Eulerian tour algorithms for data visualization and the PairViz package

Catherine Hurley
NUI Maynooth

R.W. Oldford
U. Waterloo

July 8 2009 UseR!
Graphics: Effect Ordering

- Packages: `seriation`, `gclus`, `corrgram`
- Example: PCP Flea data

Standard order

Correlation order
Pairviz: relationship ordering

- Statistical graphics are about comparisons between variables, cases, groups, models

Flea data: correlation order
A graph model

- Build a graph where nodes are statistical objects
- Edges are relationships

Example:

<table>
<thead>
<tr>
<th>Node</th>
<th>Vis</th>
<th>Edge</th>
<th>Vis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Boxplot</td>
<td>Two groups</td>
<td>CI for mean diff</td>
</tr>
<tr>
<td>Var</td>
<td>Hist</td>
<td>Two vars</td>
<td>Scatterplot</td>
</tr>
<tr>
<td>2 vars</td>
<td>Scat</td>
<td>4-d space</td>
<td>Dynamic scat</td>
</tr>
<tr>
<td>Model</td>
<td>Resid</td>
<td>2 Models</td>
<td>PCP</td>
</tr>
</tbody>
</table>
Example: planned comparisons

Mice in 5 diet groups, response is lifetime
Nodes are treatments, edges are planned comparisons
Weights are p-values

Planned comparisons of diets
Reducing calories and protein increases lifetime
Graph Traversal

- Traverse all nodes: hamiltonian path

- Traverse all edges: eulerian path

- Use gclus, seriation: hamiltonian paths on complete graphs

- PairViz: eulerian paths
Graph Structures

- **Complete graph**: all comparisons are interesting
- **Edge-weighted graphs**: low weight edges are more interesting
- **Bipartite graph**: eg only treatment-control comparisons are of interest

Weight edges by 1-corr, eulerian follows low weight edges
Graph Structures - cont’d

• Hypercube graph

Cube for factorial experiment

000
001
010
011
100
101
110
111

or model selection:
Each node in G is a predictor subset
edge: add/drop predictor

• Line graph

transform G to L(G)

e.g. Each node in G is a var, each node in L(G) is var pair, edge is 3-d transition
Algorithms - Complete graph

- Closed eulerian path exists when each node has odd number of vertices: ie for $K_{2n+1}$
- Hamiltonian decomposition of graph
  - into hamiltonian cycles: eulerian for $K_{2n+1}$
  - into hamiltonian paths: approx eulerian for $K_{2n}$
- Classical algorithm: `hpaths`
- WHam: `weighted_hpaths`: pick best for $H_1$, best orientaton and order for others.
• Recursive algorithm: \texttt{eseq}:

• Start with eulerian on $K_n$, append edges to get eulerian on $K_{n+2}$
• Eulerian graph: connected, all nodes have even number of edges

• Otherwise, add edges, pairing up odd nodes
  Chinese postman does this in optimal way

• Classical algorithm (Hierholzer; Fleury)

• Our version GrEul, (etour) follows weight increasing edges
Algorithms comparison

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Graph</th>
<th>Hamiltonians</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>eseq</td>
<td>complete</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>hpaths</td>
<td>complete</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>weighted_hpaths</td>
<td>complete</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>eulerian</td>
<td>connected</td>
<td>no</td>
<td>optional</td>
</tr>
</tbody>
</table>

Complete-no weights

- Etour 9: prefers low vertices
- Eseq 9: prefers low edges
- hpaths 9: 4 hamiltonians
Algorithms: complete, weighted

Eurodist: 21 European cities

Algorithm eseq: Eurodist edge weights

Weighted etour on Eurodist

Weighted hamiltonians on Eurodist

ignores weights

Starts in Geneva

hamiltonian decomp, with increasing path lengths
Example: model selection

**Mammal sleep data**

- $Y = \log$ brain wt.
- Predictors $A = $ non dreaming sleep, $B = $ dreaming sleep, $C = \log$ body wt, $D = $ life span

- Hypercube graph represents possible moves in a stepwise regression algorithm
- Graph $Q_n$ is hamiltonian, and eulerian for even $n$
- Edge weights: change in SSE

- Eulerian starting with full model
- All models with $C$ are good
- Bar chart: change in SSE

Sleep data: Model residuals.
More variables

Sleep data: 10 vars (nodes)
45 edges
Eulerian has length 50

Using outlying index from scagnostics package for eulerian traversal
zoom on first half of display
More variables-cont’d

Reduce the graph
NN graph: eliminate edges with outlier index < .2

Reduces graph from 10 to 5 nodes, and 45 to 5 edges
Other nodes have no edges
IN CONCLUSION..

• **Pairviz** package: relationship ordering for data visualisation

• Current version: algorithms presented here

• Thanks to **graph, igraph**

• Work in progress: ordering dynamic visualisations via **ggobi**.

with Adrian Waddell, UW