

Visualizing Graphs with Node and Edge Labels

Yifan Hu*

AT&T Labs, Shannon Laboratory, 180 Park Ave., Florham Park, NJ 07932.

Abstract. When drawing graphs whose edges and nodes contain text or graphics, such information needs to be displayed without overlaps, either as part of the initial layout or as a post-processing step. The core problem in removing overlaps lies in retaining the structural information inherent in a layout, minimizing the additional area required, and keeping edges as straight as possible. This paper presents a unified node and edge overlap removal algorithm that does well at solving this problem.

1 Introduction

Most existing graph layout algorithms for undirected graph treat nodes as points and edges as straight lines. In practice, both nodes and edges often contain labels or graphics that need to be displayed. Naively incorporating these can lead to labels that overlap (see, e.g., Figure 1), causing the information of some labels to occlude that of others. It is therefore important to remove such overlaps either by taking into account label sizes during layout or as a post processing step. Assuming that the original layout captures significant structural information such as clusters, the goal of any layout that avoids overlaps should be to retain the “shape” of the layout based on point nodes.

The simplest is to scale up the drawing [31] while preserving label size until the labels no longer overlap. This has the advantage of preserving the shape of the layout, but is often impractical due to inconveniently large drawings. Label overlap removal is typically a trade-off between preserving the shape, limiting the area, and keep edges straight, with scaling at one extreme.

1.1 Related work: node label overlap removal

Many techniques to avoid label overlaps have been devised. One approach is to make the node size part of the model of the layout algorithm. For hierarchical layouts, the node size can be naturally incorporated into the algorithm [11, 34]. For symmetric layouts, various authors [5, 17, 29, 36] have extended the spring-electrical model [9, 12] to take into account node sizes, usually as increased repulsive forces. Node overlap removal can also be built into the stress model [27] by specifying the ideal edge length to avoid overlaps along the graph edges. Such heuristics, however, do not always remove all of the node overlaps, and have to be augmented with a post-processing step.

* e-mail: yifanhu@research.att.com

An alternative approach is to remove node overlaps as a post-processing step after the graph is layed-out. Here the trade-off between layout size and preserving the graph’s shape is much more explicit. A number of algorithms have been proposed. For example, the Voronoi cluster busting algorithm [15,30] works by iteratively forming a Voronoi diagram from the current layout and moving each node to the center of its Voronoi cell until no overlaps remain. The idea is that restricting each node to its corresponding Voronoi cell should preserve the relative positions of the nodes. In practice, because of the number of iterations often required and the use of a rectangular bounding box (the latter might be improved by using something like α -shapes [10] to provide a more accurate boundary), the nodes in the final drawing can be homogeneously distributed, bearing little resemblance to the original layout.

Another group of post-processing algorithms is based on maintaining the orthogonal ordering [32] of the initial layout as a way to preserve its shape. A force scan algorithm and variants were proposed [18,21,29,32] based on these constraints. Marriott et al. [8,31] presented a quadratic programming algorithm which removes node overlaps while minimizing node displacement and keeping the orthogonal ordering. Each of these algorithms has a separate pass in both the horizontal and vertical directions. In practice, this asymmetry often results in a layout with a distorted aspect ratio [13].

While preserving orthogonal ordering is certainly important, it is more general to preserve relative proximity relations [20]. Gansner and Hu [13] proposed a node overlap removal algorithm that used a proximity graph as a scaffolding to maintain the proximity relations, and employed a sparse stress model to remove node overlaps. The algorithm was demonstrated to be scalable for large graphs. It produces layout that closely resembles the original layout, but is free of node overlaps, and takes comparatively little additional area.

1.2 Related work: edge label overlap removal or placement

Edge label overlap removal or placement has also been studied by many authors. One of the often studied problems, though not the focus of this paper, is that of edge label placement (ELP), where the geometry of the graph, other than the labels, is assumed to be fixed. The problem is finding the best placement positions for the edge labels to minimize label overlaps, as well as making the association between edges and their corresponding labels unambiguous. This is highly related to the map labeling problem [4,25,33]. Kakoulis and Tollis [23,24,26] proved that the ELP problem is NP hard, and presented an algorithm for labeling the edges of hierarchical drawings. The algorithm was also used for graph drawings of other styles [6,7].

A number of algorithms were proposed in drawing graphs without edge label overlaps. Castell et al. [3] presented a procedure for drawing state charts, with node and edge labels. Edge label overlaps are avoided as part of the layout process by truncating long label strings and allocating enough space vertically between layers. Binucci et al. [1] present MILP models to compute optimal drawings without edge label overlaps, and an exact algorithm and several heuristics

to compute such drawings with minimum area. Klau and Mutzel [28] proposed a branch and cut algorithm which computes optimally labeled orthogonal drawings.

In this paper we study the problem of visualizing undirected graphs with both node and edge labels. We assume that an aesthetic, symmetric layout based on point nodes is available. The task is to adjust the node position so that node labels and edge labels can be placed without overlaps. In doing so, we want to retain the “shape” of the layout. To minimize the additional area required, we allow edges to be bent, but strive to draw them as straight as possible. We present (Section 3) a unified overlap removal algorithm based on a proximity graph of the position of node and edge labels in the original layout. Using this graph as a guide, it iteratively moves the labels to remove overlaps, while keeping the relative positions between them as close to those in the original layout as possible, and edges as straight as possible. The algorithm is similar to the stress model [27] used for graph layout, except that the stress function involves only a sparse selection of all possible node pairs. It is an extension to the node overlap removal algorithm of Gansner and Hu [13] to deal with both node and edge labels. We evaluate our algorithm on two graphs from applications. Finally, Section 4 presents a summary and topics for further study.

2 Background

We use $G = (V, E)$ to denote an undirected graph, with V the set of nodes (vertices) and E edges. We use $|V|$ and $|E|$ for the number of vertices and edges, respectively. We let x_i represent the current coordinates of vertex i in Euclidean space.

The aim of graph drawing is to find x_i for all $i \in V$ so that the resulting drawing gives a good visual representation of the information in the graph. Two popular methods, the spring-electrical model [9, 12], and the stress model [27], both convert the problem of finding an optimal layout to that of finding a minimal energy configuration of a physical system. We describe here the stress model in more detail, as we shall use a similar model for the purpose of label overlap removal in Section 3.

The stress model assumes that there are springs connecting all nodes of the graph, with the ideal spring length equal to the graph theoretical distance between nodes. The energy of this spring system is

$$\sum_{i \neq j} w_{ij} (\|x_i - x_j\| - d_{ij})^2, \quad (1)$$

where d_{ij} is the graph theoretical distance between vertices i and j , and w_{ij} is a weight factor, typically $1/d_{ij}^2$. The layout that minimizes the above stress energy is an optimal layout of the graph. A robust technique to find an optimal of this model is stress majorization, where the cost function (1) is bounded by a series of quadratic functions from above, and the process of finding an optimum becomes that of solving a series of linear systems [14].

In the stress model, the graph theoretical distance between all pairs of vertices has to be calculated, leading to quadratic complexity in the number of vertices. There have been attempts (e.g., [2, 14]) to simplify the stress function by considering only a sparse portion of the graph. Our experience, however, is that these techniques can fail to yield good layouts on real-life graphs. Therefore, algorithms based on a spring-electrical model employing a multilevel approach and an efficient approximation scheme for long range repulsive forces [16, 19, 35] are still the most efficient choices when laying out large graphs without consideration of the node or edge labels. It is worth pointing out that the label overlap removal algorithm proposed in this paper works with any symmetric layout, regardless of how it is generated.

3 A Unified Model for Label Overlap Removal

Our goal is to remove label overlaps while preserving the shape of the initial layout by maintaining the proximity relations [20] among the nodes. Recently we proposed an efficient algorithm PRISM [13] to remove node overlaps based on the *proximity stress model*. We first summarize this model, then describe how to extend it to solve the more general problem of overlap removal for both node and edge labels.

3.1 The PRISM algorithm for node overlap removal

We first set up a rigid “scaffolding” structure so that while vertices can move around, their relative positions are maintained. This scaffolding is constructed using an approximation to proximity graph [22], the Delaunay triangulation (DT).

We then check every edge in DT and see if there are any node overlap along that edge. Let W_i and H_i denote the half width and height of the node i , and $x_i^0(1)$ and $x_i^0(2)$ the current X and Y coordinates of this node. If i and j form an edge in the DT, we calculate the *overlap factor* of these two nodes,

$$t_{ij} = \max \left(\min \left(\frac{W_i + W_j}{|x_i^0(1) - x_j^0(1)|}, \frac{H_i + H_j}{|x_i^0(2) - x_j^0(2)|} \right), 1 \right). \quad (2)$$

For nodes that do not overlap, $t_{ij} = 1$. For nodes that do overlap, such overlaps can be removed if we expand the edge by this factor. Therefore we want to generate a layout such that an edge in the proximity graph has the ideal edge length close to $t_{ij}\|x_i^0 - x_j^0\|$. In other words, we want to minimize the following stress function

$$\sum_{(i,j) \in E_P} w_{ij} (\|x_i - x_j\| - d_{ij})^2 \quad (3)$$

Here $d_{ij} = s_{ij}\|x_i^0 - x_j^0\|$ is the ideal distance for the edge $\{i, j\}$, s_{ij} is a scaling factor related to the overlap factor t_{ij} (see (4)), $w_{ij} = 1/d_{ij}^2$ is a scaling factor,

and E_P is the set of edges of the proximity graph. We call (3) the *proximity stress model*.

Because DT is a planar graph, which has no more than $3|V| - 3$ edges, the above stress function has no more than $3|V| - 3$ terms. Furthermore, because DT is rigid, it provides a good scaffolding that constrains the relative position of the vertices and helps to preserve the global structure of the original layout.

It is important that we do not attempt to remove overlaps in one iteration by minimizing the above model with $s_{ij} = t_{ij}$ to avoid high localized stress causing deviation to the original layout, hence we damp the overlap factor by setting

$$s_{ij} = \min(t_{ij}, s_{\max}) \quad (4)$$

and try to remove overlaps a little at a time. Here $s_{\max} > 1$ is a number limiting the amount of overlap we are allowed to remove in one iteration. Typically $s_{\max} = 1.5$.

After minimizing (3), we arrive at a layout that may still have node overlaps. We then regenerate the proximity graph using DT and calculate the overlap factor along the edges of this graph, and redo the minimization. This forms an iterative process that ends when there are no more overlaps along the edges of the proximity graph. At this stage there may still be node overlaps between nodes that do not constitute an edge of the DT, so a scan-line algorithm [8] is used to find all overlaps, and the proximity graph is augmented with additional edges, where each edge consists of a pair of nodes that overlap. We then re-solve (3). This process is repeated until the scan-line algorithm finds no more overlaps. We called the algorithm PRISM (PRoximity Stress Model).

Figure 1 shows the result of using PRISM on graph representing the routes of Olympic torch relay from summer 1936 to summer 2008. This graph is taken from the GD'08 contest. Nodes represent countries, and for a country that has hosted the Olympics, the year (e.g., "S2008" means summer 2008) that the Olympics was hosted is also included in the node label. The edges are color codes, represented different years and seasons the relays took place. Warm colors are used for summer games and cool color winter games. Edge labels are used to highlight the year that edge was traversed. In the figure, node overlaps are removed using PRISM. However, due to the presence of edge labels, there are a lot of overlaps among edge labels and between node and edge labels. As a result, it is very difficult to follow the relay routes, even with the help of the coloring scheme. This and other applications call for an effective algorithm for removal node as well as edge label overlaps.

3.2 The penalized proximity stress model for node and edge label overlap removal

As Figure 1 demonstrates, when there are both node and edge labels, simply removing node overlaps is not sufficient. Instead, we need an algorithm to take into account the overlaps between both node and edge labels.

the solution of sparse linear systems relating to the Laplacian of the proximity graph. In the case of our model, the linear system to be solved repeatedly is

$$(L_w + \rho L_e) x = L_{w,d} x^0 \quad (6)$$

where x^0 is the current layout, and x is the new layout that improves the stress and penalty terms. The weighted Laplacian matrix L_w has elements

$$(L_w)_{ij} = \begin{cases} \sum_{\{i,l\} \in E_P} w_{il}, & i = j \\ -w_{ij}, & i \neq j \end{cases}$$

The matrix L_e comes from penalty term and is defined as

$$(L_e)_{ij} = \begin{cases} w_i, & i = j, i \in V_e \\ 0.25 \sum_{k \in N(i) \cap V_e} w_k, & i = j, i \notin V_e \\ -0.5w_i, & i \in V_e, \{i, j\} \in E_a \\ -0.5w_j, & j \in V_e, \{i, j\} \in E_a \\ 0.25w_k, & \{i, k\} \in E_a, \{j, k\} \in E_a, k \in V_e \end{cases}$$

with $N(i)$ the set of neighboring nodes of i in the expanded graph G_a , and matrix $L_{w,d}$ has elements

$$(L_{w,d})_{ij} = \begin{cases} \sum_{\{i,l\} \in E_P} w_{il} d_{il} / \|y_i - y_l\|, & i = j \\ -w_{ij} d_{ij} / \|y_i - y_j\|, & i \neq j \end{cases}$$

The addition of matrix L_e to L_w makes the overall matrix on the left hand side of (5) somewhat denser, by introducing at most $6|E|$ extra entries if every edge has an edge label. However, as long as the original graph is sparse, (6) is still a sparse system, and can be solved using conjugate gradient algorithm efficiently. Algorithm 1 gives a detailed description of the ePRISM algorithm.

Algorithm 1 ePRISM: an algorithm for node and edge label removal

Input: coordinates for each node and edge labels, x_i^0 , and bounding box width and height $\{W_i, H_i\}$, $i = 1, 2, \dots, |V_a|$.

repeat

 Form a proximity graph G_P of x^0 by Delaunay triangulation.

 Find the overlap factors (2) along all edges in G_P .

 Solve the penalized proximity stress model (5) for x . Set $x^0 = x$.

until (convergence)

repeat

 Form a proximity graph G_P of x^0 by Delaunay triangulation.

 Find all node overlaps using a scan-line algorithm. Augment G_P with edges from node pairs that overlap.

 Find the overlap factor (2) along all edges of G_P .

 Solve the proximity stress model (3) for x . Set $x^0 = x$.

until (convergence)

Remove any remaining overlaps using PRISM.

Unlike PRISM, where the only objective is to remove overlaps, in ePRISM, we have the dual objectives of removing overlaps as well as keeping edges straight. Therefore, while in PRISM the lack of overlaps is the measure of convergence, here, in the two main loops of the ePRISM algorithm, we define convergence as $\|x - x_0\|/\|x_0\| < \epsilon$, with $\epsilon = 0.005$, and limit the number of iterations to 1000. Because of the dual objectives, it is possible that after the two main loops have converged, some overlaps may still exist, for example among edge labels on multiple edges between the same pair of nodes. Therefore we use PRISM algorithm to remove any overlaps that may still remain. We set the penalty parameter ρ to 4.

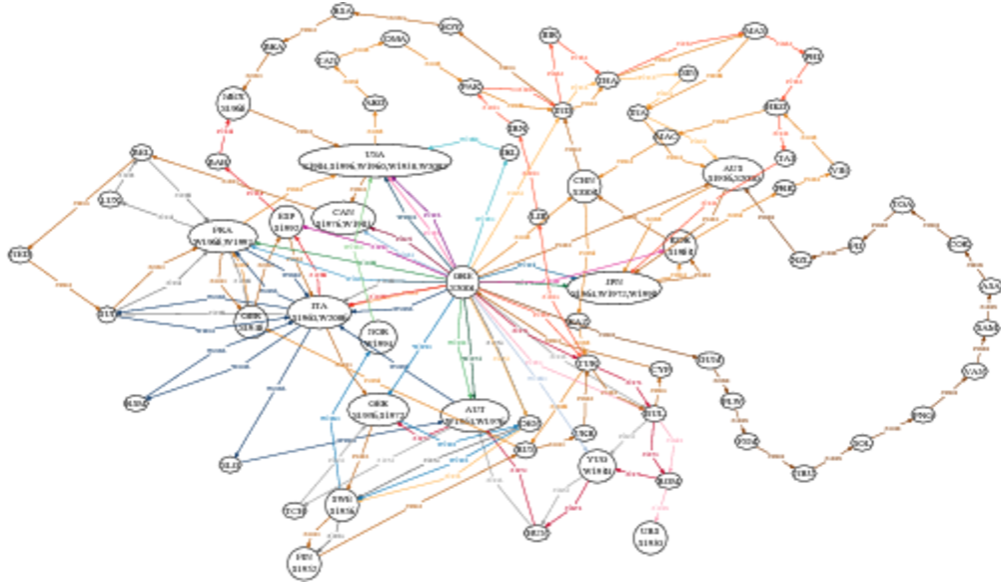


Fig. 3. The Olympic torch relay graph after node and edge label overlaps are removed using ePRISM.

The ePRISM algorithm has a similar computational complexity to PRISM, when the latter is applied to the expanded graph G_a . However we found that ePRISM tends to take more iterations to converge in practice, due to the difficulty in resolving the conflicting objective of removing label overlaps and keeping edges straight.

Figure 3 gives the result of ePRISM on the Olympic torch relay graph. Clearly, the majority of edges are now straight, and those edges that are not straight have milder bends. Some of these bends are unavoidable, such as when there are multiple edges between the same pair of nodes. In general, compared with Figure 2, Figure 3 is more pleasing to look at and easier to follow due to the great reduction of edge bends.

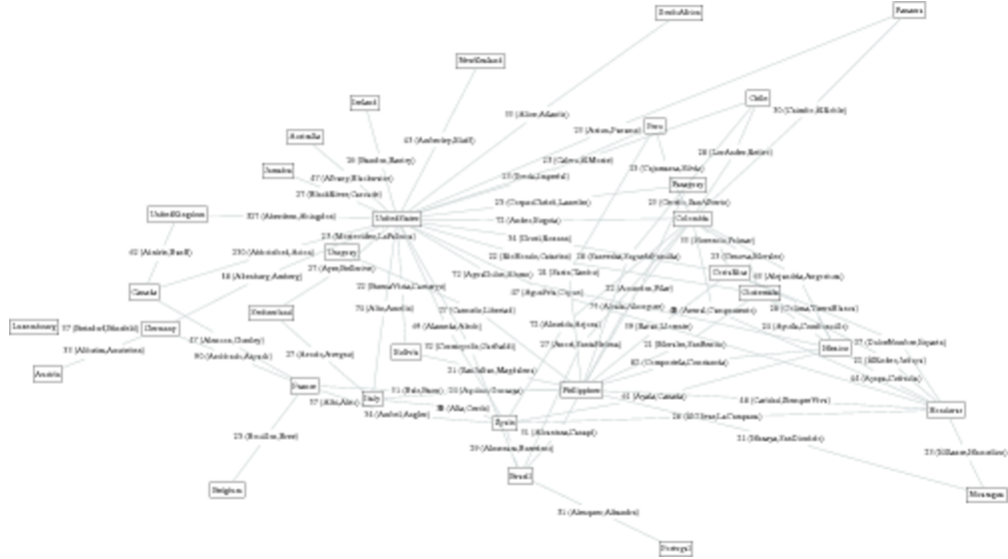


Fig. 4. Visualization of countries sharing same city names. Two countries sharing more than 20 city names are linked by an edge, with the number of shared names, as well as two such names, shown as edge label. Node and edge label overlaps are removed using ePRISM.

Finally, we apply ePRISM to another application. Here, we visualize countries and how they relate. Two countries are related if there are more than 20 cities with the same name that exist in both countries. From the drawing, we can see that United States is the country with the most connections to other countries, reflecting the fact that most people in the country emigrated from another country in the last three centuries, and that cities in the United States are often named by the immigrants using names of cities in their home countries. The United States and the United Kingdom have the greatest number of cities with the same name (327), followed by Canada (250), Germany (86), and Italy (74). The United States also shares city names with a large group of countries in South America, seen on the right hand side of the figure. Canada, on the other hand, has the United States, the United Kingdom and France as its three closest connected countries, reflecting the historical tie of Canada to these three countries.

4 Conclusions and Future Work

The main contribution of this paper is a new algorithm for removing both node and edge overlaps, based on a penalized proximity stress model. The algorithm is shown to produce layouts displaying both node and edge labels without overlaps, and is aesthetic, with mostly straight edges.

For future work, we would like to investigate a combination of this algorithm and edge label placement (ELP) techniques, so that those edge labels that can be placed within the current geometry without introducing ambiguity or overlaps are dealt with separately to those that have to be treated using the algorithm proposed in this paper.

Acknowledgments

The author would like to thank Emden Gansner and Stephen North for helpful discussions.

References

1. B., C., D., W., Liotta, G., Nonato, M.: Orthogonal drawings of graphs with vertex and edge labels. *Computational Geometry: Theory and Applications* 32(2), 71–114 (2005)
2. Brandes, U., Picj, C.: An experimental study on distance based graph drawing. In: GD'08: Proceedings of the Symposium on Graph Drawing (2008 (to appear))
3. Castelló, R., Mili, R., Tollis, I.G.: An algorithmic framework for visualizing stat-charts. In: GD'00: Proceedings of the 8th International Symposium on Graph Drawing. pp. 139–149. Springer-Verlag, London, UK (2001)
4. Christensen, J., Marks, J., Shieber, S.: An empirical study of algorithms for point-feature label placement. *ACM Transaction On Graphics* 14, 203–232 (1995)
5. Chuang, J.H., Lin, C.C., Yen, H.C.: Drawing graphs with nonuniform nodes using potential fields. In: Proc. 11th Intl. Symp. on Graph Drawing. LNCS, vol. 2012, pp. 460–465. Springer-Verlag (2003)
6. Dogrusöz, U., Kakoulis, K.G., Madden, B., Tollis, I.G.: Edge labeling in the graph layout toolkit. In: GD '98: Proceedings of the 6th International Symposium on Graph Drawing. pp. 356–363. Springer-Verlag, London, UK (1998)
7. Dogrusöz, U., Kakoulis, K.G., Madden, B., Tollis, I.G.: On labeling in graph visualization. *Information Science* 177(12), 2459–2472 (2007)
8. Dwyer, T., Marriott, K., Stuckey, P.J.: Fast node overlap removal. In: Proc. 13th Intl. Symp. Graph Drawing (GD '05). LNCS, vol. 3843, pp. 153–164. Springer (2006)
9. Eades, P.: A heuristic for graph drawing. *Congressus Numerantium* 42, 149–160 (1984)
10. Edelsbrunner, H., Mücke, E.P.: Three-dimensional alpha shapes. *ACM Trans. on Graphics* 13(1), 43–72 (1994)
11. Friedrich, C., Schreiber, F.: Flexible layering in hierarchical drawings with nodes of arbitrary size. In: Proceedings of the 27th conference on Australasian computer science (ACSC2004). pp. 369–376. Australian Computer Society (2004)
12. Fruchterman, T.M.J., Reingold, E.M.: Graph drawing by force directed placement. *Software - Practice and Experience* 21, 1129–1164 (1991)
13. Gansner, E.R., Hu, Y.F.: Efficient node overlap removal using a proximity stress model. In: GD'08: Proceedings of the Symposium on Graph Drawing (2008 (to appear))
14. Gansner, E.R., Koren, Y., North, S.C.: Graph drawing by stress majorization. In: Proc. 12th Intl. Symp. Graph Drawing (GD '04). LNCS, vol. 3383, pp. 239–250. Springer (2004)

15. Gansner, E.R., North, S.C.: Improved force-directed layouts. In: Proc. 6th Intl. Symp. Graph Drawing (GD '98). LNCS, vol. 1547, pp. 364–373. Springer (1998)
16. Hachul, S., Jünger, M.: Drawing large graphs with a potential field based multilevel algorithm. In: Proc. 12th Intl. Symp. Graph Drawing (GD '04). LNCS, vol. 3383, pp. 285–295. Springer (2004)
17. Harel, D., Koren, Y.: A fast multi-scale method for drawing large graphs. *J. Graph Algorithms and Applications* 6, 179–202 (2002)
18. Hayashi, K., Inoue, M., Masuzawa, T., Fujiwara, H.: A layout adjustment problem for disjoint rectangles preserving orthogonal order. In: Proc. 6th Intl. Symp. Graph Drawing (GD '98). LNCS, vol. 1547, pp. 183–197. Springer (1998)
19. Hu, Y.F.: Efficient and high quality force-directed graph drawing. *Mathematica Journal* 10, 37–71 (2005)
20. Hu, Y.F., Koren, Y.: Extending the spring-electrical model to overcome warping effects (2008), preprint
21. Huang, X., Lai, W.: Force-transfer: A new approach to removing overlapping nodes in graph layout. In: Proc. 25th Australian Computer Science Conference. pp. 349–358 (2003), URL citeseer.ist.psu.edu/564050.html
22. Jaromczyk, J.W., Toussaint, G.T.: Relative neighborhood graphs and their relatives. *Proc. IEEE* 80, 1502–1517 (1992)
23. Kakoulis, K.G., Tollis, I.G.: An algorithm for labeling edges of hierarchical drawings. In: GD '97: Proceedings of the 5th International Symposium on Graph Drawing. pp. 169–180. Springer-Verlag, London, UK (1997)
24. Kakoulis, K.G., Tollis, I.G.: On the edge label placement problem. In: GD '96: Proceedings of the Symposium on Graph Drawing. pp. 241–256. Springer-Verlag, London, UK (1997)
25. Kakoulis, K.G., Tollis, I.G.: A unified approach to labeling graphical features. In: SCG '98: Proceedings of the fourteenth annual symposium on Computational geometry. pp. 347–356. ACM, New York, NY, USA (1998)
26. Kakoulis, K.G., Tollis, I.G.: On the complexity of the edge label placement problem. *Computational Geometry Theory Application* 18(1), 1–17 (2001)
27. Kamada, T., Kawai, S.: An algorithm for drawing general undirected graphs. *Information Processing Letters* 31, 7–15 (1989)
28. Klau, G.W., Mutzel, P.: Combining graph labeling and compaction. In: GD '99: Proceedings of the 7th International Symposium on Graph Drawing. pp. 27–37. Springer-Verlag, London, UK (1999)
29. Li, W., Eades, P., Nikolov, N.: Using spring algorithms to remove node overlapping. In: Proc. Asia-Pacific Symp. on Information Visualisation. pp. 131–140 (2005)
30. Lyons, K.A., Meijer, H., Rappaport, D.: Algorithms for cluster busting in anchored graph drawing. *J. Graph Algorithms and Applications* 2(1) (1998)
31. Marriott, K., Stuckey, P.J., Tam, V., He, W.: Removing node overlapping in graph layout using constrained optimization. *Constraints* 8(2), 143–171 (2003)
32. Misue, K., Eades, P., Lai, W., Sugiyama, K.: Layout adjustment and the mental map. *J. Vis. Lang. Comput.* 6(2), 183–210 (1995)
33. do Nascimento, H.A.D., Eades, P.: User hints for map labeling. *Journal of Visual Languages and Computing* 19(1), 39–74 (2008)
34. Sugiyama, K., Tagawa, S., Toda, M.: Methods for visual understanding of hierarchical systems. *IEEE Trans. Systems, Man and Cybernetics SMC-11*(2), 109–125 (1981)
35. Walshaw, C.: A multilevel algorithm for force-directed graph drawing. *J. Graph Algorithms and Applications* 7, 253–285 (2003)

36. Wang, X., Miyamoto, I.: Generating customized layouts. In: GD '95: Proceedings of the Symposium on Graph Drawing. LNCS, vol. 1027, pp. 504–515. Springer (1995)