

SHORT REPORT

High-throughput pedigree drawing

Ville-Petteri Mäkinen^{*1,2}, Maija Parkkonen², Maija Wessman^{2,3}, Per-Henrik Groop^{2,4}, Timo Kanninen⁵ and Kimmo Kaski¹

¹Laboratory of Computational Engineering, Helsinki University of Technology, Finland; ²Folkhälsan Institute of Genetics, Folkhälsan Research Center, Biomedicum Helsinki, Finland; ³Department of Clinical Chemistry, University of Helsinki, Finland; ⁴Department of Medicine, Division of Nephrology, Helsinki University Central Hospital, Finland; ⁵Biocomputing Platforms Ltd, Espoo, Finland

Family trees have long been a valuable visual tool for geneticists in identifying clusters of inherited traits and genotypes. As more data are collected, drawing the graphs by hand becomes impractical and, for this reason, we have developed the pedigree software CraneFoot. It can process any family graph with minimal computational cost by making a pedigree transformation that enables the use of a linear node positioning algorithm. The program is designed for automated drawing to printed media and efficient visual classification of genetically interesting families from large data sets. It also incorporates a robust pedigree topology check with detailed error messages.

European Journal of Human Genetics (2005) 13, 987–989. doi:10.1038/sj.ejhg.5201430;
published online 4 May 2005

Keywords: pedigree; visualisation; database

Introduction

Pedigrees can be divided into three categories: Acyclic rooted graphs, cyclic but perfectly drawable graphs, and nondrawable cyclic graphs. The drawability of a pedigree depends on how well a given set of aesthetic goals can be met by a two-dimensional presentation.

The first category is the easiest and efficient visualisation methods have been available since the late 1970s. Among the first were Wetherell and Shannon,¹ who compared two naive algorithms and a more sophisticated solution and discussed the related aesthetic criteria. Soon after, Reingold and Tilford² responded by presenting a node positioning algorithm that could produce otherwise perfect layouts except for a minor defect in subtree positioning. A decade later, Walker II³ finally introduced an efficient algorithm that also adjusted the subtrees correctly.

Large families are likely to contain cycles (multiple matings, parents with common ancestors) and belong to one of the remaining categories. The second category was covered by Tores and Barillot,⁴ who presented an interval graph interpretation of perfectly drawable pedigrees. For the third category, no exact solution exists and time-consuming combinatorial optimisation is required. By contrast, our program transforms any pedigree to an acyclic graph before drawing, thus the node positioning for the first category alone needs to be solved.

Aesthetic criteria

For an optimal visualisation of ordered trees, the basic rules have been defined in the literature:

- (1) Nodes should not overlap each other.
- (2) Straight lines from children to their respective parents should not cross.
- (3) Nodes in the same generation should be placed on a straight line and the lines should be parallel.
- (4) The parents should be centred over their children.

*Correspondence: V-P Mäkinen, Laboratory of Computational Engineering, Helsinki University of Technology, Innopoli 2, Tekniikantie 14, PO Box 9203, FIN-02015 Hut, Finland.

Tel: +358 9 451 5725; Fax: +358 9 451 4830;

E-mail: vmakine2@lce.hut.fi,

URL: <http://www.lce.hut.fi/~vmakine2/cranefoot>

Received 24 December 2004; revised 24 March 2005; accepted 29 March 2005; published online 4 May 2005

- (5) (a) A subtree should be drawn the same way regardless of its position. (b) After reversing the node order, the new drawing should be a reflection of the original.

Aesthetics 4 and 5 are important for viewing quality, but in certain cases they prevent a maximally compact drawing given the other criteria, as demonstrated by Reingold and Tilford.² In fact, a drawing satisfying all five criteria may not be as narrow as possible since the overall width is affected, to some extent, by the traversal order of the tree. Fortunately, the phenomenon has little practical relevance.

Current pedigree drawing programs are motivated by the need to display every link and to project the result on a flat surface. For example, PED 4.2⁵ achieves compact drawings, but violates Aesthetics 2 and 4. Another solution titled CoPE⁶ allows batch processing of pedigrees, but cannot handle nondrawable graphs. A third example named Pedigraph⁷ displays a vertical flow chart, a different approach from the traditional drawings. In this case, failure to fulfil Aesthetic 2 and 4 leads to impaired readability of large and complex pedigrees. On the other hand, populations that have many founders and bottlenecks may be better visualised by this type of diagram.

Pedigree transformation

Figure 1 depicts a planar but nondrawable family graph. If the parents and children are placed on parallel lines according to Aesthetic 2, intersecting lines cannot be avoided. Interestingly, every crossing link involves a father–mother connection, hence Aesthetic 2 is fulfilled by omitting such links. One way to recover the lost information is to draw the problematic spouses more than once. As a side effect, the parent nodes of a child now constitute a mating unit with an important condition:

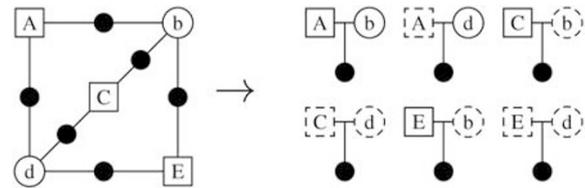


Figure 1 A family with a complex coupling structure (left). The labelled nodes are ancestors and the black dots represent their subtrees. A fragmented representation of the same family (right). Duplicates are indicated by dashed line.

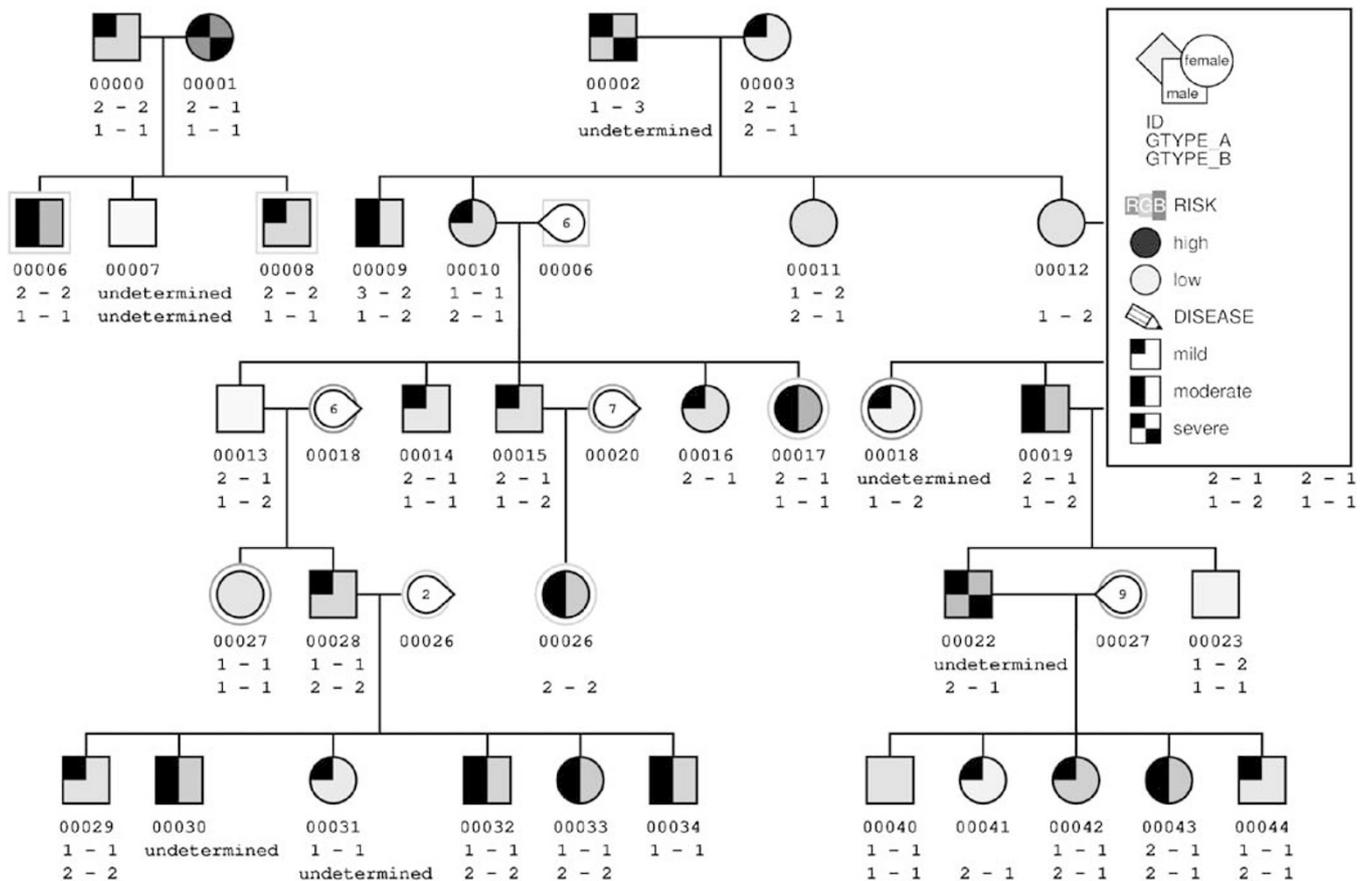


Figure 2 A highly inbred random pedigree. The node legend is overlaid for compactness.

Every child is linked to exactly one mating unit and any pedigree is reduced to a forest of rooted trees.

The drawings produced by the duplicate transformation and Walker II's algorithm³ satisfy all the five aesthetics, as illustrated in Figure 2 for a simulated pedigree. Furthermore, the node-positioning algorithm is able to take into account the individual node widths, a useful feature if you wish to print additional textual information such as medical history below the nodes. The symbols and line art in the figure follow the recommendations by Bennet *et al*⁸ although the unconventional visualisation algorithm together with the five aesthetics make some of the rules inapplicable.

Discussion

Duplicate transformation is, to our knowledge, the easiest way of making full size drawings of very large and complex pedigrees. The aesthetic quality is competitive also for simpler graphs and, owing to the improvement by Buchheim *et al*,⁹ the node positioning can be done in linear time, clearly outperforming other methods. For these reasons, the approach is ideal for situations such as interactive applications that may have strict restrictions on response time.

The program CraneFoot is currently being used in active research by the Finnish Diabetic Nephropathy Study (FinnDiane), and as the pedigree visualisation module of the data management system BCOS by Biocomputing Platforms Ltd. It is designed for large sets of pedigrees, with special attention to usability and reliability. The program also gives detailed reports on erroneous topology and provides an intuitive way of extracting visualisation information from pedigree files. To facilitate easy viewing, the results are collected in a single PostScript document

with a table of contents. Further development includes a graphical user interface, more choices over the node positioning algorithm and new visual features.

Acknowledgements

This work was conducted with the support of the graduate school of Electrical and Communication Department at Helsinki University of Technology, Jenny and Antti Wihuri Foundation, Folkhälsan Research Center and Academy of Finland Grants for MW (No. 00213) and KK (No. 209286). The Finnish Diabetic Nephropathy Study (FinnDiane) was supported by grants from the Folkhälsan Research Foundation, Samfundet Folkhälsan, the Research Funds of the Helsinki University Central Hospital, the Wilhelm and Else Stockmann Foundation, the Sigrid Juselius Foundation, European Commission (Contract No. QLG2-CT-2001-01669), and the Liv och Hälsa Foundation.

References

- 1 Wetherell C, Shannon A: Tidy drawings of trees. *IEEE Trans Software Eng* 1979; **SE-5**: 514–520.
- 2 Reingold EM, Tilford JS: Tidier drawings of trees. *IEEE Trans Software Eng* 1981; **SE-7**: 223–228.
- 3 Walker JQ II: A node-positioning algorithm for general trees. *Software Pract Exp* 1990; **20**: 685–705.
- 4 Tores F, Barillot E: The art of pedigree drawing: algorithmic aspects. *Bioinformatics* 2001; **17**: 174–179.
- 5 Plendl HJ: Stammbäume zeichnen mit PED 4.2. *Medizinische Genetik* 1998; **1**: 50–51.
- 6 Brun-Samarq L, Gallina S, Philippi A *et al*: CoPE: a collaborative pedigree drawing environment. *Bioinformatics* 1999; **15**: 345–346.
- 7 Garbe JR, Da Y: Pedigraph 2.0, a software tool for the graphing and analysis of large complex pedigrees, 2004, Abstract book, p. 242, ADSA-ASAS-PSA Joint Annual Meeting, St Louis, July 25–29.
- 8 Bennet RL, Steinhaus KA, Urich SB *et al*: Recommendations for standardized human pedigree nomenclature. *Am J Hum Genet* 1995; **56**: 745–752.
- 9 Buchheim C, Junger M, Leipert S: Improving Walker's algorithm to run in linear time. *Lecture Notes Comput Sci* 2002; **2528**: 344–353.