Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in Discrete Mathematics 44 (2013) 251–256

Electronic Notes in DISCRETE MATHEMATICS

www.elsevier.com/locate/endm

# Arc-based integer programming formulations for three variants of proportional symbol maps

Rafael G. Cano [a,1,4], Cid C. de Souza [a,2,4], Pedro J. de Rezende [a,3,4] and Tallys Yunes [b,4]

[a] *Institute of Computing, University of Campinas, Campinas, Brazil*

[b] *School of Business Administration, University of Miami, Coral Gables, USA*

**Abstract**

Proportional symbol maps are a cartographic tool that employs scaled symbols to represent data associated with specific locations. The symbols we consider are opaque disks, which may be partially covered by other overlapping disks. We address the problem of creating a suitable drawing of the disks that maximizes one of two quality metrics: the total and the minimum visible length of disk boundaries. We study three variants of this problem, two of which are known to be NP-hard and another whose complexity is open. We propose novel integer programming formulations for each problem variant and test them on real-world instances with a branch-and-cut algorithm. When compared with state-of-the-art models from the literature, our models significantly reduce computation times for most instances.

*Keywords:* Symbol Maps, Integer Programming, Computational Geometry.

[1] Rafael G. Cano is supported by FAPESP grant 2012/00673-2.

[2] Cid C. de Souza is partially supported by CNPq grants 302804/2010-2, 477692/2012-5, 473867/2010-9, FAPESP grant 07/52015-0, and a grant from FAEPEX/UNICAMP.

[3] Pedro J. de Rezende is partially supported by CNPq grants 483177/2009-1, 473867/2010-9, 477692/2012-5, FAPESP grant 07/52015-0, and a grant from FAEPEX/UNICAMP.

[4] Email: {rgcano,cid,rezende}@ic.unicamp.br, tallys@miami.edu



http://dx.doi.org/10.1016/j.endm.2013.10.039

# 1 Introduction

Proportional symbol maps are a cartographic tool to visualize data associated with specific locations (e.g. earthquake magnitudes and city populations). Symbols whose area is proportional to the numerical values they represent are placed at the locations where those values were collected. Although symbols can be of any geometric shape, opaque disks are the most frequently used and, for that reason, the focus of our study. Portions of a disk may not be visible when overlapping occurs and, when large portions of a disk are covered, it is difficult to deduce its size and the location of its center. Therefore, the way in which the disks are drawn affects the amount and quality of information that can be inferred from a symbol map.

Let $S$ be a set of $n$ disks and $\mathcal{A}$ be an *arrangement*, which is the subdivision defined by the boundaries of the disks in $S$ (Fig. 1). We denote the sets of arcs and faces of $\mathcal{A}$ by $R$ and $F$, respectively. A *drawing* of $S$ is a subset of the arcs and vertices of $\mathcal{A}$ that is drawn on top of the filled interiors of the disks in $S$. Cabello et al. [1] define two types of drawings that are suitable for symbol maps, namely, *physically realizable drawings* and *stacking drawings*.

A drawing $\mathcal{D}$ is physically realizable if for every face $f \in F$, there exists an order among the disks that contain $f$ such that: $(i)$ an arc $r$ on the boundary of a disk $d_r$ is visible in $\mathcal{D}$ if and only if $d_r$ is above all disks that contain $r$ in their interior; and $(ii)$ the orders associated with distinct faces do not contradict each other. Informally, this definition states that a drawing is physically realizable if it can be constructed from whole symbols, cut out from sheets of paper. The disks can be interleaved and warped, but cannot be cut. A stacking drawing is a restriction of a physically realizable drawing in which there exists a total order among all disks in $S$, i.e. it is a drawing that corresponds to the disks being stacked up in layers, starting with the ones on the bottom layer. Fig. 2 shows examples of both types of drawings.

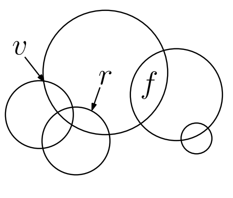


Figure 1: An arrangement with vertex $v$ (intersection of boundaries), arc $r$, and face $f$.

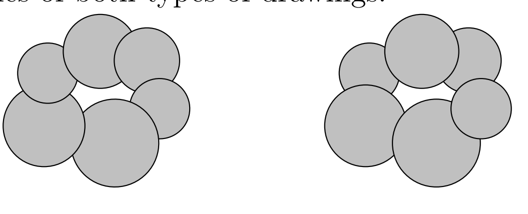

Figure 2: A physically realizable drawing that is not a stacking drawing (left) and a stacking drawing (right).

According to Cabello et al. [1], two metrics can be considered to determine the quality of a drawing: the minimum visible boundary length of any disk and the total visible boundary length over all disks. The *Max-Min* and *Max-Total* problems consist in maximizing the former and the latter values, respectively. Combining these two metrics with the two types of drawings yields four optimization problems. As in [3], we refer to the physically realizable drawing problem as PRDP and to the stacking drawing problem as SDP.

Cabello et al. [1] describe a greedy algorithm to solve the Max-Min SDP in $O(n^2 \log n)$ time. They also show that both variants of PRDP are NP-hard. The computational complexity of the Max-Total SDP remains open. Kunigami et al. [3] propose integer linear programming (ILP) models to solve both versions of PRDP and the Max-Total SDP. Their models are based on two sets of binary variables: an arc variable $x_r$ for each $r \in R$ (to indicate whether $r$ is visible in the solution) and an ordering variable $w_{ij}$ for each pair of disks $i, j \in S$ (to indicate the relative order between $i$ and $j$). The authors also present decomposition techniques to reduce the size of input instances.

We propose novel ILP models for the three problem variants studied in [3] (Sect. 2). Our formulations are in terms of arc variables only, thus reducing the dimension of the resulting polyhedra and the execution times. We show that these models are projections of the ones described in [3] (Sect. 3). We implement and test a branch-and-cut algorithm on real-world benchmark instances. A comparison with the formulations from [3] shows that our algorithm significantly improves computation times in several cases (Sect. 4).

## 2 Integer Linear Programming Models

Given an arc $r \in R$, we denote by $\ell_r$ the length of $r$ and by $d_r$ the disk in $S$ whose boundary contains $r$. In addition, for each arc $r$ (face $f$), let $S_r$ ($S_f$) denote the set of disks that contain arc $r$ (face $f$) in their interior.

For each $r \in R$, we define a binary variable $x_r$ that is equal to 1 if $r$ is visible in the drawing, and equal to 0 otherwise. For the Max-Total problem, the objective is to maximize $\sum_{r \in R} \ell_r x_r$. As in [3], for the Max-Min problem the objective is to maximize an additional real variable $z$, which is added to the model together with the following constraints:

$$z \leq \sum_{r \in R \,:\, d_r = i} \ell_r x_r, \qquad \forall\, i \in S. \tag{1}$$

We first consider the constraints required by SDP. Note that when an arc $r$ is visible in a drawing, it induces an order among $d_r$ and every disk in $S_r$ ($d_r$ must be drawn above every disk $i \in S_r$). We define an *induced order* *graph* $G_O = (V, E)$ as a directed multigraph with a vertex $v_i \in V$ for every disk $i \in S$, and a directed edge $e_{rj} = (v_{d_r}, v_j)$ for each arc $r \in R$ and disk $j \in S_r$. An example is shown in Fig. 3 (left and center). Let $\mathcal{C}$ be the set of all directed cycles in $G_O$. Now, given a cycle $C \in \mathcal{C}$, let $R_C$ be the set of arcs that give rise to the edges of $C$, i.e. $R_C = \{r : e_{rj} \in C,\ \text{for some } j \in S\}$. For every cycle $C \in \mathcal{C}$, we cannot have all arcs from $R_C$ visible in a solution because they induce a cyclic order among the corresponding disks. Thus, the following constraints must be satisfied:

$$\sum_{r \in R_C} x_r \leq |C| - 1, \qquad \forall\, C \in \mathcal{C}. \tag{2}$$

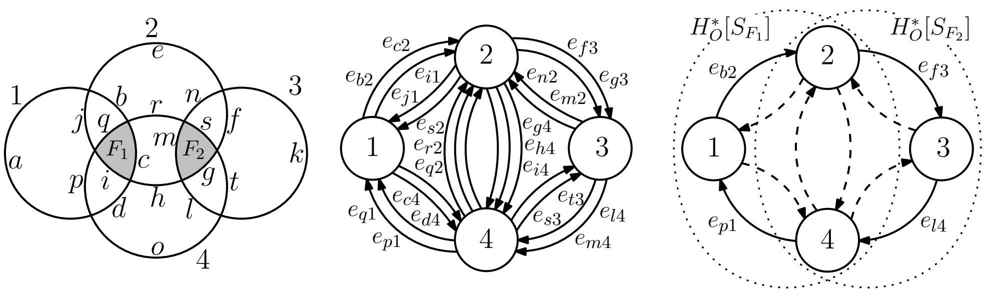


Figure 3. An arrangement with four disks (left), its graph $G_O$ (center) and the face-transitive closure of the cycle $\{e_{p1}, e_{b2}, e_{f3}, e_{l4}\}$ (right).

We now turn our attention to PRDP, which allows some cycles to occur. We determine the allowable ones as follows. Let $H_O$ be an arbitrary subgraph of $G_O$. Given a set of disks $X \subseteq S$, denote by $H_O[X]$ the subgraph of $H_O$ induced by the vertices $\{v_i : i \in X\}$. Also, let $F_X \subseteq F$ be the set of faces that are contained in at least one of the disks in $X$. We define the *face-transitive closure* $H^*_O$ of $H_O$ as its minimal supergraph such that for each face $f \in F_X$, $H^*_O[S_f]$ is transitively closed. Let $C \in \mathcal{C}$ be a cycle in $G_O$ and $C^*$ be its face-transitive closure. The arcs in $R_C$ can all be visible in a solution if and only if all subgraphs $C^*[S_f]$ are acyclic. Let $\widehat{\mathcal{C}}$ be the set of cycles in $G_O$ that do not satisfy this condition. One such cycle is depicted in Fig. 3 (right). Its face-transitive closure is obtained by adding the dashed edges to the original cycle. For PRDP, we must replace (2) by the following constraints:

$$\sum_{r \in R_C} x_r \leq |C| - 1, \qquad \forall\, C \in \widehat{\mathcal{C}}. \tag{3}$$

## 3 Polyhedral Study

We refer to the models presented in Section 2 as *arc models* (AM) and to the models proposed in [3] as *graph orientation models* (GOM). Given a model $M$, we denote by $M$-PRD and $M$-SD the specific variants designed to solve PRDP and SDP, respectively. Also, let $P_M$ and $\widetilde{P}_M$ be the polyhedra defined by the convex hull of all integer feasible solutions of $M$ and by the linear relaxation of $M$, respectively. Proofs are omitted due to space limitations.

**Proposition 3.1** *The polyhedra $\widetilde{P}_{\text{AM-SD}}$, $\widetilde{P}_{\text{AM-PRD}}$, $P_{\text{AM-SD}}$ and $P_{\text{AM-PRD}}$ are the projections of $\widetilde{P}_{\text{GOM-SD}}$, $\widetilde{P}_{\text{GOM-PRD}}$, $P_{\text{GOM-SD}}$ and $P_{\text{GOM-PRD}}$ onto the $x$-space, respectively.*

**Proposition 3.2** *The dimension of $P_{\text{AM-SD}}$ and $P_{\text{AM-PRD}}$ is $|R|$.*

**Proposition 3.3** *Given an arc $r \in R$, the inequality $x_r \geq 0$ always defines a facet of $P_{\text{AM-SD}}$ and $P_{\text{AM-PRD}}$. The inequality $x_r \leq 1$ defines facets of both polyhedra if and only if $S_r = \emptyset$.*

**Proposition 3.4** *Let $G_R$ be a graph with a vertex $v_r$ for each arc $r \in R$ and an edge $\{v_r, v_s\}$ for each pair of arcs $r$ and $s$ such that $d_r \in S_s$ and $d_s \in S_r$. Given a maximal clique $K$ in $G_R$, the inequality $\sum_{r\,:\,v_r \in K} x_r \leq 1$ defines a facet of $P_{\text{AM-SD}}$ and $P_{\text{AM-PRD}}$.*

We conclude this section mentioning that, generally, inequalities (2) and (3) are not facet-defining. However, they can be strengthened by lifting procedures, whose description we leave for the full version of this paper.

## 4 Computational Results

We assess the effectiveness of our algorithms using a set of 28 instances[5] generated from data on the population of cities from several countries. For all of them, we apply the decomposition techniques from [3]. The separation routines for inequalities (2) and (3) are based on a procedure described by Grötschel et al. [2]. The algorithms were implemented in C++ and compiled with gcc 4.4.3. We used CGAL 3.5.1 to build the arrangements and CPLEX 12.4 to solve the integer programs. The experiments were run on an Intel Xeon X3430, 2.40GHz CPU with 8GB RAM.

Table 1 summarizes the results for the hardest instances, showing the number of disks and arcs in each one and the execution times for models AM and

[5] Available at www.ic.unicamp.br/~cid/Problem-instances/Symbol-Maps.

GOM for each problem variant. Columns *Speedup* show the ratio between the execution times of the two models. We use a time limit of five hours for each ILP. For the Max-Total SDP and PRDP, the arc model performs better for all instances, especially for the hardest ones (reported here), for which it presents a (geometric) average speedup of 8.8. For the remaining instances, both problems can be solved in less than two minutes, and the arc model achieves a speedup of 2.4. For the Max-Min PRDP, it performs better for all but three instances. Still, it achieves an average speedup of 1.8.

| Instance | $\|S\|$ | $\|R\|$ | Max-Total SDP | | | Max-Total PRDP | | | Max-Min PRDP | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | AM | GOM | Speedup | AM | GOM | Speedup | AM | GOM | Speedup |
| France | 135 | 3230 | 435 | 5039 | 11.6 | 369 | 5034 | 13.6 | 3319 | >5h | >5.4 |
| Greece | 102 | 3482 | 495 | 5755 | 11.6 | 546 | 5475 | 10.0 | 15722 | >5h | >1.1 |
| Italy | 300 | 4366 | 16 | 86 | 5.4 | 14 | 86 | 6.1 | 1662 | 1890 | 1.1 |
| Japan | 150 | 3544 | 17 | 249 | 14.6 | 17 | 219 | 12.9 | 403 | 2216 | 5.5 |
| Portugal | 150 | 5070 | 39 | 304 | 7.8 | 40 | 306 | 7.7 | 85 | 28 | 0.3 |
| USA (West) | 87 | 3717 | 124 | 709 | 5.7 | 123 | 709 | 5.8 | 678 | >5h | >26.5 |

Table 1: Results for the six hardest instances. Times are given in seconds.

# 5 Conclusion

We study three variants of a proportional symbol maps problem and propose novel ILP models in terms of arc variables only. We show that these formulations are projections of another model previously described in the literature. When compared with state-of-the-art models, our formulations significantly reduce computation times for most instances. Future research directions include the study of new facet-defining inequalities and the development of branching techniques based on geometric properties of the problem.

# References

[1] Cabello, S., H. Haverkort, M. van Kreveld and B. Speckmann, *Algorithmic aspects of proportional symbol maps*, Algorithmica **58** (2010), pp. 543–565.

[2] Grötschel, M., M. Jünger and G. Reinelt, *On the acyclic subgraph polytope*, Mathematical Programming **33** (1985), pp. 28–42.

[3] Kunigami, G., P. J. de Rezende, C. C. de Souza and T. Yunes, *Generating optimal drawings of physically realizable symbol maps with integer programming*, The Visual Computer **28** (2012), pp. 1015–1026.