Visual Analysis of Network Centralities

Tim Dwyer[*]
Monash University, Melbourne, VIC, Australia

Seok-Hee Hong[†]
NICTA (National ICT Australia) & School of IT, University of Sydney, NSW, Australia

Dirk Koschützki[‡]
Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Falk Schreiber[§]
Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Kai Xu[¶]
NICTA (National ICT Australia), Sydney, NSW, Australia

Abstract

Centrality analysis determines the importance of vertices in a network based on their connectivity within the network structure. It is a widely used technique to analyse network-structured data. A particularly important task is the comparison of different centrality measures within one network. We present three methods for the exploration and comparison of centrality measures within a network: 3D parallel coordinates, orbit-based comparison and hierarchy-based comparison. There is a common underlying idea to all three methods: for each centrality measure the graph is copied and drawn in a separate 2D plane with vertex position dependent on centrality. These planes are then stacked into the third dimension so that the different centrality measures may be easily compared. Only the details of how centrality is mapped to vertex position are different in each method. For 3D parallel coordinates vertices are placed on vertical lines; for orbit-based comparison vertices are placed on concentric circles and for hierarchy-based comparison vertices are placed on horizontal lines. The second and third solutions make it particularly easy to track changing vertex-centrality values in the context of the underlying network structure. The usability of these methods is demonstrated on biological and social networks.

Keywords: network analysis, centralities, visualisation, graph drawing, biological networks, social networks.

1 Introduction

Network analysis methods support the study of structural properties in networks. One important method is centrality analysis which determines the relative importance of vertices in a network based on their connectivity within the network structure. It is particularly useful in analysing social and biological networks. In social network analysis, a methodology which uses graph-theoretic concepts to analyse and understand the structure and behaviour of social networks (Wasserman & Faust 1994), centrality is a well known individual-level network analysis method. It measures the importance or prominence of the actors in a social network. For instance, in a research collaboration network, one can identify the most prominent and influential researchers in a particular research area. In a citation network, one can identify the most important and influential papers as cited by other scientists.

In the life sciences centrality measures help scientists to understand the underlying biological processes and have been successfully applied to different biological networks. Central vertices in protein-protein interaction networks are often functionally important and the removal of such vertices is related to lethality (Jeong, Mason, Barabási & Oltvai 2001). In metabolic networks metabolites with highest degree, i.e. with the highest number of neighbours, may belong to the oldest part of the metabolism (Fell & Wagner 2000) and are main metabolites in well-known pathways (Wuchty & Stadler 2003). Closeness centrality, a centrality that ranks vertices depending on the sum of their shortest paths to all other vertices is another method for identifying central metabolites in metabolic networks (Ma & Zeng 2003). Wuchty and Stadler applied different types of centralities to metabolic, protein-protein-interaction and domain-sequence networks to identify central network elements (Wuchty & Stadler 2003).

There are many different centrality measures (Bonacich 1972, Freeman 1977, Freeman 1979, Freeman, Borgatti & White 1991, Koschützki, Lehmann, Peeters, Richter, Tenfelde-Podehl & Zlotowski 2005, Newman 2003, Wasserman & Faust 1994) which can be used to analyse networks, and choosing the "right" measure for a specific problem is usually a difficult task. For example, in biological networks correlations between specific centrality measures and functionally important properties have been shown for some networks (Jeong et al. 2001, Wuchty & Stadler 2003). However, it has also been shown that the degree of a vertex alone is not sufficient to distinguish lethal proteins from viable ones (Wuchty 2002); that in protein networks there is no relation between network connectivity and robustness against amino-acid substitutions (Hahn, Conant & Wagner 2002); that different centrality measures for metabolic networks identify different sets of central metabolites (Ma & Zeng 2003); and

[*]e-mail: Tim.Dwyer@infotech.monash.edu.au
[†]e-mail: shhong@it.usyd.edu.au
[‡]e-mail: koschuet@ipk-gatersleben.de
[§]e-mail: schreibe@ipk-gatersleben.de
[¶]e-mail: KaiKevin.Xu@nicta.com.au

This paper is a result of the NICTA VALACON (Visualisation and Analysis of Large and Complex Networks) Workshop on Biological Networks organised by Seok-Hee Hong on November 2004 in Sydney, Australia. This work has been partially supported by the German Ministry of Education and Research (BMBF) under grant 0312706A.

that for biological network analysis several centrality measures have to be considered (Koschützki & Schreiber 2004, Wuchty & Stadler 2003). Furthermore, usually there is not enough data about the functional properties of network elements available to determine "good" centrality measures and centrality analysis is often used to simply explore a biological network to discover potentially important parts. A common approach to analysing such networks is therefore to compare different centralities within one network, build a hypothesis concerning discovered interesting "central" elements and test this hypothesis by experimental approaches.

In Table 1 and Figures 1 and 2 typical methods for the comparison of different centrality measures are shown using a protein-protein interaction (PPI) network as an example: correlation (Kendall's correlation coefficient τ), scatterplot and parallel coordinates (Inselberg & Dimsdale 1990). This PPI network is based on *Mus musculus* (mouse) data from the DIP-database (Salwinski, Miller, Smith, Pettit, Bowie & Eisenberg 2004), release 26-01-2005 and shows the largest connected component of the network.

	$\mathcal{C}_e$ eccentricity	$\mathcal{C}_c$ closeness	$\mathcal{C}_\lambda$ eigenvector	$\mathcal{C}_d$ degree	$\mathcal{C}_r$ rwbetweenness
$\mathcal{C}_e$	1.00000	0.85319	0.52707	0.35398	0.36017
$\mathcal{C}_c$	0.85319	1.00000	0.58717	0.42786	0.46054
$\mathcal{C}_\lambda$	0.52707	0.58717	1.00000	0.44396	0.43527
$\mathcal{C}_d$	0.35398	0.42786	0.44396	1.00000	0.74693
$\mathcal{C}_r$	0.36017	0.46054	0.43527	0.74693	1.00000

Table 1: Kendall's correlation coefficient τ for the centrality ranks of the *M. musculus* PPI-network

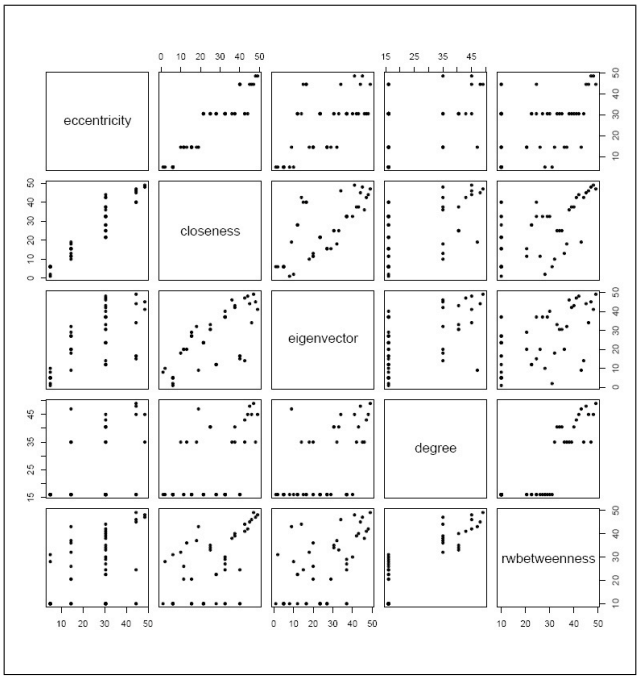

Figure 1: Scatterplot matrix of the centrality rank of the *Mus musculus* PPI-network. To the right and top of each square are vertices with high centrality. The different centrality measures used in this example are explained in Section 2

These commonly used methods for comparing centralities in biological networks have several disadvantages. Correlation coefficients only describe how well different centrality measures correlate without showing which parts of the networks are similar or different. Scatterplots and parallel coordinates allow the comparison of the centrality of single vertices and show a general trend in correlation, but it is difficult to investigate larger sets of vertices even if colours are used. Furthermore, such methods do not support the investigation of the network structure to explore, for example, the distribution of centrality values or to find clusters of vertices with high centrality.

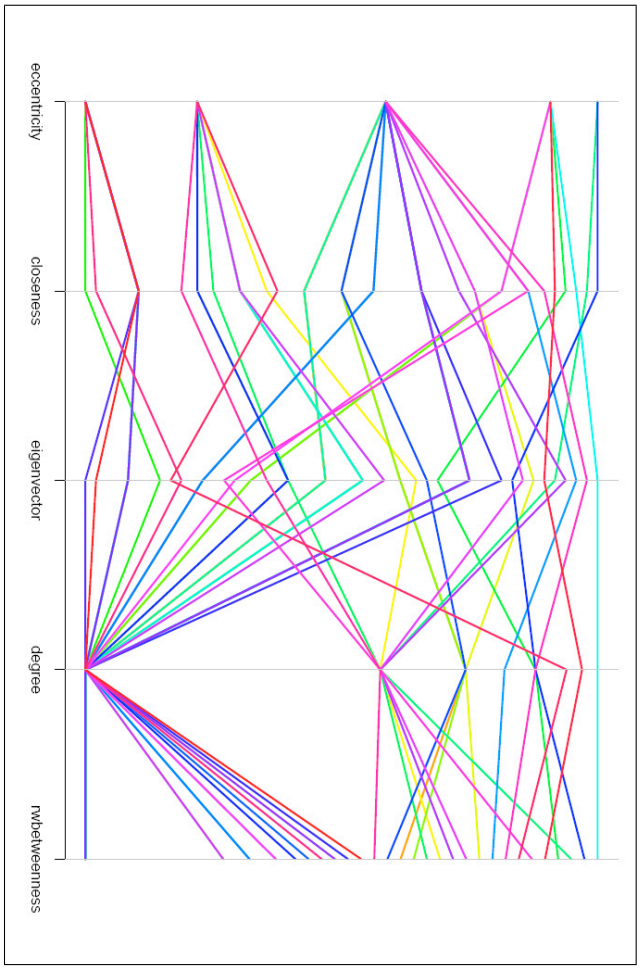

Figure 2: Parallel coordinates of the centrality rank for the same PPI network analysed in Figure 1

There are several approaches supporting such a structure-centred analysis. Centrality values of vertices can be visualised using the size of the vertices or by constraining their positions to fixed distances from the centre or the bottom of the drawing, see for example the Visone (Brandes & Wagner 2004) and Pajek systems (Batagelj & Mrvar 2004). To compare different centrality measures within one network such methods could be combined, e.g. by representing one centrality with different vertex sizes and another centrality with the positions of the vertices constrained to centrality-dependent distances from the centre. However, this solution is restricted to the comparison of only a few centralities. Further on, it does not show very well how the centrality value of a vertex changes from one centrality measure to the next.

The exploration and comparison of different centralities within one network is important, however, existing methods inefficiently support such tasks. This paper deals with three methods to assist scientists in the exploration and comparison of centrality measures within a single network. These three methods are based on a common underlying idea: for each centrality measure the graph is copied and drawn in a separate 2D plane such that the vertex positions depend on their centrality values. These planes are

then stacked into the third dimension. If the centrality value of a vertex does not change its position in the next plane with respect to the position in the previous plane it is stable. Changes of centrality values can be additionally emphasised by edges between the planes. The result is that different centrality measures can be easily compared.

This general idea of stacking 2D network information into the third dimension, also called $2\frac{1}{2}$D visualisation, has been demonstrated to be useful in several applications (Brandes & Corman 2002, Brandes, Dwyer & Schreiber 2004, Dwyer & Eades 2002, Koike 1993), but to the best of our knowledge, $2\frac{1}{2}$D techniques have not previously been applied to the study of network centrality. The details of how centrality is mapped to vertex position are different in our three methods. For *3D parallel coordinates* vertices are placed on vertical lines; for *orbit-based comparison* vertices are placed on concentric circles and for *hierarchy-based comparison* vertices are placed on horizontal lines. The second and third solutions make it particularly easy to track changing vertex-centrality values in the context of the underlying network structure.

This paper is organised as follows: in Section 2 we define the graph model on which we operate and introduce five centrality measures which are used as examples. Section 3 discusses the visualisation methods. These methods are applied to a biological and a social network in Section 4. Finally, Section 5 contains general discussion.

2 Definitions

In this section we define the graph model, introduce five centrality measures and discuss the comparison of such centrality measures.

2.1 Graphs

An undirected *graph* G consists of a finite set V of *vertices* and a set E of two-element subsets of V called *edges*. An edge $e = \{u, v\} \in E$ connects two vertices u and v. The vertices u and v are said to be *incident* with the edge e and *adjacent* to each other. The number of vertices and edges of G is given by $n = |V|$ and $m = |E|$, respectively. The set of all vertices which are adjacent to u is called the *neighbourhood* of u. A graph is called *loop-free* if no edge connects a vertex to itself. An *adjacency matrix* A of a graph is an $(n \times n)$ matrix, where $a_{ij} = 1$ if and only if $\{i, j\} \in E$ and $a_{ij} = 0$ otherwise. The adjacency matrix of any undirected graph is symmetric.

The *degree* $\deg(v)$ of a vertex v is the number of its incident edges. Let $(e_1, \ldots, e_k)$ be a sequence of edges in a graph G. This sequence is called a *walk* if there are vertices $v_0, \ldots, v_k$ such that $e_i = \{v_{i-1}, v_i\}$ for $i = 1, \ldots, k$. If the edges e_i are pairwise distinct and the vertices v_i are pairwise distinct the walk is called a *path*. The *length* of a walk or path is given by the number of involved edges $k = |(e_1, \ldots, e_k)|$. A *shortest path* between two vertices u, v is a path with minimal length. The *distance* $\mathrm{dist}(u, v)$ between two vertices u, v is the length of a shortest path between them. Two vertices u, v of a graph are called *connected* if there exists a walk from vertex u to vertex v. If any pair of different vertices of the graph is connected, the graph is called *connected*. A *random walk* between u and v is a walk which starts at vertex u, chooses uniformly at random one of the incident edges of the current vertex until it finally reaches the target v.

In the remainder of this paper we consider only undirected, loop-free, connected, non-trivial (at least two vertices and one edge) graphs. This restriction is required to be able to use all centrality measures defined in the following paragraph. Note that several centralities can easily be expanded to cover directed or unconnected graphs. Even an extension for weighted edges is possible.

2.2 Centralities

A centrality is a function $\mathcal{C}$ which assigns every vertex $v \in V$ of a given graph G a value $\mathcal{C}(v) \in \mathbb{R}$. As we are interested in the ranking of the vertices of G we choose the convention that a vertex u is more important than another vertex v iff $\mathcal{C}(u) > \mathcal{C}(v)$.

Degree An obvious order of the vertices of a graph can be established by sorting them according to their degree. The corresponding centrality measure *degree-centrality* ($\mathcal{C}_d$) is defined as $\mathcal{C}_d(v) := \deg(v)$. Freeman gives a long list of references to the usage of degree-centrality in social network analysis (Freeman 1979). For biological network analysis, for example, proteins with high degree-centrality have been shown to be important in PPI-networks (Jeong et al. 2001).

Eccentricity This and the two subsequent centrality definitions operate on the concepts of paths within the given graph. A simple definition uses the distance between vertices. The *eccentricity* ecc of a vertex u is defined as $\mathrm{ecc}(u) := \max_{v \in V} \mathrm{dist}(u, v)$ and the *eccentricity-centrality* ($\mathcal{C}_e$) as $\mathcal{C}_e(u) := \frac{1}{\mathrm{ecc}(u)}$. The reciprocal of $\mathrm{ecc}(u)$ is used to ensure that more central vertices have a higher value of $\mathcal{C}_e$, since such central vertices are the ones with the smallest eccentricity value. An application of eccentricity within the biological context is shown by Wuchty and Stadler (Wuchty & Stadler 2003).

Closeness In contrast to eccentricity, closeness-centrality uses not the maximum distance between the vertex of interest and all other vertices but the sum of the distances of this vertex and all other vertices. The *closeness-centrality* $\mathcal{C}_c(u)$ is defined as $\mathcal{C}_c(u) := \frac{1}{\mathrm{sumdist}(u)}$ with $\mathrm{sumdist}(u) = \sum_{v \in V} \mathrm{dist}(u, v)$. Closeness-based centrality measures are used in social (Freeman 1979) and biological (Wuchty & Stadler 2003) network analysis.

Random Walk Betweenness Within networks a communication between two vertices u, v may be visible to a third vertex w if this vertex lies in the path of the communication between u and v. To measure the centrality of a vertex the ability to observe communication is a feasible approach. Different methods to model communication are conceivable, e.g., over shortest paths, paths with maximum flow and random walks. All of these are potential models for *betweenness* (Freeman 1977, Freeman et al. 1991, Newman 2003). Newman's random walk approach models information transmission and therefore matches problems often modelled in biological networks. For the *random-walk betweenness centrality* ($\mathcal{C}_r$) the centrality of a vertex w is equal to the number of times that a random walk from u to v goes through w, averaged over all u and v.

Bonacich's Eigenvector Centrality A different approach to order the vertices of a graph was suggested by Bonacich (Bonacich 1972). It is based on the assumption that the value of a single

vertex is determined by the values of the neighbouring vertices. In contrast to the previous measures not only the position of a vertex within the graph is considered but also the centrality values of its neighbours. Bonacich suggested the following definition: $\mathcal{C}_\lambda(u) := \sum_{v \in N(u)} \mathcal{C}_\lambda(v)$. Considering the adjacency matrix representation of the graph this is equivalent to $\mathcal{C}_\lambda(v_i) := \sum_{j=1}^{n} a_{ij}\mathcal{C}_\lambda(v_j)$. This leads directly to the well known problem of eigenvector computation $\lambda S = AS$ and the eigenvector of the largest eigenvalue is the *eigenvector-centrality* ($\mathcal{C}_\lambda := S$) (Bonacich 1972).

2.3 Correlation and Ordering

Given a graph and a set of centrality measures our method for the exploration and comparison of the different centralities is based on the stacking of copies of the graph, one for each centrality, into the third dimension. An important task is to compute an appropriate ordering of these copies. We want to order them such that those which are similar with respect to the correlation, i.e., which have a correlation close to 1.0, are close to each other. This is similar to Keim's one-dimensional ordering of dimensions (Keim 2000). The correlation $\tau(\mathcal{C}_i, \mathcal{C}_j)$ between the centralities $\mathcal{C}_i$ and $\mathcal{C}_j$ is our similarity measure with large values (close to 1.0) meaning high similarity, whereas low or negative values mean dissimilarity. In detail we use Kendall's correlation coefficient τ_b as this correlation coefficient is known to behave better than other coefficients (e.g. Spearman's ρ) in the case of ties, e.g., values which are equal for several objects. The definition and the details of the computation are beyond the scope of this paper and can be found in the literature (Lienert 1973, Press, Teukolsky, Vetterling & Flannery 1992).

Let $\mathcal{C}_1, \ldots, \mathcal{C}_k$ be the centralities under consideration. The optimal ordering of the centralities and therefore the corresponding copies of the graph is a permutation $\{\pi(1), \ldots, \pi(k)\}$ of the centralities such that $\sum_{i=1}^{k-1} \tau(\mathcal{C}_{\pi(i)}, \mathcal{C}_{\pi(i+1)})$ is maximal. Note that the computation of an optimal ordering is NP-hard. In our example with five centralities we can compute the optimal ordering, for larger sets of different centralities several ordering heuristics exists.

3 Visualisation

In this section we discuss briefly the overarching visualisation concept and present our three visualisation approaches.

3.1 General Visualisation Method

The common idea behind the different visualisation approaches is that there is a copy of the graph for each centrality. Each copy is drawn in a separate 2D plane such that central vertices are easily detectable. The planes are then stacked into the third dimension in an optimal ordering and an arrangement of vertices within planes is computed with the goal of supporting comparison of centralities.

3.2 3D Parallel Coordinates-based Comparison

Parallel coordinates techniques have been used successfully for the visualisation of multi-dimensional or multivariate data (Inselberg & Dimsdale 1990). However, from our experiments with biological and social networks we observed that vertices in the network can frequently share the same centrality value. Note that in standard two-dimensional parallel coordinates as in Figures 2 it is rather difficult to show each vertex separately, in particular where two or more vertices have the same centrality value.

This motivates the use of the additional dimension to display all the vertices which have the same centrality value in a horizontal line simultaneously. The main idea of our 3D parallel coordinates is very simple. Each two dimensional plane contains the information for a particular centrality and several centralities are stacked in the third dimension. For each centrality we use horizontal lines to place vertices with the same centrality values. See Figures 4 for an example.

More specifically, for each centrality measure we use a vertical line and several horizontal lines within the plane. That is, the vertical line represents the range of centrality values from the lowest centrality value to the highest. For each centrality measure all vertices belonging to exactly the same centrality value are placed on a horizontal line to separate the vertices. The ordering of planes is decided using the method described in Section 2.3. Note that two vertices may overlap if their centrality values are very close; however this problem can be solved by adjusting the size of the vertices.

Next, we add inter-plane edges between two adjacent planes, where each plane represents one centrality measure, such that the same vertices of the graph are connected by inter-plane edges. This helps to trace vertices between two different planes. Note that to reduce visual complexity it is possible to show edges only when its centrality difference is greater than a given threshold.

Finally, we need to decide the ordering of the vertices on each horizontal line, i.e. the ordering of vertices which have the same centrality value. We want an ordering which minimises the total edge length of inter-plane edges between the two planes, as the use of long edges will increase the visual complexity and thus decrease ease of comparison. Similarly, the ordering should avoid edge crossing. This problem can be solved easily considering the position of the same vertex in the previous plane. For tie-breaking, we use the difference in height.

Overall, the time complexity of comparison based on the 3D parallel coordinates is linear if the centrality values for each centrality measure and the optimal ordering of the different centrality measures are given. Thus, this method may scale well for large networks.

3.3 Orbit-based Comparison

The objective of our graph drawing algorithm for orbit-based comparison is to find coordinates for each vertex of the graphs $G_1, \ldots, G_k$ such that:

1. all vertices of graph G_i have the same z-coordinate (plane constraint);
2. the order of z-coordinates corresponds with the computed order of the graphs, see Section 2.3;
3. all vertices of graph G_i are constrained to lie on concentric circles (orbits) depending on the centrality value of the vertex. That is, each vertex v is assigned an *orbital constraint* with radius $r = f(\mathcal{C}(u))$ where f is typically a partitioning such that there are only a small number of distinct radii;
4. the centre of the concentric circles for all graphs has the same x- and y-coordinate;

5. all vertices of graph G_i have x- and y-coordinates such that the distance of a vertex to its neighbours is as close to the optimal distance d as possible; and

6. each vertex of G_i should be as close as possible to the same vertex of the adjacent graphs.

A formulation for the "stress" of a graph layout is (Borg & Groenen 1997, Gansner, Koren & North 2004):

$$\sum_{i<j} w_{ij}(\parallel X_i - X_j \parallel - d_{ij})^2 \qquad (1)$$

where X_i and X_j are the positions of the i^{th} and j^{th} vertices of a graph G, d_{ij} is the ideal distance between these vertices (typically a function of the graph theoretic distance between them) and w_{ij} is a normalisation constant (for example: d_{ij}^{-2}). A "good" layout is said to be one that minimises this stress function.

Since the plane constraints are constant and independent of the layout within planes, our layout scheme with plane and orbital constraints can be reduced to a 2D layout problem. That is, we seek to find an arrangement for the union graph $G = \bigcup_{i=1}^{k} G_i$ and we augment the edges of G with a set of inter-plane edges $\{\{u, v\}|$ where u and v are the equivalent vertices in adjacent levels G_i and $G_{i+1}\}$. The fifth objective in the list above is met by setting the ideal length $d = 0$ for such inter-plane edges. By restating this "stress" function in polar coordinates we can precisely define the orbital constraints of this new layout problem:

$$\sum_{i<j} w_{ij}(\parallel r_i(\cos\theta_i, \sin\theta_i) - r_j(\cos\theta_j, \sin\theta_j) \parallel - d_{ij})^2 \qquad (2)$$

where r_i and r_j are constants corresponding to the radii of the orbit constraints for vertices i and j respectively. While Kamada and Kawai (Kamada & Kawai 1989) showed that an approximate solution for (1) could be found by iteratively fixing all but one vertex and solving the resulting convex quadratic form, and Gansner *et al.* (Gansner et al. 2004) were able to bound the global function with a quadratic form using functional majorisation, the cyclical nature (and hence non-convexity) of (2) means that no such simplification is readily apparent. For this reason Brandes *et al.* (Brandes, Kenis & Wagner 2003) chose a simulated annealing solution for a similar radial layout problem. Instead, we have had reasonable success adapting a naïve Fruchterman and Reingold (Fruchterman & Reingold 1991) force-directed method.

A standard force-directed method works by computing a "force" vector $\vec{f}$ for each vertex u from the sum of attractive forces between u and all v where there exists an edge $\{u, v\}$ and repulsive forces between u and all other vertices in G. Each vertex is then moved by a small amount in the direction of this force, for example a vertex at position p is moved to $p' = p + c\vec{f}$ where c is a constant. The process repeats iteratively until the total length of all force vectors for an interaction falls below some threshold. We augment the standard force-directed layout method with orbital constraints by moving vertices only by the projection of $\vec{f}$ on the orbital constraint arc. That is, to satisfy the orbital constraint centred about o with radius r the above vertex is placed at $o + r|\vec{op'}|$, see Figure 3.

To aid comparison of centralities it is more important that inter-plane edge lengths are minimised than the length of edges within each plane. Thus,

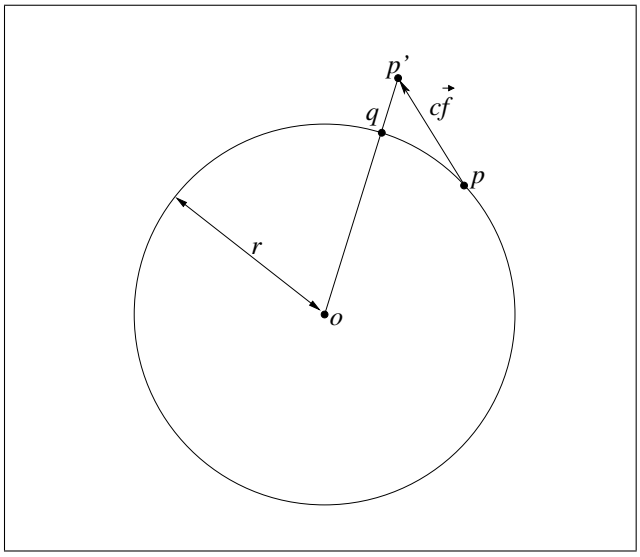

Figure 3: Calculating the new position q of a vertex at position p due to force f subject to an orbital constraint of radius r

we run the force-directed placement to completion in several stages. First, inter-plane attractive forces are activated; then forces within each plane are added, one plane at a time. This tends to avoid local minima of (2) with long inter-plane edges. Also, repulsive forces between vertices on different planar levels are not considered.

3.4 Hierarchy-based Comparison

In this section we describe a centrality comparison method based on a drawing algorithm which uses a hierarchy to display centrality values. More specifically, we use horizontal lines to represent centrality values. As with the previous methods, a drawing in each plane displays different centrality measures, and the ordering of the planes is decided based on the method described in Section 2.3.

To draw each graph in a plane by displaying centrality as hierarchy we first divide the vertices into layers depending on their centrality value. That is the vertices in the upper layers have higher centrality values than the vertices in the lower layers. Then we place each vertex in the layer on a horizontal line. Note that this method is different to the well-known *Sugiyama* method (Battista, Eades, Tamassia & Tollis 1999, Sugiyama, Tagawa & Toda 1981) for drawing hierarchical graphs as a layered drawing, in the sense that the layering is not based on a traversal of a directed graph and that there may exist edges between vertices in the same layer.

Inside each layer, we can define a few different ordering methods to order the vertices. For example, we can choose an ordering of vertices such that the vertices are ordered from the left to the right based on their centrality value in a decreasing way. Further, it may be possible to choose a different ordering of vertices in the layer with different optimisation criteria. For example, one may choose an ordering which minimises the length of edges between each plane. This problem can be easily solved as in the 3D parallel coordinates method. Alternately, one may choose an ordering which minimises the edge crossings in each plane. This problem is relatively well-studied in graph drawing literature, it is NP-hard even if there are only two layers (Eades & Whitesides 1994). However, there are several fast heuristics such as the *Median* and the *Barycenter* method available (Battista

et al. 1999).

To reduce edge crossings between to adjacent layers within one plane we use a modification of the Barycenter method. More specifically, the position of a vertex is averaged over the positions of all the neighbours on upper layers, not necessarily just one level above. This considers edges which span more than two layers. Alternatively, these edges can be handled by introducing dummy vertices, as in the traditional *Sugiyama* method.

Finally, to reduce the total number of edge crossings in each plane, we perform a *layer-by-layer sweep* heuristic (Battista et al. 1999), that is, the algorithm sweeps from the top layer to the bottom layer, and then sweeps from the bottom layer to the top layer. We had reasonable success when this process was repeated ten times. Note that we can further combine these two different optimisation criteria. A simple heuristic is to use a variation of the Barycenter method, now also considering the position of the neighbour in the previous plane.

This hierarchy-based comparison method can be implemented in linear time, as the Barycenter heuristic can be implemented to run in linear time. Thus it can scale better than the orbit-based comparison method for large networks. Examples of this method are shown in Figures 6 and 7.

Again, as in the previous methods, we use filtering for displaying inter-plane edges to reduce cognitive load. That is we display the inter-plane edges only when the difference between the two centrality values of the same vertices differ more than some given threshold.

4 Implementation and Results

We implemented the presented methods in GEOMI (Geometry for Maximum Insight), a visual analysis tool for the visualisation and analysis of large and complex networks such as social networks, biological networks, scale-free networks and dynamic networks (Ahmed, Dwyer, Forster, Fu, Ho, Hong, Koschützki, Murray, Nikolov, Taib, Tarassov & Xu 2005). GEOMI is based on WilmaScope (Dwyer & Eckersley 2004). We used it for the analysis of the *Mus musculus* PPI network introduced in Section 1 and a social network based on Padgett's Florentine families marital relation data (Wasserman & Faust 1994). The PPI network from the DIP-database consists of 49 vertices and 54 edges. The social network has 16 vertices and 40 edges.

Figures 4-7 show some layouts generated using the three methods. In Figure 4 the biological network is presented with 3D parallel coordinates as described in Section 3.2. Each vertex of the network represents a protein and each edge a protein-protein-interaction. Each vertical column in the layout represents a specific centrality measure. If vertices have the same centrality value within one centrality measure they are placed on a horizontal line. The similarity to common 2D parallel coordinates makes this representation easy to understand. If viewed from the front the picture would look like a conventional 2D parallel coordinate visualisation as shown in Figure 2.

Even though it is easier to compare single vertices in the 3D representation than in common 2D parallel coordinates, there is still the disadvantage that centrality values and network structure cannot be explored within one representation. The orbit-based and the hierarchy-based comparisons overcome this disadvantage. Figure 5 shows the different centralities of the biological network with the orbit-based method described in Section 3.3. Each plane shows the network with the vertices placed on different orbits depending on their centrality values. If the orbit of a vertex changes between adjacent planes an edge is shown. Additionally in the small window the plane highlighted by the water-layer is displayed in a two dimensional representation. By moving the water-layer up and down it is easy to navigate through the different centralities and to see how the centrality values of a vertex change (move in and out) or are stable (a stable vertex has nearly the same x- and y-coordinates in each layer). In Figure 6 the biological network is presented with the hierarchy-based comparison described in Section 3.4.

As an example of social-network analysis we used Padgett's Florentine families marital relation data (Wasserman & Faust 1994). Each vertex in the network represents a family in 15th century Florence, Italy and each edge represents marital relations between the families. The 16 families are chosen for analysis from a larger collection of 116 leading Florentine families due to their historical prominence, such as the Medicis and Strozzis. Figure 7 shows a visual comparison of centralities for Padgett's Florentine families marital relation data. From the drawing, it is easy to see that Medici always belongs to the top rank in any centrality measure. Clearly, one can see the marital relations between the families in one plane and then compare the difference between different centrality measure.

5 Conclusion

We have demonstrated three different methods for the analysis and visualisation of centrality measures of a network. All methods represent the information in three dimensions and two of them allow the simultaneous visualisation of both the network structure and the ranking of vertices based on the centrality measures. All three visualisation methods support the exploration of different centrality measures for the same network. In this paper we presented the comparison of centrality measures, however, the visualisation methods could be also used to represent other multidimensional numerical data within an underlying network.

Our methods work well with networks of up to a hundred vertices. An important question is the scalability of our approach to larger networks with thousands of vertices. Larger networks result in very dense visualisations which are difficult to read and understand. To deal with such networks abstraction methods have to be applied before any of the methods described may show a useful result. The definition of such abstraction methods should be a focus for further work. Another issue is a formal user study to compare the different proposed methods with each other and with well-known 2D techniques.

References

Ahmed, A., Dwyer, T., Forster, M., Fu, X., Ho, J., Hong, S.-H., Koschützki, D., Murray, C., Nikolov, N. S., Taib, R., Tarassov, A. & Xu, K. (2005), GEOMI: GEOmetry for Maximum Insight, *in* 'Proceedings of the 13th International Symposium on Graph Drawing (GD'05)'. to appear.

Batagelj, V. & Mrvar, A. (2004), *Pajek - Analysis and Visualization of Large Networks*, in Jünger & Mutzel (2004), pp. 77–103.

Battista, G. D., Eades, P., Tamassia, R. & Tollis, I. G. (1999), *Graph Drawing: Algorithms for the Visualization of Graphs*, Prentice-Hall.

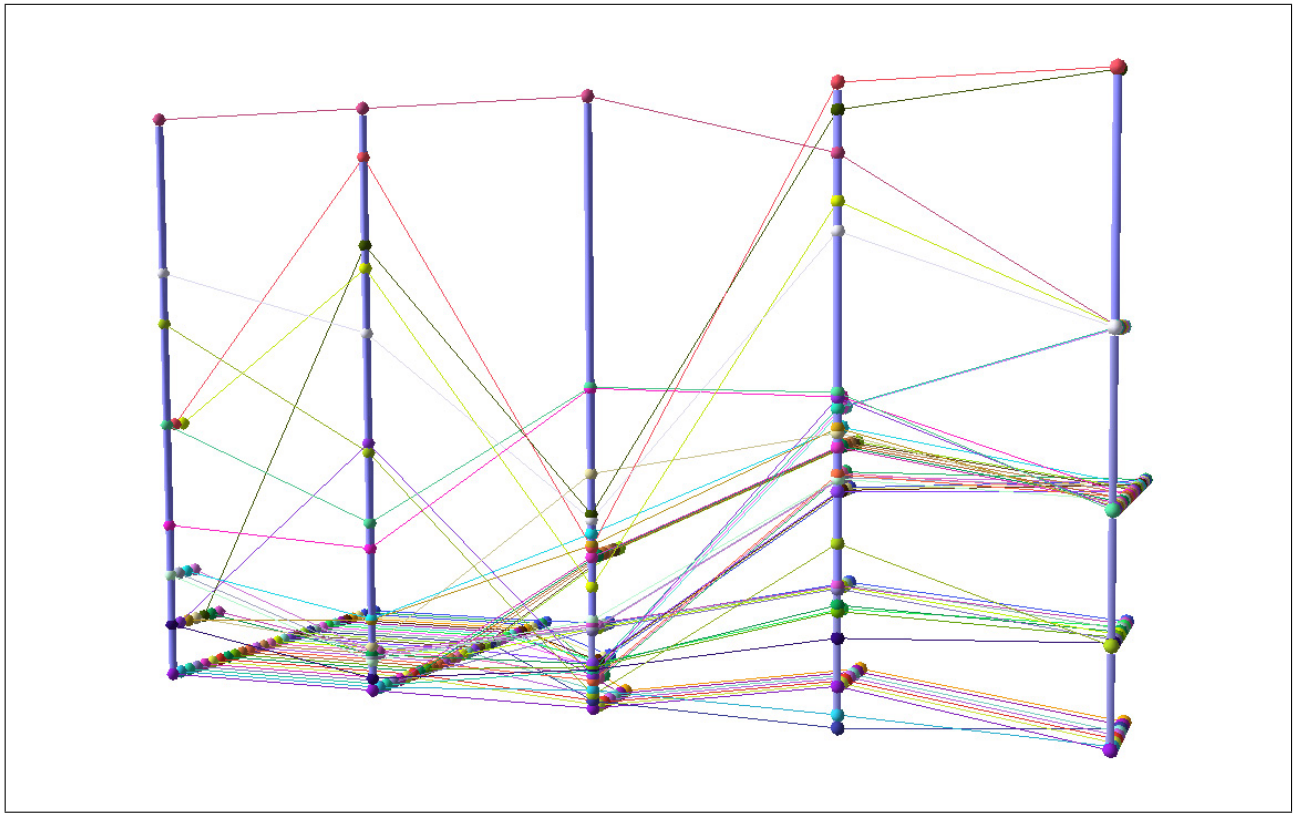

Figure 4: The network described in Section 1 in 3D parallel coordinates, a method described in Section 3.2. In particular, 3D parallel coordinates can display the exact centrality value for each vertex

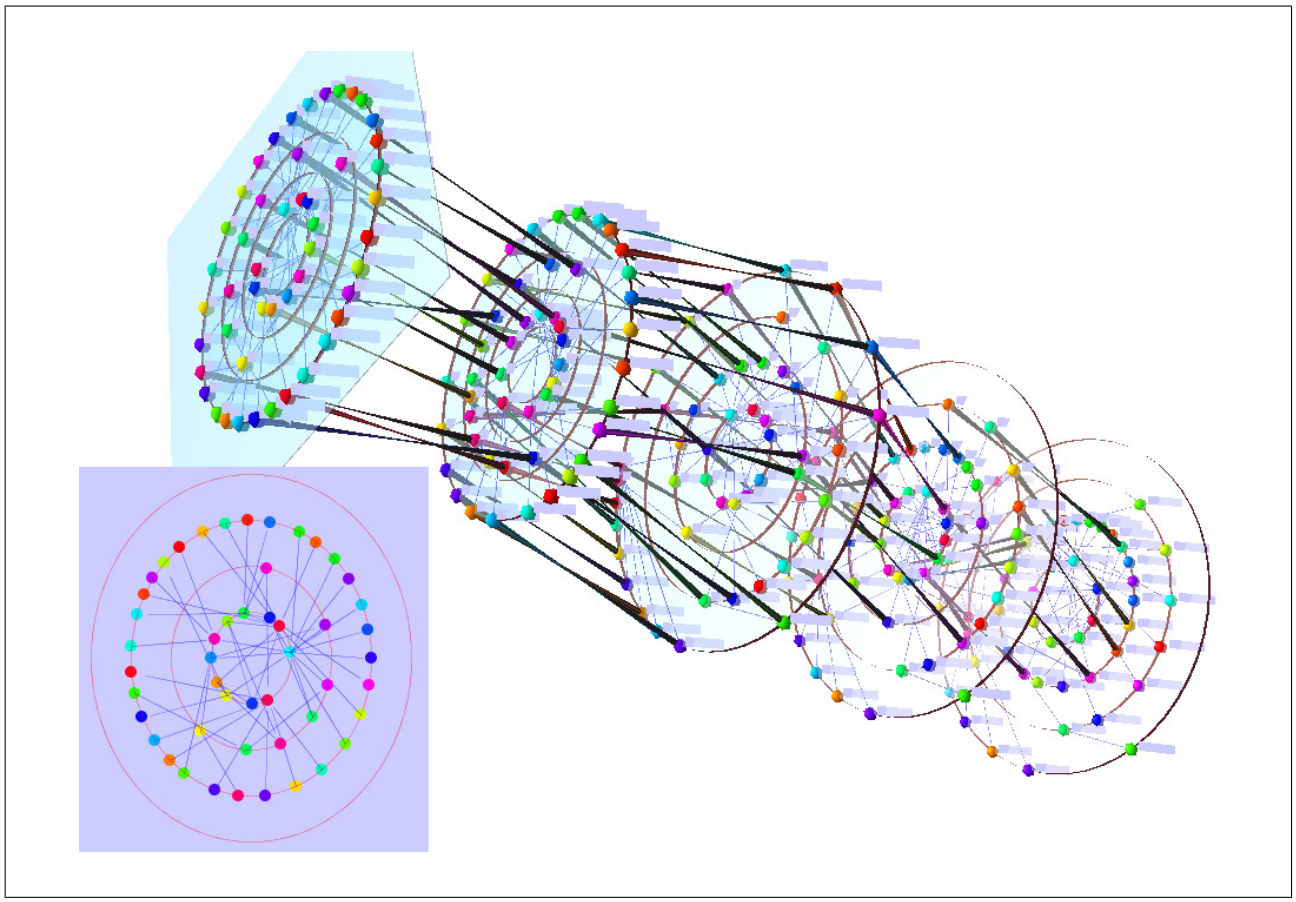

Figure 5: The same PPI network as shown in Figure 4 with the same centrality measures represented with the orbit-based comparison method. In the small window (left) the plane highlighted by the water-layer is displayed in a two dimensional representation

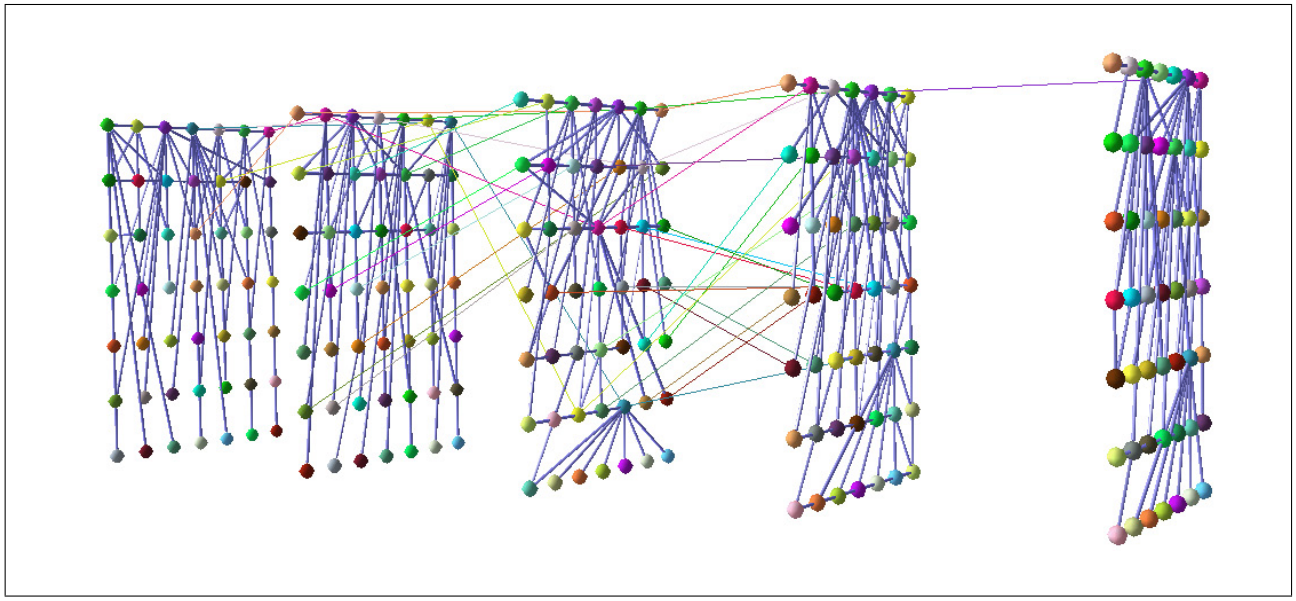

Figure 6: The network used previously with the five centrality measures displayed in a hierarchy-based comparison

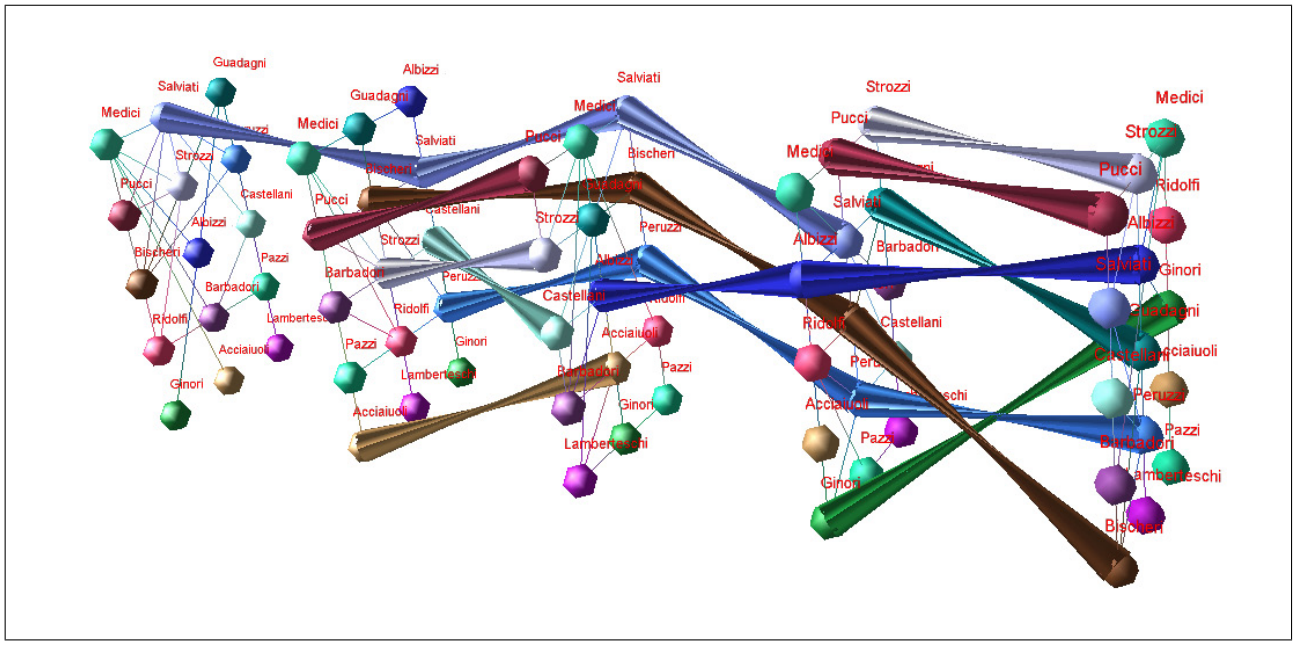

Figure 7: Padgett's Florentine families marital relation data. As in Figure 6 a hierarchy-based comparison of the centralities is used

Bonacich, P. (1972), 'Factoring and weighting approaches to status scores and clique identification', *Journal of Mathematical Sociology* **2**, 113–120.

Borg, I. & Groenen, P. (1997), *Modern Multidimensional Scaling: Theory and Applications*, Springer.

Brandes, U. & Corman, S. R. (2002), Visual unrolling of network evolution and the analysis of dynamic discourse, *in* 'Proceedings of the IEEE Symposium on Information Visualization 2002 (InfoVis '02)', pp. 145–151.

Brandes, U., Dwyer, T. & Schreiber, F. (2004), 'Visual understanding of metabolic pathways across organisms using layout in two and a half dimensions', *Journal of Integrative Bioinformatics* **1**, e0002.

Brandes, U., Kenis, P. & Wagner, D. (2003), 'Communicating centrality in policy network drawings', *IEEE Transactions on Visualization and Computer Graphics* **9**(2), 241–253.

Brandes, U. & Wagner, D. (2004), *visone - Analysis and Visualization of Social Networks*, in Jünger & Mutzel (2004), pp. 321–340.

Dwyer, T. & Eades, P. (2002), Visualising a fund manager flow graph with columns and worms, *in* 'Proceedings of the 6th International Conference on Information Visualisation (IV '02)', IEEE Computer Society Press, pp. 147–152.

Dwyer, T. & Eckersley, P. (2004), *WilmaScope - A 3D Graph Visualization System*, in Jünger & Mutzel (2004), pp. 55–75.

Eades, P. & Whitesides, S. (1994), 'Drawing graphs in two layers', *Theoretical Computer Science* **131**(2), 361–374.

Fell, D. A. & Wagner, A. (2000), 'The small world of metabolism', *Nature Biotechnology* **18**, 1121–1122.

Freeman, L. C. (1977), 'A Set of Measures of Centrality Based on Betweenness', *Sociometry* **40**(6), 35–41.

Freeman, L. C. (1979), 'Centrality in social networks: Conceptual clarification', *Social Networks* **1**, 215–239.

Freeman, L. C., Borgatti, S. P. & White, D. R. (1991), 'Centrality in valued graphs: a measure of betweenness based on network flow', *Social Networks* **13**, 141–154.

Fruchterman, T. M. J. & Reingold, E. M. (1991), 'Graph drawing by force-directed placement', *Software - Practice and Experience* **21**(11), 1129–1164.

Gansner, E. R., Koren, Y. & North, S. C. (2004), Graph drawing by stress majorization, *in* 'Proceedings of the 12th International Symposium on Graph Drawing (GD'04)', Vol. 3383 of *Lecture Notes in Computer Science*, Springer, pp. 239–250.

Hahn, M. W., Conant, G. C. & Wagner, A. (2002), Molecular evolution in large genetic networks: Connectivity does not equal importance, Technical Report 02-08-039, Santa Fe Institute.

Inselberg, A. & Dimsdale, B. (1990), Parallel coordinates: A tool for visualizing multidimensional geometry, *in* 'Proceedings of the 1st IEEE Conference on Visualization (Vis '90)', pp. 361–378.

Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. (2001), 'Lethality and centrality in protein networks', *Nature* **411**, 41–42.

Jünger, M. & Mutzel, P., eds (2004), *Graph Drawing Software*, Springer.

Kamada, T. & Kawai, S. (1989), 'An algorithm for drawing general undirected graphs', *Information Processing Letters* **31**(1), 7–15.

Keim, D. A. (2000), 'Designing pixel-oriented visualization techniques: Theory and applications', *IEEE Transactions on Visualization and Computer Graphics* **6**(1), 59–78.

Koike, H. (1993), 'The role of another spatial dimension in software visualization', *ACM Transactions on Information Systems* **11**(3), 266–286.

Koschützki, D., Lehmann, K. A., Peeters, L., Richter, S., Tenfelde-Podehl, D. & Zlotowski, O. (2005), Centrality indices, *in* U. Brandes & T. Erlebach, eds, 'Network Analysis', Vol. 3418 of *LNCS Tutorial*, Springer, pp. 16–61.

Koschützki, D. & Schreiber, F. (2004), Comparison of centralities for biological networks, *in* 'Proceedings of the German Conference on Bioinformatics 2004', Vol. P-53 of *Lecture Notes in Informatics*, Springer, pp. 199–206.

Lienert, G. A. (1973), *Verteilungsfreie Methoden in der Biostatistik*, Verlag Anton Hain.

Ma, H.-W. & Zeng, A.-P. (2003), 'The connectivity structure, giant strong component and centrality of metabolic networks', *Bioinformatics* **19**(11), 1423–1430.

Newman, M. E. J. (2003), A measure of betweenness centrality based on random walks. arXiv cond-mat/0309045.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992), *Numerical Recipes in C*, Cambridge University Press.

Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U. & Eisenberg, D. (2004), 'The database of interacting proteins: 2004 update', *Nucleic Acids Res* **32**(1), 449–451.

Sugiyama, K., Tagawa, S. & Toda, M. (1981), 'Methods for visual understanding of hierarchical system structures', *IEEE Transactions on Systems, Man and Cybernetics* **SMC-11**(2), 109–125.

Wasserman, S. & Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge Univ. Press.

Wuchty, S. (2002), 'Interaction and domain networks of yeast', *Proteomics* **2**, 1715–1723.

Wuchty, S. & Stadler, P. F. (2003), 'Centers of complex networks', *Journal of Theoretical Biology* **223**, 45–53.