Crossing Minimization Problems of Drawing Bipartite Graphs in Two Clusters

Lanbo Zheng, Le Song, and Peter Eades

National ICT Australia,
and
School of Information Technologies, University of Sydney,Australia
Email: {lzheng,lsong,peter}@it.usyd.edu.au

Abstract

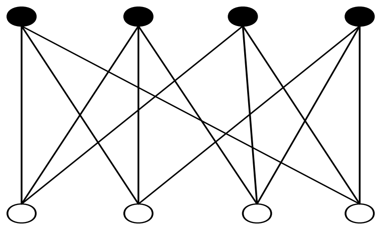
The crossing minimization problem is a classic and very important problem in graph drawing (Pach, Tòth 1997); the results directly affect the effectiveness of the layout, especially for very large scale graphs. But in many cases crossings cannot be avoided. In this paper we present two models for bipartite graph drawing, aiming to reduce crossings that cannot be avoided in the traditional bilayer drawings. We characterize crossing minimization problems in these models, and prove that they are $\mathcal{NP}$-complete.

1 Motivation

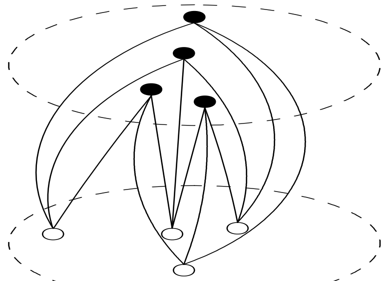
Digraphs are widely used in applications to model dependency relationships. Examples include PERT diagrams, *is-a* hierarchies, and subroutine-call graphs (Battista et.al 1999). For a given bipartite graph $G = (A \cup B, E)$, vertex set A and B are put on two parallel lines, open straight line segments are used for edges. The $2-layer$ crossing problem is, if the vertex ordering of one line is fixed, choosing a vertex ordering of the other line that minimize the number of edge crossings. This problem is critical to the readability of digraphs drawn using the widely used Sugiyama heuristic (Eades, Sugiyama 1990). Several variations of this problem are $\mathcal{NP}$-complete (Eades, Worlmald 1994, Gary, Johnson 1983). Heuristics such as Barycenter (Sugiyama et al. 1981), Median (Eades et al. 1994) and Random-Key (Nagamochi 2003) have achieved satisfactory solutions.

In the Sugiyama heuristic, since the vertices are restricted to straight lines, some kinds of crossings are impossible to avoid (see Fig.1 (a)). Readability decreases, especially for dense graphs. Our study is motivated by this problem. In our model, the two parts of the bipartite graph are drawn in two disjoint regions, as in Fig.1 (b); vertices are less constrained and crossings can be reduced.

Biedl et al. (Biedl 1998) and Föβmeier et al. (Föβmeier 1997) have investigated similar problems. Biedl proposes necessary and sufficient conditions for drawing planar bipartite graphs in a so called *HH-drawing*, and provides a linear time algorithm. Föβmeier proposes a drawing model called *two-sides-free model* in which one part of the vertex set is fixed in a row, and the other part can be drawn without restrictions above or below the row. They proved that it is $\mathcal{NP}$-hard to determine whether a given planar bipartite graph has a such drawing. In contrast to these two papers, we consider nonplanar graphs.



(a) drawing vertices on two lines



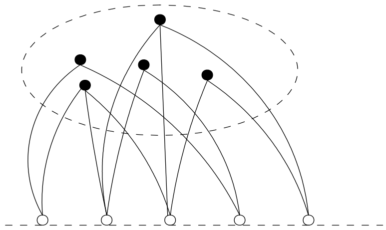
(b) drawing vertices in two regions

Figure 1: crossings in two different drawings

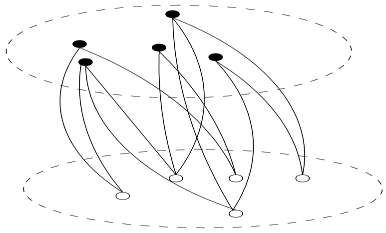
In this paper we propose two drawing models for a given bipartite graph $G = (A \cup B, E)$, as follows. A *One-Layer-One-Cluster (OLOC)* drawing of G has a horizontal line L and a region R above L. Vertices of A are drawn on L and vertices of B are drawn in R. Edges are vertically monotone. Edges may cross, but any pair crosses at most once, and edges with same common end points will not cross. (see Fig.2 (a)). A *Two-Cluster (TC)* drawing has two disjoint regions R_1 and R_2, vertex set A are drawn in R_1 and vertex set B are drawn in R_2. Edges in it have the same characters as edges in OLOC drawings (see Fig.2 (b)). In the following sections, we start from the simpler OLOC crossing problem, prove it is $\mathcal{NP}$-complete, then step to the TC crossing problem and prove it following the same way. Noticeably, our proofs are similar to Gary and Johnson's proof in (Gary, Johnson 1983), but they are necessary and important and the results provide a direction for our future researches.

2 One-Layer-One-Cluster Model

In the *One-Layer-One-Cluster* model, vertex set A can be placed freely on a line L, and vertex set B can be set freely in a region R. Our aim is to find an *OLOC* drawing with a minimum number of edge crossings. We prove it is an $\mathcal{NP}$-complete problem. In order to simplify the proof, we consider region R as the half space above L.



(a) a One-Layer-One-Cluster drawing



(b) a Two-Cluster drawing

Figure 2: two drawing models in this paper

We state this problem as a decision problem in the usual complexity theory style as follows:

OLOC CROSSING MINIMIZATION(OLOCCM)

Instance: A bipartite graph $G = (A \cup B, E)$, a positive integer K.

Question: Is there a *One-Layer-One-Cluster* drawing G' of G that has at most K edge crossings?

Theorem 2.1 *OLOC Crossing Minimization is* $\mathcal{NP}$*-complete.*

Proof: It is easy to see that OLOCCM $\in \mathcal{NP}$ since a nondeterministic algorithm need only to find an OLOC embedding for G and check in polynomial time that whether the crossing number of this embedding is at most K. To prove OLOC Crossing Minimization is $\mathcal{NP}$-Complete, we transform Optimal Linear Arrangement (Gary, Johnson 1976) to it.

OPTIMAL LINEAR ARRANGEMENT(OLA)

Instance: Graph $G = (V, E)$, positive integer K

Question: Is there a one-to-one function $f : V \rightarrow \{1, 2, \ldots, |V|\}$ such that $\sum_{(u,v)\in E} |f(v) - f(u)| \leq K$?

For an instance $G = (V, E)$, K of OLA where $V = \{v_1, v_2, \ldots, v_n\}$, we construct an instance $G' = (V_1, V_2 \cup V_3, E_1 \cup E_2 \cup E_3), K'$ of OLOCCM as follows:

$V_1 = \{u_i : 1 \leq i \leq n\}$,
$V_2 = \{w_i : 1 \leq i \leq n\}$,
$V_3 = \{z\}$,
$E_1 = \{|E|^2 \quad \textit{copies of} \quad (u_i, w_i) : 1 \leq i \leq n\}$,
$E_2 = \{|E|^2 \quad \textit{copies of} \quad (u_i, z) : 1 \leq i \leq n\}$,
$E_3 = \{(u_i, w_j) : i > j \quad \textit{and} \quad (v_i, v_j) \in E\}$,
$K' = |E|^2(K - |E|) + (|E|^2 - 1)$.

Here, part A is V_1, and part B is $V_2 \cup V_3$. The edge set includes two edge sets $E_1 \cup E_2$ of multiple edges, and one edge set E_3 of simple edges. Without loss of generality, we consider multiple edges as one edge, but maintain the number of edges. We must show that the answer for OLA is yes if and only if the answer for OLOCCM is yes. Before this, we need to characterize G'.

Notice that, each vertex $u_i \in V_1$ connects two multiple edges from E_1 and E_2 respectively. We denote the union of these edge sets by $arch_i$ because of the shape they naturally generate (see Fig. 3).

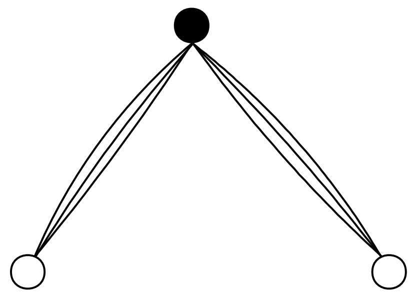
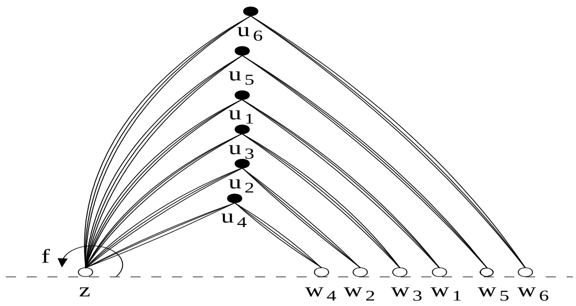
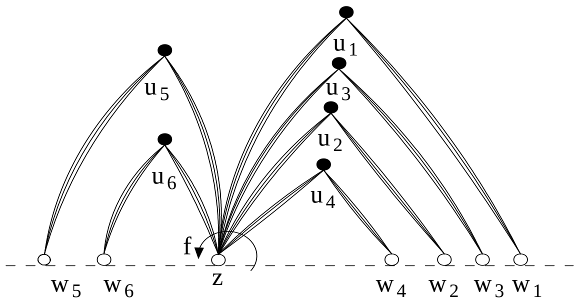


Figure 3: an arch



(a) single tower



(b) double tower

Figure 4: two kinds of tower structure (edge set E_3 is omitted)

To finish our proof we need the following definition, lemmas and corollary.

definition 2.1 *An embedding of* $G' = (V_1, V_2 \cup V_3, E_1 \cup E_2 \cup E_3)$ *is a tower if there are no crossings between arches.*

For a tower embedding, if all vertices in V_2 are on the same side of z, we call it a single tower embedding (see Fig 4 (a)), otherwise we call it a double tower embedding (see Fig 4 (b)).

Suppose first that the desired ordering function f exists. We need to find an embedding of G' with at most K' crossings.

Noticeably, all edges in E_2 are connected to z, they naturally generate a cyclic order around z. We let f be the order in counterclockwise around z, and the following proofs are based on this relationship.

Lemma 2.1 *There exists a single tower embedding* H' *of* G' *with at most*

$$\sum_{(v_i,v_j)\in E}(|f(v_i) - f(v_j)| - 1)|E|^2 + (|E|^2 - 1)$$

crossings.

Proof: Suppose that H' is a single tower embedding of G'. There are two kinds of crossings in H': one is

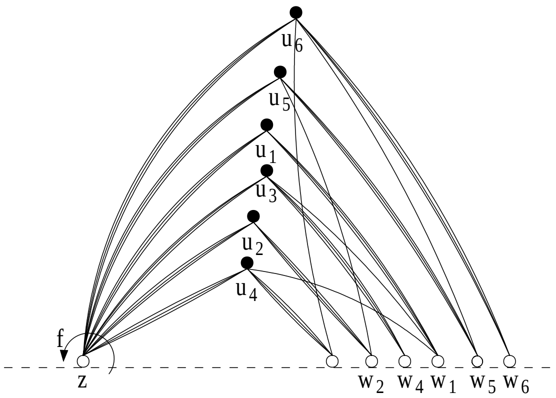


Figure 5: crossings in a single tower embedding

crossings between simple edges and arches, the other is crossings between simple edges. For the first kind, the number of crossings is $\sum_{e\in E_3} C_e|E|^2$ where C_e is the number of arches that e crosses. We can route all simple edges of E_3 in the shortest paths through the dual graph of $H' - E_3$, then each edge (u_i, w_j) crosses $(|f(v_i) - f(v_j)| - 1)$ arches resulting in $(|f(v_i) - f(v_j)| - 1)|E|^2$ crossings. For the second part, the number of crossings between simple edges is at most $|E|^2 - 1$. So there is a H' that has at most $\sum_{(v_i,v_j)\in E}(|f(v_i) - f(v_j)| - 1)|E|^2 + (|E|^2 - 1)$ crossings (see Fig. 5). □

Since $\sum_{(v_i,v_j)\in E} |f(v_i) - f(v_j)| \leq K$, then the number of crossings of H' is at most K'. So H' is a desired embedding for G'.

Corollary 2.1 *If H' is a single tower embedding of G', then it has at least*

$$\textstyle\sum_{(v_i,v_j)\in E}(|f(v_i) - f(v_j)| - 1)|E|^2$$

crossings.

Proof: The monotonicity requirement implies that no edge crosses the line L. Note that L together with H divides the plane into a number of regions. A curve from u_i to w_j, if it does not cross L, must traverse at least $(|f(v_i) - f(v_j)| - 1)|E|^2 - 1$ regions. Thus (u_i, w_j) crosses at least $(|f(v_i) - f(v_j)| - 1)|E|^2$ edges. So the number of edge crossings in H' are at least $\sum_{(v_i,v_j)\in E}(|f(v_i) - f(v_j)| - 1)|E|^2$. □

Now, suppose G' has an embedding with less than K' crossings, we need to show the desired function f exists for G. To prove this, we need the following lemma.

Lemma 2.2 *If there is an embedding H of G' with K' crossings, then there is a single tower embedding H' of G' with K' crossings.*

Proof: Suppose that H is an embedding of G' with K' crossings. Note that if two arches cross, then they generate at least $|E|^4$ crossings which is larger than K', thus H must be a tower. If it is a single tower then H' is H. If it is a double tower, then we consider the following transformation of H to form H'. We move the vertices on one side of z to the other side of it while keeping the cyclic order of edges (u_i, z) around z (Fig. 4 (a) and (b) indicates the transformation clearly). Since the cyclic order is not changed, this transformation does not increase the number of edge crossings and H becomes a single tower embedding H' with K' edge crossings. □

Now recalling corollary 2.1, we have:

$$\sum_{(v_i,v_j)\in E} (|f(v_i) - f(v_j)| - 1)|E|^2 \leq$$

$$K' = (K - |E|)|E|^2 + (|E|^2 - 1)$$

which implies that

$$\sum_{(v_i,v_j)\in E} |f(v_i) - f(v_j)| \leq K$$

Thus there is a desired one to one function f for G as well. □

From above result, we conclude that minimizing crossings of a *One-Layer-One-Cluster* drawing is intractable and approximation algorithms are needed to find a satisfactory solution.

3 Two-Cluster Model

In the *Two-Cluster* model, vertex sets A and B are drawn in two disjoint regions R_1 and R_2. In each region, vertices can be set freely. Our aim is to find a *TC* drawing for a given bipartite graph with minimum number of crossings. This problem is also $\mathcal{NP}$-complete. To simplify our proof, we assume R_1 and R_2 are two parts of a plane divided by a borderline l.

We state this problem as the usual complexity theory style.

TC CROSSING MINIMIZATION(TCCM)

Instance: A bipartite graph $G = (A \cup B, E)$, a positive integer K.

Question: Is there a *Two-Cluster* drawing G' of G that has at most K edge crossings?

Theorem 3.1 *TC Crossing Minimization is $\mathcal{NP}$-complete.*

Proof: It is easy to see that TCCM $\in \mathcal{NP}$, and we also transform Optimal Linear Arrangement to it.

Let $G = (V, E), V = (v_1, v_2, \ldots, v_n)$, a positive integer K be an instance of OLA. We construct an instance $G' = (V_1 \cup V_2, V_3 \cup V_4, E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5), K'$ of TCCM as follows:

$V_1 = \{u_i : 1 \leq i \leq n\}$,
$V_2 = \{z_1\}$,
$V_3 = \{w_i : 1 \leq i \leq n\}$,
$V_4 = \{z_2\}$,
$E_1 = \{|E|^2 \ \ copies \ \ of \ \ (u_i, z_2) : 1 \leq i \leq n\}$,
$E_2 = \{|E|^2 \ \ copies \ \ of \ \ (w_i, z_1) : 1 \leq i \leq n\}$,
$E_3 = \{|E|^2 \ \ copies \ \ of \ \ (u_i, w_i) : 1 \leq i \leq n\}$,
$E_4 = \{|E|^4 \ \ copies \ \ of \ \ (z_1, z_2)\}$,
$E_5 = \{(u_i, w_j) : i > j \ \ and \ \ (v_i, v_j) \in E\}$,
$K' = (K - |E|)|E|^2 + (|E|^2 - 1)$.

Here part A consists of $V_1 \cup V_2$, part B consists of $V_3 \cup V_4$, edge set consists of four edge sets $E_1 \cup E_2 \cup E_3 \cup E_4$ of multiple edges and an edge set E_5 of simple edges. We need to prove that there is an embedding of G' with the number of crossings less or equal to K' if and only if there is a one to one function f for G that $\sum_{(v_i,v_j)\in E} |f(v_i) - f(v_j)| \leq K$.

Before this, we will characterize G'.

Notice that, every three multiple edges $(u_i, z_2), (w_i, z_1), (u_i, w_i)$ from $E_1 \cup E_2 \cup E_3$ form a closed region with edge (z_1, z_2). We denote the union of these three edges by $track_i$. To topologically simplify the following proof, we assume each pair of vertices (u_i, w_i) are always on the same side of edge (z_1, z_2) since we can always change the shape of

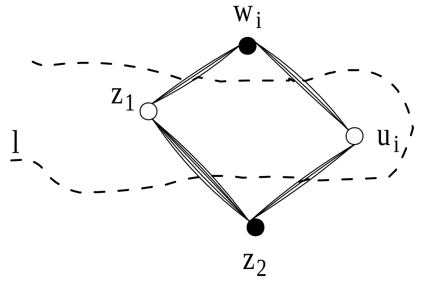
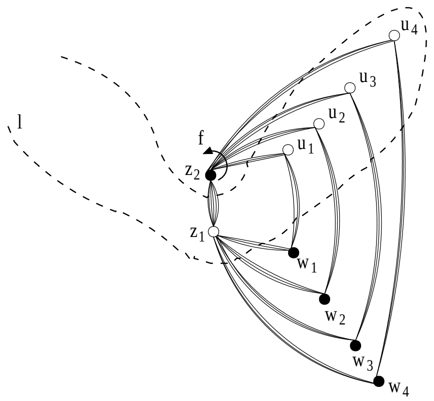
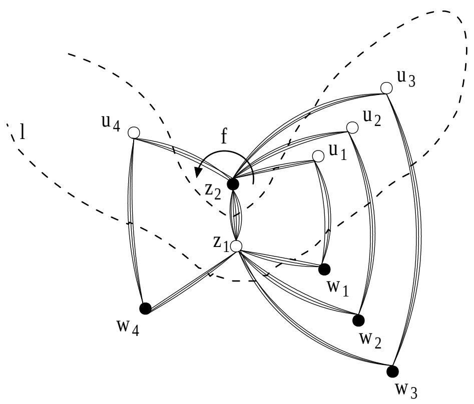


Figure 6: a region formed by $track_i$ and (z_1, z_2)



(a) single shell



(b) double shell

Figure 7: two kinds of shell embedding(edges in E_5 are omitted)

borderline l to depart A from B (see Fig. 6). Then there is no crossing either between tracks and (z_1, z_2) or between simple edges and (z_1, z_2) because of this restriction.

To finish our proof we need the following definition, lemmas and corollary.

definition 3.1 *An embedding of G' is a shell if there are no crossings between tracks.*

In a shell, if all tracks are on the same side of (z_1, z_2) we call it a single shell embedding (see Fig 7 (a)), otherwise we call it a double shell embedding (see Fig 7 (b)).

Suppose first that the desired ordering function f exists, we need to find a desired embedding for G' with less than K' crossings.

Notice that, in a shell all edges in E_1 are connected with z_2, they naturally generate a cyclic order around z_2. We map f to the order in counterclockwise around z_2, then the following proofs are based on this relationship.

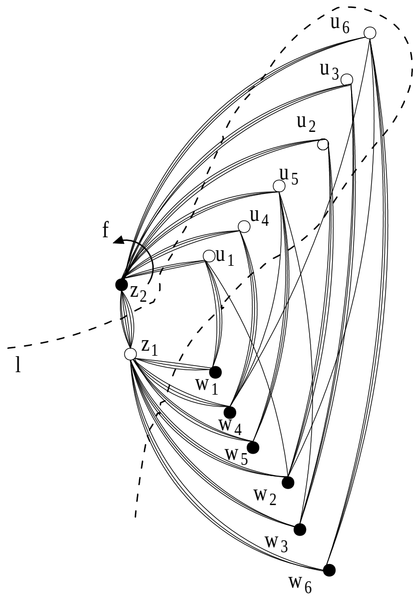


Figure 8: crossings in a single shell embedding

Lemma 3.1 *There is a single shell embedding of G' with at most*

$$\sum_{(v_i,v_j)\in E}(|f(v_i)-f(v_j)|-1)|E|^2+(|E|^2-1)$$

crossings.

Proof: Suppose that H' is a single shell embedding of G'. Crossings in H' are of two kinds: one is crossings between simple edges and tracks, the other is crossings between simple edges. For the first kind, the number of crossings is $\sum_{e\in E_5} C_e|E|^2$ where C_e is the number of tracks that e crosses. We can route all simple edges of E_5 in the shortest paths through the dual graph of $H' - E_5$. Noticeably, there is no edge crosses (z_1, z_2) in H', we consider there is no edge of the dual graph crosses it too. Then each edge (u_i, w_j) crosses $(|f(v_i) - f(v_j)| - 1)$ tracks resulting in $(|f(v_i) - f(v_j)| - 1)|E|^2$ crossings. For the second kind, the number of crossings between simple edges is at most $|E|^2 - 1$. So there is a H' that has at most $\sum_{(v_i,v_j)\in E}(|f(v_i) - f(v_j)| - 1)|E|^2 + (|E|^2 - 1)$ crossings (see Fig. 8). □

Since $\sum_{(v_i,v_j)\in E} |f(v_i) - f(v_j)| \leq K$, then the number of crossings of H' is at most K'. So H' is a desired embedding for G'.

Corollary 3.1 *If H' is a single shell embedding of G', then it has at least*

$$\sum_{(v_i,v_j)\in E}(|f(v_i) - f(v_j)| - 1)|E|^2$$

crossings.

Proof: Similar to corollary 2.1, every track and (z_1, z_2) form a closed region, traversing an edge in the dual graph of $H' - E_5$ corresponds to crossing a track, hence minimizing crossings with tracks corresponds to selecting shortest paths in the dual graph. So the number of edge crossings in H' will be not less than $\sum_{(v_i,v_j)\in E}(|f(v_i) - f(v_j)| - 1)|E|^2$.□

Now, suppose there is an embedding of G' with crossings at most K', we need to find a desired function for G as well. To prove this, we need the following lemma.

Lemma 3.2 *If there is an embedding H of G' with K' crossings, then there is a single shell embedding H' of G' with K' crossings.*

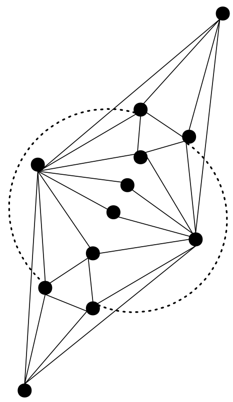


Figure 9: dashed lines can only be drawn as arcs to minimize the crossings

Proof: Suppose that H is an embedding of G' with K' crossings. Note that if there are two tracks cross, then they will generate at least $|E|^4$ crossings which is larger than K', thus H must be a shell. If it is a single shell then it's easy. If it is a double shell, then we consider the following transformation. Since tracks in H are on different sides of (z_1, z_2), we move the tracks on one side of (z_1, z_2) to the other side of it while keeping the cyclic order of edges (u_i, z_2) around z_2 (Fig. 7 (a) and (b) indicates the transformation clearly). Since the order is not changed, this transformation dose not increase the number of edge crossings and H becomes a single shell embedding H' with K' edge crossings. □

Recalling corollary 3.1, we have:

$$\sum_{(v_i,v_j)\in E} (|f(v_i) - f(v_j)| - 1)|E|^2 \leq$$
$$K' = (K - |E|)|E|^2 + (|E|^2 - 1)$$

which implies that

$$\sum_{(v_i,v_j)\in E} |f(v_i) - f(v_j)| \leq K$$

So there is a desired one to one function f for G as well. □

So since many crossings in *Two-Cluster* drawings can be reduced, it is still hard to find a TC drawing with minimum crossings.

4 Conclusion

With a little modification, the previous two proofs can be used to prove similar problems where edges are straight lines. However, for some graphs, crossings that can be avoided by using arcs cannot be avoided by straight lines (see Fig. 9 (Nagamochi 2004)).

Furthermore, straight line drawings need very large area (Biedl et al. 1998). In future we plan to concentrate on drawings presented by arcs or polylines.

Our aim is to find more efficient layouts with less crossings; this paper is just the start of our research, our plan is to focus on optimization methods to generate drawings with a small number of crossings automatically, and use them to solve problems in practical graphs such as social networks.

References

J. Pach and G. Tòth. (2000), Which crossing number is it anyway?, *J. Combin. Theory Ser. B*, **80**, 225–246.

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

P. Eades, K. Sugiyama. (1990), How to draw a directed graph, *Journal of Information Processing*, **13**(4), 424–437.

P. Eades and N. C. Wormald. (1994), Edge crossing in drawing bipartite graphs, *Algorithmica*, **11**, 379–403.

M. R. Gary and D. S. Johnson. (1983), Crossing number is NP-complete, *SIAM J. Algeraic and Discrete Methods*, **4**, 312–316.

K. Sugiyama, S. Tagawa, M. Toda. (1981), Methods for Visual Understanding of Hierarchical System Structures, *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-11**, 109–125.

H. Nagamochi. (2003), An Improved Approximation to the One-sided Bilayer Drawing, *in* Proceedings of the 11th International Symposium on Graph Drawing (GD '03), *Lecture Notes in Computer Science*, **2912**, 406–418, Springer.

T. Biedl, M. Kaufmann and P. Mutzel. (1998), Drawing Planar Partitions II: HH-Drawings, Technical Report RRR 11- 98, RUTCOR, Rutgers University.

U. F*öβ*meier and M. Kaufmann. (1997), Nice Drawings for Planar Bipartite Graphs, *Proceedings of the 3th Italian Conference on Algorithms and Complexity*, March **12-14**, 122–134.

M. R. Garey, D. S. Johnson and L. J. Stockmeyer. (1976), Some simplified NP-complete graph problems, *Journal of Theoretical Computer Science*, **1**, 237–267.

H. Nagamochi, (2004), by private comunication.