06] or the so called Timeline Trees [BBD08] can be used (see Figure 17 a and b). Timeline trees show the hierarchy on one side and the time sequences on the other side of the view. The treemap repre-
sentation directly shows the hierarchic structure and time-variation in one combined view. This allows for an easy comparison of the time-developments across the hierarchy. However, the comparison is affected by different node sizes and difficult for small nodes. Therefore, a specific treemap layout preserving the aspect ratio has been developed [DHKS05, SKM06]. Timeline Trees assign the same space to all nodes. The vertical positioning of time lines allows for very good comparison of the values at the same time points. The separation of the time dimension from the hierarchic structure, however, complicates the comparison of tree branches.

For visualization of dynamic data with structural changes, animated views are used. Card et al [CSP∗06] have used and extension of DOI Trees [CN02, HC04] to visualize the changes of an administration over time: A time-slider is used to control the visualized time-span. Animated graphs (see Section 3.2.2) can be employed in general. In particular, the layouts based on the Sugiyama approach [GBP04] are suitable. Alternatively, animated treemaps [GF01, TS07] or icicle/circular plots [TS08a] can be used (see Figure 17 c). When choosing the graph layout, the layout stability needs to be taken into consideration. E.g., in the treemap representations, the spiral layout [TS07] achieves a high continuity with high stability of the layout. Strip and pivot-by-middle layouts have also been shown to have higher layout suitability [BSW02]. All these layouts are preferable in spite of their higher aspect ratios in comparison to the squarified treemap. Furthermore, dynamic Voronoi treemaps [SFL10] offer both good aspect ratios and stable layouts for displaying dynamic data. Alternatively, Tu and Shen [TS07] propose also static comparison of two time points in a treemap visualization (called contrast treemap).

3.2.2. Directed and Undirected Graphs

For attribute changes only, techniques for visualization of static graphs can be combined with visualizations of individual time dependent data items (e.g., color charts [SLN05]) are used (see Figure 18a). The advantage of this approach is the large number of the available graph layouts.

In case of structural changes, time-dependent graph layouts (animated graphs) need to be employed [CBTT95, Nor96, DGK01, EHK∗03, KG06]. In animated graph visualization (in analogy to animated tree visualization), a stable graph layout, which changes minimally, is of essence. A stable graph layout preserves the mental map of the user. It enables the user to follow changes on the screen [ELMS01, DGK01] and thereby it facilitates the analysis of graph changes. In laying out dynamic graphs, there is a large difference between strategies for drawing graphs with known histories and those that need to be adjusted in real-time depending on new data streams. A paper of Frishman and Tal [FT08] addresses this particular issue by proposing an online algorithm for dynamic layout implemented on the GPU, thereby accelerating the layout computation (see Figure 18b).

Instead of animation, Brandes and Corman [BC03] use the third dimension to show the evolution on time. GraphDice [BCD∗10] uses interaction to switch between projections where time can be mapped to one dimension.

3.2.3. Compound Graphs

There are only few techniques that visualize time-varying compound graphs. They employ either animation or static data representations.

Kumar et al. [KG06] present a specific layout for animation of a node-link diagram with transparent “bubbles” for the hierarchic grouping of nodes (see Figure 19a). Frishman and Tal [FT04] present a layout which focuses on maintaining the clustered structure during the animation. The groups of nodes are displayed using bounding boxes around the groups. Reitz et al. [RPD09] use dynamic graph layouts for showing areas of interest in dynamic compound graphs.
A static approach to visualization of dynamic compound digraphs using TimeArcTrees was presented by Greilich et al. [GBD09] (see Figure 19b). They show a sequence of node-link diagrams with horizontal node alignment in a single view, thereby supporting their direct comparison. TimeRadarTrees [BD08] use radial tree layouts for the hierarchy and a sequence of circle segments for representation of the temporal change of the structure (edges) of the Digraph (see Figure 19c). This view easily gets complex for larger graphs.

4. User Interaction in Graph Visualization

Interaction helps users solving tasks connected to exploration of graphs. These tasks can be of different nature such as topology-based or attribute-based [LPS06]. Topology-based tasks include finding adjacent nodes, or determining connections between nodes. Attribute-based tasks include, e.g., searching for nodes with specific values, and finding edges of certain types. For each task, one or more interaction techniques can be employed. Standard interaction techniques such as zooming, panning, or brushing and linking [CMS99, War00] are commonly used in graph visualization. In addition, specialized techniques have been developed for interactive visual graph navigation and exploration.

Interaction and exploration are deeply inter-related. Some graph analysis systems such as Pajek [dMB05] claim to support exploratory graph analysis by chaining complex operations on graphs without showing the intermediary results. However, Ahlberg et al. describe interactions and more specifically dynamic queries [AWS92] as required to truly achieve exploration. The main reason is cognitive: exploring requires several hypothesis to be maintained in short-term memory which is very limited in capacity. Planning complex operations without feedback or using a textual syntax consumes all the short-term memory and exploration becomes impossible from short-term memory alone. Therefore, providing interactions with immediate feedback for the most frequent operations supports exploration. Other less frequent operation could still be done using more complex mechanisms, as explained in the next section on graph analysis.

The categorization of interaction techniques can be based on various criteria such as task, user intention [YKSR07] or user action [EF09]. These criteria are interrelated. For example, one task may include performing several actions, or one task may correspond to several user intentions. Moreover, one user intention can be achieved by several user actions or, vice versa, an action can suit several intentions.

We categorize interaction techniques according to stages in the Information Visualization reference model of Card et al. [CR98, CMS99] and user actions. The reference model has three stages: data, visual form (a.k.a. visual abstraction) and view. The classification criterion is whether the user action affects the data (the selection of the displayed data or
the data values), the visual display of the data (visual parameters or visual representation), or the view. Data, visualization and view manipulation can be used for interactive data exploration and navigation. This categorization follows the idea of Elmqvist and Fekete [EF09] and Bertini and Lalanne [BL09]. Please note that these three types of interaction are sometimes closely connected. For example, data manipulation may automatically lead to changes of visual parameters (e.g., data filtering can influence the graph layout, or zooming can be combined with data filtering forming a type of semantic zooming).

4.1. View Interaction
Panning and Zooming Panning and zooming allow to navigate in any direction and change the zoom-level in the view. For node-link diagrams, a specific type of panning (guided panning) has been proposed. It allows to navigate along edges of a selected node and thereby to explore the structure of the graph. It can be combined with automatic zooming on the edge and distortion of end-node position closer to the currently selected node [MCH∗09].

Magic Lenses Owing to the limited display space, showing the whole data set may lead to strong overplotting or very small (up to, unreadable) data items. Magic Lenses [BSP∗93], including distortion techniques change the representation or allocate more space to items in focused areas and thereby, improve the readability of the data of interest. They are used both for node-link and space filling graph visualization techniques. The changes can concentrate either on one area or on multiple areas of the screen. For geometric changes, the technique is called fisheye views. Interactive selection of the focus area helps to explore different parts of the data in more detail.

• Single focus Graphical fisheye views were introduced in [SB92]. So called edge lenses resolve strong overlaps of edges in the view. They displace the edges to a larger area [WCG03] (see Figure 20). This approach is especially useful for geographic-based graphs, where node repositioning is not desired and therefore, cannot help to solve edge overlap. Another approach uses filtering of interesting edges in a specified area, or moving neighbor nodes closer to a selected node relying on the graph structure [MCH∗09]. This type of node position change can be combined with geometric view distortion [TAvHS06] (see Figure 21). In node-link visualization of hierarchies, a degree-of-interest function can be used for allocating more area to more interesting parts of the tree, e.g., in DOTrees [CG02,HC04].

None-geometric magic lenses include Excentric Labels and Color Lenses. Excentric Labels [FP99,BRL09] show labels or other statistics for items contained in dense focus regions (nodes or matrix cells). The information is displayed outside the focus region with connectors linking the nodes/cells to their related label. Color Lenses [EDF10] dynamically adapt the color range of items inside the focus region to better use the screen color range when mapping values with a very large dynamic to the color of nodes or matrix cells.

• Multiple foci Multiple foci distort several view areas at the same time. It is useful for comparing various parts of the display or focusing on several items that are spread across the view. In node-link diagrams either magnification of the areas of interest [SZG∗96,TS99] or space folding (shrinking of area out of focus) can be used [MGT∗03,ERHF09] (see Figure 28 bottom right). For treemaps, the so-called balloon focus can be used for enlarging multiple items in a treemap [TS08b]. This approach keeps the form of other areas keeping relative position of items unchanged (see Figure 22).
4.2. Visual Abstraction Interaction

In these approaches, the change of the visual presentation of the data concerns adjusting the type of visual presentation and its parameters.

Most of the graph visualization system provide standard dialog boxes and widgets to change the visual abstraction parameters, including the layout technique and its various parameters. Currently, very few systems allow the interactive manipulation of layout parameters, except using indirect manipulation such as sliders, list boxes, radio buttons and check boxes. Rich visualization systems provide a large number of these indirect manipulation widgets which use an important amount of the screen real-estate and force users to search for the right widget by reading their labels and trying to make sense of them, which can be quite long and tedious. This is why several research work is devoted to providing more direct mechanism to change the parameters.

4.2.1. Changes of Visual Parameters

These techniques affect the parameters of the visual presentation. They include highlighting of items and other techniques.

Highlighting The emphasis of interesting items is a standard interaction technique. Recently, new techniques for highlighting a node and its neighborhood using hotbox and lasso selections were presented in [MJ09].

Brushing & Linking Multiple coordinated views are used to show the data from different perspectives. In these views, changes in one visualization (e.g., highlighting) are automatically transferred to the other views. For example, a matrix view coupled to a hierarchical view of the data can be used to reveal important information in the data [AvH04].

Semantic Zooming Semantic zooming combines zooming with an increasing level of detail. In particular, graph aggregation can be used for gaining a coarser view on a large graph. The semantic zooming increases the level of detail by drilling down to lower levels of aggregation of the original data [EDG08, AvH04].

4.2.2. Changes of Visual Scheme

Changes of the visual scheme cover changing of the type of data visualization either by changing the layout or by changing the visual mapping.

Layout change In node-link diagrams, layout change (adjustment) affects the positions of the data items on the screen (see Section 3). It can be performed by changing of the layout type with automatic recalculation of the new layout, by manual movement of nodes, or by adjusting the layout parameters including automatic readjustment of the layout.

When concentrating on user-defined changes to graph layouts, an approach to easy selection and layout change of nodes and subgraphs was presented in [MJ09]. Furthermore, interactive adjustment of the layout constraints was presented in [DMW09a]. For matrix visualizations, user-driven reordering of matrix representation was described in [HF06].

Change of visual representation The change of the type of data presentation, e.g., from a matrix to a node-link diagram was presented in [ZMC05, HFM07]. This change can affect the whole data view [HFM07] (see Figure 23) or only a part of it [ZMC05, HFM07]. By changing the visual representation, new insights into the data can be reached. In order to be able to follow the changes, smooth animations across transitions should be used.

4.3. Data Interaction

Data-level interaction affects the selection of the data to be displayed, or may change the data values and structure.

Some operations can be done interactively but general graph analysis system provide more sophisticated mechanisms including scripting languages or powerful macro facilities to perform more complex operations.
4.3.1. Data Filtering

These interaction techniques influence which parts of the data set are displayed. The data filtering may follow three paths.

A top down approach This approach starts from the whole graph and then constrains the part of the data set to be visualized by filtering according to criteria or by manual data selection. The disadvantage of this approach is the need to show the whole graph at the beginning, which may require higher computational time for the layout and may lead to occlusions owing to the limited screen size. The advantage is gaining an overview of the graph structure first and then concentrating on interesting parts.

A bottom up approach This approach starts from one selected node [Fur86, AF07, vHP09] and successively shows more nodes/connections on demand. There are two main methods of choosing the additional nodes/edges to be displayed: based on graph structure, or based on a degree-of-interest function. The advantage of this approach is that only the most interesting part of the data set is visualized, however it is difficult to determine the starting point for the exploration and to define the degree-of-interest function. Therefore, we consider these methods in more detail.

- **Navigation based on graph structure.** These techniques reveal/hide that part of the graph that is determined by the connections between nodes. In graphs, neighborhood traversal shows neighbor nodes of a focus node up to a certain level [HB05]. For hierarchies, several traversal methods for have been described in [EF09]. The hierarchy traversal methods include: (1) above traversal, where nodes up to a certain level are displayed; (2) below traversal, where nodes starting from a selected level are displayed; (3) level traversal, where nodes at a certain level are displayed; (4) range traversal, where nodes in a range of levels are shown; and (5) unbalanced traversal, where certain branches of a tree are visible (see Figure 24).

- **Navigation based on a degree of interest function** These methods start from a selected node, and next the edges and nodes of highest interest are shown [Fur86, vHP09]. For the determination of the interesting nodes, a specific degree of interest (DOI) function is used. Depending on the specification of the DOI function, various graph exploration paths can be followed. These DOI functions were used for building specific views on trees (DOITrees,SpaceTree) [CN02, HC04, PGB02]. In the work of Furnas [Fur86], the DOI of a node depends on the distance to the node in focus and the a priori interest in this node (e.g., according to node importance in the network, or node properties). Van Ham and Perer [vHP09] extended this function with user interest (UI), which reflects the current specific exploratory focus of the user.

A middle-out approach This method combines both bottom-up and top-down approaches. It starts with a coarsened graph (middle) and then interactively either reduces or increases the graph coarsening level by hiding visible nodes or showing additional nodes [WMC*09]. For determining
the middle coarsening level and the next interactive steps, graph algorithms are used (see Section 5).

4.3.2. Changes of data values

In these approaches, the change of the displayed data set result from direct data value manipulation. Specifically, the user can change the data values on one level or create/change graph aggregations.

Graph editing The user can interactively delete or add nodes or edges directly in the visual interface. These graph editing actions trigger adjustment of the layout, while still maintaining the layout style and, where reasonable, the current layout topology. Graph editing affects the structural properties of the graph. In particular, the changes can affect specific types of subgraphs (so-called motifs). Automatic identification and highlighting of such structural changes was presented in [vLGRS09].

Interactive graph aggregation For simplification of graphs, graph aggregation is often used. The graph aggregation can be pre-defined, or determined interactively by the user [HF06, AMA08, AMA09]. For example, Grouse-Flocks [AMA08] allows the user to add and remove aggregated nodes on demand (see Figure 25). This allows for variable views on the graph and its structure.

![Figure 25: Interactive editing of a graph hierarchy. a) Creating a new aggregation node by merging of nodes. b) Deleting an aggregation node, thereby revealing the underlying merged nodes. From [AMA08], ©2008 IEEE.](image)

5. Graph Analysis

Algorithmic graph analysis is beneficial during all stages of the visual graph analysis process. Relevant techniques allow, e.g., to reduce a large graph to a smaller graph prior to visualization, to search for specific graph structures of interest, or to find similarities and dissimilarities for generating comparative graph views. In this section, we describe a number of graph analytical approaches.

5.1. Analysis of Graph Structure

In most user tasks, the analysis of the relationships between entities in the graph and the assessment of the global graph structure plays the key role. These tasks may be effectively supported by a combination of algorithmic graph analysis and interactive visualization. The algorithmic methods allow, e.g., to calculate node/edge properties, identify clusters in the graphs, etc., the results of which are visualized interactively. In the following, we summarize the methods according to user tasks starting from more simple to more complex tasks.

Identification of important nodes In networks, some nodes play a specific role owing to their position within the network. For example, so-called hubs and authorities can be identified and visualized in the network, enabling faster analysis of the graph [OPPROG09]. The importance of nodes and edges is measured by derived quantities (i.e., network metrics) such as centrality-based measures [Fre79] and ranking-measures [WS03]. Network metrics can help the analysts to explore networks. Color coding of nodes or edges by metric values, or displaying metrics and networks in multiple linked views (as lists, scatterplots, or parallel coordinates) are used in this respect. They offer the possibility to interactively chose the metrics of interest and to filter/highlighting nodes according to these metrics [CJM04, PS06, BCD10, VMC10].

Analysis of connections between two nodes Besides focusing on single nodes, relations between two nodes can be analyzed, typically by calculation and highlighting of shortest paths between the entities. Usually, such analysis is combined with interactive selection of two entities of interest [HB05, HF07, TK08, GBD09] (see Figure 14b).

![Figure 26: Interactive graph motif search and visualization. From [vLGRS09], ©2009 held by the authors.](image)

Analysis of graph substructures In many applications, specific types of substructures (i.e., motifs) play an important role. For example, in social networks, cliques identify highly connected communities, or feed-forward motifs (substructures in form of a triangle where directed edges...
exist from nodes A to B, A to C and B to C) in biology networks indicate the functional properties of the network [Sch08]. In order to support the substructure analysis, these motifs can be calculated and visualized in the network [MM05, SS05, KSS06, vLGRS09, MJW*09] (see Figure 26). The type of structure can be interactively chosen by the user in order to support various analytical tasks.

**Analysis of graph structure on several aggregation levels** User-defined or data-driven graph aggregation can reveal relationships between groups of entities in a graph. The grouping may be based on categoric node attributes [Wat06], or on a predefined node hierarchy [AMA09]. It can also be user-specified [AMA08], on clustering results based on node properties [PS06], or depend on structural properties of the graph [vLGRS09] (see Figures 5 and 25).

**Identification of the impact of graph changes on the structural properties** In time-dependent graphs, the role of the nodes can change over time, therefore analysis and visualization of topologic properties (e.g., betweenness centrality) of selected nodes has been proposed [PD08, PRB08]. Additionally, when analyzing user-defined changes (in what-if-scenarios) the impact of node or edge deletion/addition on local substructure can be analyzed and highlighted [vLGRS09].

### 5.2. Graph Comparison

One specifically important analytical task is the examination of the similarities and differences between multiple graphs, especially focusing on structural aspects. Usually, structural differences are in the focus. Such difference may be identified by the identical node labels in both graphs, or by graph matching algorithms. After the matching, visualization is employed to explore the differences [AWW09]. There are various types of analysis which we describe next.

**One-to-one node comparison of two graphs** Probably the most common task in graph comparison is the matching of individual nodes from one graph to individual nodes of the second graph. The VisLink visualization approach [CC07] was developed to support this task. It shows both graphs on separate planes in 3D, and draws matching links between corresponding nodes (see Figure 27a). For comparison of hierarchies, a similar approach, based on drawing the two hierarchies in opposite parts of the display and linking of their leaf nodes was proposed in [HvW08] (see Figure 27b). In both cases, the visibility of matching links can be increased by edge bundling.

**One-to-many nodes comparison of two graphs** One-to-many nodes comparison concerns correspondence of one node in one graph to many nodes in another graph. Di Giacomo et al. [GDLP09] developed a system that visualizes these one-to-many connections with low overlapping of links (see Figure 27c).

---

*Figure 27: Visualization of graph comparison. a) One-to-one graph matching [CC07], ©2007 IEEE. b) One-to-one hierarchy matching [HvW08], ©2009 held by the authors. c) One-to-many graph matching [GDLP09], ©2009 Springer-Verlag Berlin Heidelberg.*
Structural differences between two graphs When analyzing structural differences between two graphs, analysts are often interested in identifying which links or parts of the graphs correspond to or differ from the other one. For the analysis of trees, the TreeJuxtaposer system supports to analyze and highlight structural differences between two trees [MGT+03] (see Figure 28). For general graphs, Fung et al. [FHK+09] use both multi-level graph views following the VisLink approach [CC07], and overlapping of two networks with highlighting of common structural parts (see Figure 29a). Archambault [Arc09] uses graph aggregation and graph filtering to reveal structural differences between two graphs (see Figure 29b).

Structural similarity among multiple graphs Structural comparison of multiple graphs is often based on their description by several graph properties such as graph size, density, connectedness etc. (see also Section 2.1). These properties can be used for exploration of large sets of graphs [FPSG10], or for the determination of structural similarity between graphs. Graph similarity may serve as an input for clustering of graphs (grouping similar graphs). Clustering helps gaining overview of types of graphs in large graph databases. Interactive combination of graph clustering and visualization of clustering results has been proposed in [vLGS09] (see Figure 30).

6. Concluding Remarks and Future Challenges

Research on visual graph analysis deals with the interrelated issues of graph drawing, graph presentation, human-computer-interaction, and analytics. This state-of-the-art report represents an encompassing overview and systematization of recent developments in this field. Many advances have been made on individual parts of visual graph analysis. On the other hand, the surveyed literature discusses many important open challenges, that researchers see in need of work. In the following, we summarize key research challenges. The discussion of the relevant topics is divided into three broad areas: graph visualization and interaction, visual analysis systems, and conceptual issues.
6.1. Graph Visualization and Interaction

Scalability issues in graph drawing There has been much interest in the development of faster layout algorithms that produce more readable layouts for large graphs, also using parallel computing, as provided e.g., by current CPUs and GPUs. It is recognized that using a combination of automatic graph layout generation and user-oriented, interactive layout steering, better layouts can be obtained. As graphs get larger, graph filtering and aggregation have been the main means of graph simplification allowing to draw them. Alternatively, the limited screen space leading to strong overplotting in large graph visualization can be avoided by drawing graphs on large screens, where specialized layouts can be applied [MGL06]. It can be foreseen that work on more sophisticated graph layouts revealing the main structures in the whole graphs, or parts thereof, will continue. In particular, user involvement in the graph layout process involving analytical expertise of the user is a promising approach and may lead to easier interpretation of the drawings. From an analytical perspective, also the understanding of the meaning of the nodes and edges, besides their global structure, is necessary. In particular, the readable/non-overlapping drawing of nodes, edges and their labels is an important issue. When displaying graphs with labels, even smaller graphs can easily lead to overcrowded displays. This topic is gaining more interest in visual analytics research.

Graph types in graph drawing In recent years, the variety of considered graph types has increased substantially. In particular, there has been a large amount of work on drawing dynamic and compound graphs. When drawing dynamic graphs, layout stability and on-line graph drawing are the main points of interest for the future research. In visual analysis, the understanding of the graph changes needs to be supported by stable layouts that preserve the mental map of the analyst thereby allowing them to follow changes on the screen [DGK01]. These layouts should be very stable for minor graph changes and, at the same time, be able to effectively show large graph changes. While a non-trivial challenge, if successfully supported it may lead to easier spotting of structural changes in the graph and thereby, more efficient and effective analysis. On-line graph drawing, where the data stream is unpredictable, poses major challenges in this respect. Compound graphs as a combined graph type, including aggregated graphs, represent a complex data type. The main analytical problem there is the understanding of both types of connections in a graph, as well as the understanding of the graph structures on multiple abstraction levels. This is a very cumbersome task, which can be supported by graph visualization systems. However, the drawing of such complex graphs is still in its infancy. In the future, also further graph types such as hypergraphs [KKS09], or graphs with overlapping sets of nodes [HD10] may become more prominent in visual graph analysis research.

Graph uncertainty Graph visualization by now mainly deals with drawing graphs with given data, largely disregarding graph uncertainty. Visualization of uncertain data is a general challenge in visual analytics. As has been shown in [GS05], the degree of data certainty affects analytical decisions. Therefore, it is an important issue in visual graph analysis. In graph visualization, various types of uncertainty can be regarded. The uncertainty can relate to the graph structure (the existence of nodes and edges between them) and/or on graph attributes (edge and node attributes). For displaying node and edge attribute uncertainty, various methods from multivariate data visualization with uncertainty (see e.g., overviews given in [PWL97, THM’05, GS06]) could be applied. However, their applicability to graph visualization needs to be studied. When dealing with structural uncertainty, there are few dedicated techniques. For example, Candidate [LRC07] uses transparency and color for conveying uncertainty in merged graphs. Therefore, it is expected that more methods will be developed in the future to address graph uncertainty issues.

Perception issues in graph visualization The understanding of graph structures in visualization strongly depends on human perception capabilities. Studies of human perception for graph drawing have recently focused on comparison of graph understanding using varying graph layouts. In graph design, studies on edge visualization have shown that the edge design has an influence on the graph reading. These various studies have given rise to new problems in graph visualization, which need to be studied in the future.
Graph Interaction Techniques In graph exploration, recently new interaction techniques for various graph types have been developed. These techniques increasingly make use of the structural properties of the graph to interactively navigate in the graph (e.g., in [TS08b, vHP09, TAS09]). This tendency supports the analytical purpose of graph visualization, as analysts can more easily examine the structural relationship between entities in the graph. In the future, this direction can be extended.

6.2. Visual Analysis Systems

Visual analysis systems In line with Keim’s visual analytics process [KAF∗08], modern visual graph analysis systems should interactively integrate data pre-processing, interactive data visualization, and building and visualizing of data models for gaining knowledge from the data. Many visual analysis techniques already include parts of this process. However, many of them rely on black box computations (e.g., automatic graph pre-processing, automatic calculation of graph similarities or cliques). To support the variable hypothesis-insight-driven analytical process, more user involvement in the process should be aimed at. The user should have full control of the type of the analysis and its parameters. As this process includes multiple loops, interactive feedback possibilities are necessary. Therefore, integrated visual analysis systems should include such features.

Integration of various data types in visual analysis

Graphs as data structures capturing relationships between entities are part of a larger set of data types examined in various applications. Usually, the analysis of graphs is undertaken in combination with analysis of related data sets, or other data sets are transformed into graphs for their analysis [CGK∗07, BMGK08]. For analysis of the various data sets as a whole, the sole focus on visual graph analysis (in particular graph exploration) without taking other relevant data into account, is not suitable. In the future, larger integrated visual analytics systems combining research results from several areas are needed.

Addressing new analytical tasks With the increasing data set sizes and their complexity, new analytical tasks arise. For example, one such task is the examination of the similarities and differences between graphs. This task builds on the examination of the structure of one graph as discussed above. Lately, several papers about visual graph comparison for both trees and general graphs have been published (see Section 5). The comparison can concern only two graphs, trying to match nodes and edges between them. It can focus on finding similar graphs for one particular graph from a large set of graphs. It can concern gaining an overview of the types of structures in a large set of graphs. It can concentrate on analyzing the similarities of whole graphs or on matching of parts of one graph to other graphs. Owing to its complexity, and the variety of the problems, it is foreseeable that the research in this area will need to continue.

Collaborative visual graph analysis For solving complex analytical tasks concerning multiple large related data sets, a collaboration of several experts is necessary. Recently, the development of collaborative visual analysis systems has received attention [BMZ∗06, Kee06, Ise07]. However, collaborative visual graph analysis is not represented prominently. Therefore, the study of collaborative systems including graph data sets would be of advantage. The specifics of graph exploration, in particular, need to be studied.

Insight provenance for visual graph analysis In Visual Analytics applications, the analytical processes are often long-running and/or distributed. To support the reproducibility, reversibility and automation of these processes, user tracking of the graph interaction steps is necessary. As a basis for tracking, a taxonomy of graph interaction techniques is necessary. The theory of interaction is a general Visual Analytics challenge [TC06]. Although several interaction taxonomies also for insight provenance have been recently introduced [GZ08, HMSA08], their applicability and the need for their adaptation to graph analysis needs to be studied. In return, specific classifications of graph interaction techniques could be developed. In this report, we have aimed to classify them for gaining a concise overview of the current state of the research. This classification, however, may not be directly applicable to user tracking applications.

Applications For analytical purposes, standard graph visualization and analysis methods need to be adapted to the specific needs of the particular application domain. For example, there are specialized systems for visualization of bio-chemical structures, shareholding structures, and many more. Designing graph visualization systems with fast adaptability to various data types, analytical tasks and application-dependent analytical processes is still a challenge. Even within one application, often, the network to be analyzed needs to be constructed from heterogeneous data sources, and the focus of interest (attributes of nodes and edges) varies dynamically. Designing such systems is obviously not trivial.

6.3. Conceptual Issues

Evaluation Evaluation of usability and user acceptability of the techniques including development of the evaluation methodologies is an important future challenge for the Visual Analytics research area [KMS∗08, TC05, TC06, LL07]. Currently, there is a broad discussion in the Visual Analytics community on the appropriate methodology for the evaluation of Visual Analytics and information visualization systems. This discussion applies also to the visual analysis of graphs. This challenge is expressed in the words of Plaisant et al. in the introduction to the special issue
of Computer Graphics and Applications [PGS09] “Assessing VA [Visual Analytics] technology’s effectiveness is challenging because VA tools combine several disparate components, both low and high level, integrated in complex interactive systems used by analysts, emergency responders, and others. ... Traditional evaluation metrics such as task completion time, number of errors, or recall and precision are insufficient to quantify the utility of VA tools, and new research is needed to improve our VA evaluation methodology.”. When concentrating on the evaluation of graph visualization techniques, several approaches have been proposed, ranging from quantitative to qualitative studies. Controlled experiments measuring accuracy and duration of user tasks have been used, for example, to compare tree visualization techniques [Kob04, AK07, ZK08]. An extension of these two main measures, the so-called cognitive load measure was used for evaluating general graph visualizations [HEH09]. Moreover, eye tracking can be employed for quantitative evaluation, e.g., for comparing graph layouts [Haa07, PSD09]. These controlled studies offer a quantitative comparison across techniques, however often suffer from the focus only on selected low level tasks. Note that the formulation of these tasks can influence the comparison result [ZK08]. A combination of quantitative and qualitative study has been performed for comparing graph layouts produced by both in a manual and in an algorithmic way [DLF+09]. The subjective user view has been used for ranking of layouts. A qualitative view on the effectiveness of visual analytics techniques can be gained by use case studies conducted by domain experts (e.g., in [PS08, MGT+03]). This method offers insights into the usability of the systems in real world scenarios, however does not allow for standardized quantitative comparison of the techniques. The choice of appropriate evaluation method and its design is still discussed in the community.

Taxonomies and benchmarks The field of visual graph analysis would profit from more elaborate taxonomies for tasks, interaction, visualization techniques, measures for quality, and benchmarks for comparing the new techniques. They would support both the design and development of visual analytic systems and their evaluation. Although several taxonomies and sample data sets exist, a more broader scope of theory and data aspects is needed owing to the large set of problems/tasks in visual analysis of graphs.

Acknowledgements

We thank Vyara Ivanova, Anna Mitkova, Melanie Görner, and Robert Rehner for their helpful comments and suggestions. We are grateful to the anonymous reviewers for their constructive and useful comments. We thank all authors and copyright holders of the original figures for agreeing to their reproduction in this paper. This work was partially supported by the German Federal Ministry of Economics and Technology within the THESEUS project (http://www. theseus-programm.de/). It was also partially supported by the German Research Foundation (DFG) within the project Visual Feature Space Analysis as part of the Priority Program on Scalable Visual Analytics (SPP 1335).

References


[BBLB09] Bachmaier C., Brandenburg F. J., Brunner W., Lovász L.: Cyclic leveling of directed graphs. Lecture Notes in Computer Science 5417 (2009), 348–359. 8

submitted to COMPUTER GRAPHICS Forum (1/2011).


