Comparison sequences for visualization: applications and algorithms

Catherine Hurley NUI Maynooth

joint work with Wayne Oldford (Waterloo)

catherine.hurley@nuim.ie

December 3nd 2008

NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

Introduction

- Statistical graphics
  - are about comparisons
  - comparisons between variables, cases, groups, models or models.
- Formalize this:
  - build a graph whose nodes are statistical objects.
  - Edges connect objects to be compared.
  - Weight edges to reflect importance of comparison
- Graph traversal informs the construction, and the layout (in space and time) of statistical graphics

Talk is about graphs and graphics, more applications than algorithms

Topics

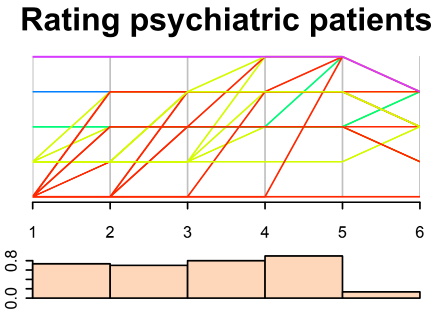
Parallel coordinate plots
Table plots
Interaction plots
Model comparison
Dynamic Scatterplots
Cable plots

⟷

Graphs
Complete Graphs
Hamiltonians and Eulerians
Bipartite Graphs
Hypercube graphs
Product Graphs
Line graphs
3d and 4d transition graphs

PairViz

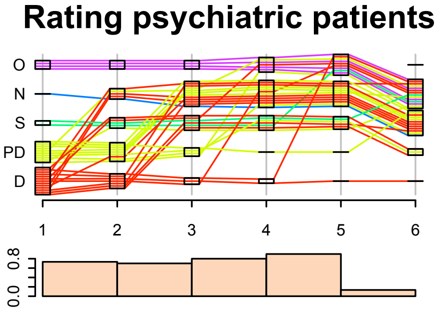
Comparing Raters



- Diagnoses data (Fleiss) contains psychiatric diagnoses of 30 patients provided by 6 raters.
- Depression (1), Personality disorder (2), Schizophrenia (3), Neurosis (4), O= Other (5).
- Barchart: proportion agreement

Comparing Raters

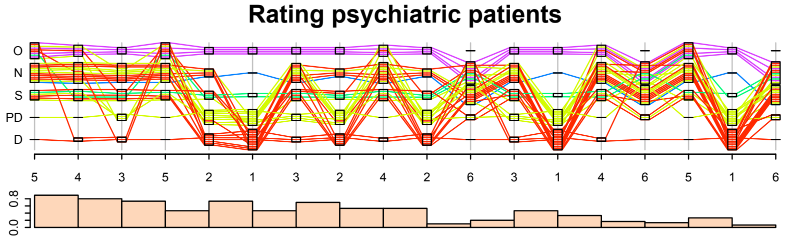
Improved display:



- Spread out values
- fatten axes,
- and add boxes

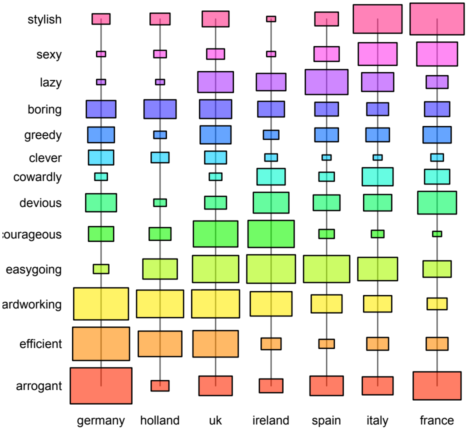
- Can compare marginal distributions and
- Can follow patients

Comparing All Raters



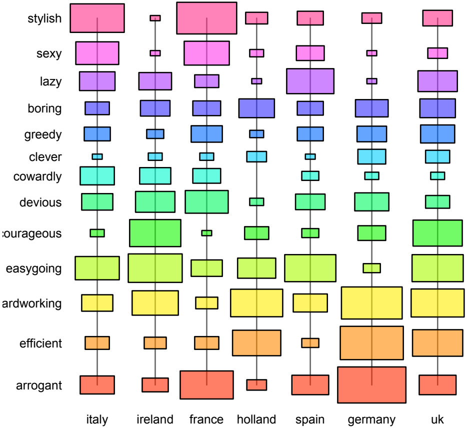
- All pairs of raters appear adjacently at least once, and
- the pairs are ordered in such a way that the raters whose agreement is higher tend to appear first.
- Good agreement between raters 3, 4 and 5.
- Rater 6 has low agreement with all other raters, but especially with rater 1.

Comparing barcharts



- UK survey: which of 13 characteristics they would associate with nationals of other European countries.
- Data are percents of respondants. (Der and Everitt.)
- Table plot, with both countries and characteristics sorted.
- Ireland's ratings are quite similar to those of the UK, less so to its other neighbour Spain

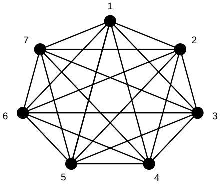
Comparing barcharts



- The second version shows a different ordering of countries.
- Dissimilar countries are adjacent, facilitates different comparisons
- Need 4 orderings to see all countries adjacently.

Graphs: nodes, edges and weights

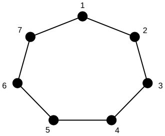
- n variables, cases, factor levels, boxplots: identify with nodes of graph
- visualisation: requires graph traversal
- All possible pairings are of interest: place an undirected edge between each pair of nodes
- Graph is complete, K_n



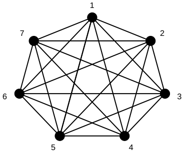
- Dissimilarity measure: edge weight

Hamiltonian and Eulerian paths

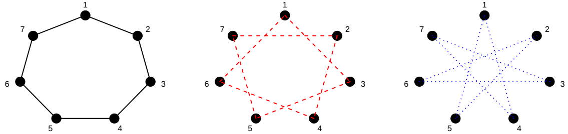
Hamiltonian path gives a permutation of vertices



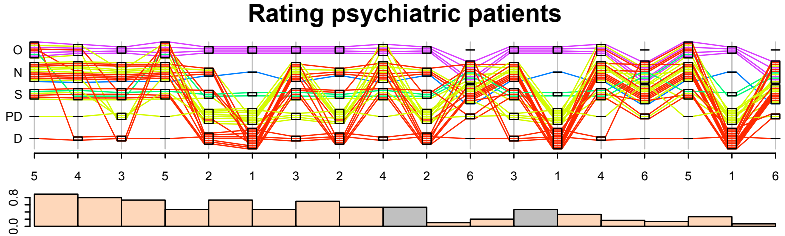
Eulerian path visits all edges



Hamiltonian decomposition: an eulerian tour composed of edge-distinct hamiltonian cycles

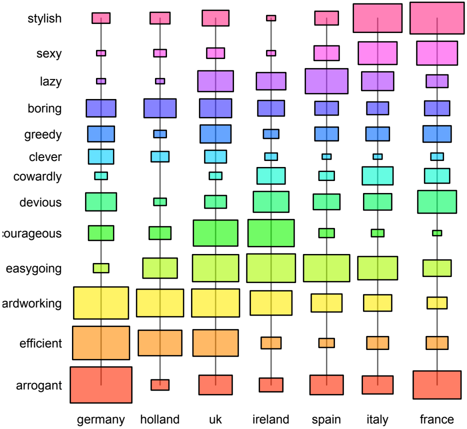


Revisit: Comparing All Raters



- Graph K_6 is not eulerian, because all nodes are odd.
- K_6^e, augmented version of K_6 which is eulerian.
- Duplicate edges 1-3, 2-4 and 5-6 (omitted).
- Modification of classical algorithm for weight-decreasing eulerians (etour in PairViz).

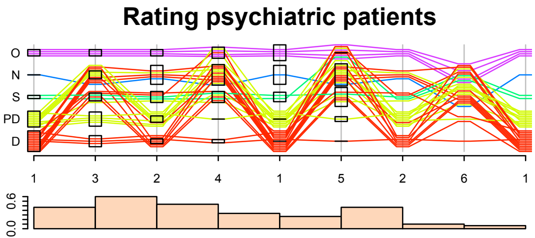
Revisit: Comparing barcharts

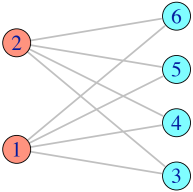


- Country ordering is a hamiltonian path on K_7.
- Graph is eulerian and also decomposes into 3 hamiltonian cycles
- Here we show hamiltonian paths.
- Require 4 hamiltonian paths to visit all edges, which is not a decomposition.
- weighted_hpaths in PairViz

Bipartite graphs

Suppose two are expert raters and we wish to compare others to them:



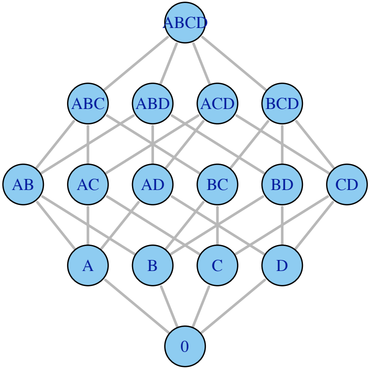


- Graph is eulerian: Use 132415261.
- Graph $K_{m,n}$ is eulerian if m and n are both even.
- Other applications: m responses, n predictors, where only response-predictor relationships are of interest.

Model comparison and hypercube graph

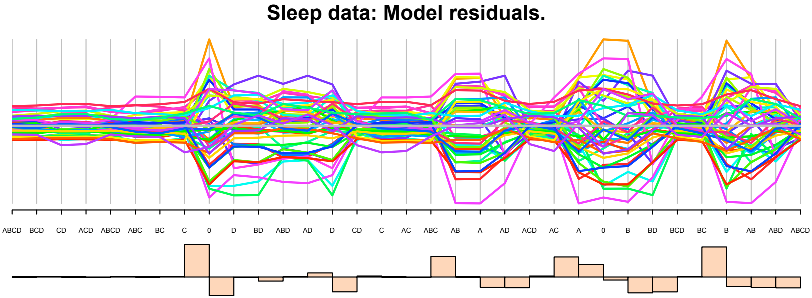
Model selection with n predictors

- Hypercube graph represents possible moves in a stepwise regression algorithm
- Example with $n = 4$
- Graph Q_n is hamiltonian, and eulerian for even n



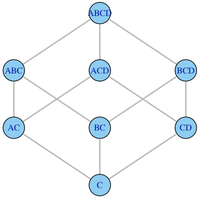
Sleep data

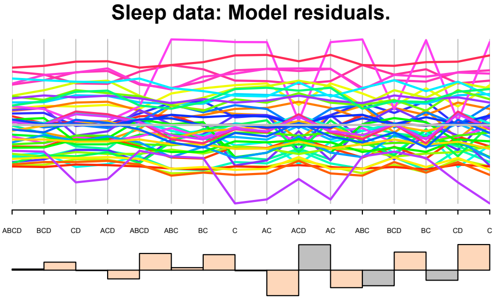
- Y= log brain wt. Predictors A= non dreaming sleep, B=dreaming sleep, C=log body wt, D=life span
- Eulerian starting at full model. Bars show change in SSE.
- All models with C give good results



Reduce model space

- Drop intercept
- Or, show only models with C
- Graph Q_3 is not eulerian: all nodes are odd



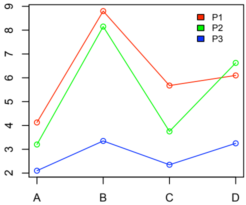


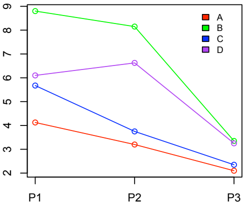
Add edges: ACD-AC, ABC-BCD, BC-CD (grey)

Open eulerian path from ABCD to C

Interaction plots

Data: survival time of 48 rats, each given one of four treatments A, B, C, or D and one of three poisons P1, P2, or P3 (Box and Cox)

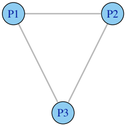




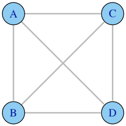
Check for parallelism of profiles

Graphs

Main effects:

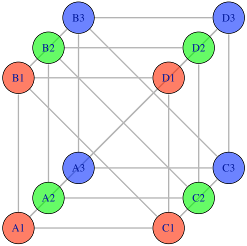


Poisons: K_3



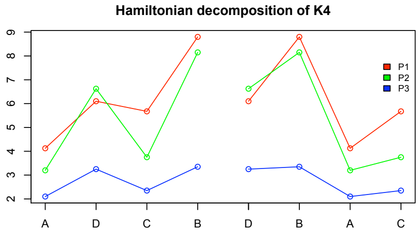
Treatments: K_4

Interactions:

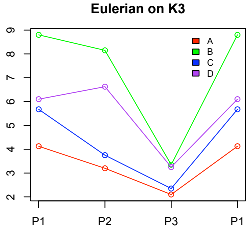


Cartesian product graph $K_3 \times K_4$

Revised interaction plots



Double crossing in first set of profiles gives stronger impression of interaction.
Second set of profiles: long line segments connecting treatments B and A gives impression of parallelism

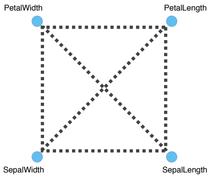


Survival times for P3 are low, regardless of treatment

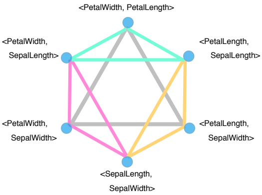
Alternatively: reduce tilt by subtracting average profile, for easier vertical comparisons.

Scatterplot transitions

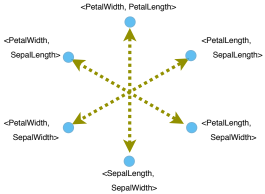
Example $p = 4$, Iris data



K_4 Complete graph



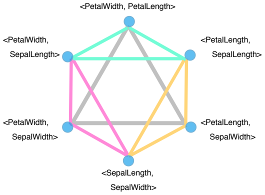
$L(K_4)$ 3d transitions graph Eulerian graph, Hamiltonian decomposition



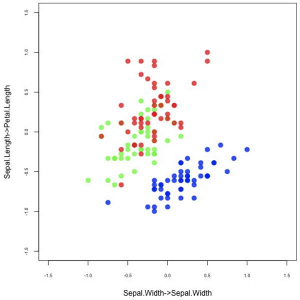
$\overline{L(K_4)}$ 4d transitions graph

Visualizing scatterplot transitions

Example $p = 4$, Iris data



$L(K_4)$ 3d transitions graph
Eulerian graph, Hamiltonian decomposition

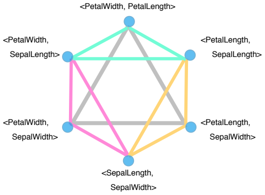


The yellow route

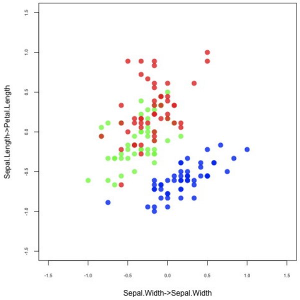
Choice of edge weights.... scagnostics?

Visualizing scatterplot transitions

Example $p = 4$, Iris data



$L(K_4)$ 3d transitions graph
Eulerian graph, Hamiltonian decomposition

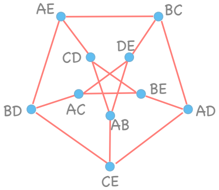


The yellow route

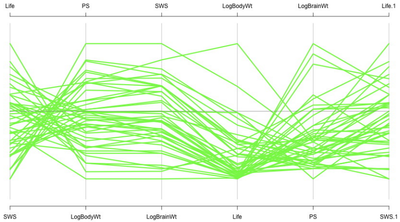
Choice of edge weights.... scagnostics?

Another example

Five variables of sleep data



4d transition graph for p=5

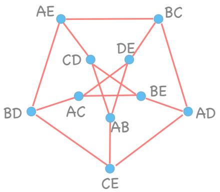


- Static display is hamiltonian cycle on variables
- Movie visits outer cycle of the transition graph and
- transitions to another hamiltonian cycle on the variables

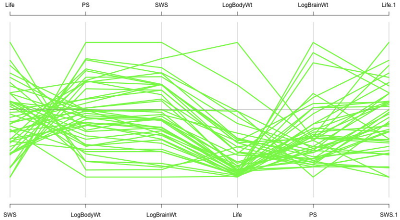
Graphs provide a map for navigating through high dimensional space

Another example

Five variables of sleep data



4d transition graph for p=5



- Static display is hamiltonian cycle on variables
- Movie visits outer cycle of the transition graph and
- transitions to another hamiltonian cycle on the variables

Graphs provide a map for navigating through high dimensional space

PairViz

What's available

Parallel coordinate plots
Table plots
Interaction plots
Model comparison

Dynamic Scatterplots
Cable plots

⟷

Graphs
Complete Graphs
Hamiltonians and Eulerians

Bipartite Graphs
Hypercube graphs
Product Graphs
Line graphs
3d and 4d transition graphs

Graph traversals:

- eseq - eulerians on complete graphs
- hpaths- hamiltonian decompositions on complete graphs
  also with weights
- eulerian- eulerians on connected graph, using weights

Graphics:

- guided_pcp, table_plot, mc_plot (for multiple comparisons)

Conclusions: Graphs and Graphics

- A graph structure underlies many statistical graphics
- Leads to improved understanding, improved graphics
- A roadmap for exploring high-dimensional spaces
- and perhaps even a GUI.