Graph traversals and visual ordering: Eulerians, Hamiltonians and pairwise comparisons

Catherine Hurley NUI Maynooth

joint work wtih Wayne Oldford (Waterloo)

catherine.hurley@nuim.ie

June 26 2008

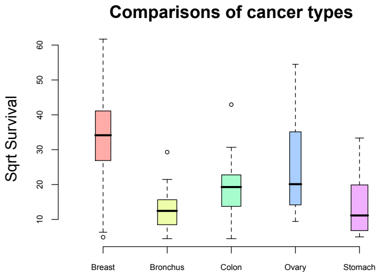
NUI MAYNOOTH

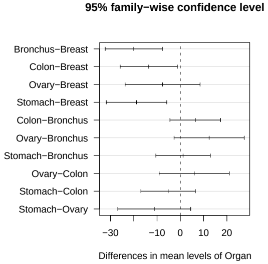
Graph traversals and visual ordering: Eulerians, Hamiltonians and pairwise comparisons

- Comparison of treatment groups
  → A new multiple comparison display
- Visual ordering as graph traversal
  → Eulerians and hamiltonians
- Parallel coordinates
  → guided by scagnostics

Comparison of treatment groups

Vit. C treated cancer patients: Cameron and Pauling 1978



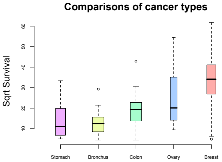


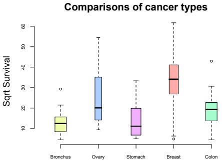
- Easy to visually compare adjacent groups
- not so easy for distant groups

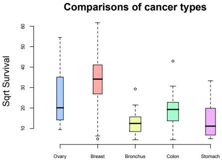
- Which pairs are significantly different?
- 95% Tukey HSD comparisons

Response is sq. root of survival times

Comparison of treatment groups





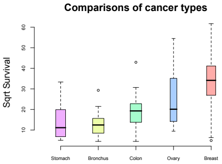


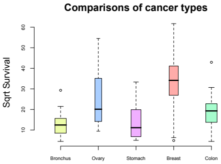
Require 3 sequences for all pairwise comparisons.

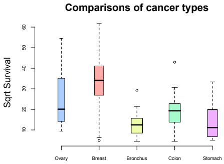
Note there is duplication: Breast-Ovary and Bronchus-Colon are in first and third plots

Theory

Comparison of treatment groups





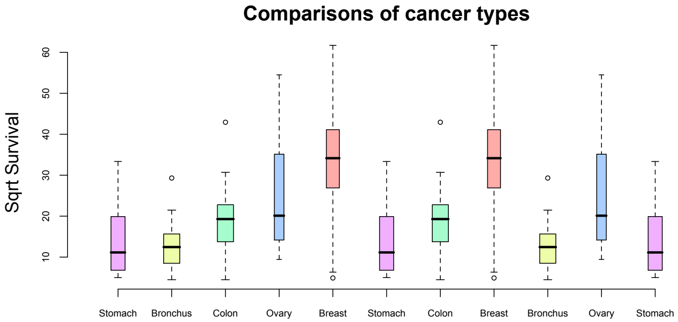


Require 3 sequences for all pairwise comparisons.

Note there is duplication: Breast-Ovary and Bronchus-Colon are in first and third plots

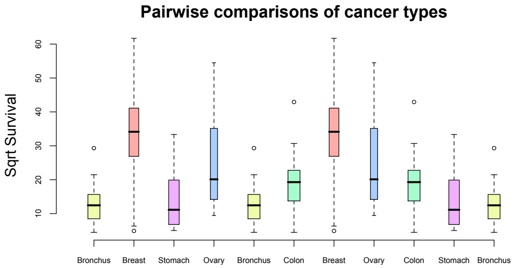
Comparison of treatment groups

More compactly: Glue the sequences in the first two plots together, append an extra ‘Stomach’.



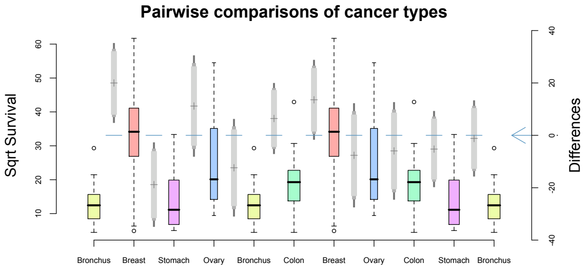
◂ Theory1

New pairwise comparison display



- Rearrange boxplots so significantly different means appear on lhs.

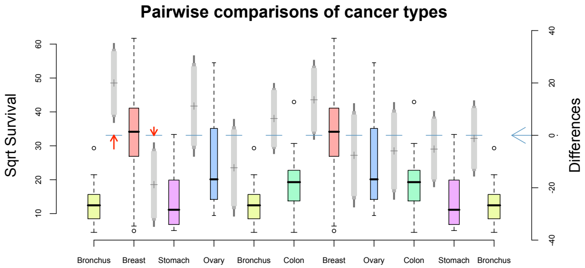
New pairwise comparison display



- Rearrange boxplots so significantly different means appear on lhs.
- Overlay 99% (HSD) CIs ($\mu_{left} - \mu_{right}$)

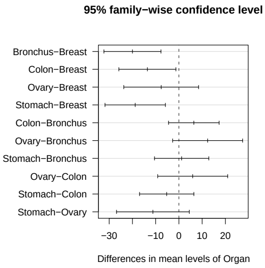
Theory2

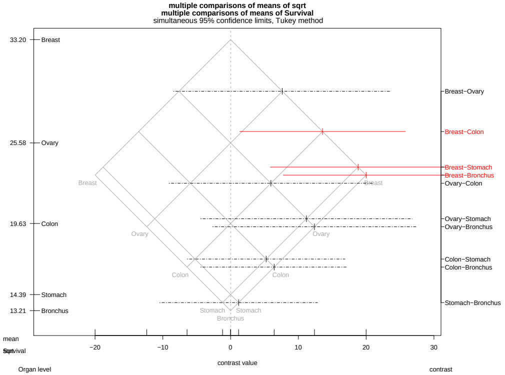
New pairwise comparison display



- Rearrange boxplots so significantly different means appear on lhs.
- Overlay 99% (HSD) CIs ($\mu_{left} - \mu_{right}$)
- Red arrow: significantly different comparisons
- Simple yet informative

Improvement on..???

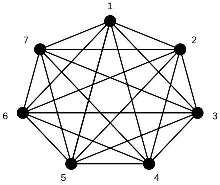




Hsu, Periggia (1994), Heiberger and Holland (2006)

Graphs: nodes, edges and weights

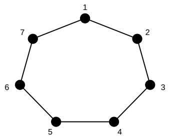
- n variables, cases, factor levels, boxplots: identify with nodes of graph
- visualisation: requires graph traversal
- All possible pairings are of interest: place an undirected edge between each pair of nodes
- Graph is complete, K_n



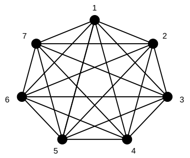
- Dissimilarity measure: edge weight

Hamiltonian and Eulerian paths

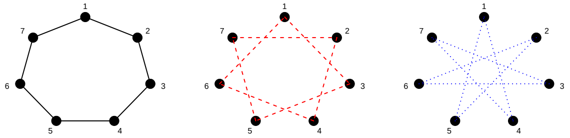
Hamiltonian path gives a permutation of vertices



Eulerian path visits all edges



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



Classical results: Euler paths- existence

- Eulerian tour (closed path) exists when every vertex is even. ie for K_{2m+1}

▸ Example: K_5

- Eulerian path (open) exists when two vertices are odd. *Augment K_{2m} with extra edges to achieve this.*

Which eulerian?

- How many?
  - K_7: about 130 million choices
  - K_{21} has more than 3.4×10^{184} discounting cyclic permutations

Online Encyclopedia of Integer Sequences (Sloane 2004)

- Prefer eulerians where low-weight edges (interesting comparisons) occur early on.
- Standard algorithm follows unused edges until all are visited. Our version (GrEul) picks low-weight edges.

Example: GrEul

Hierholzer 1873

Classical results: Hamiltonian Decompositions

K_n can be decomposed as follows:

- For $n = 2m + 1$, into either
  - m hamiltonian cycles, or
  - m hamiltonian paths and an almost-one factor Example: K_5
- For $n = 2m$ into either
  - m hamiltonian paths, or
  - $m - 1$ hamiltonian cycles and a 1-factor (or perfect matching).

Lucas-Walecki (1892) Alspach(1990)

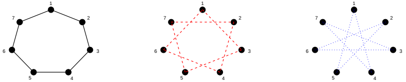
Which hamiltonian?

- Depends on question of interest.
- Sort nodes, eg by median
- Find shortest or lowest-weight path: (TSP)
- Choice of weights?
  - How interesting is the comparison between treatements? or the relationship between variables?

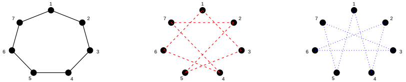
Which hamiltonian decomposition?

How many?

- K_7: 2 canonical forms



120 like this



and 840 like this

- K_{11}: 45,000+ canonical forms

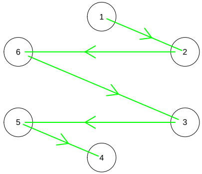
Colburn (1982)

Lucas-Walecki construction: gives one canonical form ▸ SkipLW

Hamiltonian decomposition algorithm

– for decomposition into hamiltonian cycles

- When n is even $n/2 - 1$ edges must be visited twice
- Lucas-Walecki construction (1892)
- Construction: n even

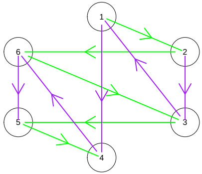


1 2 6 3 5 4

Hamiltonian decomposition algorithm

– for decomposition into hamiltonian cycles

- When n is even $n/2 - 1$ edges must be visited twice
- Lucas-Walecki construction (1892)
- Construction: n even



1 2 6 3 5 4

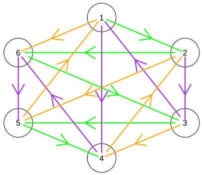
2 3 1 4 6 5

Hamiltonian decomposition algorithm

– for decomposition into hamiltonian cycles

- When n is even $n/2 - 1$ edges must be visited twice
- Lucas-Walecki construction (1892)
- Construction: n even

same as zig-zag method used in Wegman (1990)



1 2 6 3 5 4

2 3 1 4 6 5

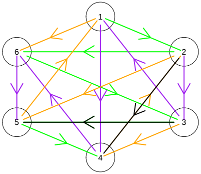
3 4 2 5 1 6

Hamiltonian decomposition algorithm

– for decomposition into hamiltonian cycles

- When n is even $n/2 - 1$ edges must be visited twice
- Lucas-Walecki construction (1892)
- Construction: n even

same as zig-zag method used in Wegman (1990)



1 2 6 3 5 4

2 3 1 4 6 5

3 4 2 5 1 6

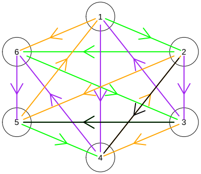
black edges- visited twice

Hamiltonian decomposition algorithm

– for decomposition into hamiltonian cycles

- When n is even $n/2 - 1$ edges must be visited twice
- Lucas-Walecki construction (1892)
- Construction: n even

same as zig-zag method used in Wegman (1990)



black edges- visited twice

1 2 6 3 5 4
2 3 1 4 6 5
3 4 2 5 1 6

- Construction: n odd

7 1 2 6 3 5 4
7 2 3 1 4 6 5
7 3 4 2 5 1 6 7

Hamiltonian decomposition algorithm

– for weighted graphs

Goal: a decomposition where weights increase: first hamiltonian has lowest weights, 2nd has next lowest weights etc.

- Greedy algorithm:
  - Start with Lucas-Walecki construction
  - WHam: use TSP for first hamiltonian, using weights, vary cycle order, direction and contact point in others.

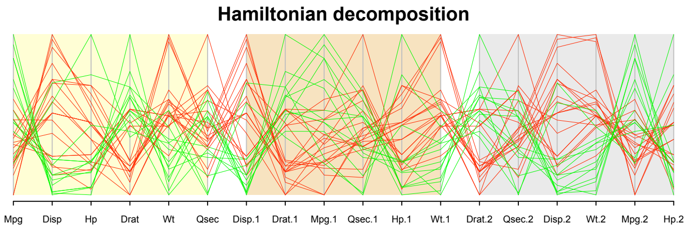
- Or:
  - Or, peripatetic TSP: k-best edge-disjoint hamiltonians
  - use other seriation as alternatives to TSP

Applications

- Pairwise comparison of treatments
- Parallel coordinates
- Interaction plots
- Star glyphs of multivariate data

Parallel coordinates

mtcars data from R: 6 variables

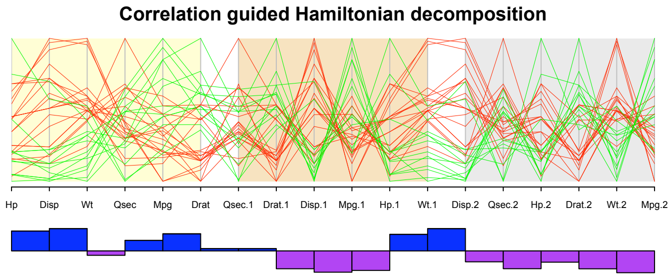


- Shows all pairs of variables adjacently.

Panel colors - three hamiltonian paths. Line color -transmission type.

Parallel coordinates

mtcars data from R: 6 variables

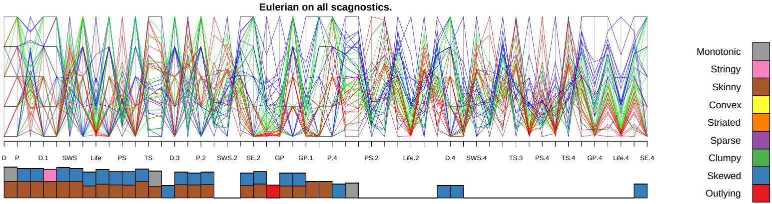


- Shows all pairs of variables adjacently.
- WHam: use correlation to choose decomposition
- Add correlation guide.

Panel colors - three hamiltonian paths. Line color -transmission type.

Parallel coordinates- more variables?

sleep data- 10 variables, 62 species



- Eulerian has 49 edges - use GrEul to follow interesting edges first.
- Barchart shows panel scagnostics

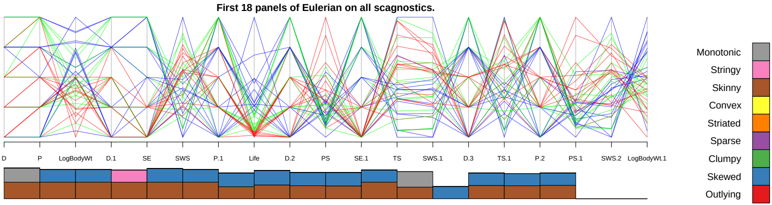
  scagnostics package, Hofmann et al.

- Lots of skinniness, skewness

Brain and body weight log transformed, colour by life expectancy
Use index values of 0.7 or more.

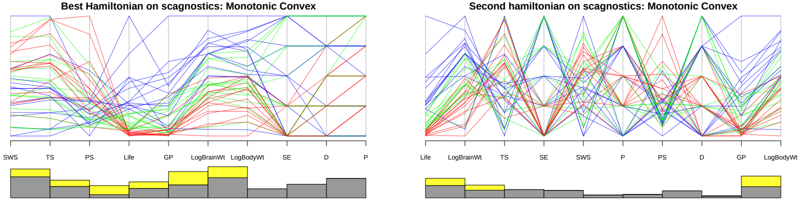
Parallel coordinates- more variables?

sleep data- 10 variables, 62 species



- Zoom on first 18 panels- captures 'interesting" relationships
- Lots of skinniness, skewness

Parallel coordinates- hamiltonian decomposition

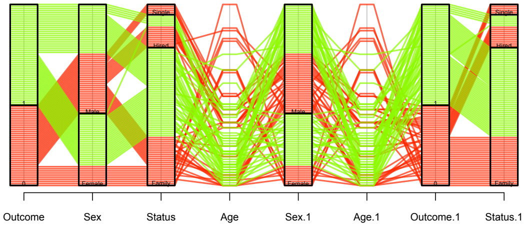


- Hamiltonians that chase "interesting" relationships-here correlational structure
- WHam: first two (of 5) hamiltonians

Monotone (grey) + convexity (yellow)

Categorical data

The Donner Party- 1846-47, Sierra Nevada



- Categorical variables: spread out uniformly within bars, along axis
- Double axis
- All pairwise relationships, and p(survival | x,y)

Concluding remarks

- Other applications: PCP-categorical, star glyphs, interaction plots
- Wegman(1990) - LW hamiltonian path algorithm in parallel coordinate displays
- Bailey et al (2003)- Hamiltonian cycles, in DOE

★★★★

- Software EulerViz R-package
- Uses TSP(Hahsler et al), scagnostics (Hofmann et al)

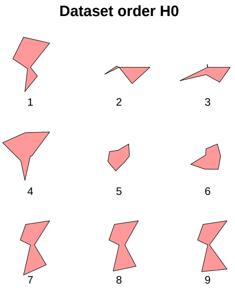
★★★★

- Further work... better algorithms?
- other types of graphs eg bipartite?
- Next talk....

Cars data

- Task: visually cluster cases

Default ordering of variables.

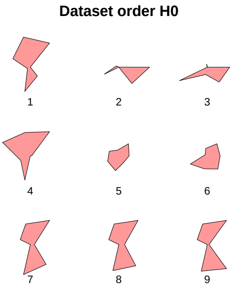


789 look similar, and to 1?
Other groups: 23, 56
4 on its own

Cars data

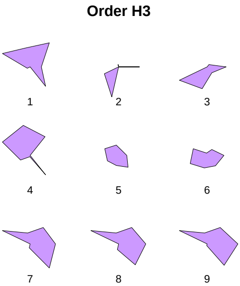
- Task: visually cluster cases

Default ordering of variables.



789 look similar, and to 1?
Other groups: 23, 56
4 on its own

Another hamiltonian

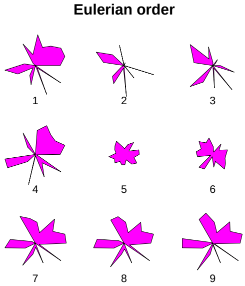


14 look similar
23 look different

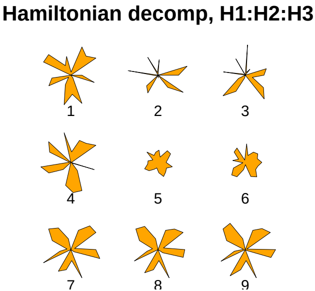
Conclusions are order dependent!

Cars data

Eulerian order



Another hamiltonian



Groups: 789,23,56,14

Less shape variation between orderings.
Conclusions are less order dependent!