Optimizing Pedigree Drawing Using Interval Graph Theory

Frédéric Tores[1] **Emmanuel Barillot**[1,2]
tores@infobiogen.fr manu@infobiogen.fr

[1] INFOBIOGEN, 7 rue Guy Moquet, BP8, 94800 Villejuif, France
[2] GÉNÉTHON, 1 rue de l'Internationale, 91000 Evry, France

Keywords: pedigree drawing, interval graph, interval graph sandwich, graph augmentation, graph diminut

1 Introduction

A pedigree is a set of individuals that are related by four types of relations: mate, parent, child and sib. Conventionally, in a pedigree drawing: (1) Mates are linked to a mating unit. (2) Sibs are linked to a sibship unit. (3) The sibship is linked to its parental mating (4) people from the same generation are drawn on the same horizontal line, and the older generations are at the top (see Figure 1).

Naive strategies to pedigree drawing may lead to poor readability of the representation because of numerous link crossings. Pedigrees including consanguinity loops, individuals with multiples mates or several related families are particularly problematic to draw neatly. To our knowledge, none of the existing pedigree drawing softwares draws pedigree perfectly in all possible cases.

We propose to use interval graph theory to find a perfect representation of the pedigree, that is a representation with no link crossing, if such a representation exists. If not, we propose to use line-crossing elimination in directed graph to find the best layout.

2 Method and Results

We have defined the rules of readability to whom a pedigree drawing should conform to be perfectly meaningful: (a) No overlap is allowed between individuals. (b) Mates must be adjacent. (c) Sibs must be adjacent, but orphan spouses may be inserted within a sibship. (d) Parents are above their child sibship. (e) There are no link crossing. A pedigree for which a drawing verifying these five rules and the four conventions given in introduction exists is said to be a perfectly drawable pedigree (PDP).

Starting from a pedigree, we now define the following set V of vertices: a vertex per individual, a vertex per mating, and a vertex per sibship. We have shown that establishing if a pedigree is PDP is

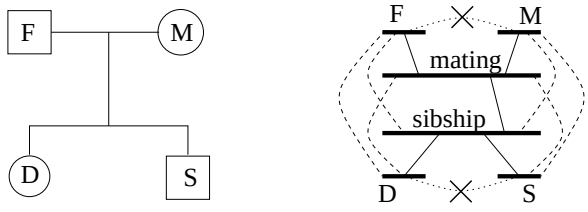

Figure 1: Example of simple pedigree drawing and its associated interval graph. Right, solid lines are mandatory edges, dotted with a cross are forbidden, dashed have been added to get an interval graph.

equivalent to establishing if there exists an interval graph[1] on V that respect some constraints directly derived from rules (a) to (e). Indeed, each of these rules corresponds in fact to requiring some edges and/or forbidding some other edges in the interval graph. For exemple, rule (d) implies that the vertex corresponding to a sibship should be adjacent to the parental mating vertex.

We have therefore introduced the sets E^+ of required edges and E^- of forbidden edges. The interval graph $I(V, E)$ should be chosen so that $E^+ \subset E \subset V \times V \setminus E^-$, which means that the PDP problem is equivalent to the interval graph sandwich problem (finding an interval graph whose set of edges is comprised between two nested sets). A polynomial solution is known for this problem, for which we refer the reader to [1, 2]. The pedigree is PDP if, and only if, a sandwich interval graph I exists. In such a case, a perfect drawing can be deduced from a realisation of the interval graph I (that is from a set of intervals consistent with I). Finding such a realisation is linear in time [3, 4].

If the pedigree is not PDP, one solution is to remove some required links and/or to add some forbidden ones until we get an interval graph (minimal graph diminution or augmentation). There also exists exact methods and heuristics to reorder the vertices and thus to minimize the number of line crossing (see [5]). Unfortunatly, exact methods have been shown to be NP-hard, even in simplest cases. In both cases, we propose to use a genetic algorithm in conjunction with interval graph or with a direct method as a heuristic to reduce the number of line crossing.

3 Conclusion

The problem of optimizing the drawing of a pedigree is not trivial. When a perfect solution exists, it can be found in a polynomial time using the interval graph sandwich theory. We are now implementing such a method in the CoPE software for pedigree drawing [6] and we are also studying the case of pedigrees that are not PDP.

We thank Frédéric Guyon for stimulating discussions on pedigree drawing optimization. We are also grateful to Laurence Brun-Samarcq, Guy Vaysseix and Philippe Dessen for their help in this work.

References

[1] Golumbic, M. C. and Shamir, R., Interval graphs, interval orders and the consistency of temporal events, *Lecture Notes in Computer Science*, 601:32–42, Springer-Verlag, 1992.

[2] Golumbic, M.C. and Shamir, R., Complexity and algorithms for reasoning about time a graph-theoretic approach, *J. ACM*, 40(5):1108–1133, 1993.

[3] Booth, K. S. and Lueker, G. S., Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms, *J. Comput. and Sys. Sci.*, 13:335–379, 1976.

[4] Rose, D. J., Tarjan, R. E. and Lueker, G. S., Algorithmic aspects of vertex elimination on graphs, *SIAM J. Computing*, 5:266–283, 1976.

[5] Di Batista, G., Eades, P., Tamassia, R. and Tollis, I.G., *Graph Drawing: Algorithms for the Visualization of Graphs*, Prentice-Hall, Upper Saddle River, 1999.

[6] Brun-Samarcq, L., Gallina, S., Philippi, A., Demenais, F., Vaysseix, G. and Barillot, E., Cope: a collaborative pedigree drawing environment, *Bioinformatics*, 15:345–346, 1999.

[1] A graph is an interval graph if, and only if, it exists for each vertex an interval on the real line such that two vertices are adjacent if and only if their corresponding intervals intersect. It can be shown that a graph is an interval graph if, and only if, it contains no chordless cycles bigger than three (it is said chordal) and no asteroidal triplets (an asteroidal triplet is a triplet of non–adjacent vertices for which it exists between each two of them a path that has no vertex adjacent to the third).