NEW BOUNDS ON THE BARYCENTER HEURISTIC FOR BIPARTITE GRAPH DRAWING

XIAO YU LI AND MATTHIAS F. STALLMANN

Abstract. The barycenter heuristic is often used to solve the NP-hard two-layer edge crossing minimization problem. It is well-known that the barycenter heuristic can give solutions as bad as $\Omega(\sqrt{n})$ times the optimum, where n is the number of nodes in the graph. However, the example used in the proof has many isolated nodes. Mäkinen [10] conjectured that a better performance ratio is possible if isolated nodes are not present. We show that the performance ratio for the barycenter heuristic is still $\Omega(\sqrt{n})$ even for connected bipartite graphs. We also prove a tight constant ratio for the barycenter heuristic on bounded-degree graphs. The performance ratio is $d-1$, where d is the maximum degree of a node in the layer that can be permuted.

1. Introduction.

The two-layer edge crossing minimization problem was first proposed by Harary and Schwenk [6, 7]: Given a bipartite graph $G = (V_0, V_1, E)$, embed the nodes of *layer* V_i on the line $y = i$ so that the number of crossings when the edges are drawn as straight lines is minimized. Since the number of crossings depends only on the permutation π_i of the nodes on each layer i (and not on their absolute position), we can say that the objective is to find the *presentation* $\langle G, \pi_0, \pi_1 \rangle$ of G that minimizes crossings.

Gary and Johnson [5] proved that the two-layer edge crossing minimization problem is NP-hard. More recently, Eades, McKay, and Wormald [2] proved that the problem is NP-hard even if the permutation of nodes on one layer is fixed. The problem of determining the minimum cardinality set of edges whose removal allows G to be drawn with no crossings is also NP-hard, whether or not the order on one of the layers is fixed [3].

This paper concentrates on the *fixed-layer bipartite crossing minimization* problem: Given a bipartite graph $G = (V_0, V_1, E)$ and a permutation π_0 of V_0, find π_1, a permutation of V_1, that minimizes the number of crossings of $\langle G, \pi_0, \pi_1 \rangle$. Since an efficient exact algorithm is unlikely, we are interested in analyzing the *performance ratio* of a heuristic h, the least upper bound of $h(G, \pi_0)/opt(G, \pi_0)$, where $h(G, \pi_0)$ is the number of crossings reported for the permutation of V_1 produced by h (given a specific permutation π_0 of V_0) and $opt(G, \pi_0)$ is the smallest number of crossings achievable by any permutation of V_1.

The two most common heuristics employed for this problem are the median heuristic and the barycenter heuristic. Both sort the nodes of V_1 based on the x-coordinates of the neighbors in V_0 of each node. The barycenter heuristic sorts nodes by the mean position of their neighbors and median uses the median (with a slight modification — see [1]).

The performance ratios for the barycenter and median heuristics — least upper bounds of $bary(G, \pi_0)/opt(G, \pi_0)$ and $med(G, \pi_0)/opt(G, \pi_0)$, respectively — appear to differ from each other dramatically. Eades and Wormald [4] showed that the performance ratio for the barycenter heuristic is $\Theta(\min(\sqrt{n}, n/\delta))$, where $n = |V_0| + |V_1|$, and δ is the minimum degree of a node. They also proved a performance ratio of 3 for the median heuristic, specializing to $1 + \alpha$, where α approaches 0 as density (the ratio $|E|/|V_0| \cdot |V_1|$) approaches 1.

In contrast to the theoretical bounds favoring the median heuristic, experiments have shown that the barycenter heuristic performs significantly better on a variety of graph types, especially sparse graphs [8, 9].

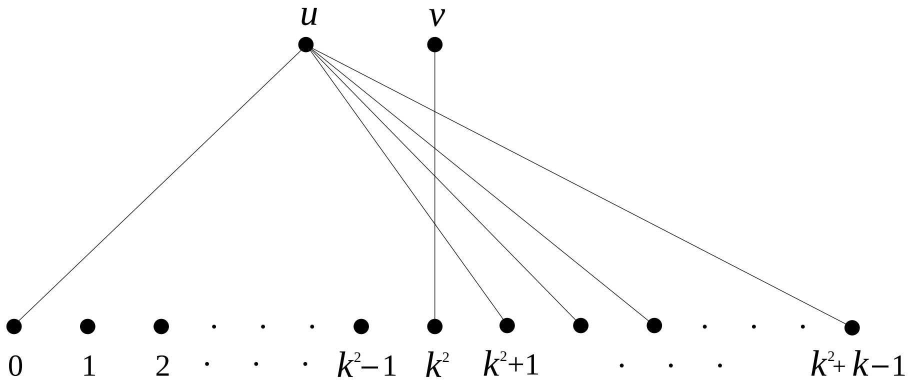

FIG. 1.1. *Worst case for the barycenter heuristic on disconnected graphs.*

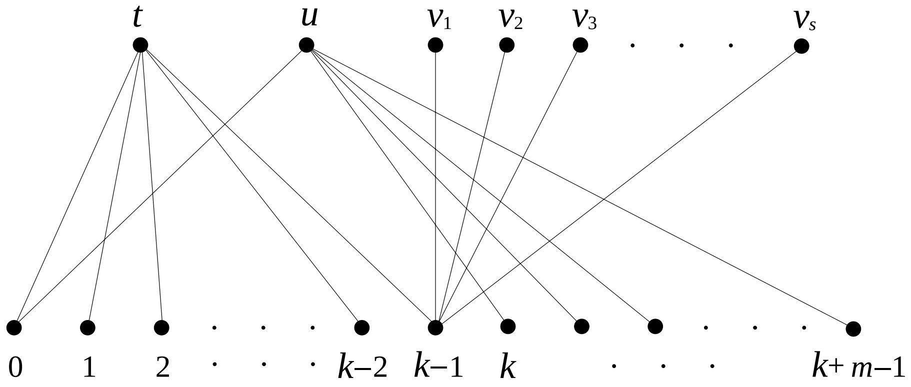

FIG. 2.1. *Worst case for the barycenter heuristic on connected graphs*

Fig. 1.1 shows the example used in [1] to illustrate the lower bound on barycenter performance ratio, a graph with $n = k^2 + k$ nodes for which $bary(G, \pi_0)/opt(G, \pi_0) = \sqrt{n}$. This graph, like the one originally used in [4], is not connected — in fact, there are $k^2 - 1$ isolated nodes in V_0.

Mäkinen [10] conjectured that a much better performance ratio (possibly a constant, as in the case of the median heuristic) can be achieved if the presence of the isolated nodes is ignored while calculating the average position. In the following section, we prove that the performance ratio for the barycenter heuristic is still $\Omega(\sqrt{n})$ even for a connected bipartite graph. Section 3 shows that when the degree of the nodes in V_1 is bounded by d, the performance ratio for the barycenter heuristic is $d - 1$ whether or not the graph is connected (or has isolated nodes).

2. The Worst-Case Performance Ratio. Consider the connected graph in Fig. 2.1. The $k + m$ nodes in V_0 are labelled the same as their x-coordinates, from 0 to $k + m - 1$. The x-coordinates of nodes in V_1 are determined by their barycenter values. Let $avg(x, \pi_0)$ be the mean index of the neighbors of x in π_0. Assuming $avg(u, \pi_0) \leq k - 1 = avg(v_1, \pi_0)$ we have $bary(G, \pi_0) = k - 1 + sm$. The optimum solution, with u to the right of v_s, has $opt(G, \pi_0) = k - 1 + s$ crossings. If we let $s = k - 1$ the ratio $bary(G, \pi_0)/opt(G, \pi_0) = (m + 1)/2$.

The size of m is constrained by the fact that $avg(u, \pi_0) \leq avg(v_1, \pi_0) = k - 1$.

This means

$$\left(\sum_{i=k}^{k+m-1} i\right)/(m+1) \leq k-1$$

or

$$m^2 + m \leq 2k - 2$$

and we are able to achieve $m \in \Omega(\sqrt{k})$. Since the number of nodes $n = k + m$ and the ratio is $(m+1)/2$, it is easy to see that the ratio is $\Omega(\sqrt{n})$ (in fact, the constant is somewhat larger than 1). Therefore we have the following.

PROPOSITION 1. *The performance ratio for the barycenter heuristic is* $\Theta(\sqrt{n})$ *even when the input graph is required to be connected.* □

The node u in the example has an arbitrarily large degree, however. In the next section, we prove that for bipartite graphs where the maximum degree of the nodes in V_1 is d, the performance ratio $bary(G, \pi_0)/opt(G, \pi_0) \leq d - 1$ and this bound is tight if graphs with isolated nodes are permitted.

3. Tight Performance Ratio for Bounded-Degree Graphs.

3.1. The Upper Bound. An upper bound on the performance ratio of a heuristic for the fixed-layer crossing minimization problem is obtained by considering the maximum ratio involving the edges of two arbitrary nodes in V_1 [1, 4]. If t and u are two nodes of V_1, define $c(t, u, \pi_0)$ to be the number of crossings among edges incident to t and u if π_0 is fixed and π_1 places t to the left of u. The following theorem says that the barycenter placement of any two nodes in V_1 never yields more than $d - 1$ times as many crossings as the reverse arrangement, where d is the maximum degree of the two nodes. An upper bound of $d - 1$ on the performance ratio follows directly.

THEOREM 2. *Let t and u be two arbitrary nodes in V_1 and let $avg(t, \pi_0) \leq avg(u, \pi_0)$. Let $c(t, u, \pi_0) > 0$. Then $c(t, u, \pi_0)/c(u, t, \pi_0) \leq \max(d_t, d_u) - 1$.*

Proof. Let $v_1, \ldots, v_j$ be t's neighbors in V_0, listed in left to right order, and let $w_1, \ldots, w_k$ be u's neighbors listed in the same way. Since what matters about these neighbors is their position wrt π_0, we use the name of a node interchangeably with its position. That is, $v_1 < v_2 < \cdots < v_j$ and $w_1 < w_2 < \cdots < w_k$.

If $v_1 < w_1$ the leftmost edge of t does not cross any edges of u when t is to the left of u, so $c(t, u, \pi_0) \leq (j-1)k$. When t is to the right of u, the edge to v_1 will cross all k of u's edges, so $c(u, t, \pi_0) \geq k$. The ratio will be $\leq j - 1$, giving the desired result.

A symmetric situation occurs when $v_j < w_k$.

In the remaining case we have $v_1 \geq w_1$ and $v_j \geq w_k$. Assume for now that $j \geq k$ — the argument is symmetric otherwise, as will become apparent later.

Let $w_1, \ldots, w_m$ be the neighbors of u that are $\leq v_1$. If u has no additional neighbor w_{m+1}, then $avg(u, \pi_0) \leq w_m \leq v_1 \leq avg(t, \pi_0)$. Since the barycenter placement of t to the left of u is contradicted if $avg(u, \pi_0) < avg(t, \pi_0)$, equality must hold. But then $avg(u, \pi_0) = w_m$ means $m = k = 1$ ($v_1 = avg(t, \pi_0)$ means $j = 1$) and we have $v_1 = w_1$. No crossing exists between tv_1, the sole edge from t, and uw_1, the sole edge from u, and $c(t, u, \pi_0) = c(u, t, \pi_0) = 0$, making the theorem vacuously true.

From here on we assume that u has at least one additional neighbor w_{m+1}. Since $j \geq k$, there must also be a v_{m+1}.

Case 1. $w_k > v_{m+1}$ — see Fig. 3.1.

We count the number of crossings. If t is to the left of u, at most the leftmost m of u's edges can cross all j of t's edges. The edges from u to $w_{m+1}, \ldots, w_{k-1}$ do not

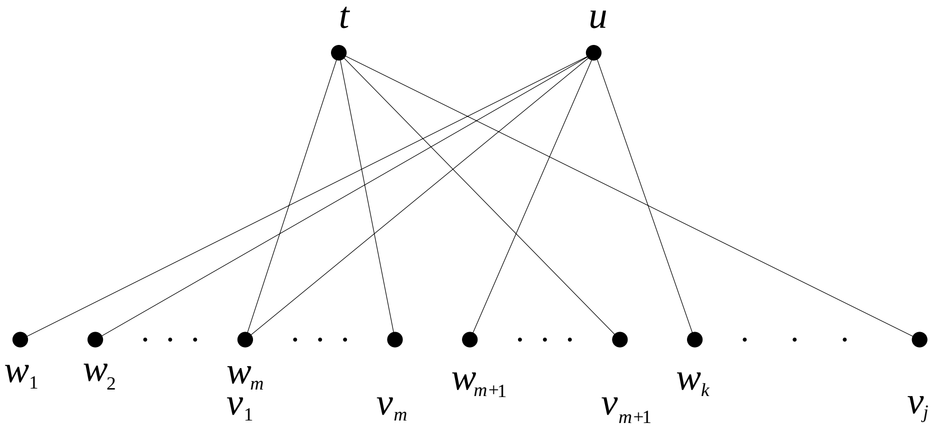

FIG. 3.1. *Case 1 in the proof of Theorem 2:* $w_k > v_{m+1}$

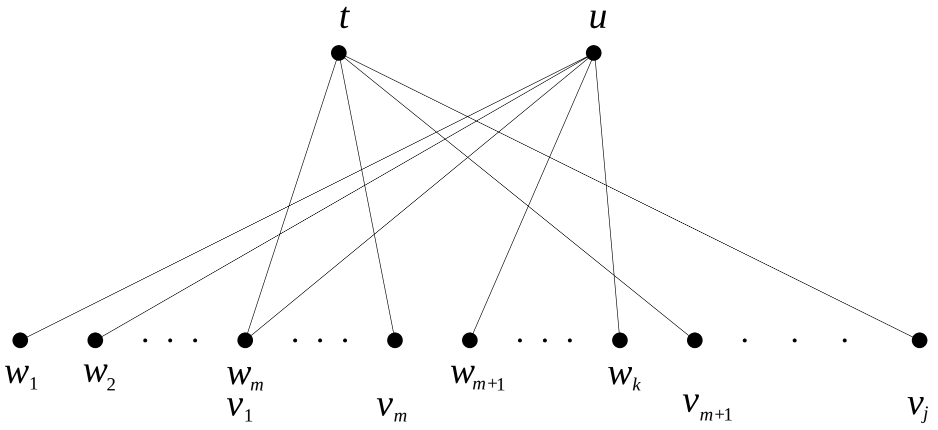

FIG. 3.2. *Case 2 in the proof of Theorem 2:* $w_k \leq v_{m+1}$

cross tv_1, giving them each at most $j-1$ crossings. Finally, the edge uw_k does not cross any of the edges from t to $v_1, \ldots, v_{m+1}$, and therefore crosses at most $j-m-1$ edges. This means $c(t,u,\pi_0) \leq jm + (j-1)(k-m-1) + j - m - 1 = (j-1)k$.

If t is to the right of u, the edge uw_k crosses the edges from t to $v_1, \ldots, v_{m+1}$ and the edges $uw_{m+1}, \ldots, uw_{k-1}$ all cross tv_1. This means $c(u,t,\pi_0) \geq (m+1) + (k-m-1) = k$ and the ratio $c(t,u,\pi_0)/c(u,t,\pi_0) \leq j-1$ as desired.

Case 2. $w_k \leq v_{m+1}$ — see Fig. 3.2.

Let $\text{mean}\{x_i \mid 1 \leq i \leq p\}$ denote the average (mean) of $x_1, \ldots, x_p$. Clearly $\text{mean}\{w_i \mid 1 \leq i \leq m\} \leq \text{mean}\{v_i \mid 1 \leq i \leq m\}$ and equality holds only if $m = 1$ and $v_1 = w_1$. Furthermore $\text{mean}\{w_i \mid m+1 \leq i \leq k\} \leq \text{mean}\{v_i \mid m+1 \leq i \leq k\}$, equality holding only if $m+1 = k$ and $v_k = w_k$. If $j > k$ or either inequality is strict then $avg(u,\pi_0) = \text{mean}\{w_i \mid 1 \leq i \leq k\} < \text{mean}\{v_i \mid 1 \leq i \leq j\} = avg(t,\pi_0)$, contradicting the barycenter placement. Otherwise $m = 1$, $j = k = 2$, $v_1 = w_1$, and $v_2 = w_2$ — see Fig. 3.3 — and there is exactly one crossing regardless of the relative placement of t and u.

In the latter part of the argument we assumed that $j \geq k$ and used the fact that at least one neighbor of u was $\leq v_1$ to obtain either the desired result or a contradiction

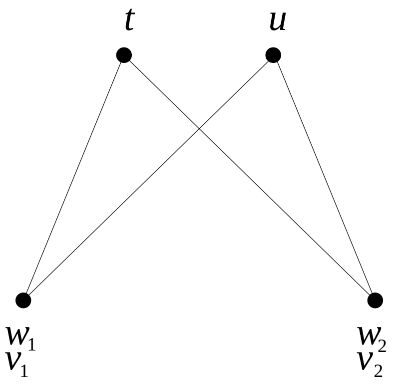

FIG. 3.3. *A special sub-case of case 2 in the proof of Theorem 2:* $w_1 = v_1$ *and* $w_2 = v_2$.

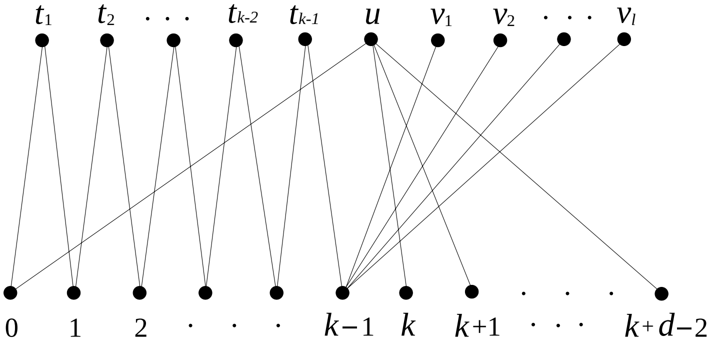

FIG. 3.4. *Worst case for the barycenter heuristic on bounded-degree bipartite graphs.*

of the barycenter placement. If $j < k$, we can use the fact that $v_j \geq w_k$ and use the m neighbors of t that are $\geq w_k$ in the same way that the previous argument used the m neighbors of u that are $\leq v_1$. The reader can verify that the argument is completely symmetric. □

Based on observations of Eades and Wormald [4] we have:

COROLLARY 3. *The performance ratio of the barycenter heuristic is* $\leq d - 1$, *where* d *is the maximum degree of any node of* V_1. □

3.2. The Lower Bound.

To prove the upper bound of $d-1$ on the performance ratio is tight for bounded-degree bipartite graphs, we illustrate a family of graphs for which the ratio $bary(G, \pi_0)/opt(G, \pi_0)$ approaches $d-1$. A generic instance is shown in Fig. 3.4.

It is easy to see that all the nodes in V_1 have a degree not greater than d. The node u is placed on the left hand side of v_1. Obviously, the optimum number of crossings is achieved by placing u on the right hand side of v_ℓ. It follows that:

$$\begin{aligned}\lim_{\ell\to\infty} \frac{bary(G,\pi_0)}{opt(G,\pi_0)} &= \lim_{\ell\to\infty} \frac{k^2 - 3k + 3 + \ell(d-1)}{2k - 3 + \ell} \\ &= d - 1.\end{aligned}$$

In conjunction with Corollary 3 we have the following.

Table 3.1

Comparison of performance ratios among the median, greedy, and barycenter heuristics, where d is the maximum degree of any node.

Method	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d>6$
median	2	2	≥ 2.5	≥ 2.5	≥ 2.67	$3-\epsilon$
greedy	2	2	2	$7/3+\epsilon$	$7/3+\epsilon$	$3-\epsilon$
barycenter (no isolated nodes)	1	≥ 1.6	≥ 1.89	≥ 2.07	≥ 2.2	$\geq 3-\epsilon$
barycenter (bound on V_1 only)	1	2	3	4	5	$d-1$

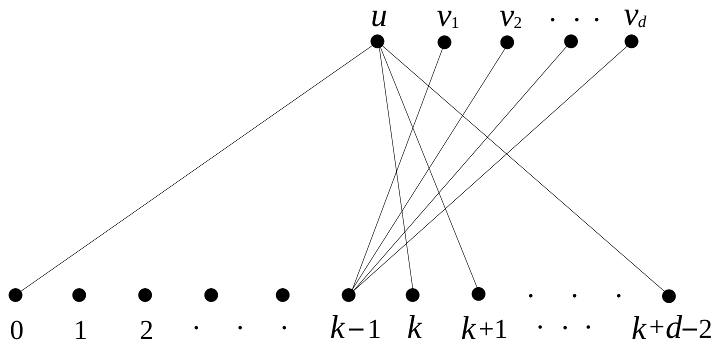

Fig. 3.5. *Worst case for the barycenter heuristic on bipartite graphs with bounds on the degree on both layers.*

Theorem 4. *The performance ratio of the barycenter heuristic $= d-1$, where d is the largest degree of any node in V_1.* □

This example requires arbitrarily large degree for node $k-1$ of the fixed layer V_0, but it does show that the bound of Corollary 3 cannot be improved easily. A simpler example can be used to show that $d-1$ is a tight bound even when d is a bound on the degree of all nodes, *provided that isolated nodes are permitted* — see Fig. 3.5. Thus, any improvement in the upper bound will have to be based on a degree bound of d on both layers and a requirement that no isolated nodes be present.

Suppose that the nodes $1,\ldots,k-2$ of V_0 in Fig. 3.5 are not isolated. For example, add an edge from each node i to a "companion" node t_i. Crossings involving the it_i edges are clearly minimized if $t_1,\ldots,t_{k-2}$ are in that left-to-right order to the left of u, which is also the barycenter order. Then the ratio of crossings when u is to the left of v_1 versus having u to the right of v_d is $(d(d-1)+k-2)/(d+k-2)$. To force the barycenter heuristic to place u to the left of v_1, we must have $k \geq (d^2-d+2)/2$, simplifying the ratio to $(3d^2-3d-2)/(d^2+d-2)$. This shows that for large fixed degree d, a lower bound on the performance ratio of the barycenter heuristic is 3, no better than the already established upper bound of the median heuristic. On the other hand, there is still room for improvement when d is small.

Table 3.1 shows known performance ratios for three heuristics whose ratios are constant for graphs with bounded degree. All but the barycenter heuristic have constant ratios regardless of node degree. A $\geq$ sign in the table means that examples exist proving a lower bound on the ratio but no corresponding upper bound has been

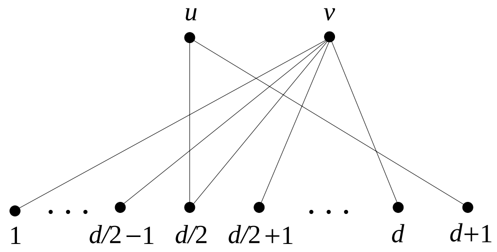

FIG. 3.6. *The minimum crossings can be achieved by placing* u *on the right of* v*. For the median method, the performance ratio is at least* $(3d-2)/d$ *when* d *is even.*

shown. For example, the observation in the previous paragraph implies that a version of the graph in Fig. 3.5 with no isolated nodes yields a ratio of 1.6 when $d = 3$. The 1.6 ratio is therefore a lower bound on the ratio for the barycenter heuristic when both sides have degree ≤ 3 and there are no isolated nodes. The only known upper bound for this situation is 2, based on Theorem 2.

Bounds for the median heuristic are due to Eades and Wormald [4] or brute-force applications of observations in [1]. Fig. 3.6 shows the general example on which the lower bounds for the median are based — the example relies heavily on the fact that the median heuristic chooses the smaller of two middle elements when degree is even (examples when d is odd appear to be no worse than those for $d-1$). The “greedy algorithm” is reported and analyzed by Yamaguchi and Sugimoto [11]. Results for the barycenter are reported here for the first time.

Although the barycenter heuristic has worse performance than the others when isolated nodes are present or the degree bound applies only to V_1, it appears from the table that it *might* outperform the others when there are no isolated nodes and a small degree bound d applies to both sides. We leave this as an open question.

Acknowledgements. We thank the anonymous referees for improving the clarity and accuracy of our presentation of the results.

REFERENCES

[1] G. Di Battista, P. Eades, R. Tamassia, and I. G. Tollis, *Graph Drawing: Algorithms for the Visualization of Graphs*, Prentice Hall, 1999.

[2] P. Eades, B. D. McKay, and N. C. Wormald, *On an edge crossing problem*, in Proc. Ninth Australian Computer Science Conference, 1986, pp. 327–334.

[3] P. Eades and S. Whitesides, *Drawing graphs in two layers*, Theoretical Computer Science, 131 (1994), pp. 361–374.

[4] P. Eades and N. C. Wormald, *Edge Crossings in Drawings of Bipartite Graphs*, Algorithmica, 11 (1994), pp. 379–403.

[5] M. R. Garey and D. S. Johnson, *Crossing Number is NP-complete*, SIAM J. Algebraic Discrete Methods, 4 (1983), pp. 312–316.

[6] F. Harary and A. J. Schwenk, *Trees with Hamiltonian Square*, Mathematika, 18 (1971), pp. 138–140.

[7] ———, *A New Crossing Number for Bipartite Graphs*, Utilitas Mathematica, 1 (1972), pp. 203–209.

[8] M. Jünger and P. Mutzel, *2–Layer Straightline Crossing Minimization: Performance of Exact and Heuristic Algorithms*, Journal of Graph Algorithms and Applications (JGAA), 1 (1997), pp. 1–25.

[9] E. Mäkinen, *Experiments on drawing 2-level hierarchical graphs*, International Journal of Computer Mathematics, 36 (1990), pp. 175–181.

[10] ———, *A conjecture on the error bound of the barycenter heuristic*, EATCS Bulletin, 70 (2000), pp. 156–158.

[11] A. Yamaguchi and A. Sugimoto, *An approximation algorithm for the two-layered graph drawing problem*, in COCOON, no. 1627 in LNCS, 1999, pp. 81–91.