

Pairwise display of high dimensional information via Eulerian tours and Hamiltonian decompositions

Wayne Oldford
Department of Statistics & Actuarial Science
University of Waterloo

2010

Every statistical graphic is a construct of display elements laid out in some arrangement. Sometimes these arrangements require an ordering of their elements. For example, a parallel coordinate layout must order its parallel axes, a multivariate star glyph its radial axes. Whenever this is the case, the display items to be ordered can be represented as nodes in a formal graph. Edges between nodes appear if and only if the corresponding display items may follow one another in the display's layout. In many statistical graphics any item can follow any other (e.g. parallel coordinates) and so a complete graph results.

Much is known about paths on complete graphs that is of immediate relevance to statistical display layout. Every ordering of the display items is a Hamiltonian path on the complete graph. A display that ensures that every item immediately follows every other item is equivalent to an Eulerian tour on the graph. If statistically meaningful weights can be attached to each edge, then the graph is a weighted graph and some paths may be preferred to others.

We illustrate how these and other graph theoretic results can be put to good use by showing how several statistical displays can be improved. Examples include star glyphs, parallel coordinate plots, and multiple comparison plots. These new displays will be illustrated on data. (N.B. All methods and graphical displays are now available from the PairViz R package.)

This is based on joint work with Catherine Hurley of the National University of Ireland, Maynooth.