Visualizing high-dimensional data:

Applying graph theory to data visualization

Wayne Oldford

based on joint work with

Catherine Hurley (Maynooth, Ireland)

Adrian Waddell (Waterloo, Canada)

WATERLOO | MATHEMATICS
STATISTICS AND ACTUARIAL SCIENCE

Challenge

- p values on each of n individuals
- modern data: n, or p, or both, can be very large

Challenge

- p values on each of n individuals
- modern data: n, or p, or both, can be very large

Challenge

- p values on each of n individuals
- modern data: n, or p, or both, can be very large

Challenge

- p values on each of n individuals
- modern data: n, or p, or both, can be very large

Challenge

- p values on each of n individuals
- modern data: n, or p, or both, can be very large

Challenge

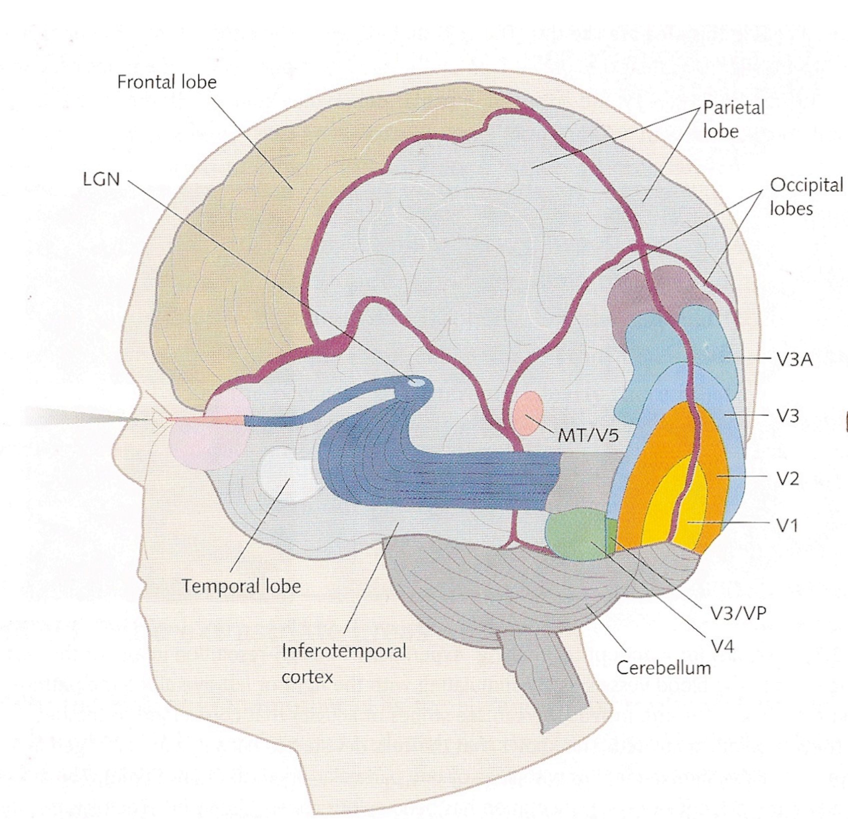
- p values on each of n individuals
- modern data: n, or p, or both, can be very large

- can have non-obvious variables, complex, unanticipated structure, ...

Data Visualization

powerful human visual system

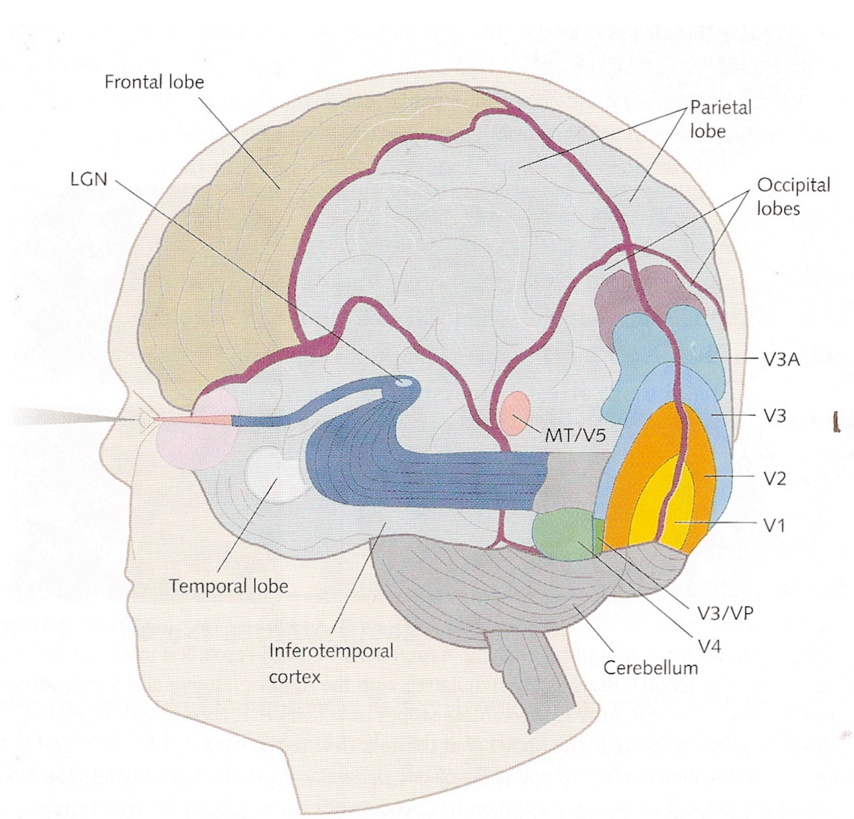
- use a variety of cues:
  - proximity, movement, shape, colour, texture, ...
- patterns, relations, like and unlike, ...
- recognition and discovery
- structure need not be anticipated



Data Visualization

powerful human visual system

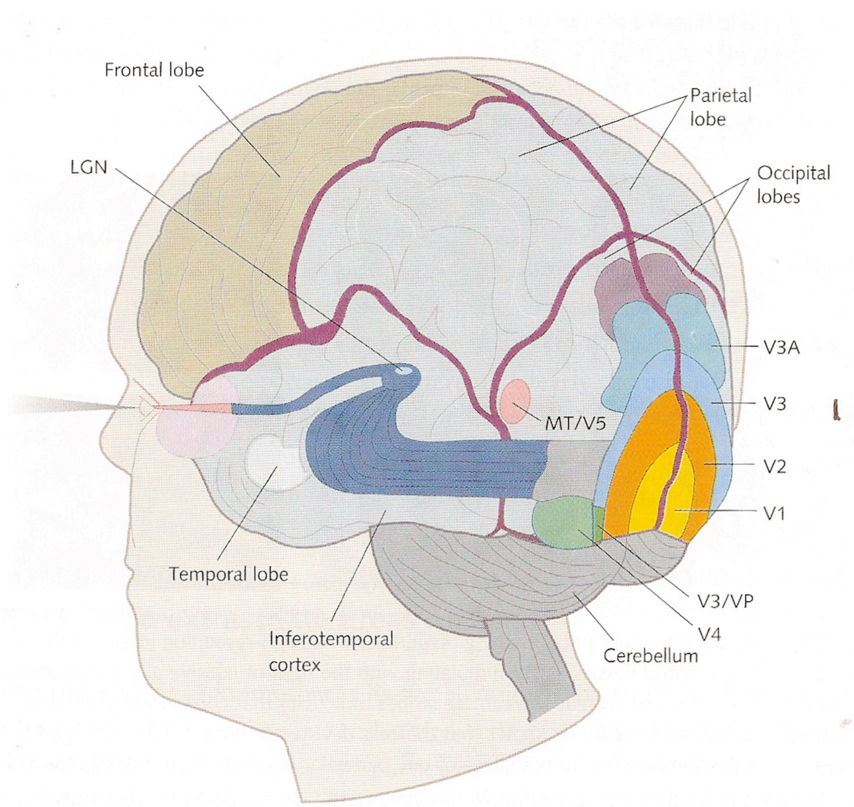
- use a variety of cues:
    - proximity, movement, shape, colour, texture, ...
- patterns, relations, like and unlike, ...
- recognition and discovery
- structure need not be anticipated



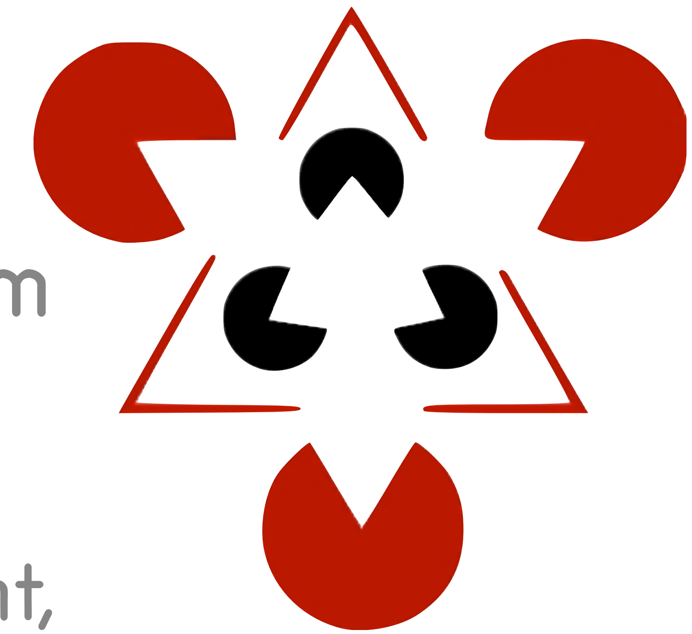
Data Visualization

powerful human visual system

- use a variety of cues:
    - proximity, movement, shape, colour, texture, ...
- patterns, relations, like and unlike, ...
- recognition and discovery
- structure need not be anticipated

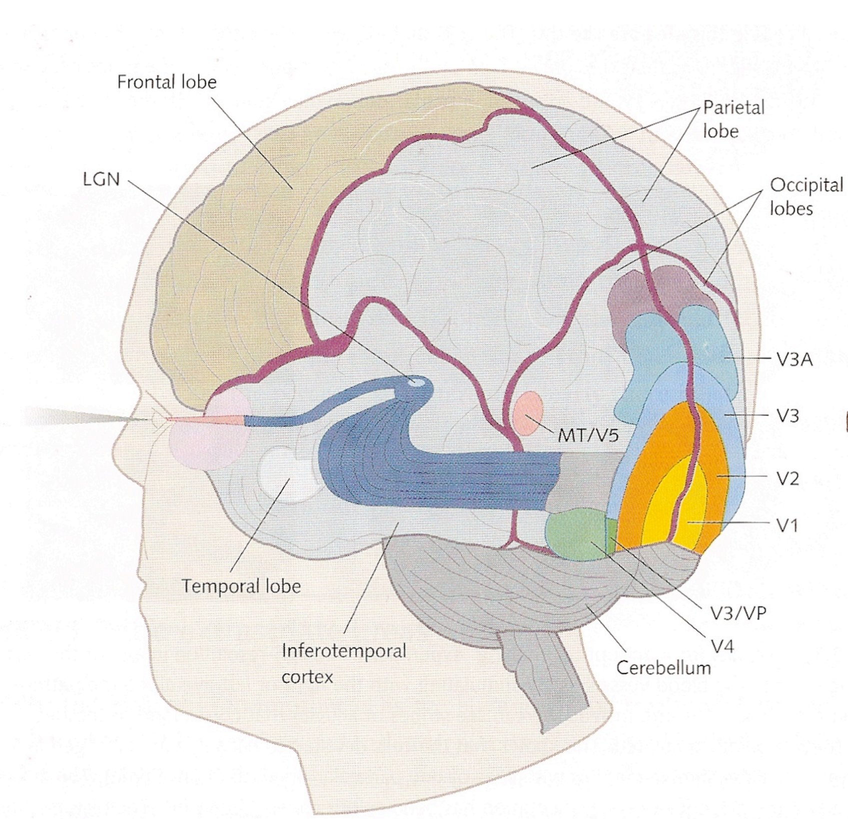


Data Visualization



powerful human visual system

- use a variety of cues:
  - proximity, movement, shape, colour, texture, ...
- patterns, relations, like and unlike, ...
- recognition and discovery
- structure need not be anticipated



Challenge

Large p

- visually, we're constrained to small p
  - locations: $p < 4$
  - use colour, shape, texture, movement, ...
- comprehension depends on only a few dimensions

  ... at a time

Challenge

Large p

- visually, we're constrained to small p
  - locations: $p < 4$
  - use colour, shape, texture, movement, ...
- comprehension depends on only a few dimensions

  ... at a time
- Approach: large number of low dimensional views

Challenge

Large p

- visually, we're constrained to small p
  - locations: $p < 4$
  - use colour, shape, texture, movement, ...
- comprehension depends on only a few dimensions

  ... at a time

- Approach: large number of low dimensional views
  - $\binom{p}{d}$ d-dimensional views, preferably highly interactive

Challenge

Large p

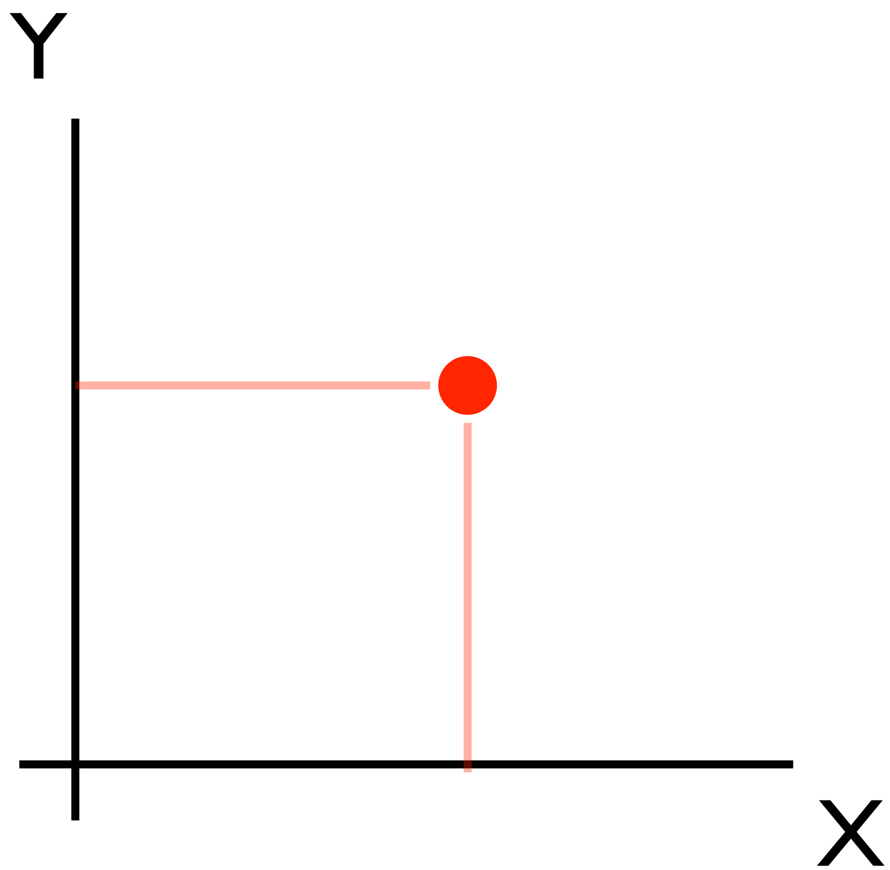
- visually, we're constrained to small p
  - locations: $p < 4$
  - use colour, shape, texture, movement, ...
- comprehension depends on only a few dimensions

  ... at a time
- Approach: large number of low dimensional views
  - $\binom{p}{d}$ d-dimensional views, preferably highly interactive
  - Which dimensions? How connected? How explored?

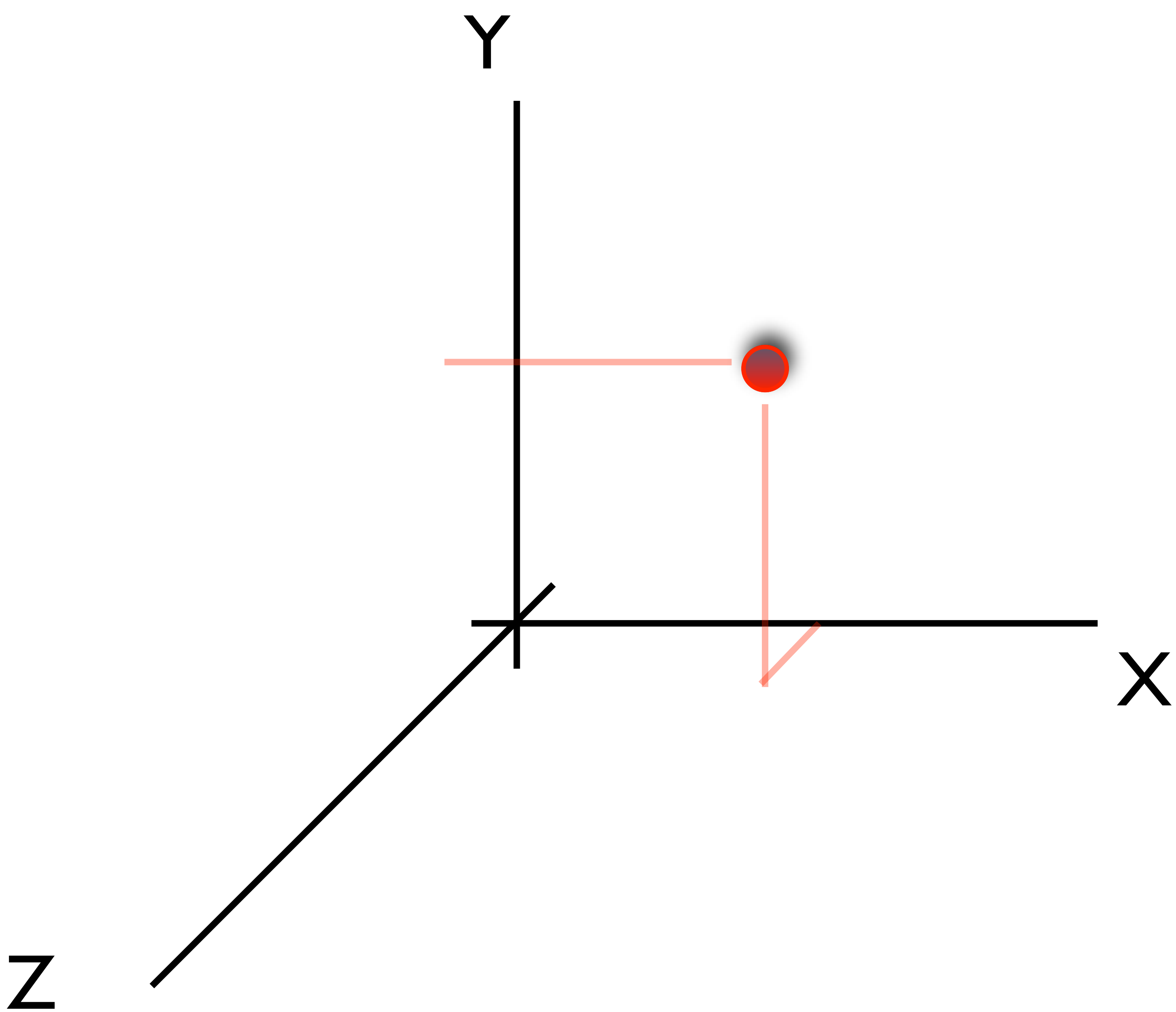
Axis systems

- Choice of coordinate axis layout
  - Orthogonal (`RnavGraph` R package)
  - Radial (`PairViz` R package)
  - Parallel (`PairViz` R package)
- Punchline
  - graph theory framework for exploratory data analysis looks very promising

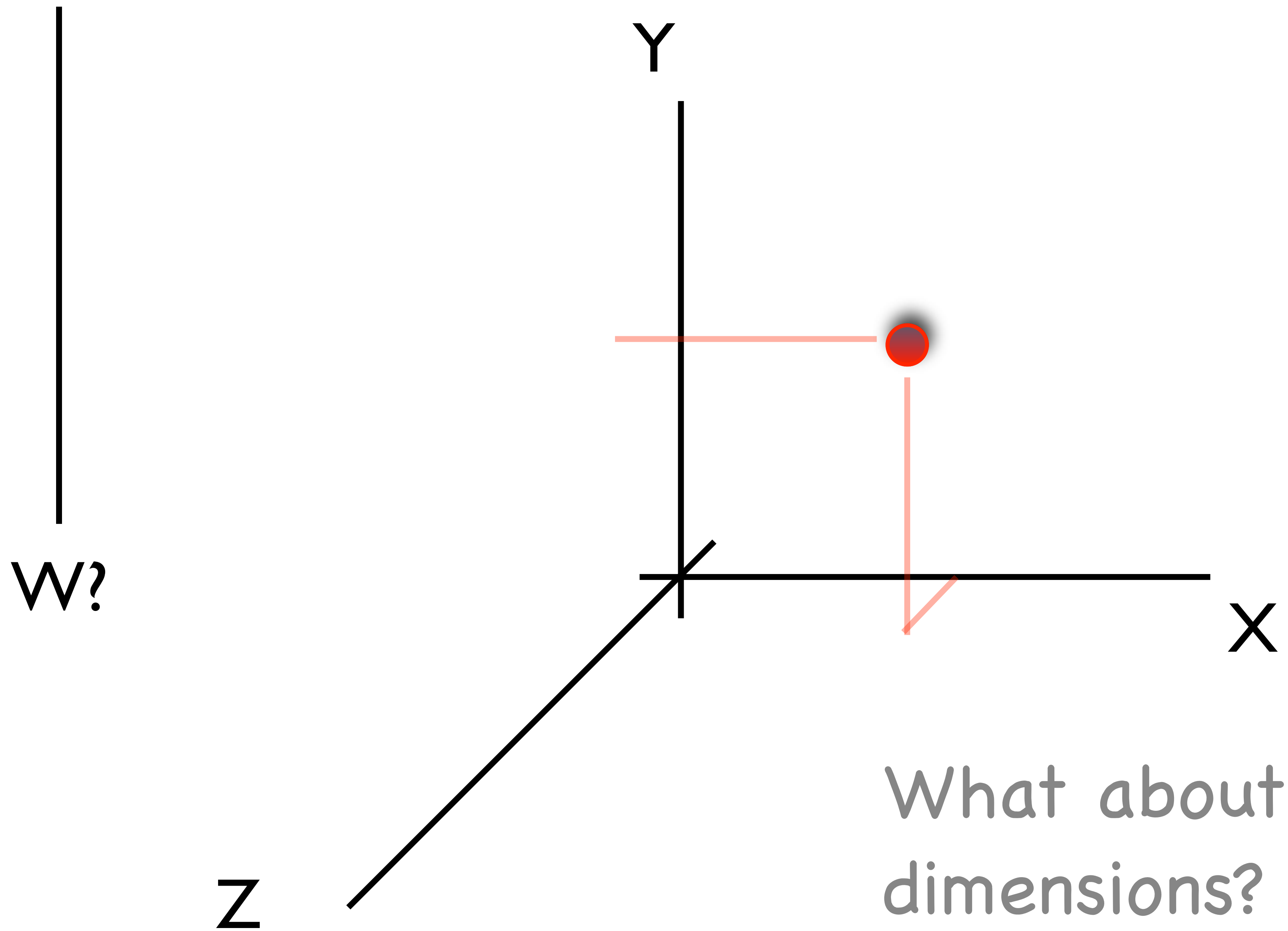
Orthogonal axes



Orthogonal axes

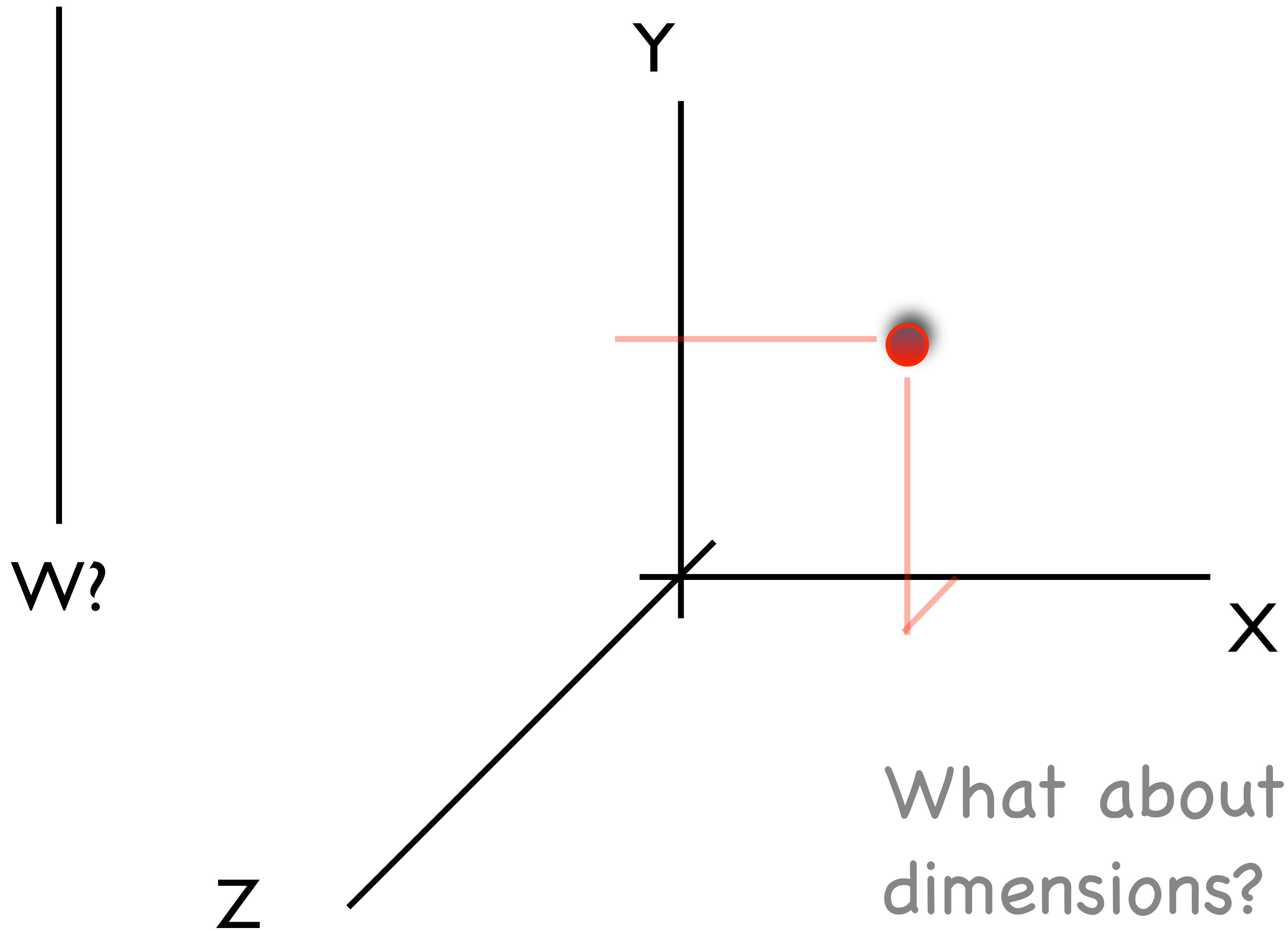


Orthogonal axes



What about more than 3 dimensions?

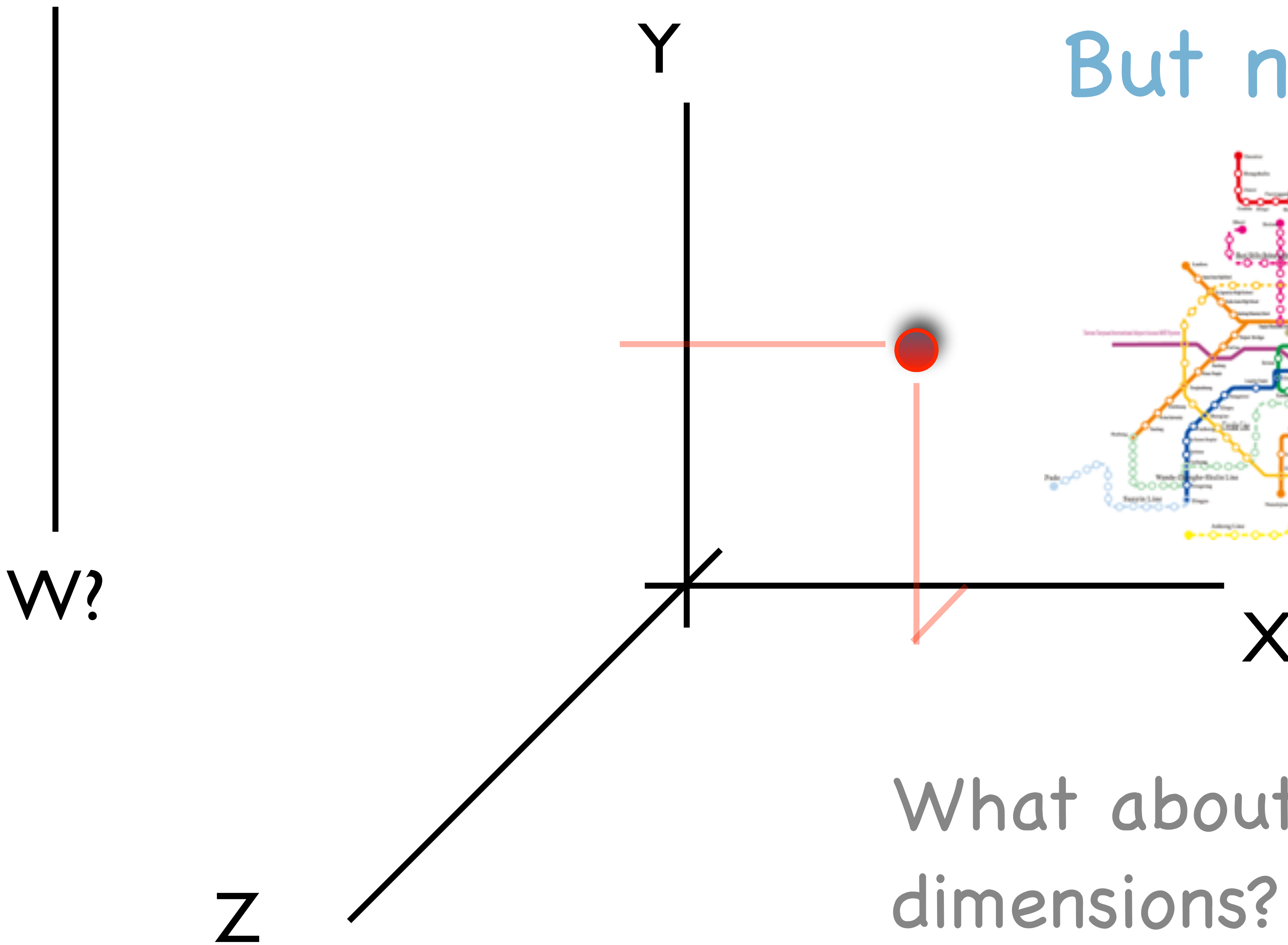
Orthogonal axes



What about more than 3 dimensions?

Travel from one space to another

Orthogonal axes



But need a Map!

What about more than 3 dimensions?

Travel from one space to another

Orthogonal axes

But need a Map!

x

z

What about more than 3 dimensions?

Travel from one space to another

Example: Italian olive oils

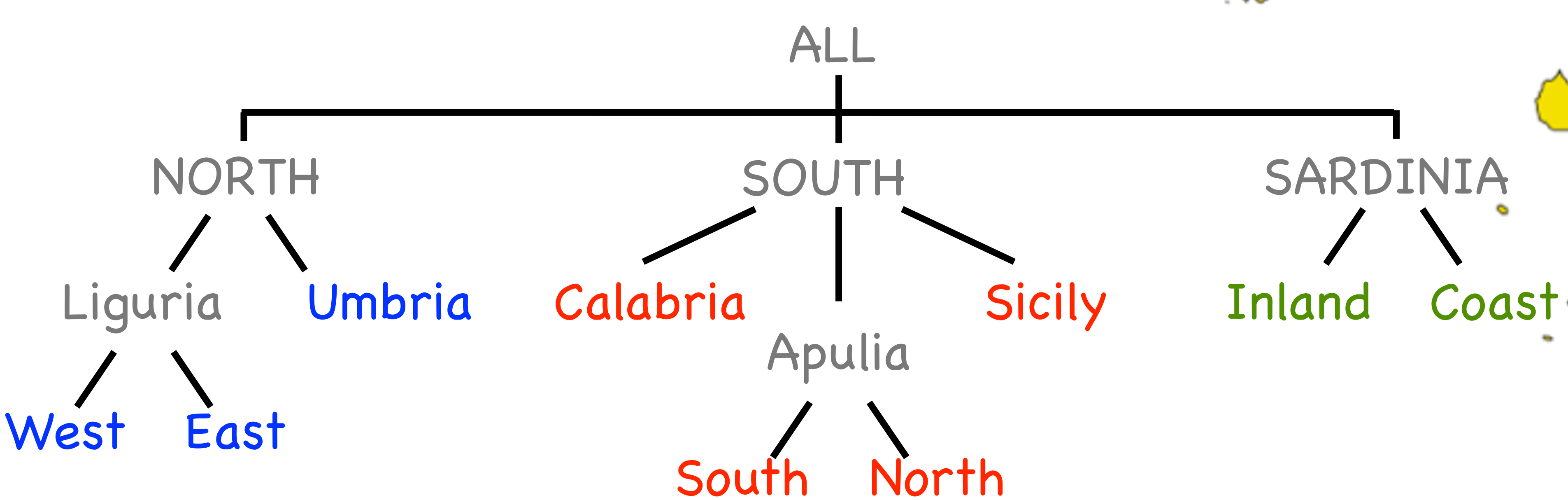
Different regions of Italy:

- NORTH (Umbria, East-Liguria, West-Liguria)
- SOUTH (Calabria, Sicily, North-Apulia, South-Apulia)
- SARDINIA (Inland, Coast)

Example: Italian olive oils

Different regions of Italy:

- NORTH (Umbria, East-Liguria, West-Liguria)
- SOUTH (Calabria, Sicily, North-Apulia, South-Apulia)
- SARDINIA (Inland, Coast)



Example: Italian olive oils

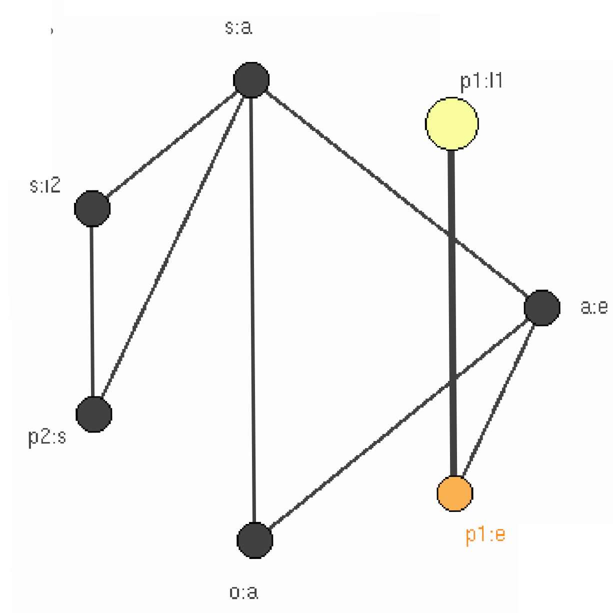
Measurements:

- n = 572 different olive samples
- concentrations of p=8 fatty acids:
  - **a**rachidic, **e**icosenoic, **l**inoleic (**l1**), **l**inolenic (**l2**), **o**leic, **p**almitic (**p1**), **p**almitoleic (**p2**), and **s**tearic.

Navigation Graphs

Connecting low-d spaces

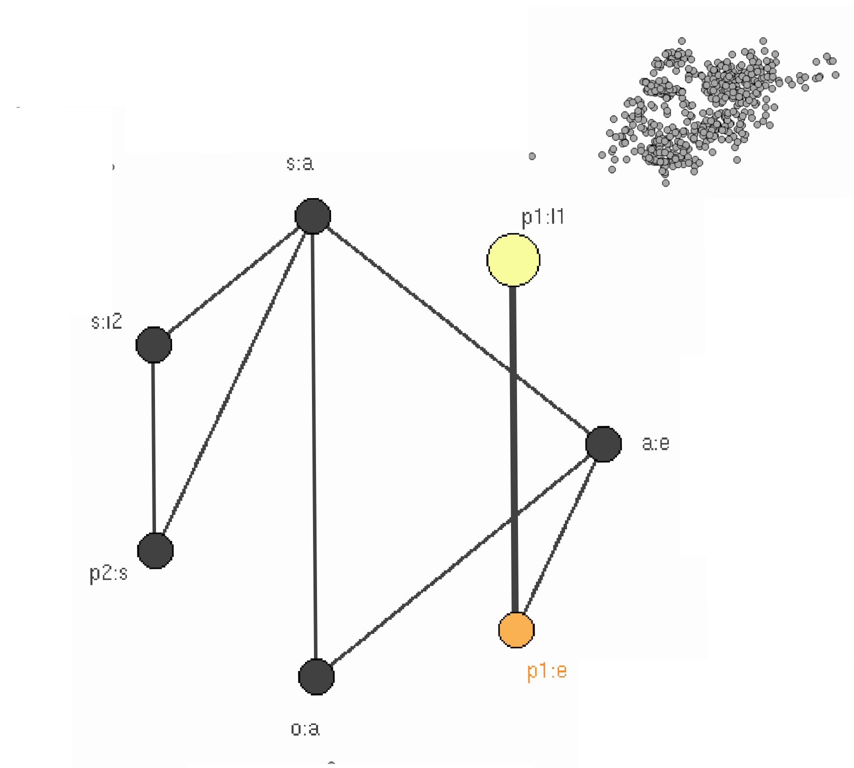
- node = variable pair
- edges connect nodes that share a variable
- could display scatterplot at each node
- edges are 3D transitions
- move from one 2D space to another through 3D (or 4D) transitions
- track/map exploration
- explore the sites!
- suggest routes



Navigation Graphs

Connecting low-d spaces

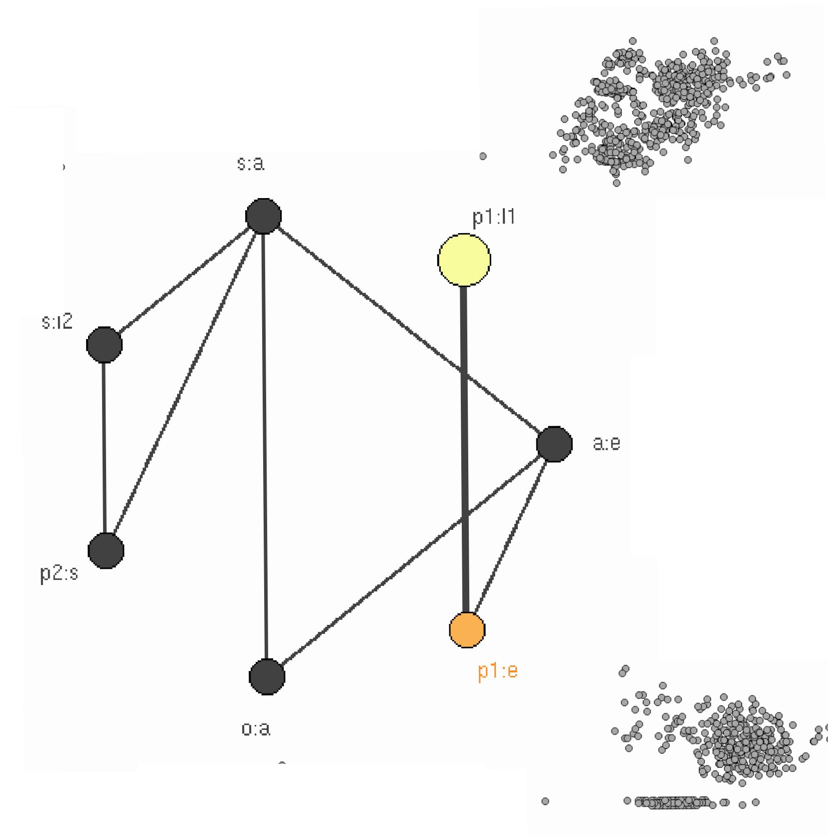
- node = variable pair
- edges connect nodes that share a variable
- could display scatterplot at each node
- edges are 3D transitions
- move from one 2D space to another through 3D (or 4D) transitions
- track/map exploration
- explore the sites!
- suggest routes



Navigation Graphs

Connecting low-d spaces

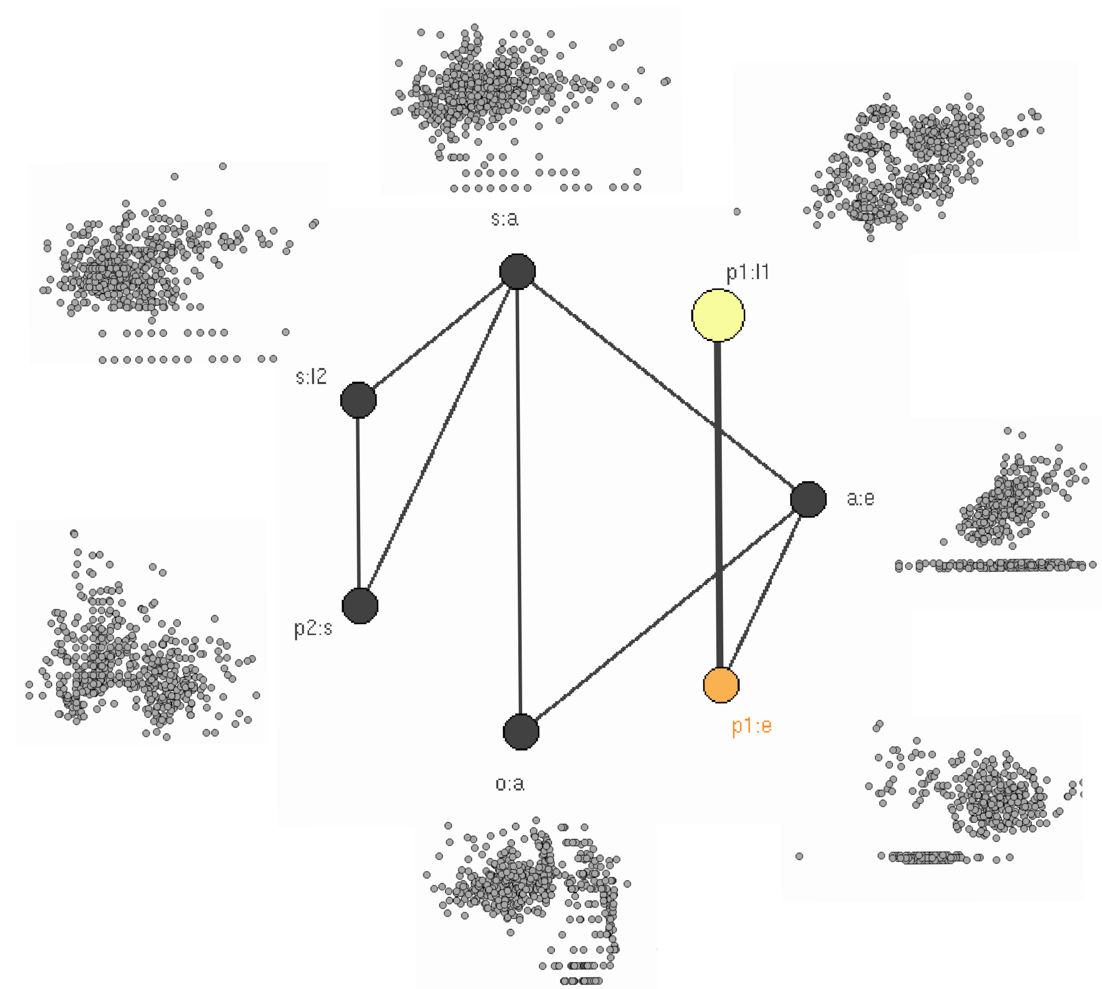
- node = variable pair
- edges connect nodes that share a variable
- could display scatterplot at each node
- edges are 3D transitions
- move from one 2D space to another through 3D (or 4D) transitions
- track/map exploration
- explore the sites!
- suggest routes



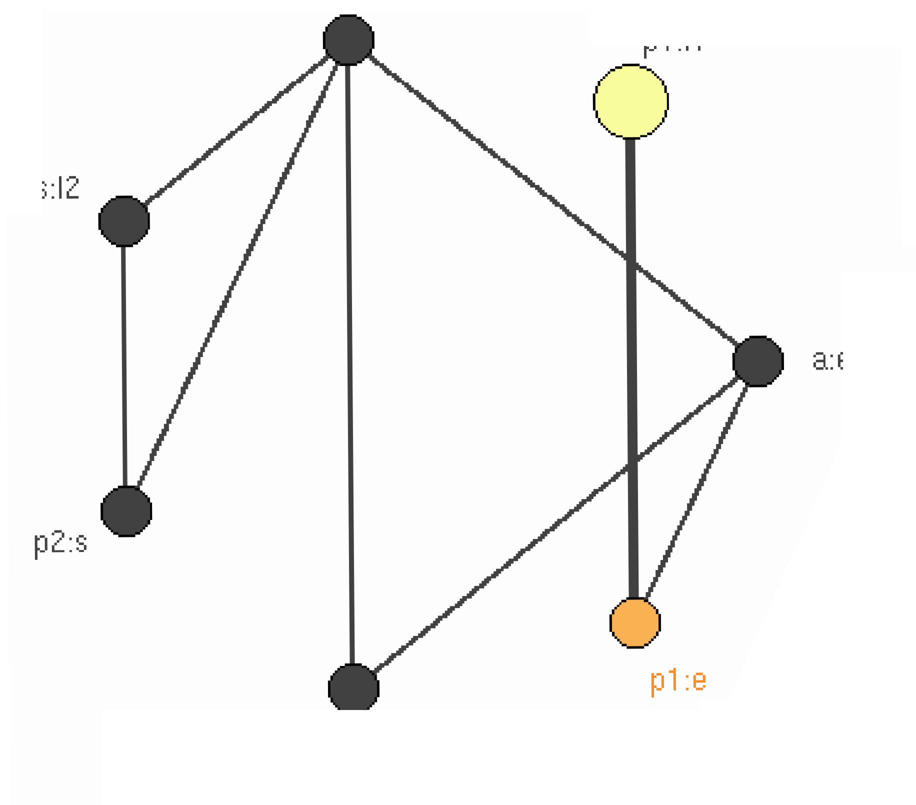
Navigation Graphs

Connecting low-d spaces

- node = variable pair
- edges connect nodes that share a variable
- could display scatterplot at each node
- edges are 3D transitions
- move from one 2D space to another through 3D (or 4D) transitions
- track/map exploration
- explore the sites!
- suggest routes



Navigation Graphs



RNavgraph

... R implementation

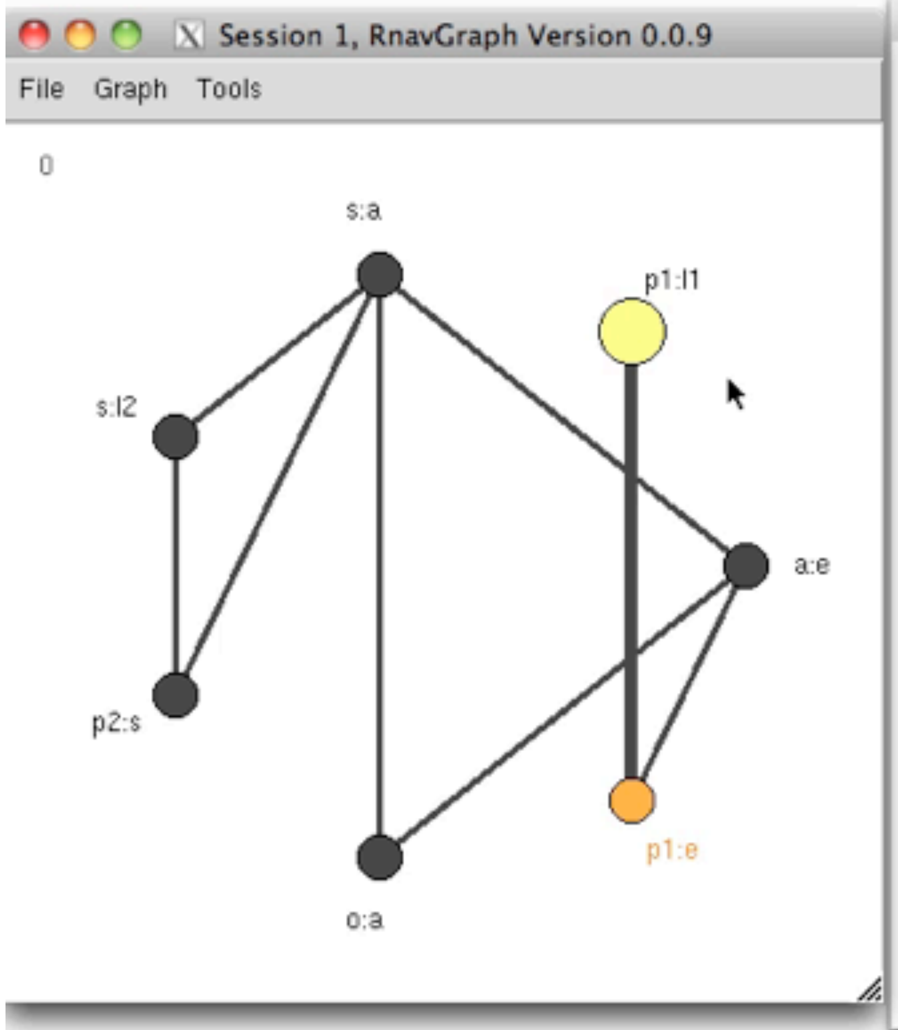
Example: Italian olive oils

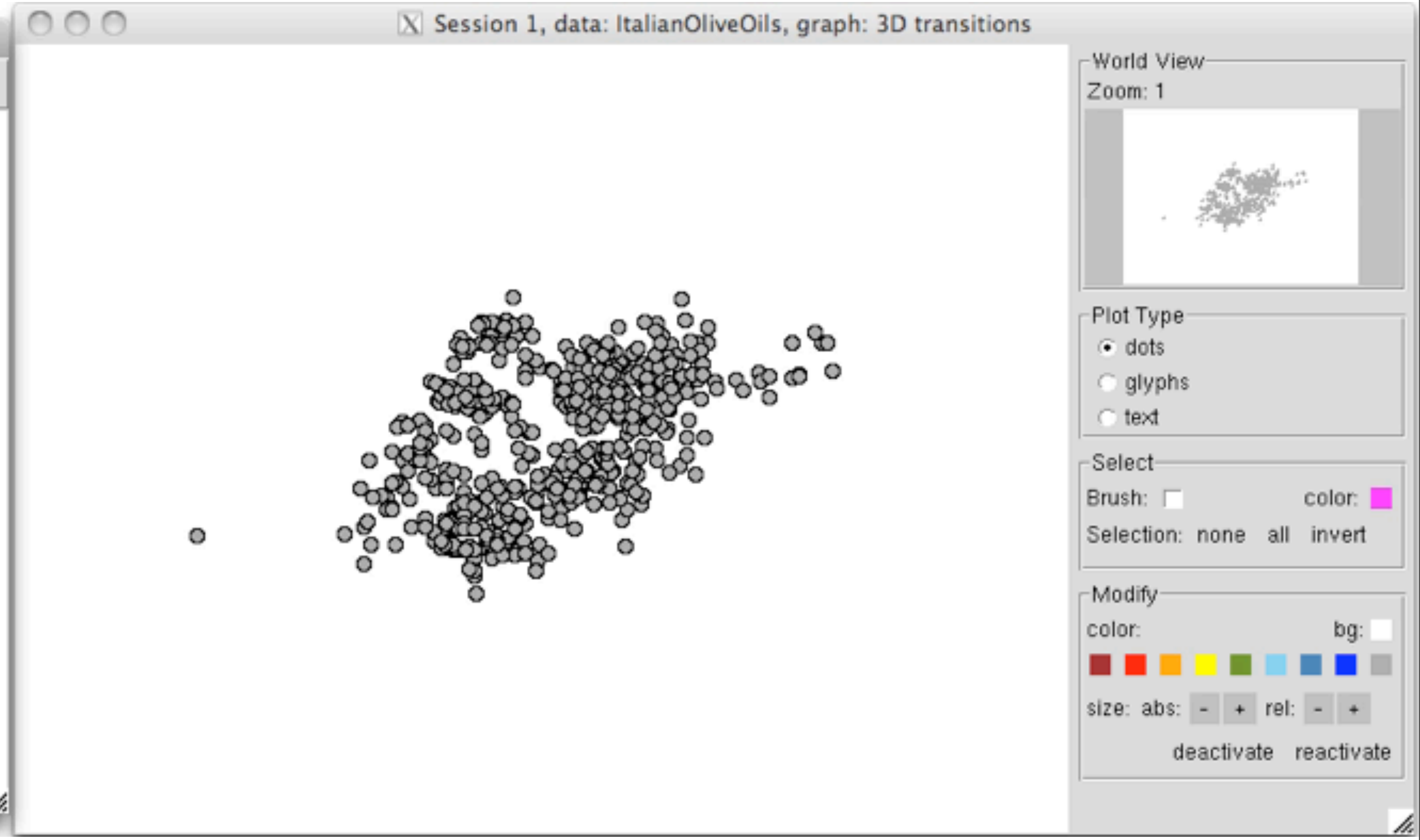
Interactive

3d transition graph

Interactive scatterplot

Example: Italian olive oils





Interactive

3d transition graph

Interactive scatterplot

Example: Italian olive oils

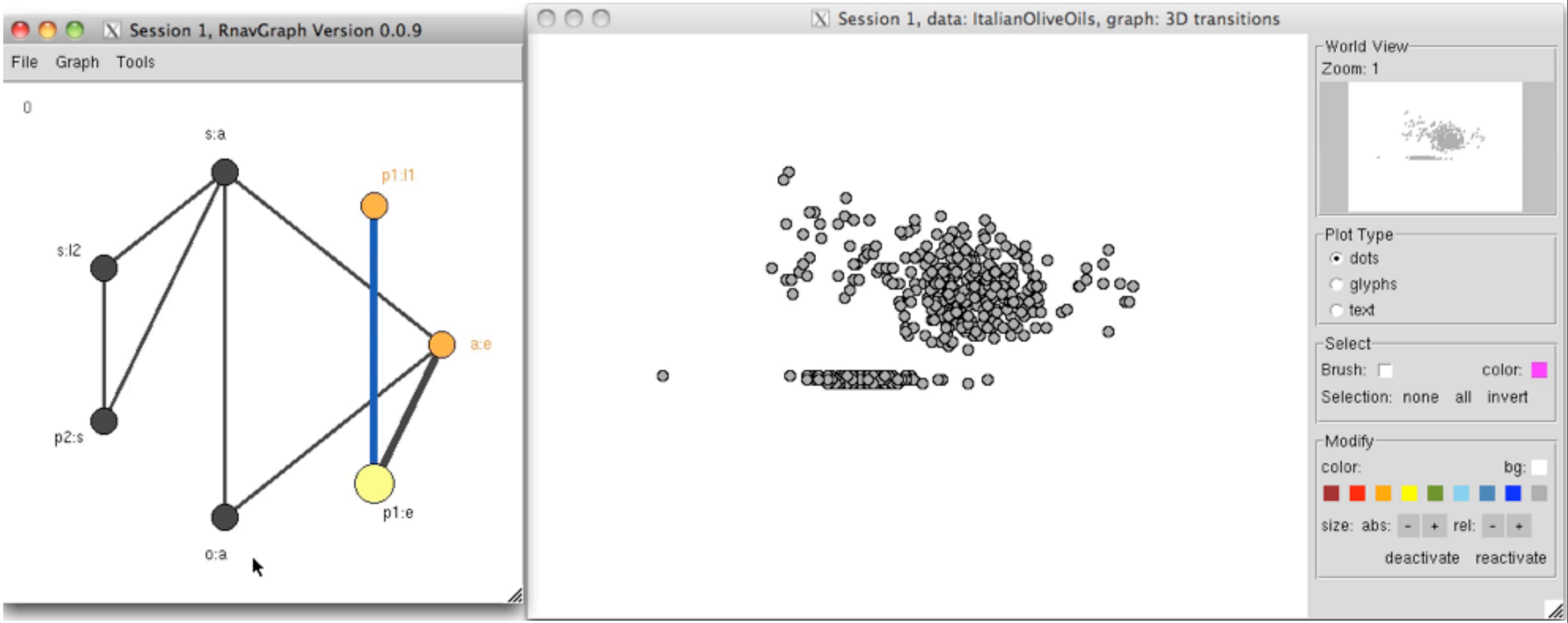
Interactive

3d transition graph

Interactive scatterplot

Move back and forth by hand

Example: Italian olive oils

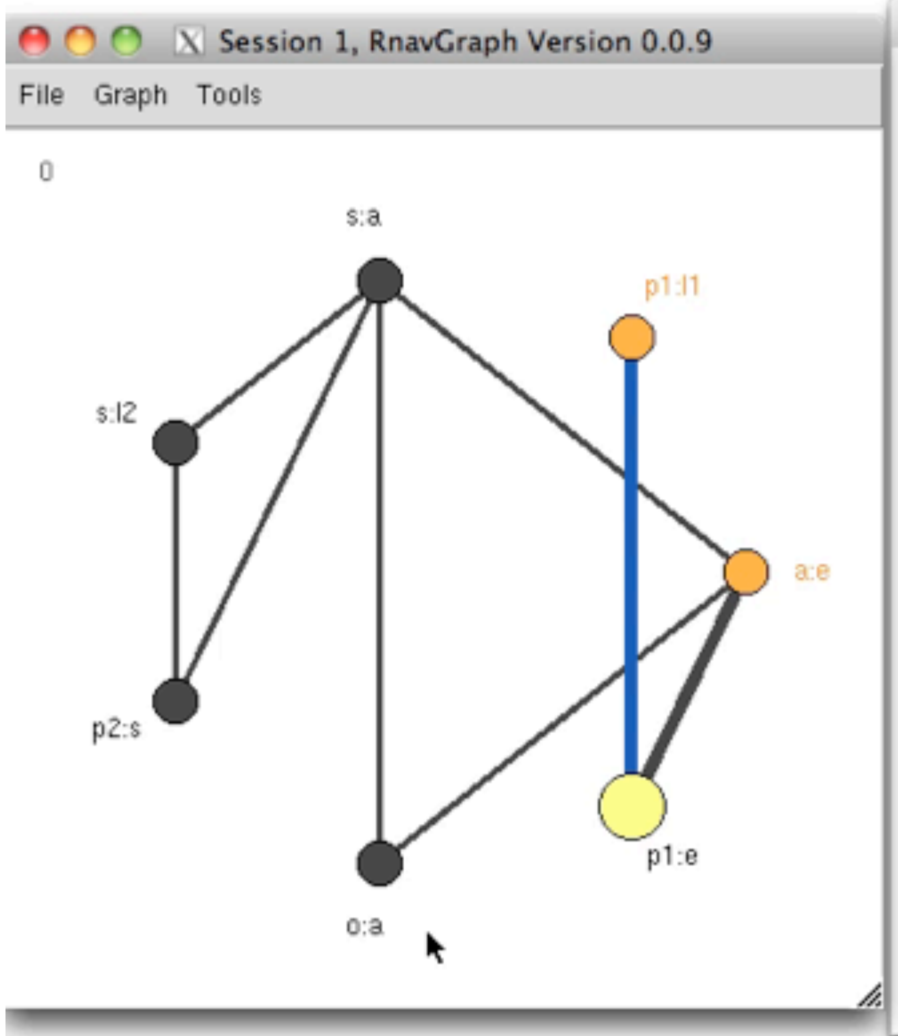


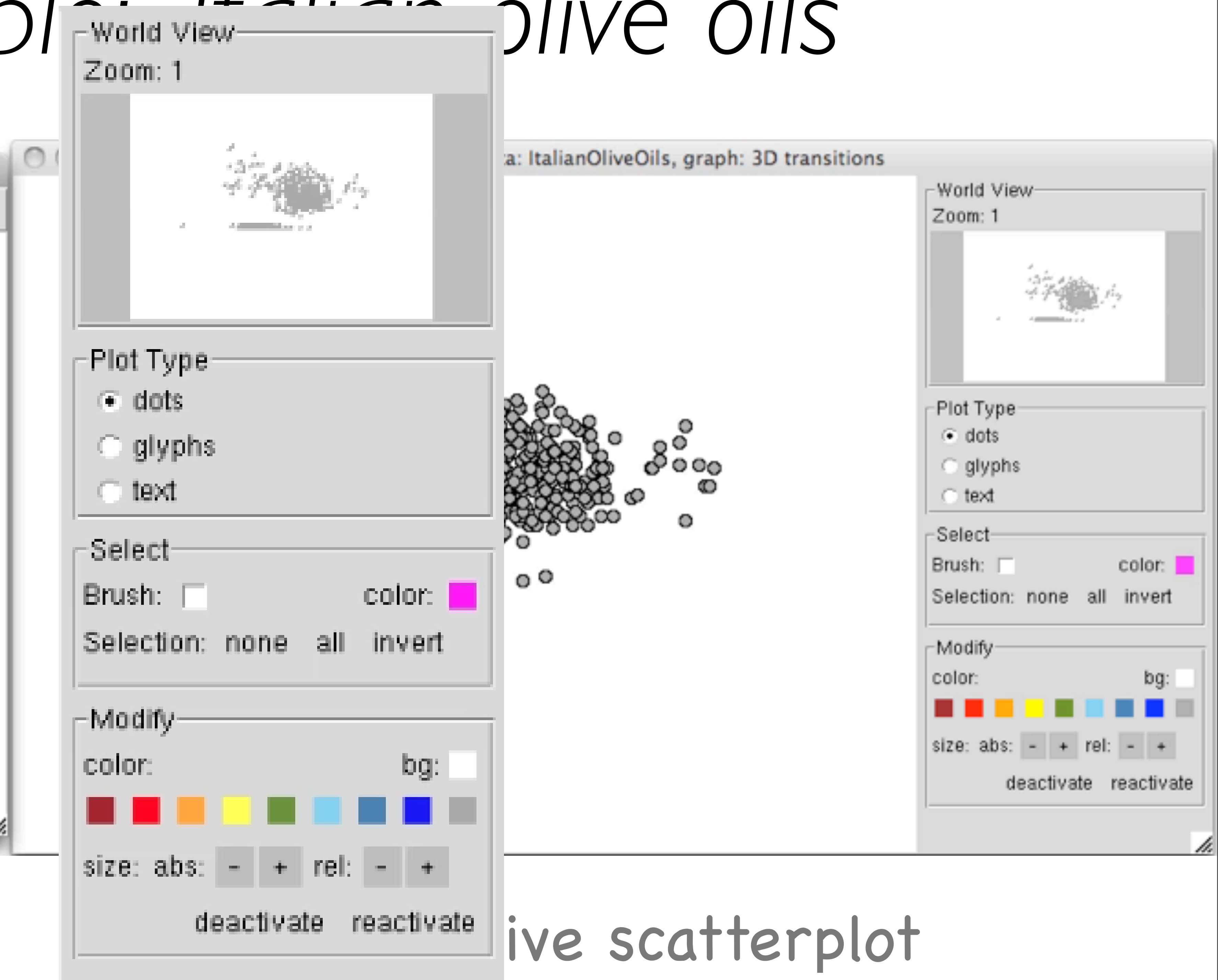
Interactive 3d transition graph

Interactive scatterplot

Move back and forth by hand

Exampl… olive oils





Interactive

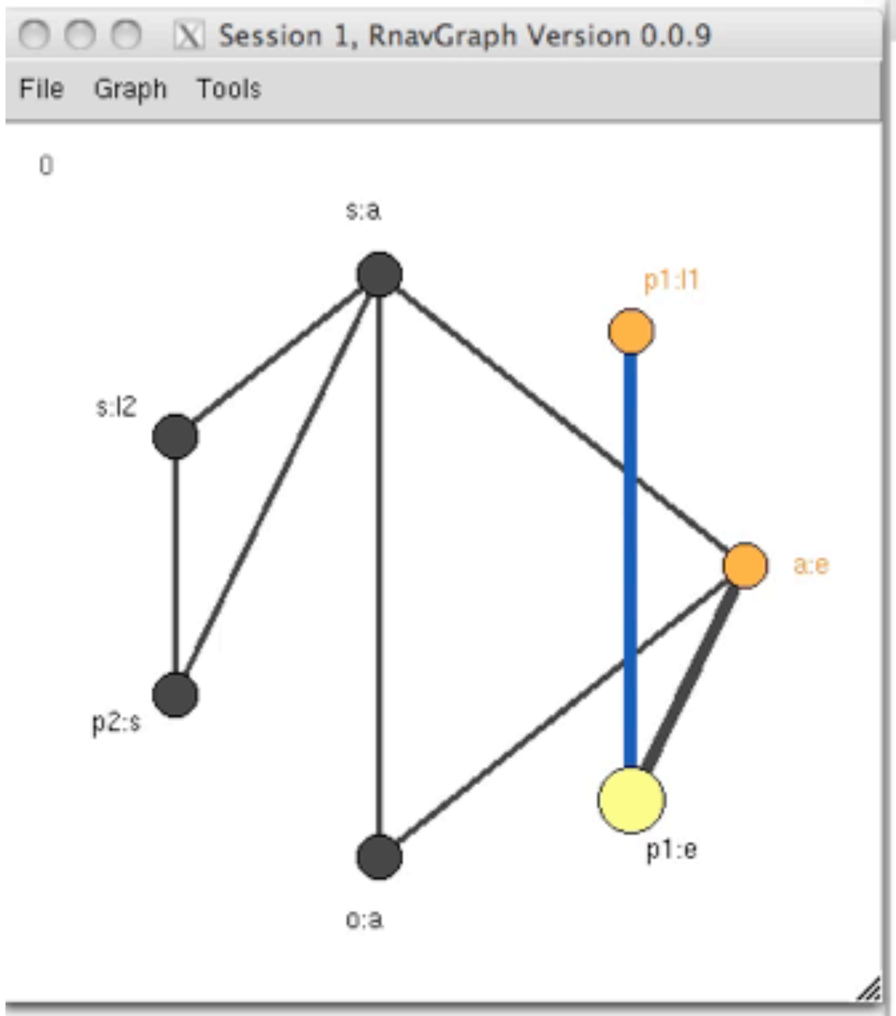
3d transition graph

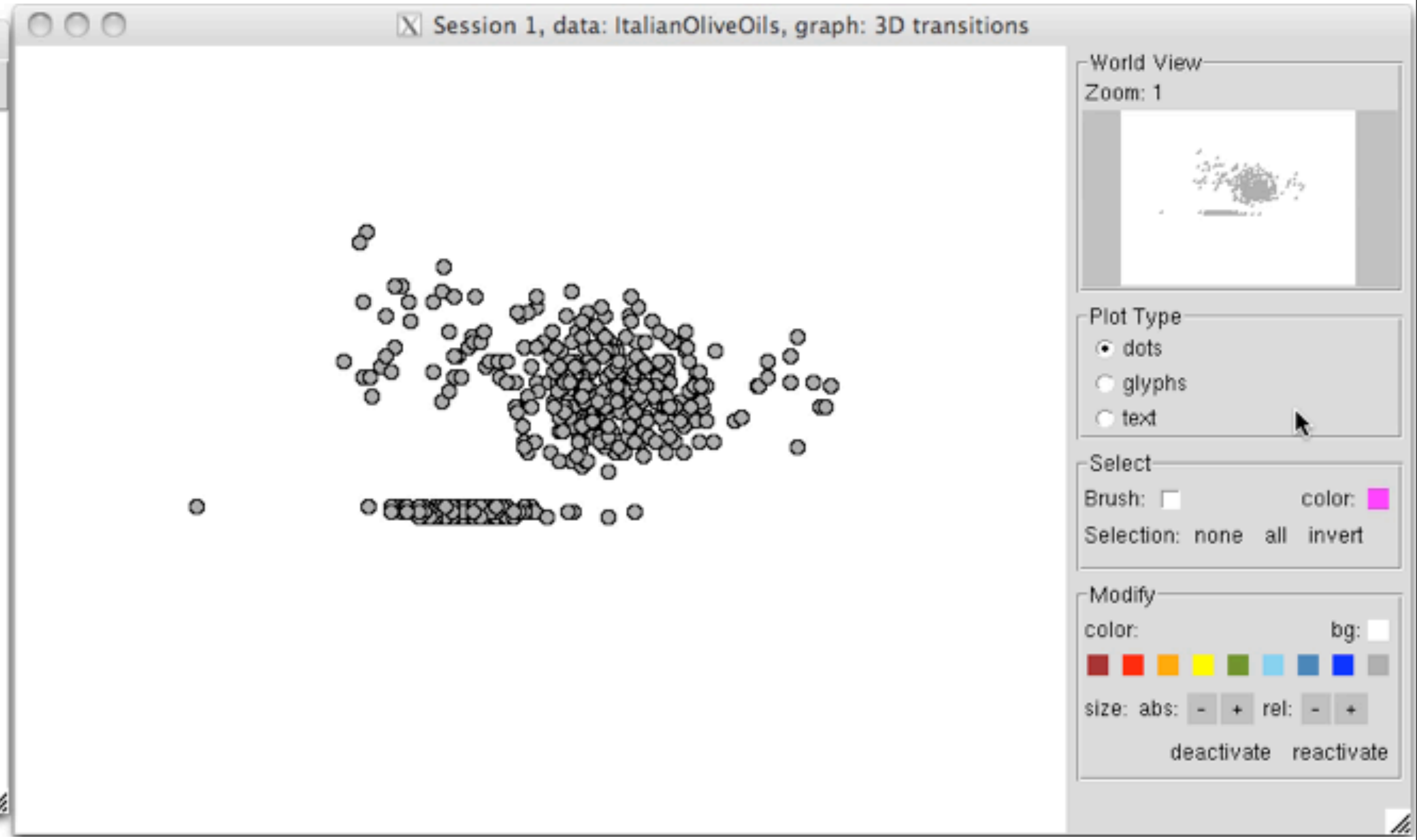
ive scatterplot

Scatterplot control panel

offers interactive features

Example: Italian olive oils





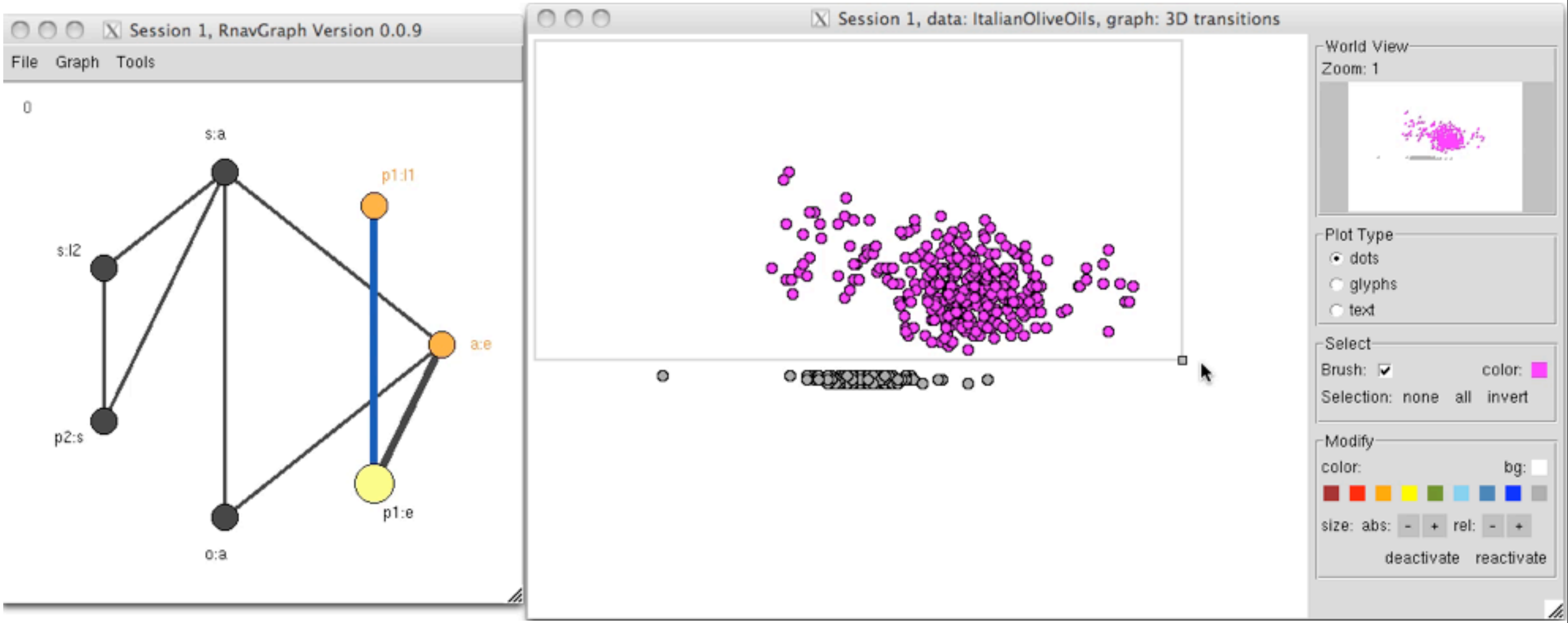
Interactive

3d transition graph

Interactive scatterplot

Brushing

Example: Italian olive oils



Interactive

3d transition graph

Interactive scatterplot

Deactivate selected points

Example: Italian olive oils

Interactive

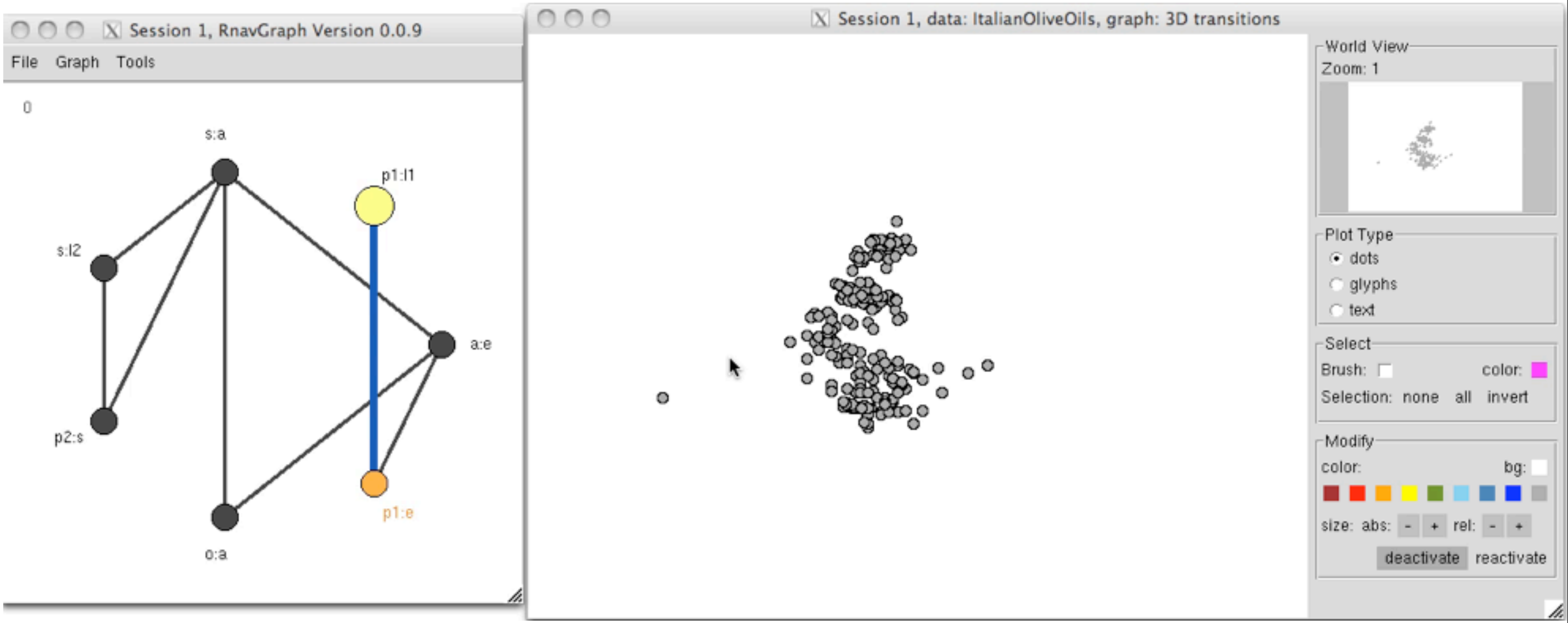
3d transition graph

Interactive scatterplot

Deactivate selected points

Return to starting position

Example: Italian olive oils



Interactive

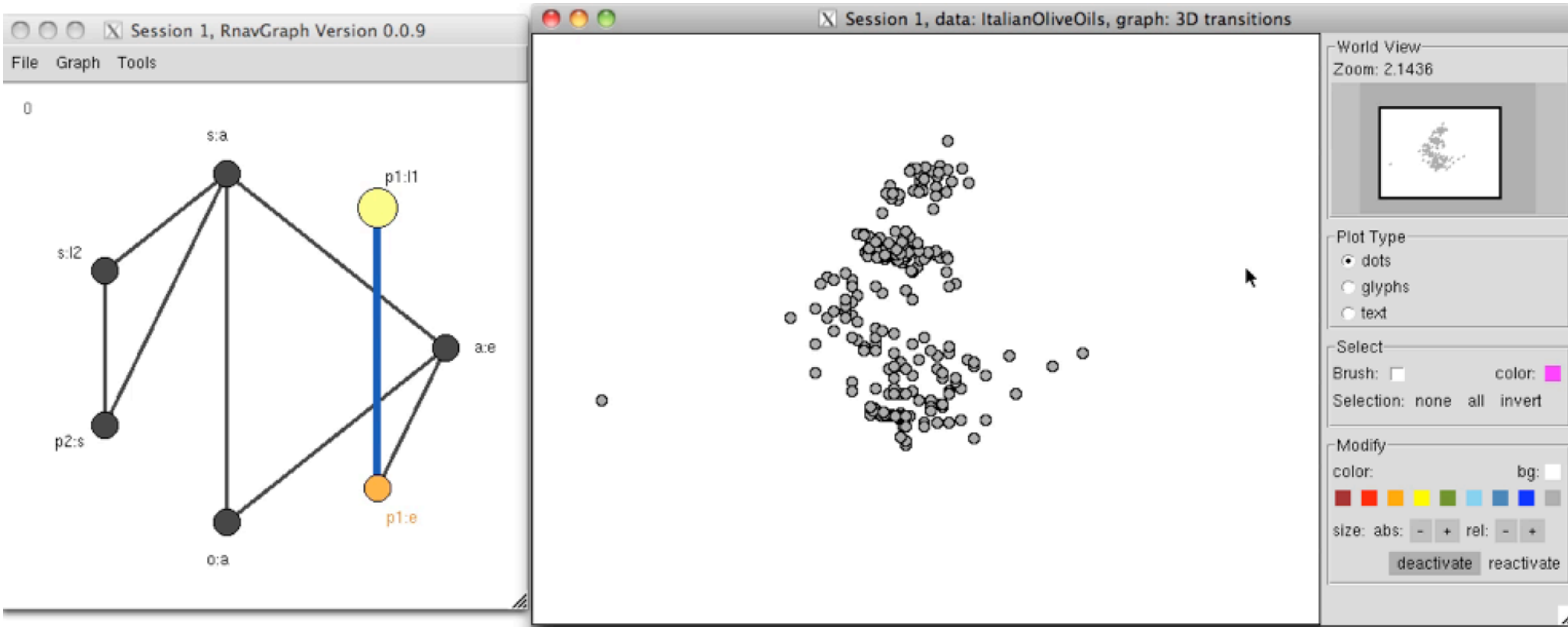
3d transition graph

Interactive scatterplot

Zoom and relocate

Note "World View" changes

Example: Italian olive oils



Interactive

3d transition graph

Interactive scatterplot

At least 3 groups;

Colour two of them.

Example: Italian olive oils

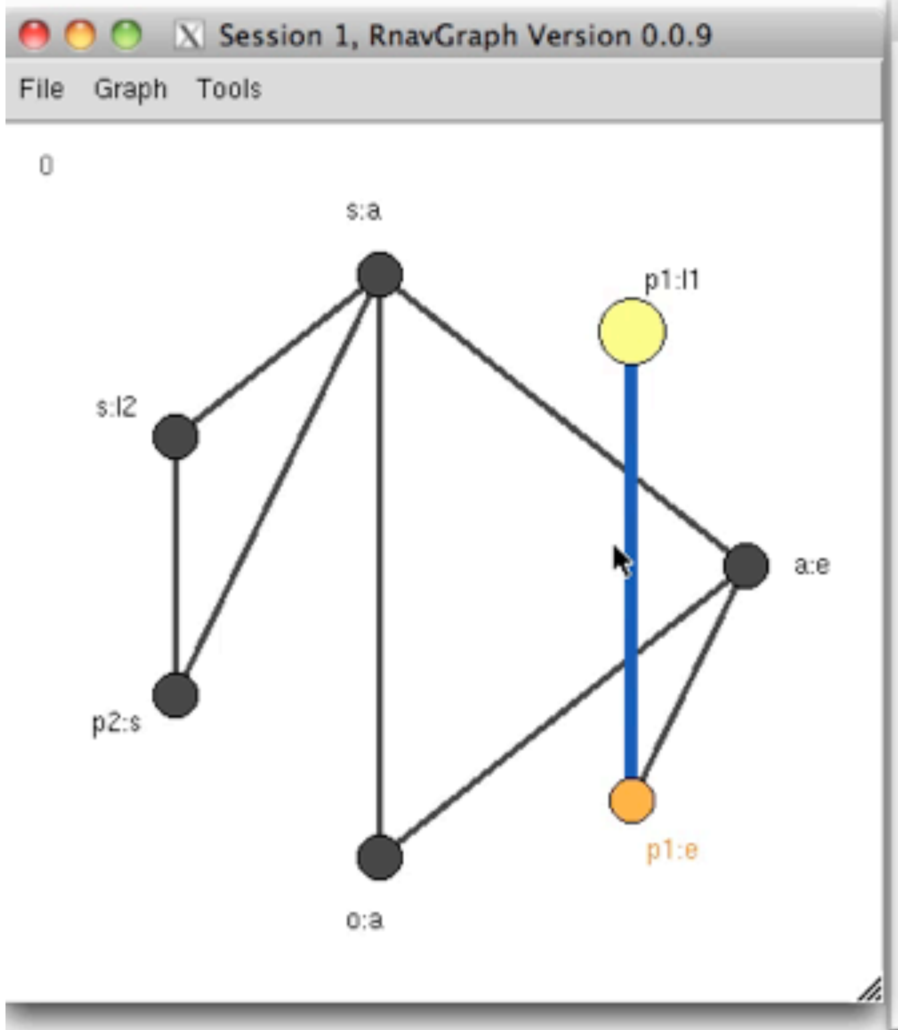
Interactive

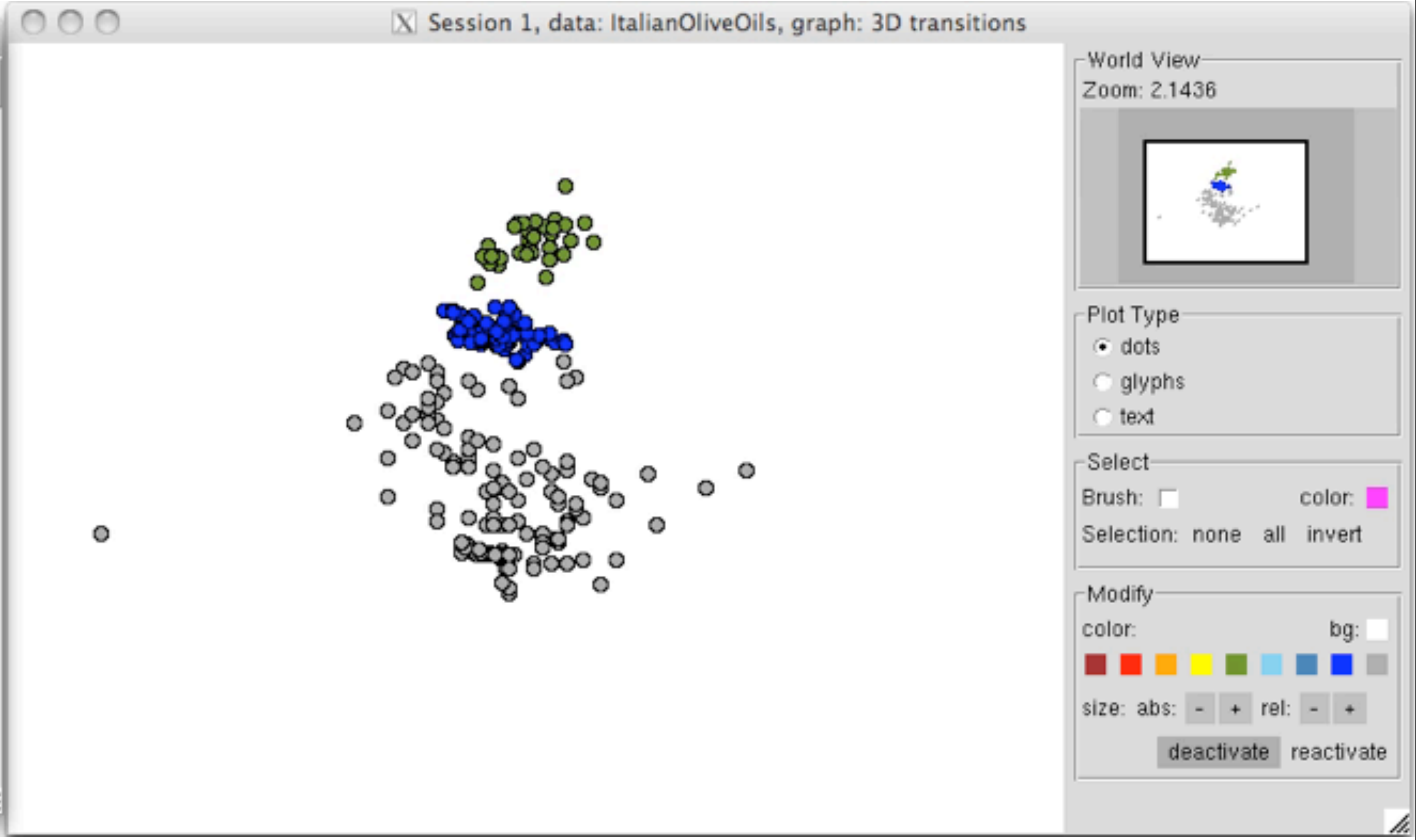
3d transition graph

Interactive scatterplot

Could also select a whole path to traverse

Example: Italian olive oils





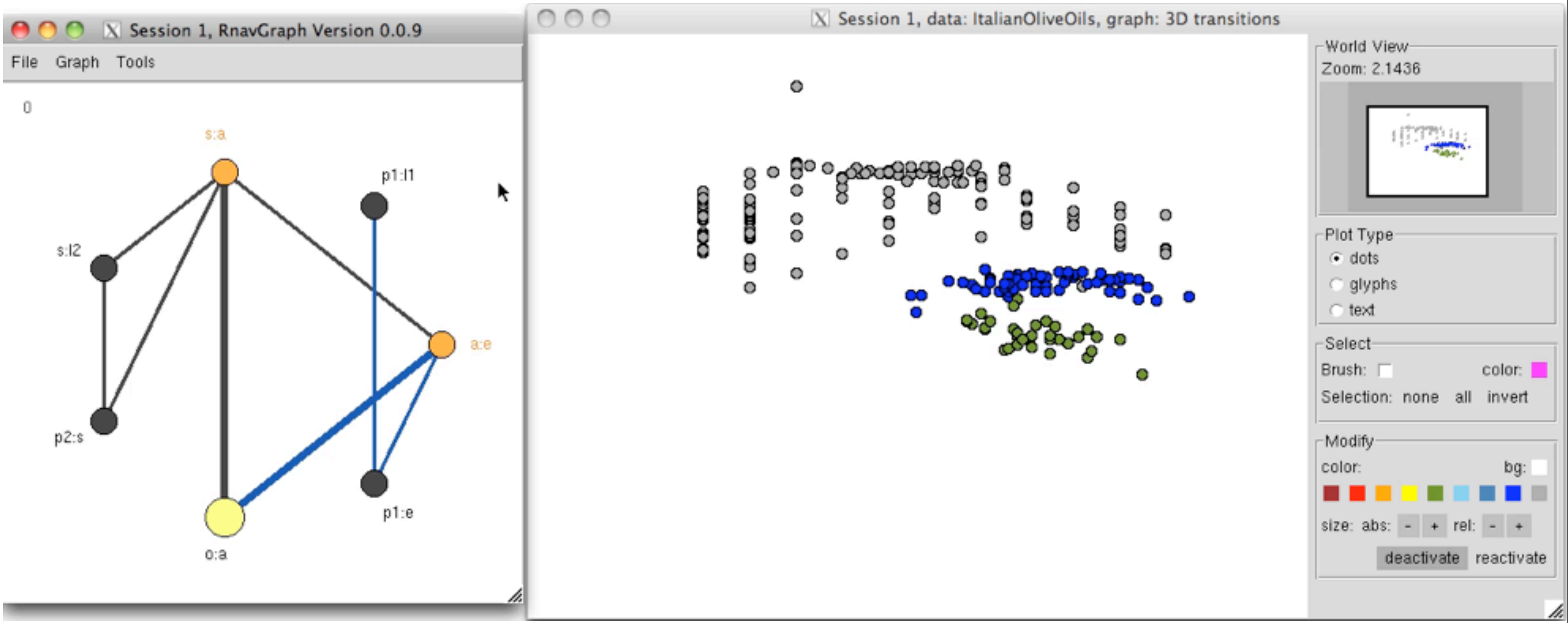
Interactive

3d transition graph

Interactive scatterplot

Could also select a whole path to traverse

Example: Italian olive oils



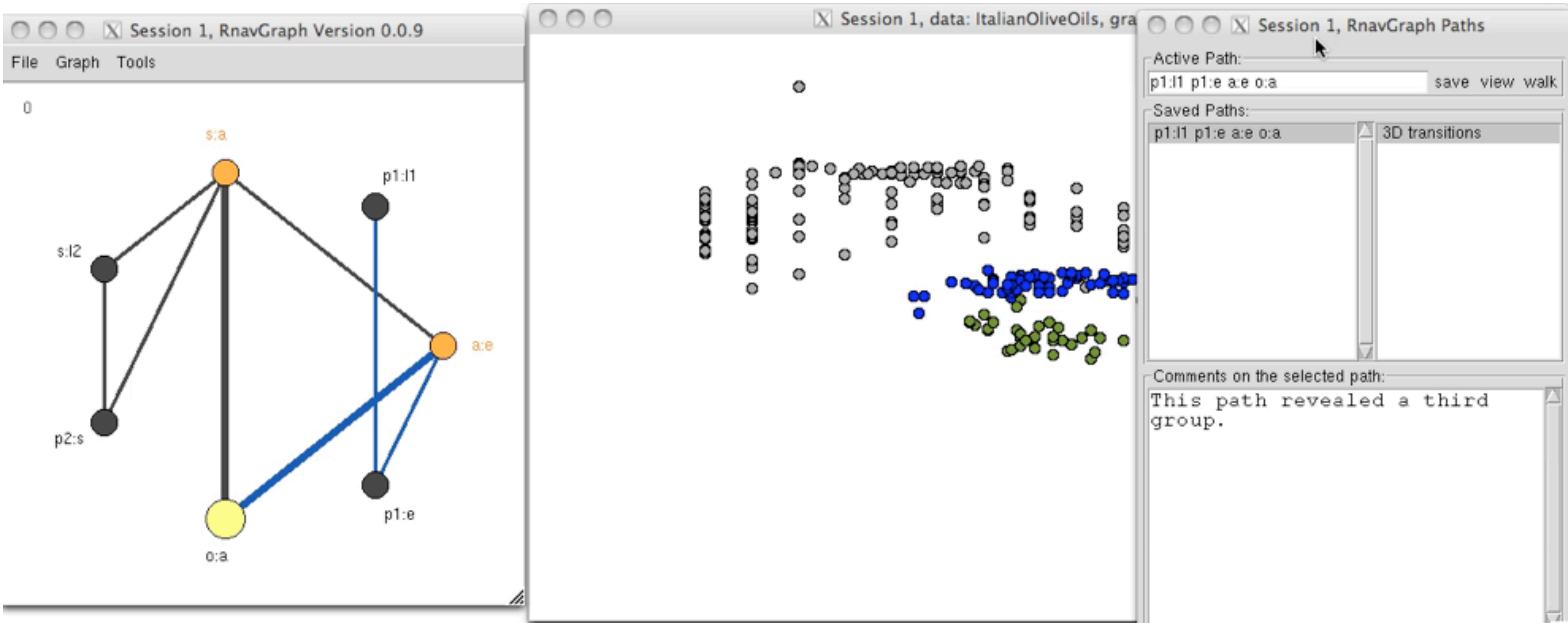
Interactive

3d transition graph

Interactive scatterplot

Paths can be saved, annotated, viewed, and walked again.

Example: Italian olive oils



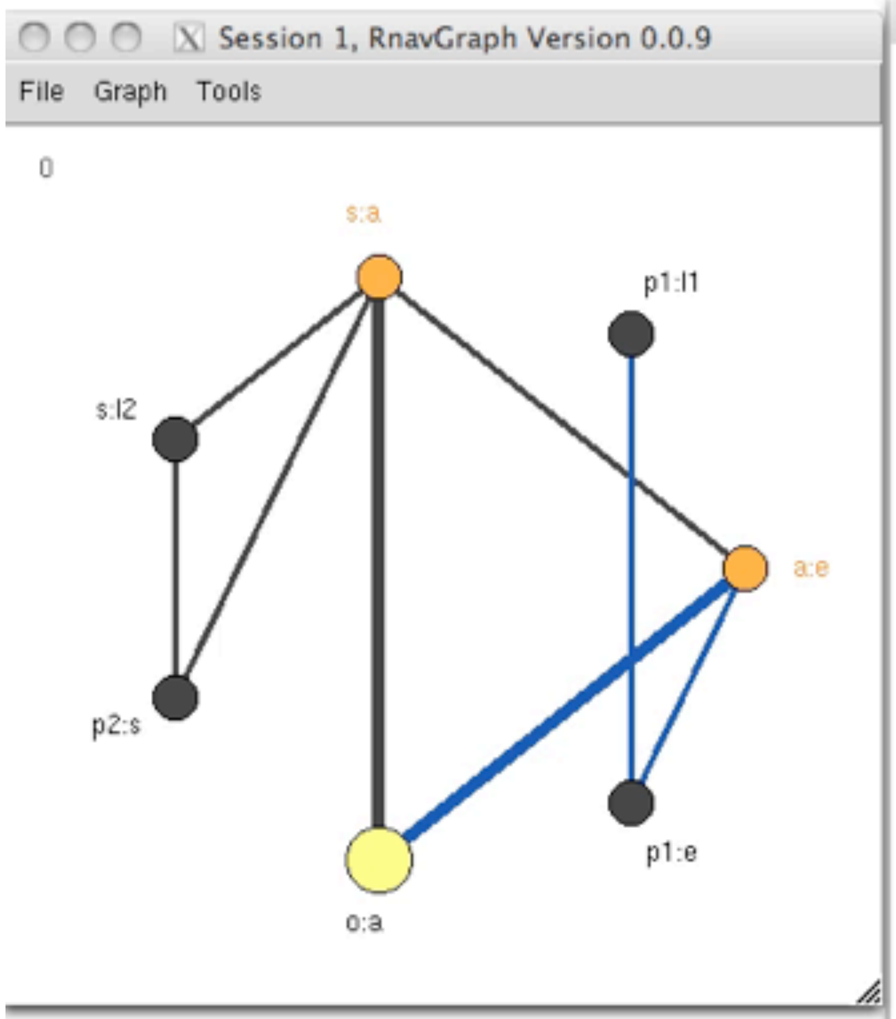
Interactive

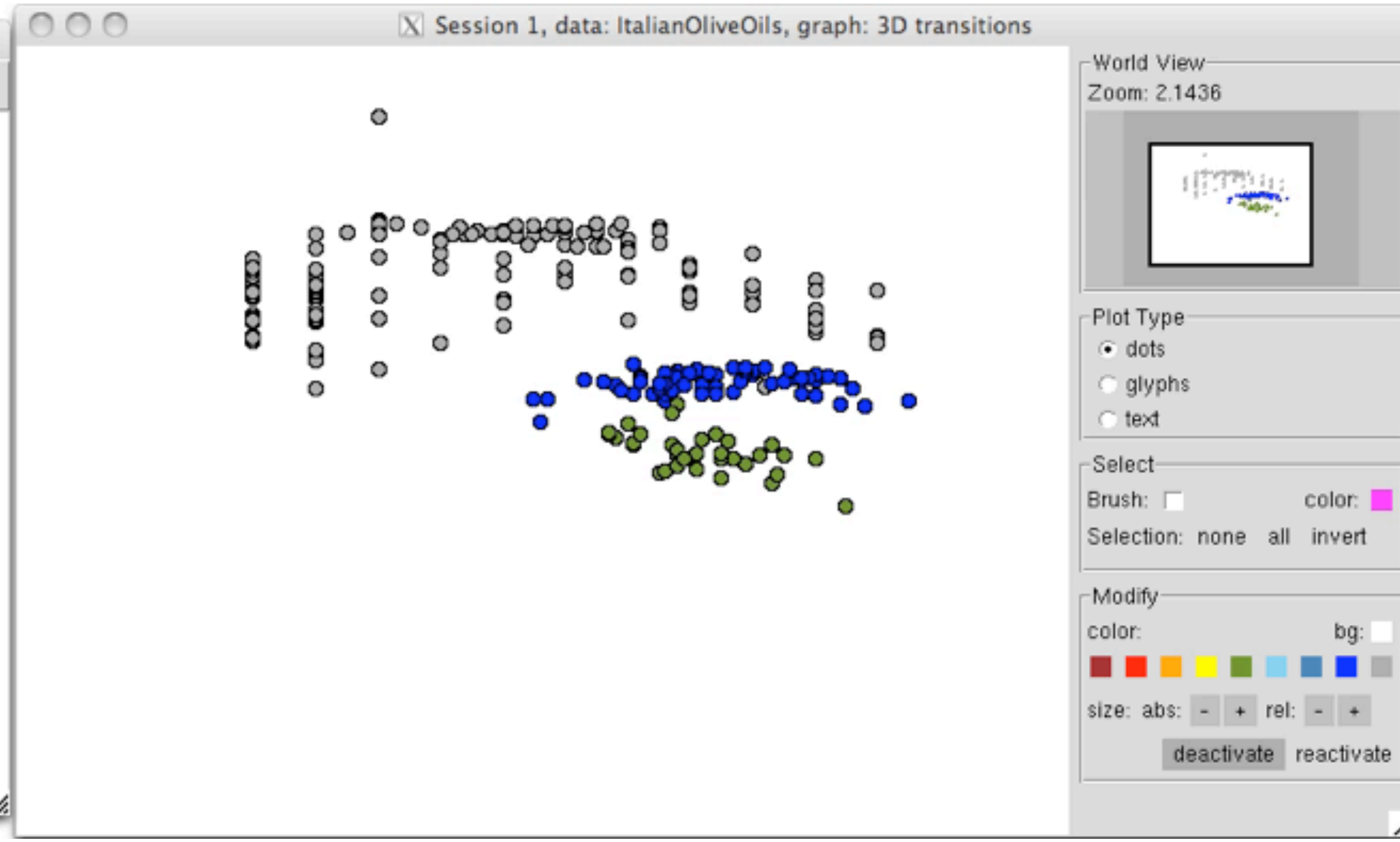
3d transition graph

Interactive scatterplot

Paths can be saved, annotated, viewed, and walked again.

Example: Italian olive oils





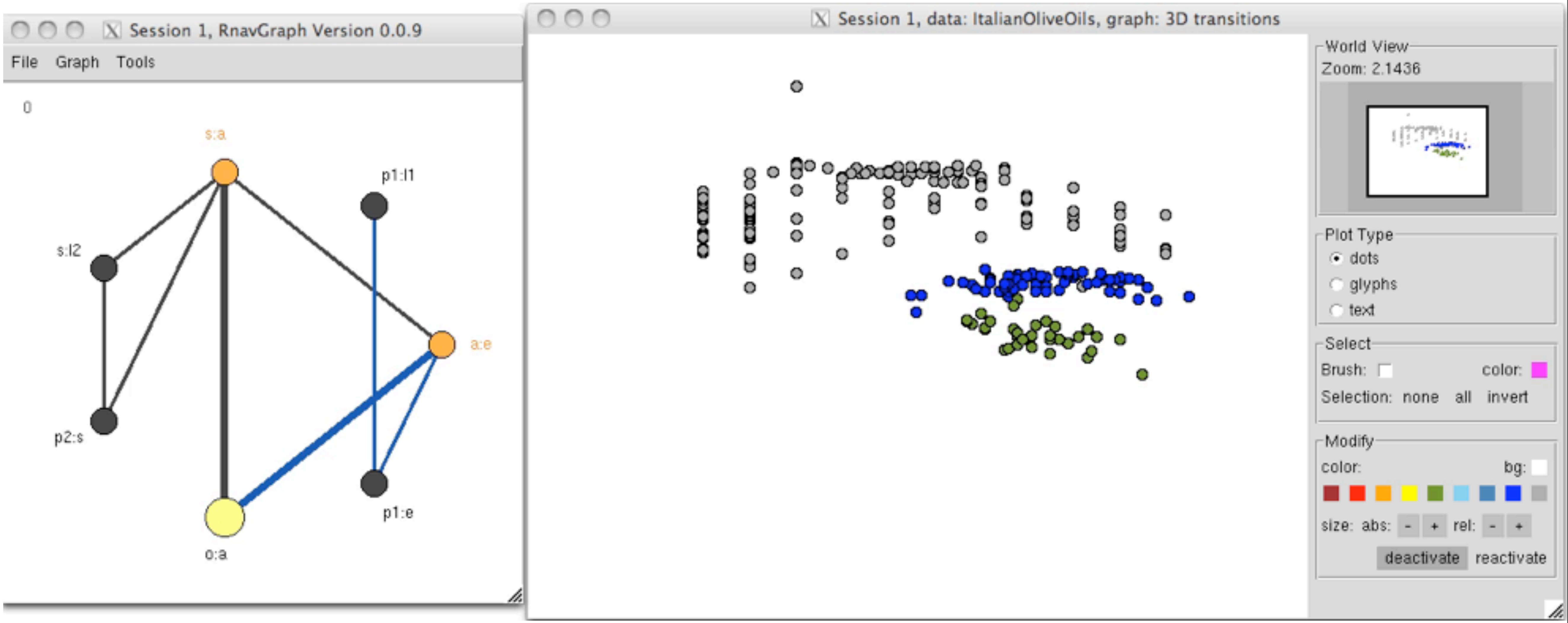
Interactive

3d transition graph

Interactive scatterplot

Appears to be a third horizontal group ... zoom etc.

Example: Italian olive oils



Interactive

3d transition graph

Interactive scatterplot

Appears to be a third horizontal group ... zoom etc.

And that outlier

Example: Italian olive oils

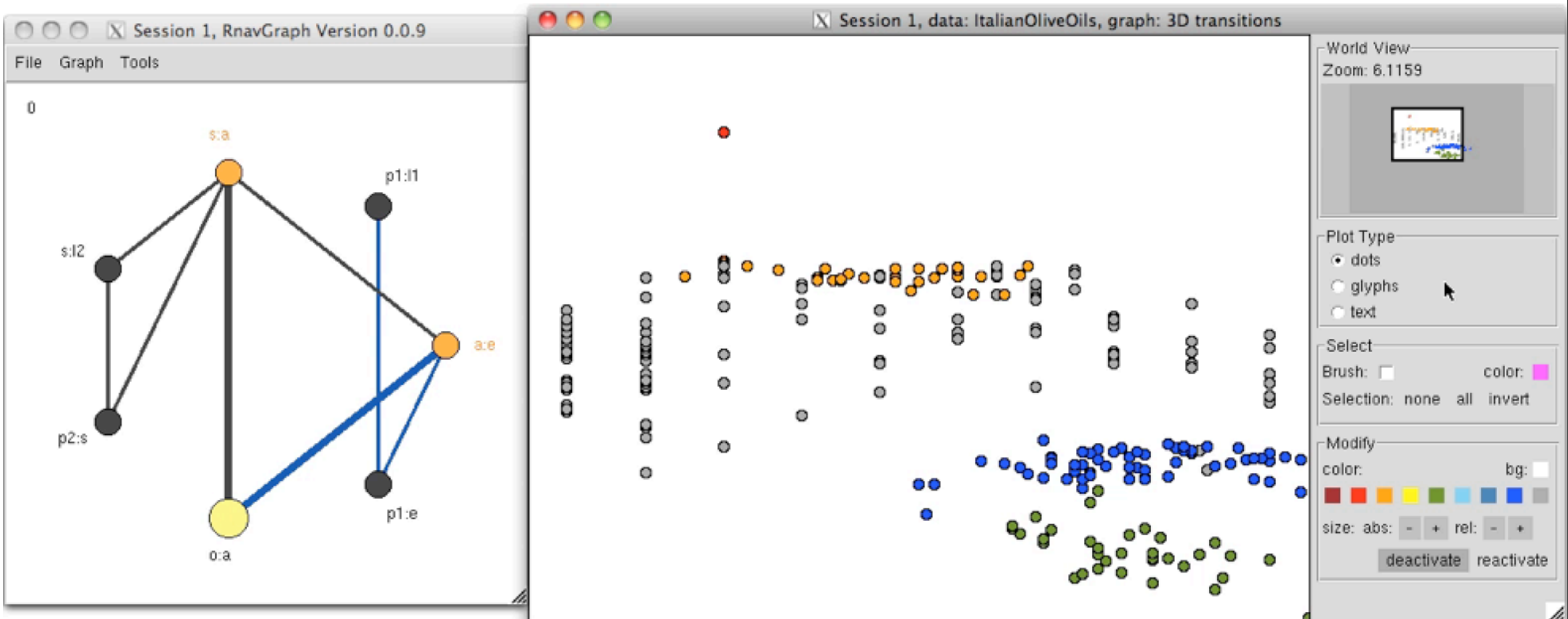
Interactive

3d transition graph

Interactive scatterplot

Colour group orange, outlier red.

Example: Italian olive oils



Interactive

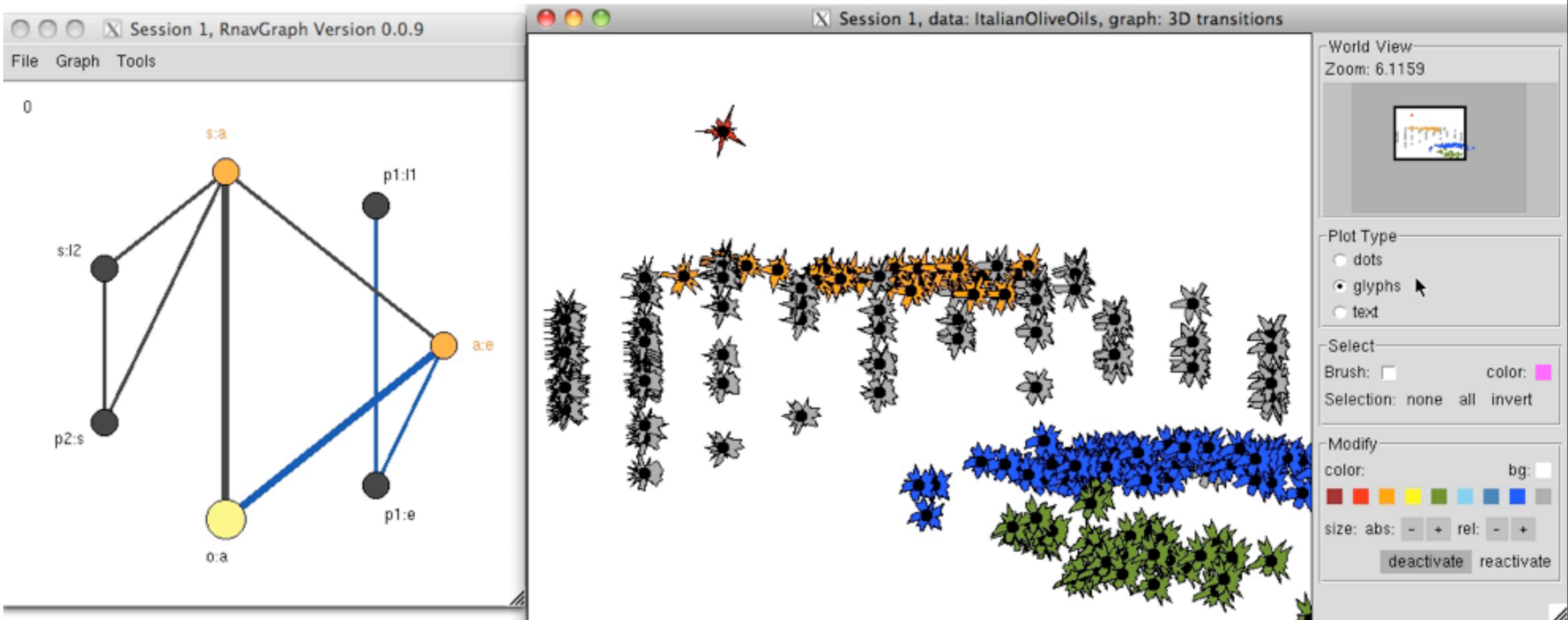
3d transition graph

Interactive scatterplot

Colour group orange, outlier red.

Can switch to glyphs

Example: Italian olive oils



Interactive

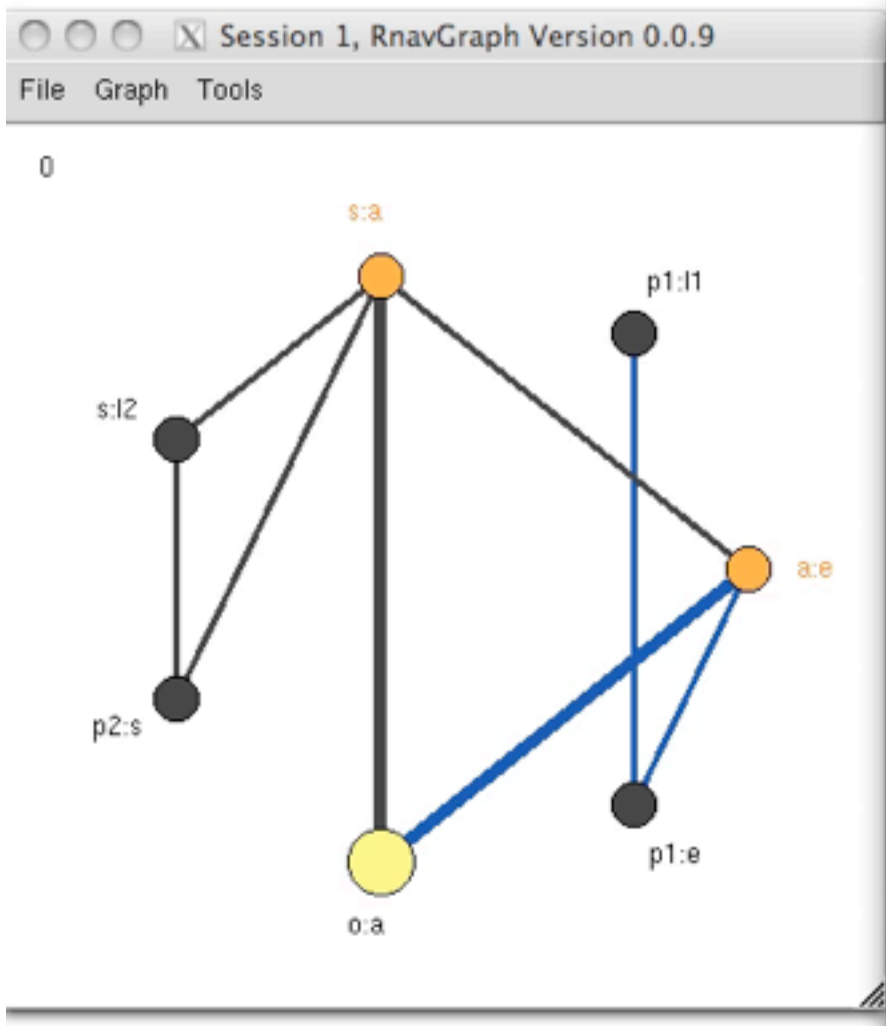
3d transition graph

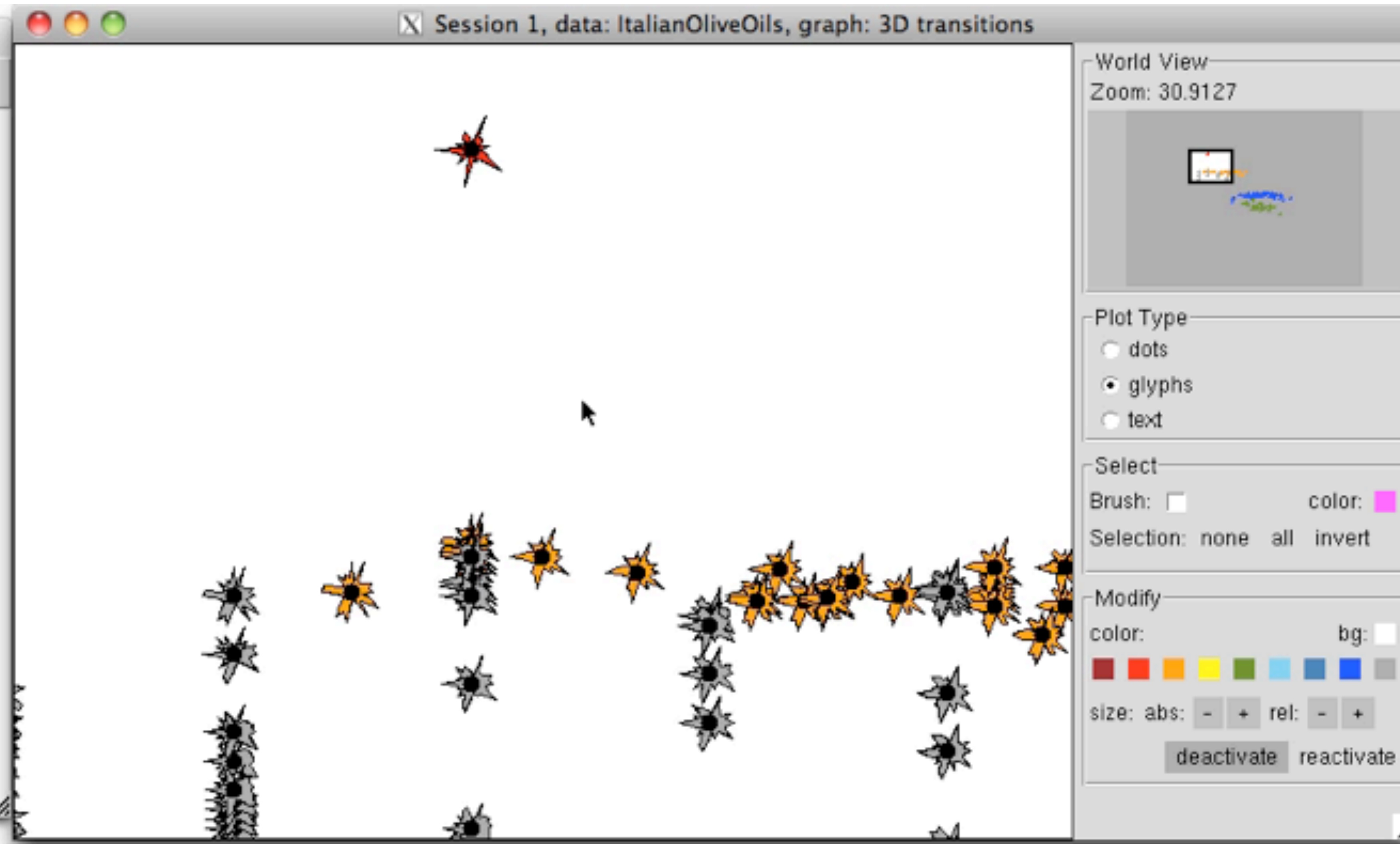
Interactive scatterplot

Colour group orange, outlier red.

Focus on a region

Example: Italian olive oils





Interactive

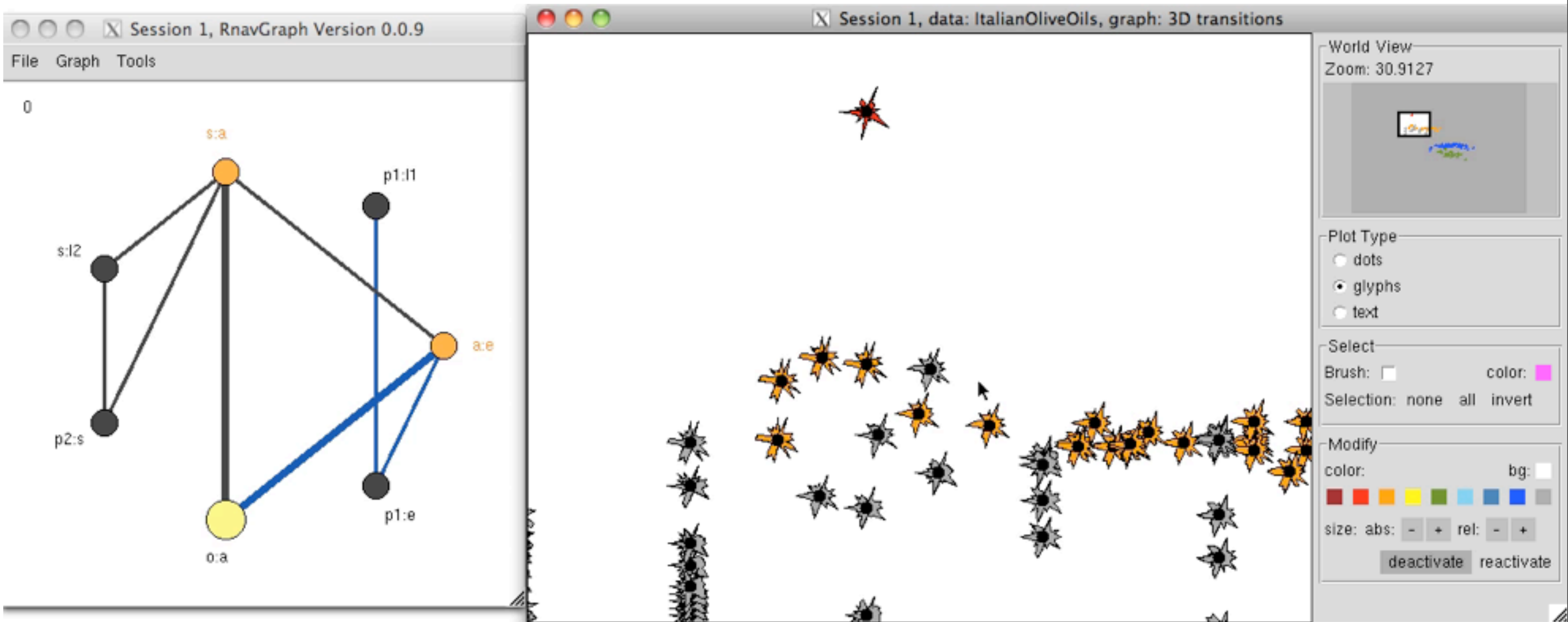
3d transition graph

Interactive scatterplot

Colour group orange, outlier red.

Move to compare shapes

Example: Italian olive oils



Interactive

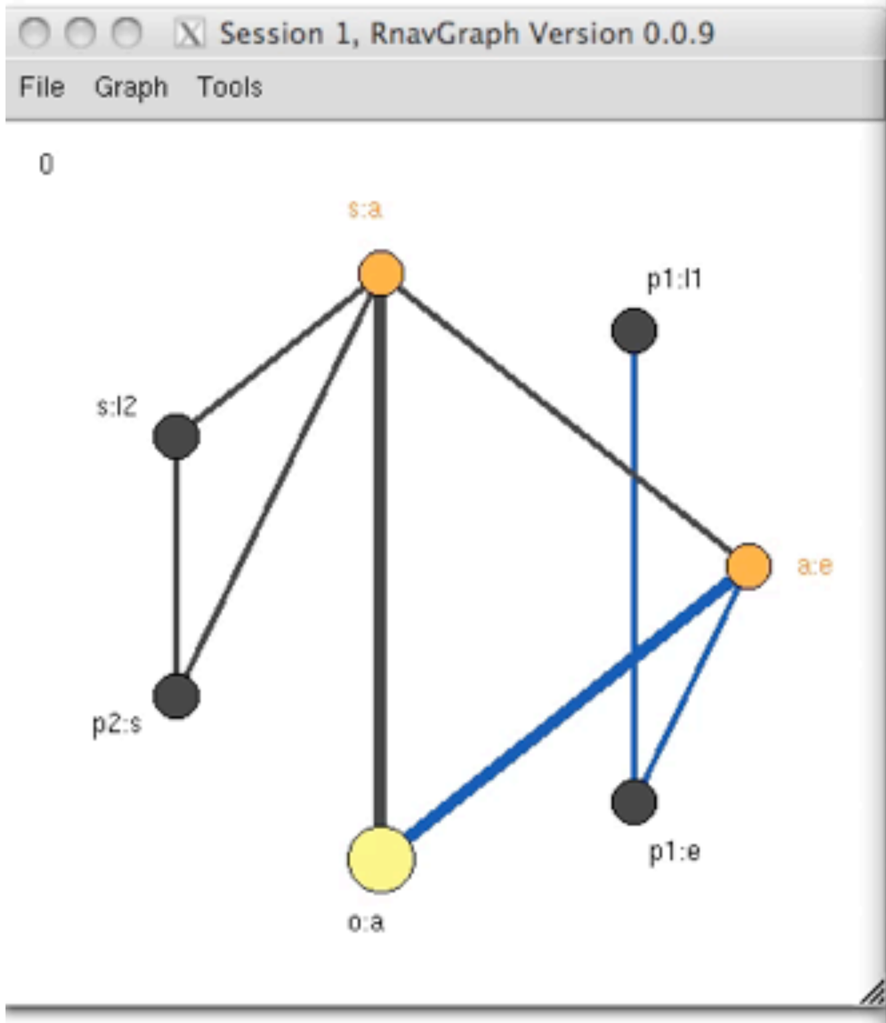
3d transition graph

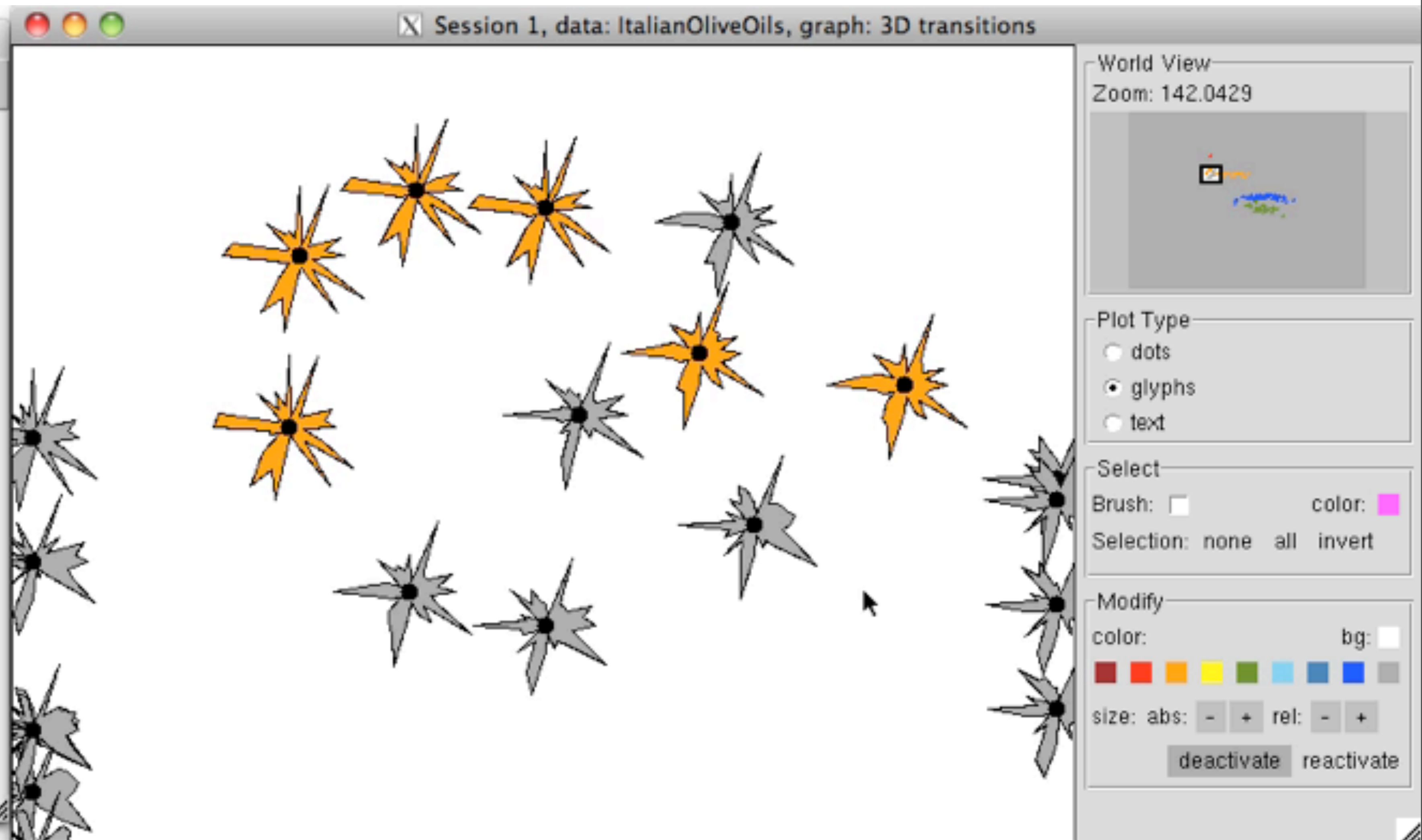
Interactive scatterplot

Colour group orange, outlier red.

Enlarge to compare shapes

Example: Italian olive oils





Interactive

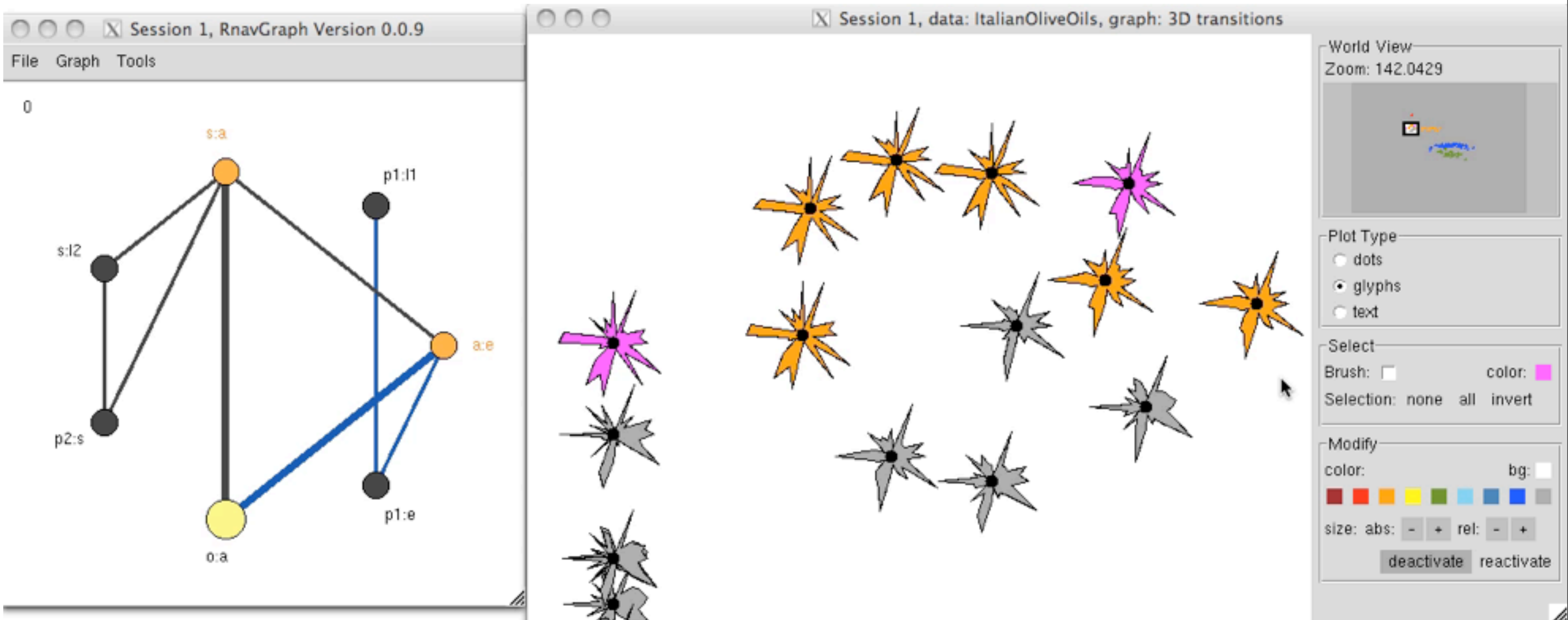
3d transition graph

Interactive scatterplot

Colour group orange, outlier red.

Identify possible orange?

Example: Italian olive oils



Interactive

3d transition graph

Interactive scatterplot

Colour group orange, outlier red.

Can actually check here

Example: Italian olive oils

Continue in this way:

- bring back deactivated points
- identify groups, reassign points
- note natural hierarchical clustering
- save grouping by colour in R

Challenge

Large p => large graphs

- p ... overall dimensionality (olive, p=8)
  - $\binom{p}{2}$... potential 2d nodes (28)
  - $\binom{p}{3}$... potential 3d edges (56)

p	5	10	20	50
$\binom{p}{2}$	10	45	190	1225
$\binom{p}{3}$	10	120	1140	19600

Challenge

Large p => large graphs

- p ... overall dimensionality (olive, p=8)
  - $\binom{p}{2}$... potential 2d nodes (28)
  - $\binom{p}{3}$... potential 3d edges (56)

p	5	10	20	50
$\binom{p}{2}$	10	45	190	1225
$\binom{p}{3}$	10	120	1140	19600

Need to start with small, but interesting, graphs

Interesting node pairs

Graph construction is actually general

- start with any graph G on the variables
- its line graph $L(G)$ will be a 3D-transition graph
- the complement of the line graph $\overline{L(G)}$ will be a 4D-transition graph

Interesting node pairs

Graph construction is actually general

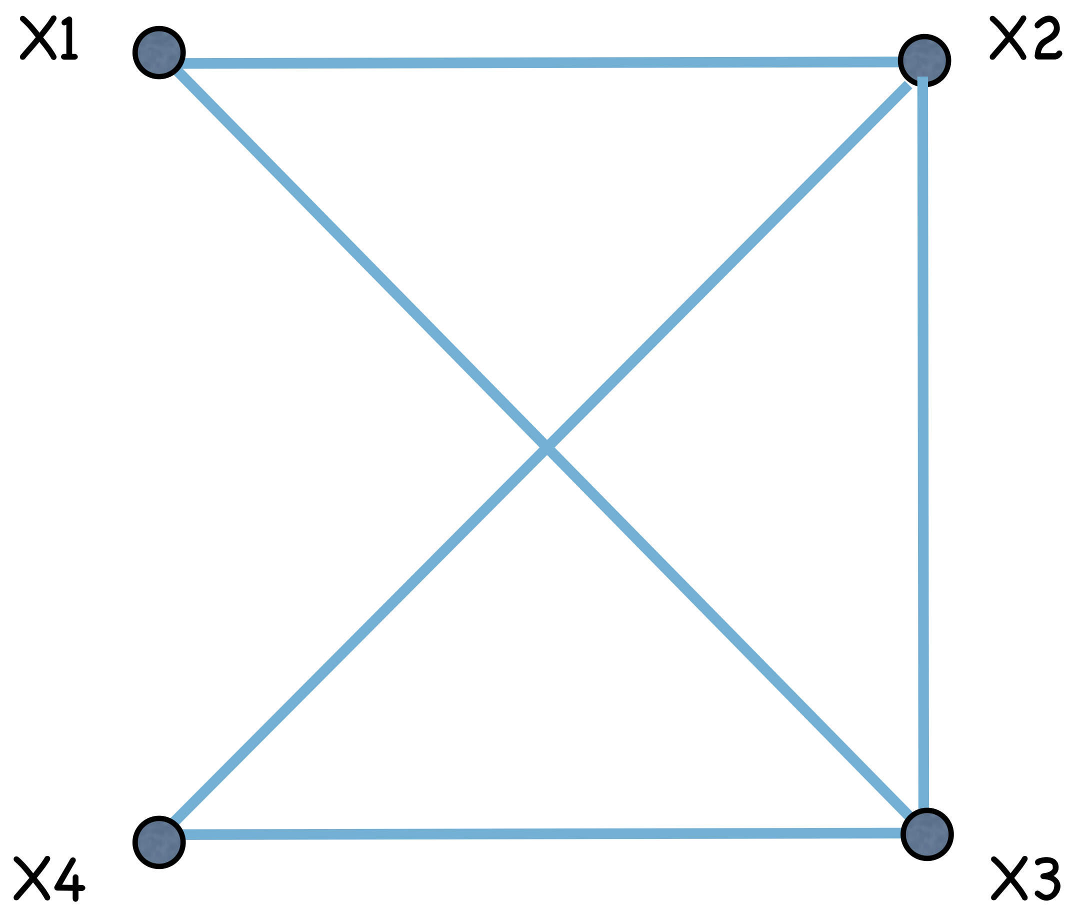
- start with **any** graph G on the variables
- its line graph $L(G)$ **will** be a 3D-transition graph
- the complement of the line graph $\overline{L(G)}$ **will** be a 4D-transition graph

Variable graph:

Only place edges between **interesting** pairs

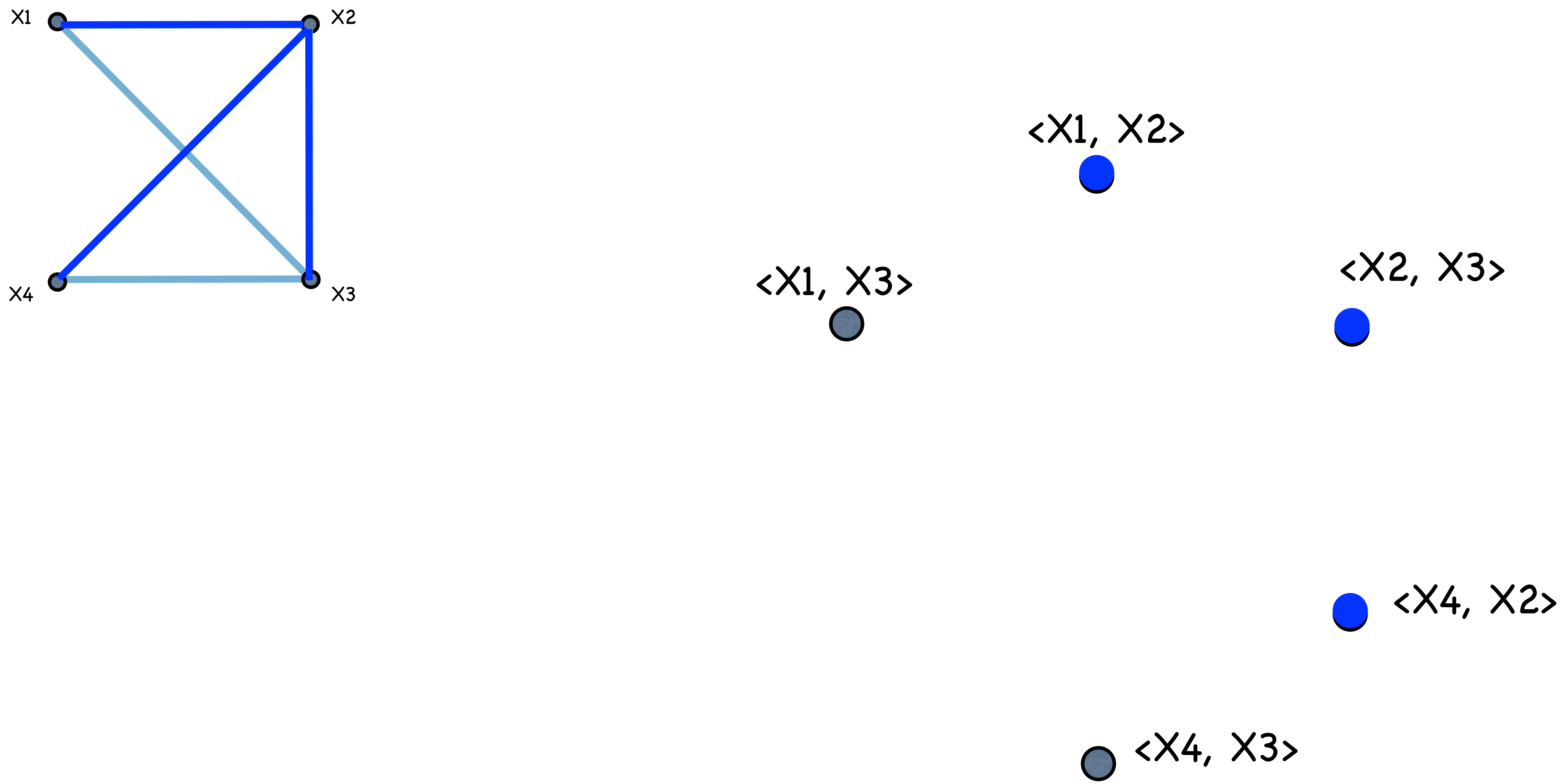
Graph construction

Construction: Line graph of the variable graph



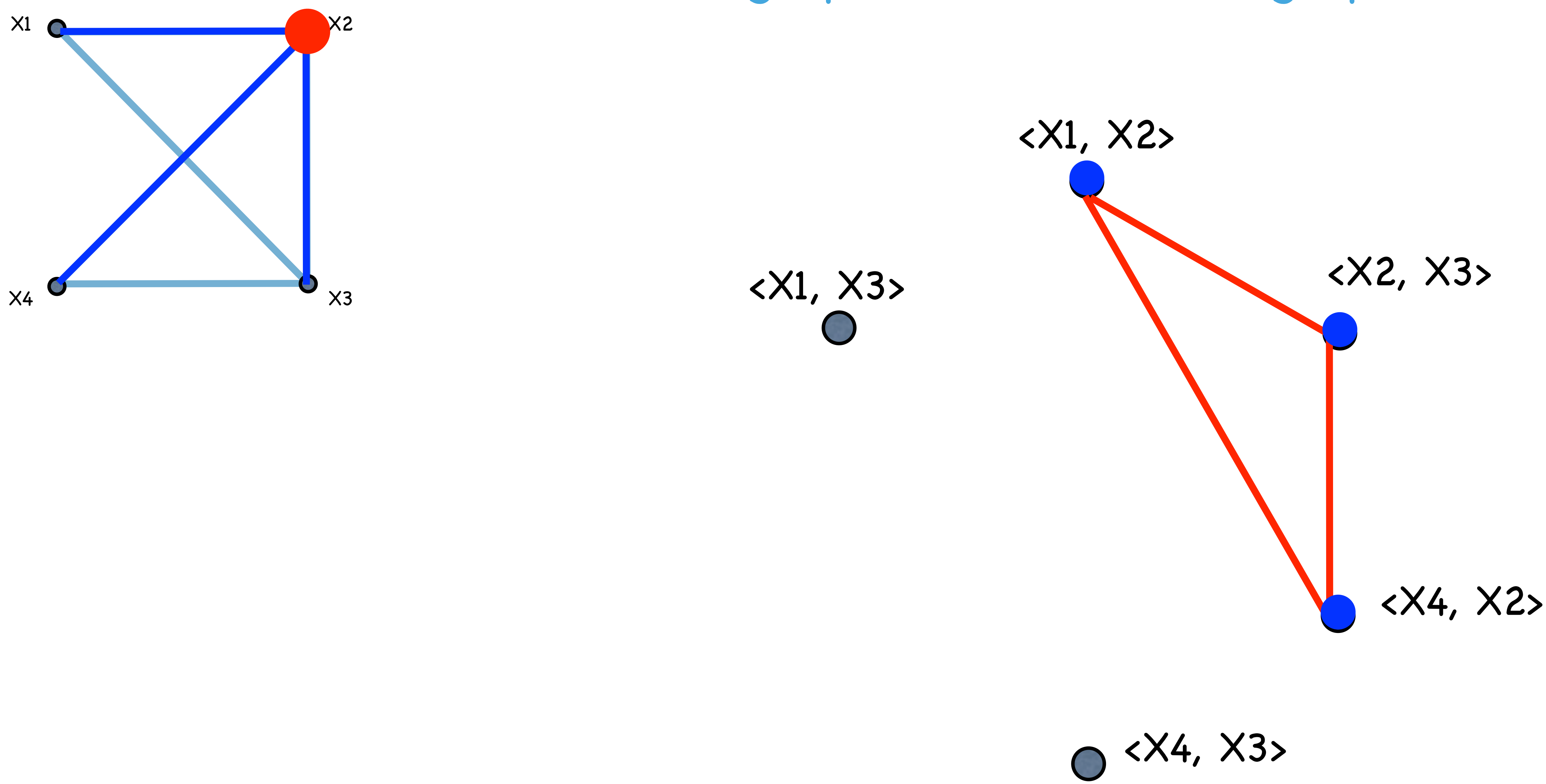
Graph construction

Construction: Line graph of the variable graph



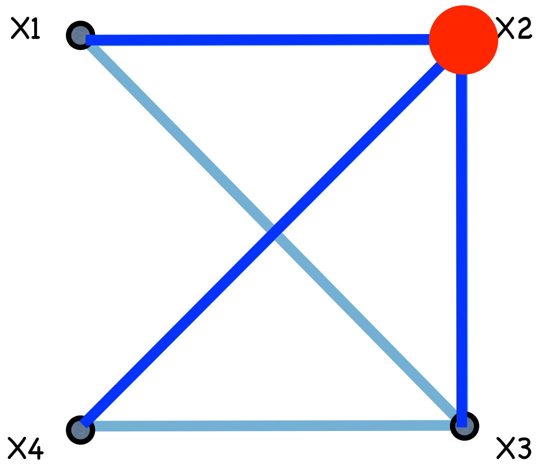
Graph construction

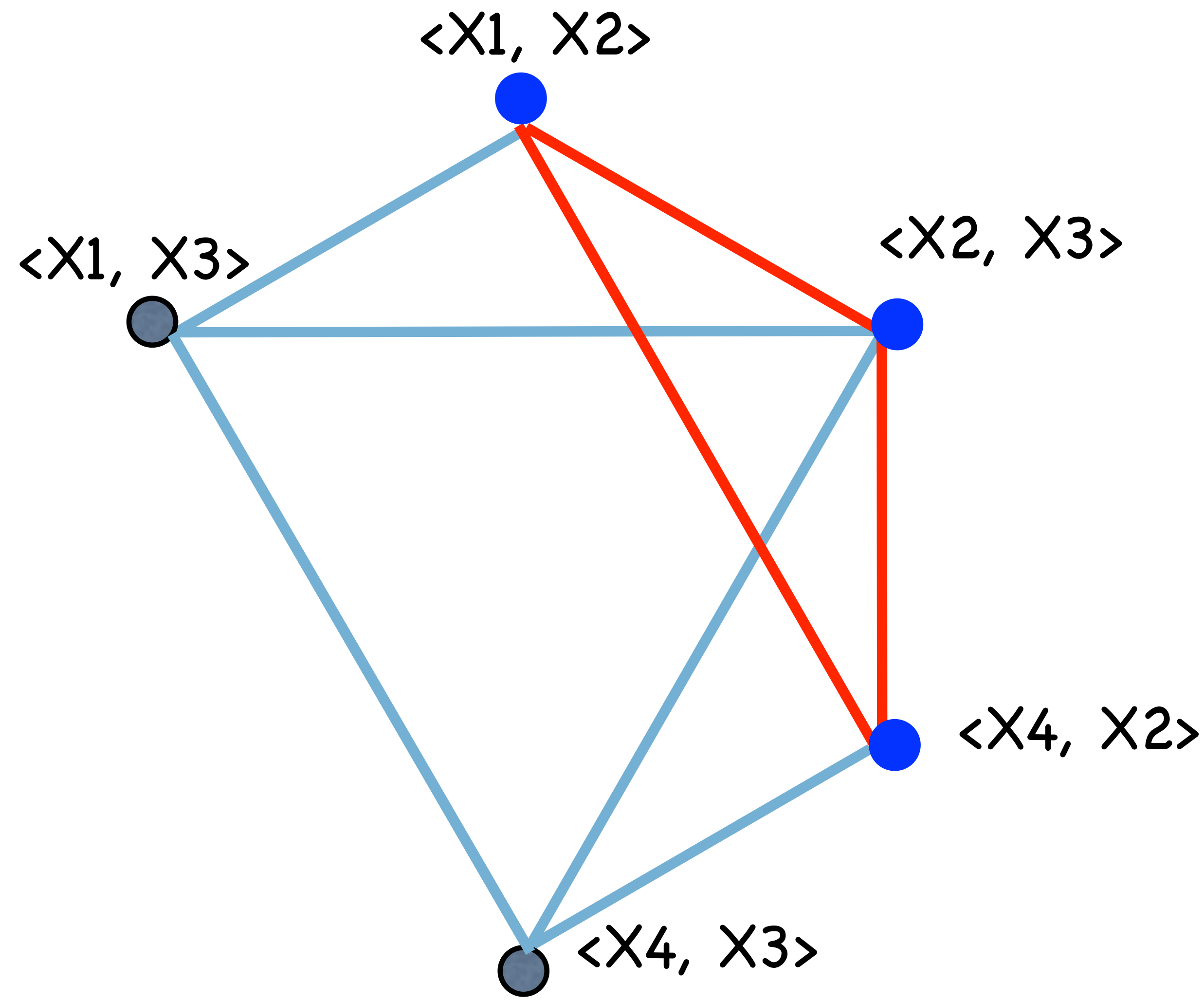
Construction: Line graph of the variable graph



Graph construction

Construction: Line graph of the variable graph

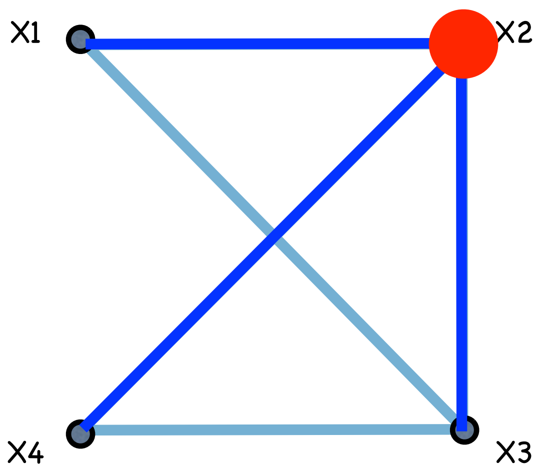




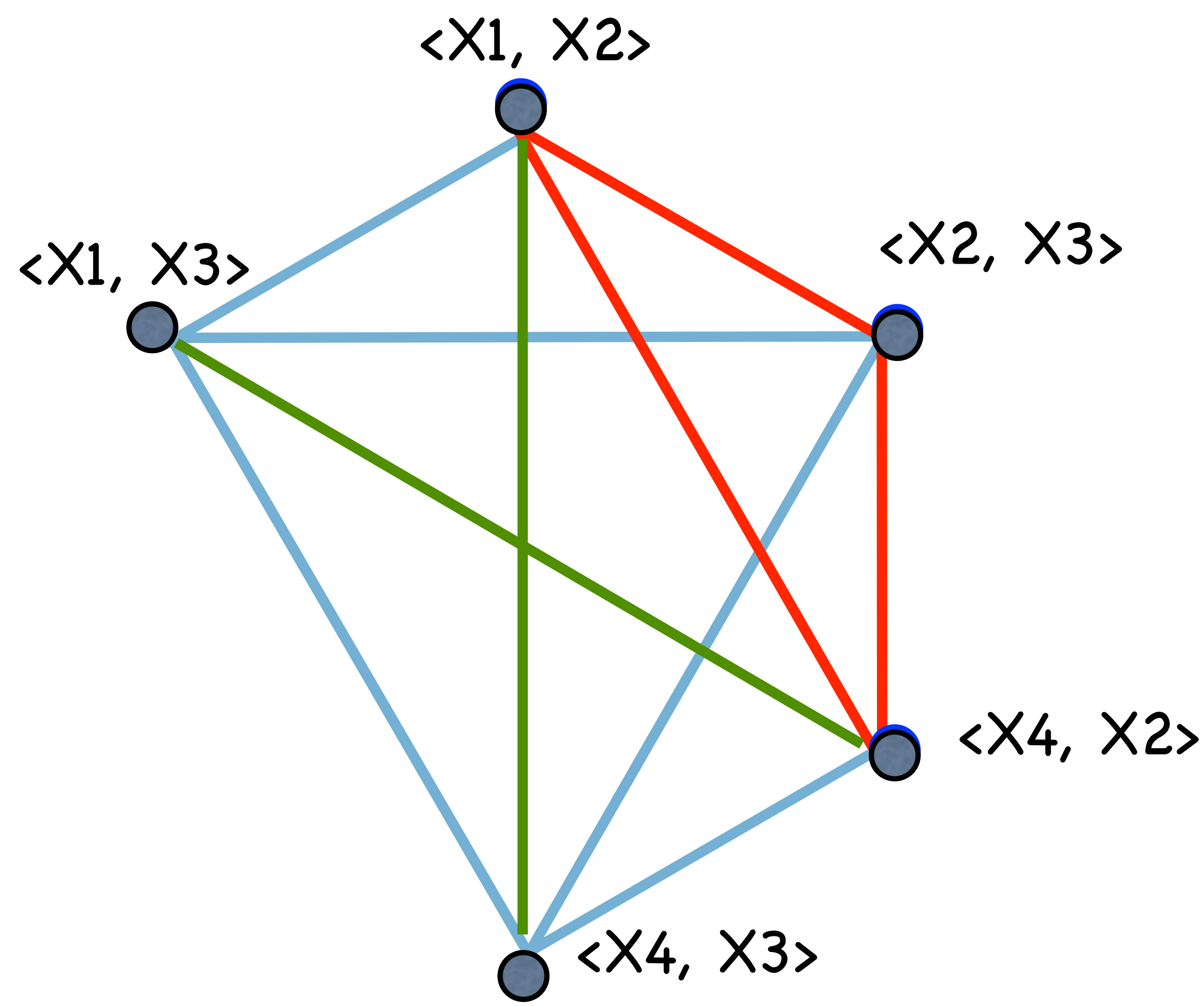
variable graph <--> line graph
<--> 3D transition graph

Graph construction

Construction: Line graph of the variable graph



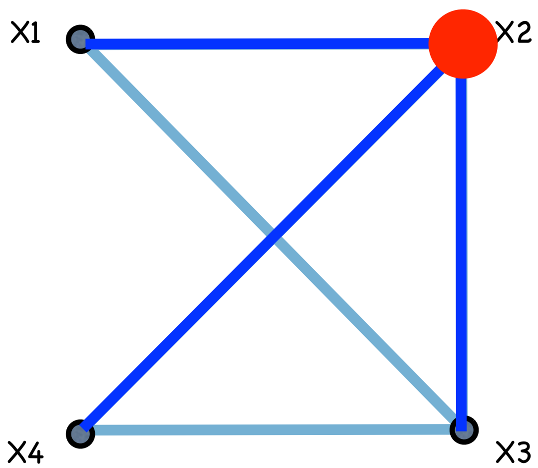
Complement(Line graph)

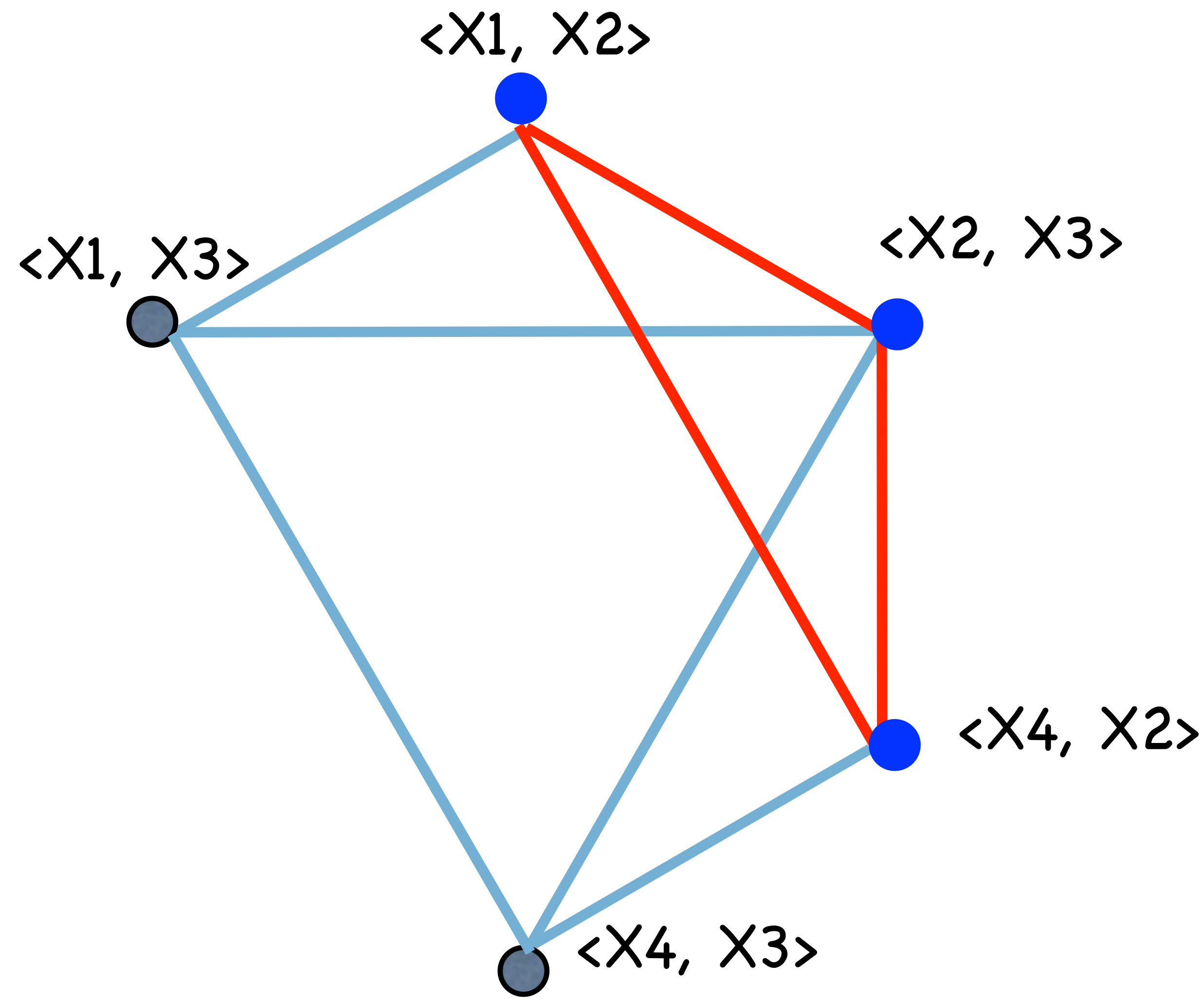


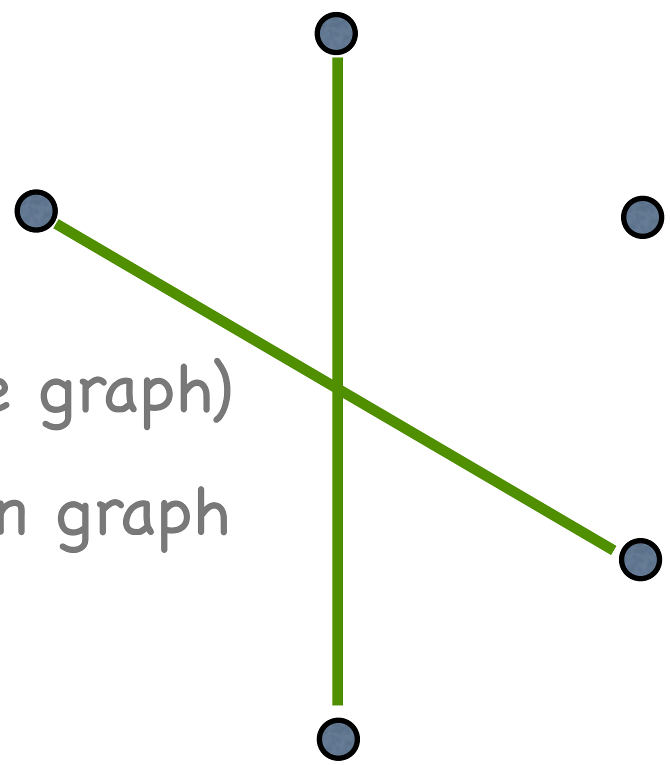
variable graph <--> line graph
<--> 3D transition graph

Graph construction

Construction: Line graph of the variable graph







Complement(Line graph)

= 4D transition graph

variable graph <--> line graph

<--> 3D transition graph

Scagnostics

Cognostics (Computer aided diagnostics)

Scagnostics ... Scatterplot cognostics

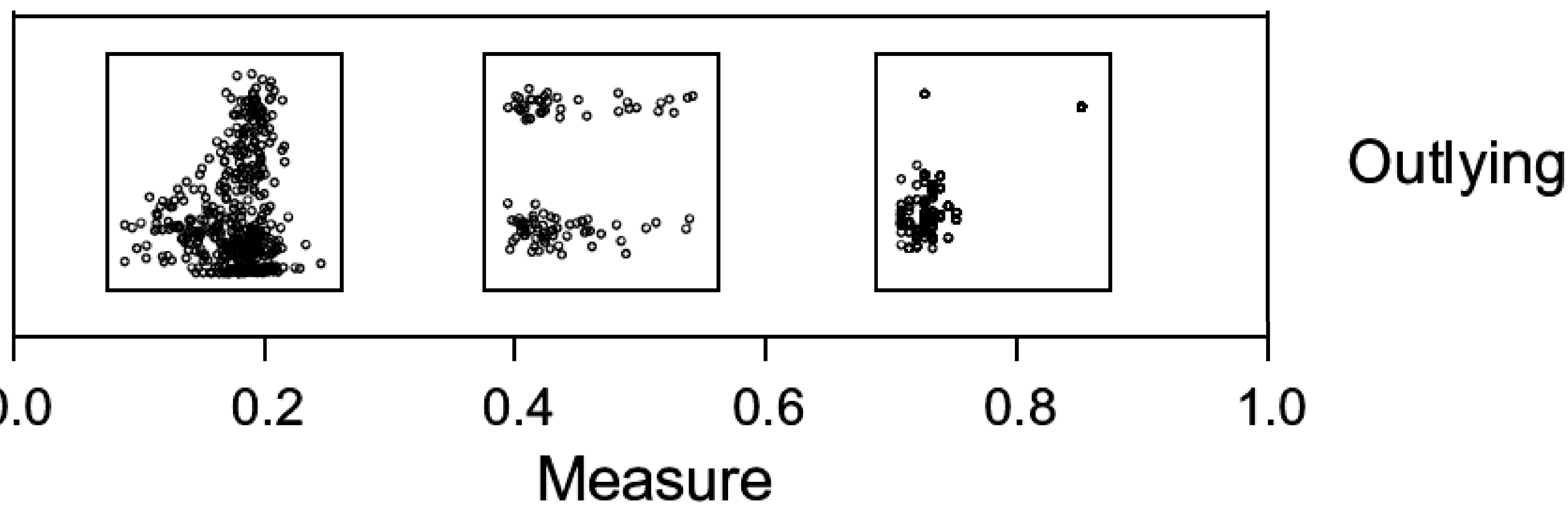
Wilkinson et al (2006) (from idea proposed by Tukey & Tukey (1985))

Scagnostics

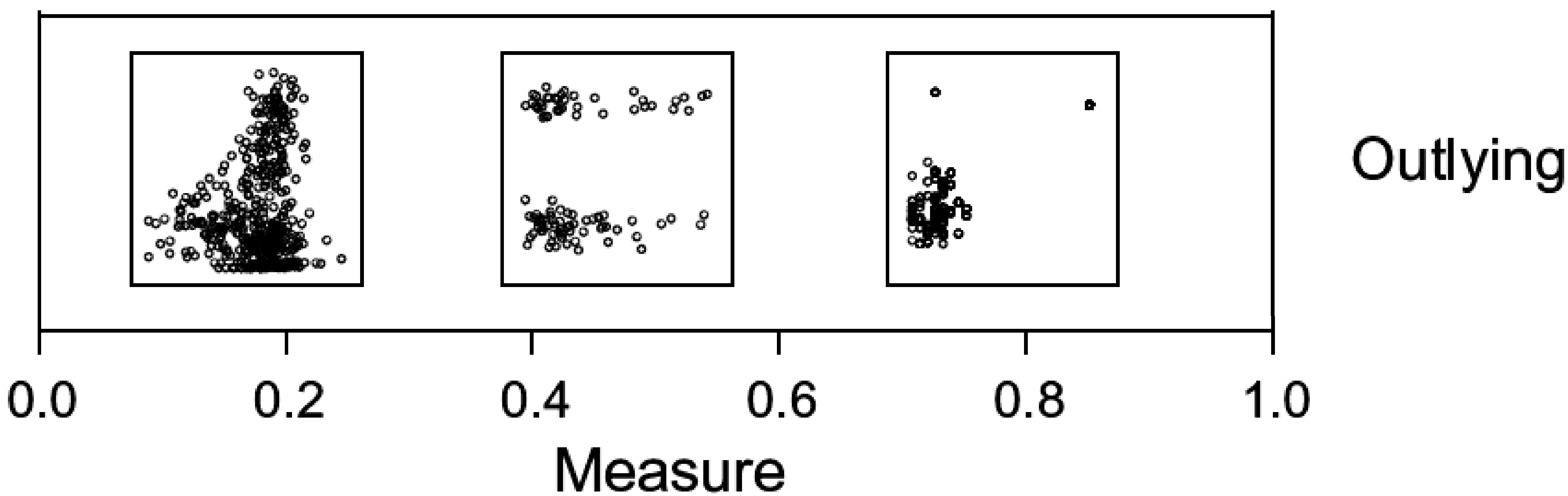
Cognostics (Computer aided diagnostics)

Scagnostics ... Scatterplot cognostics

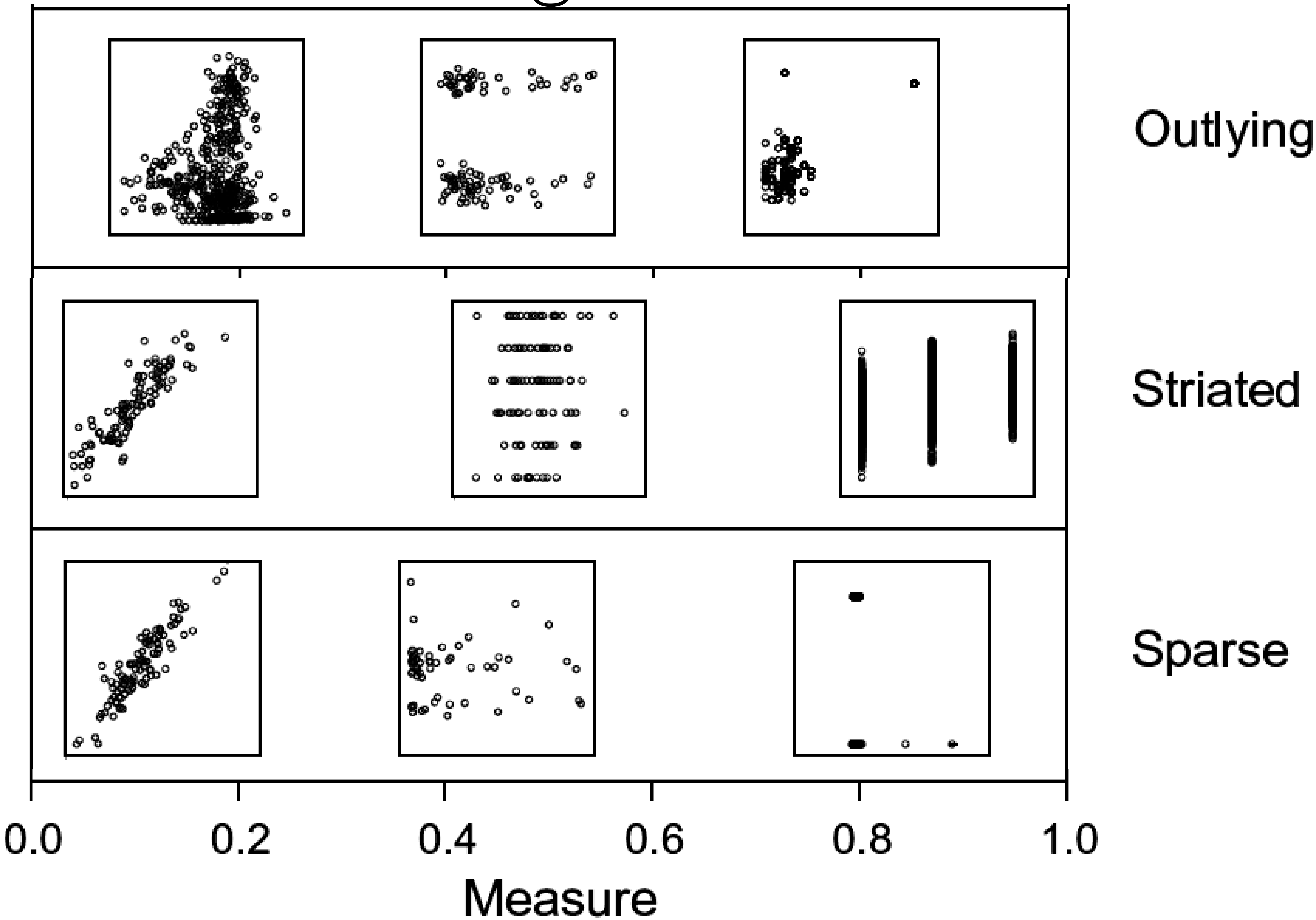
Wilkinson et al (2006) (from idea proposed by Tukey & Tukey (1985))



Scagnostics



Scagnostics

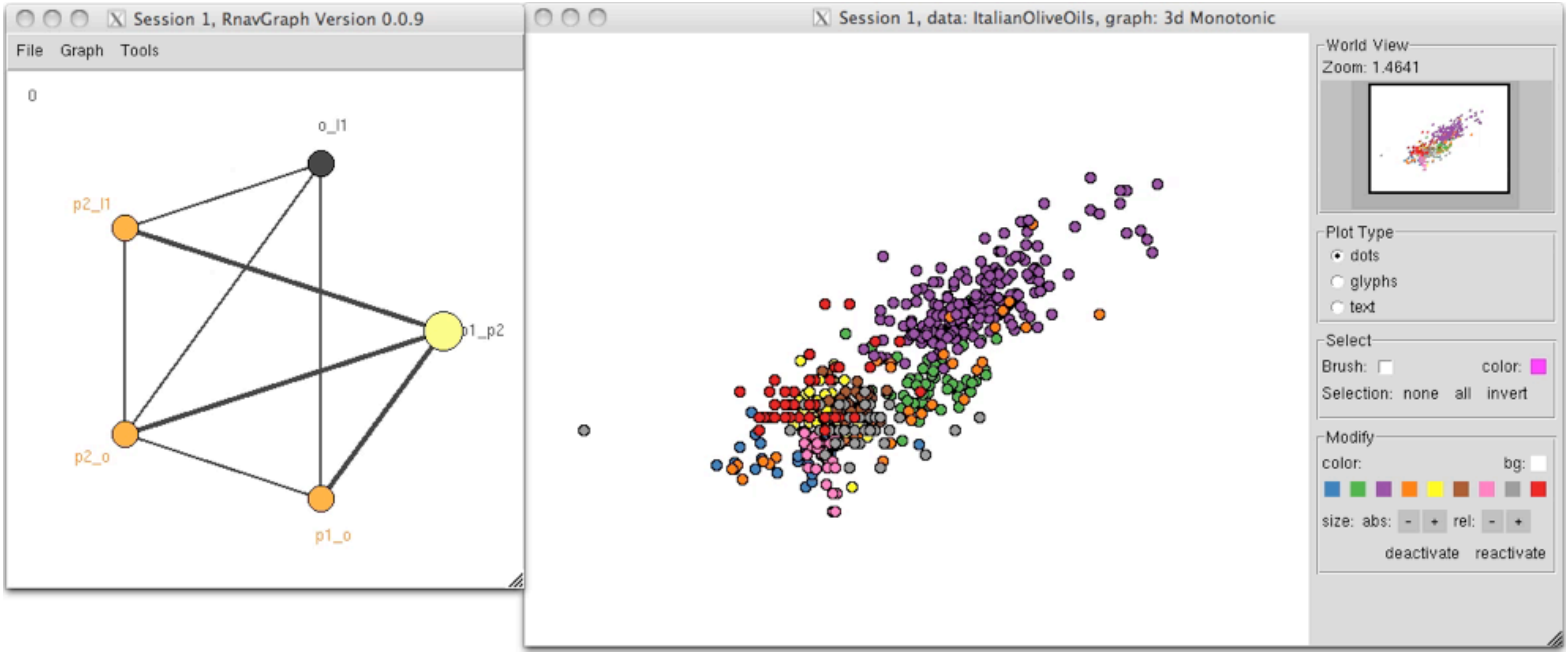


Interesting node pairs

For each scagnostic, calculate its value for **every** pair.

Use **only** those pairs with high scores in variable graph (e.g. top fraction of scores).

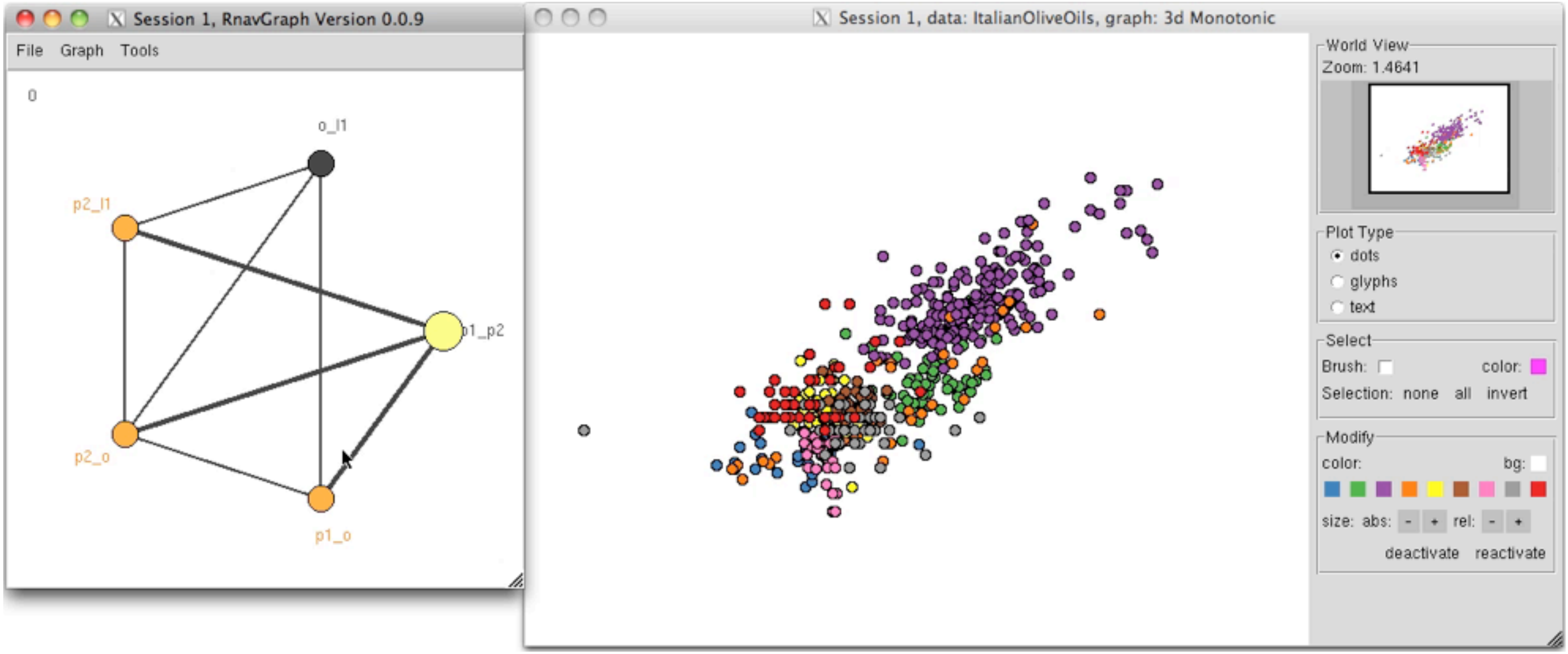
Scagnostics: Italian olive oils



3D Monotonic

Groups coloured by regions

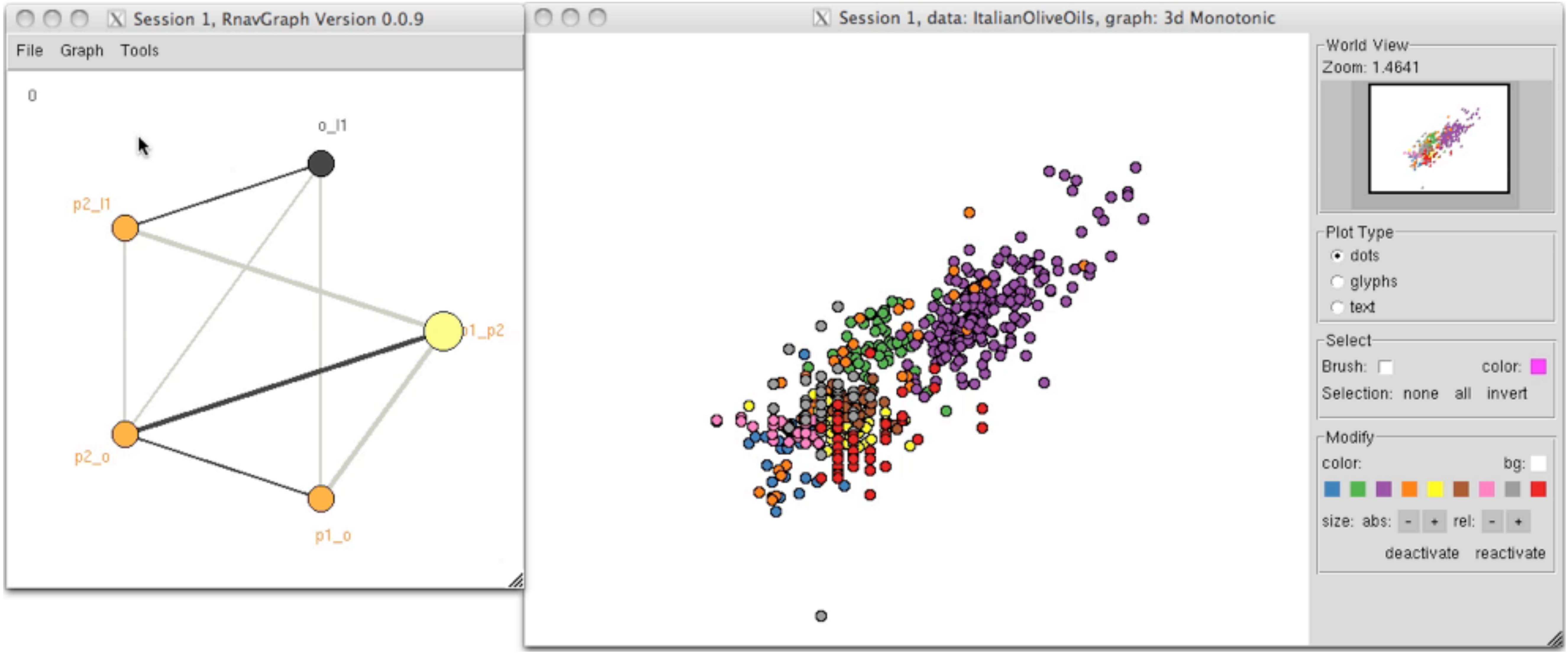
Scagnostics: Italian olive oils



3D Monotonic

Groups coloured by regions

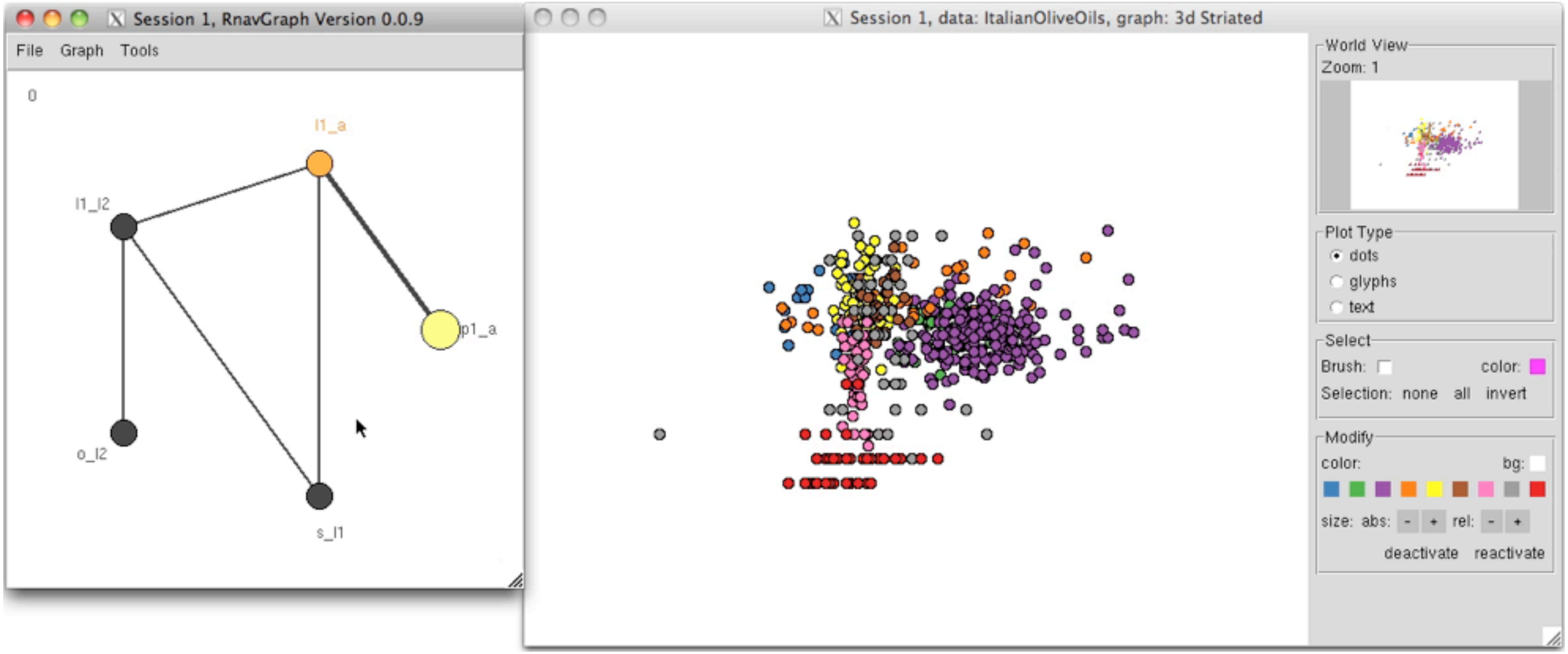
Scagnostics: Italian olive oils



Switch to 3D Striated

Groups coloured by regions

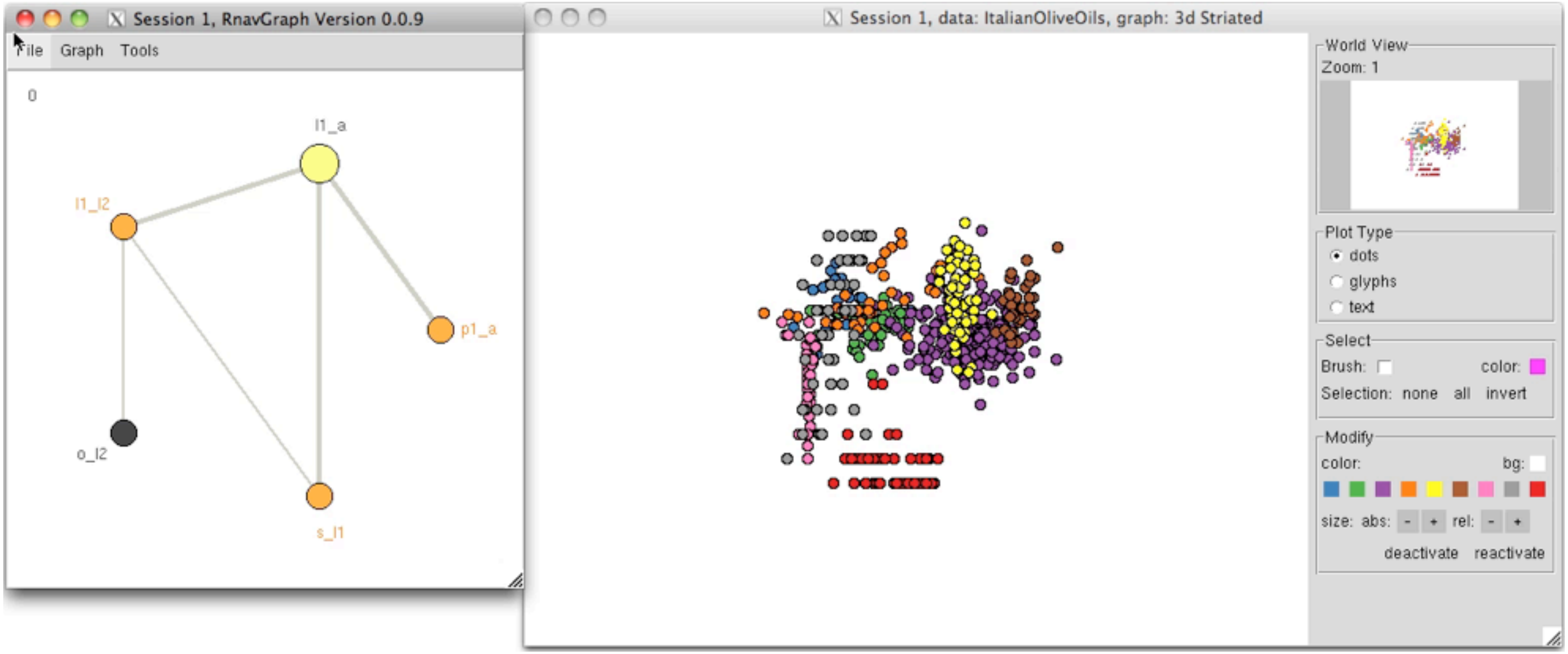
Scagnostics: Italian olive oils



3D Striated

Groups coloured by regions

Scagnostics: Italian olive oils



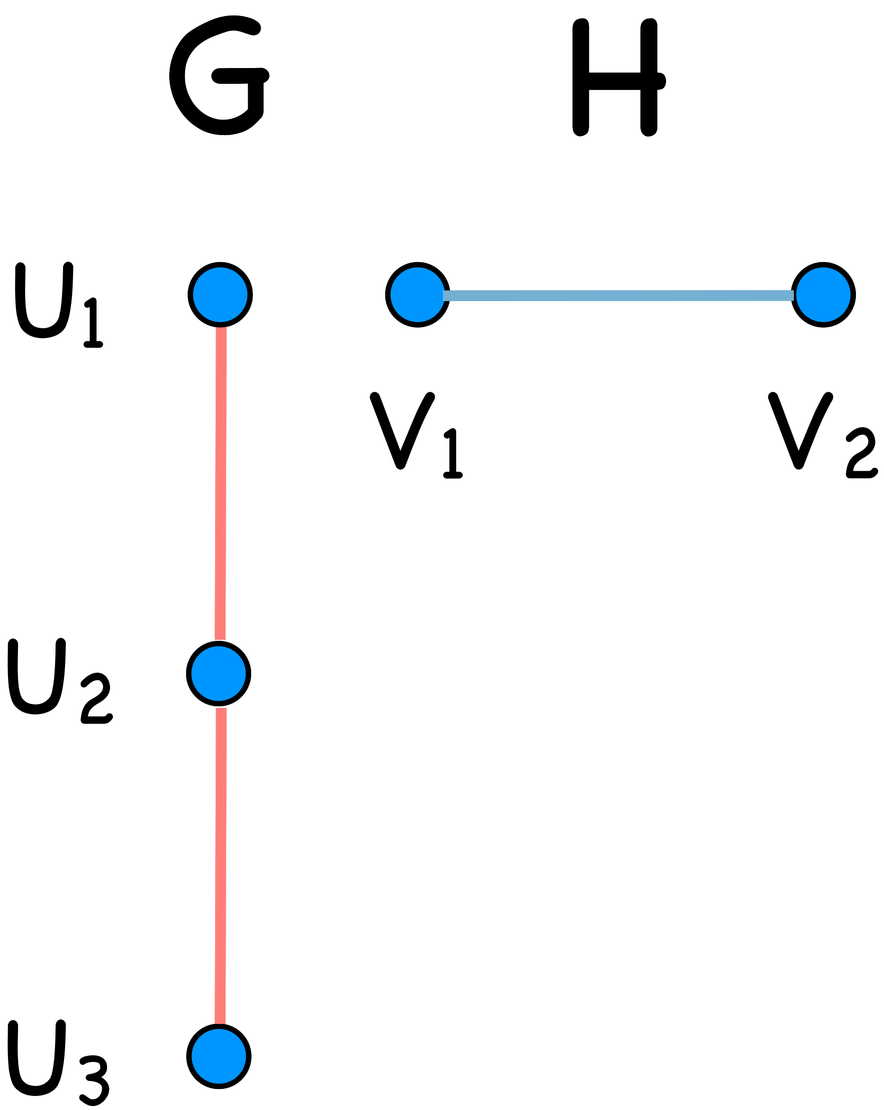
3D Non-Convex

Groups coloured by regions

Graph Products

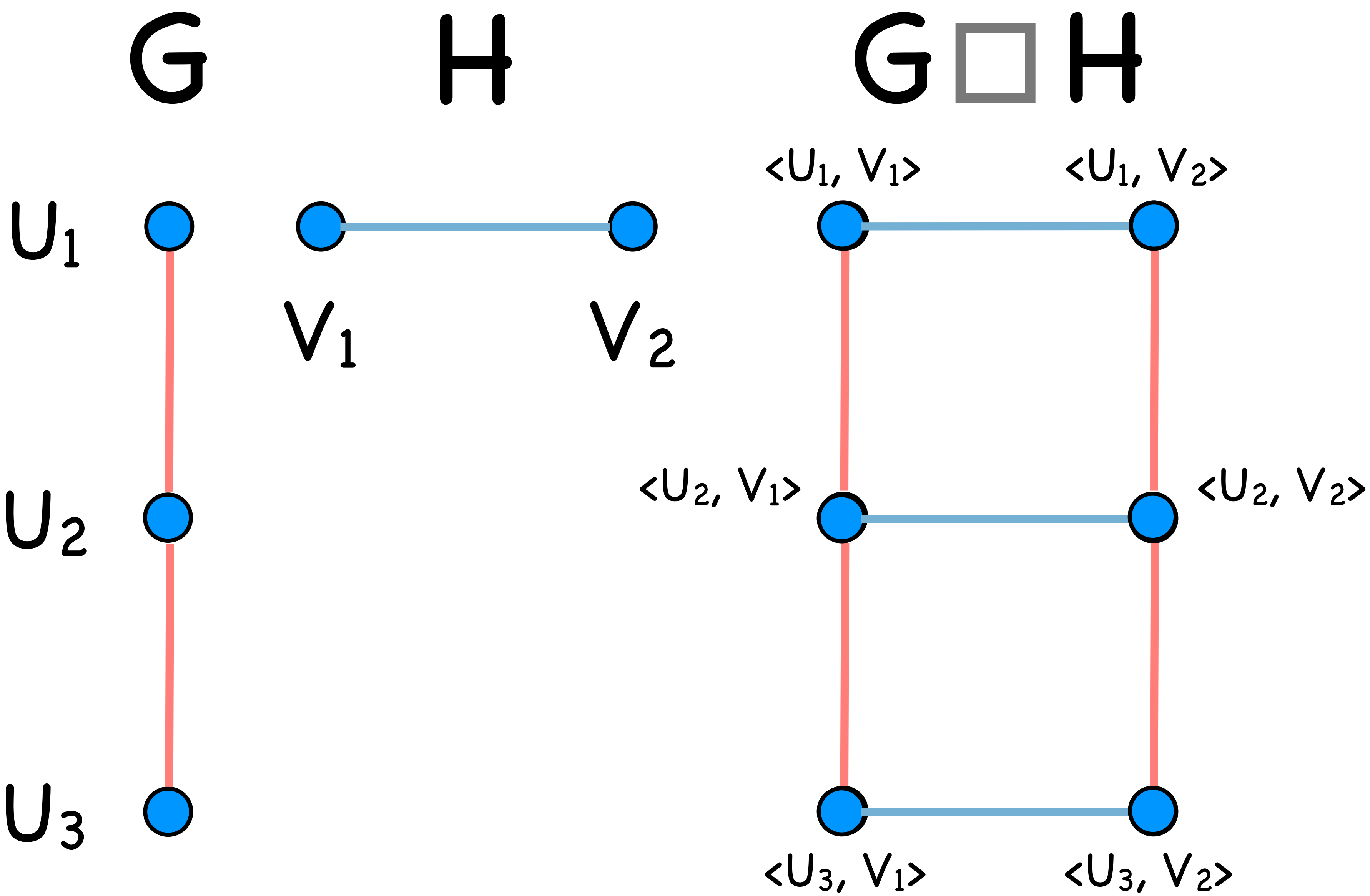
- Another general construction: graph products

Graph Products



E.g.
explanatory U (or Xs),
responses V (or Ys)

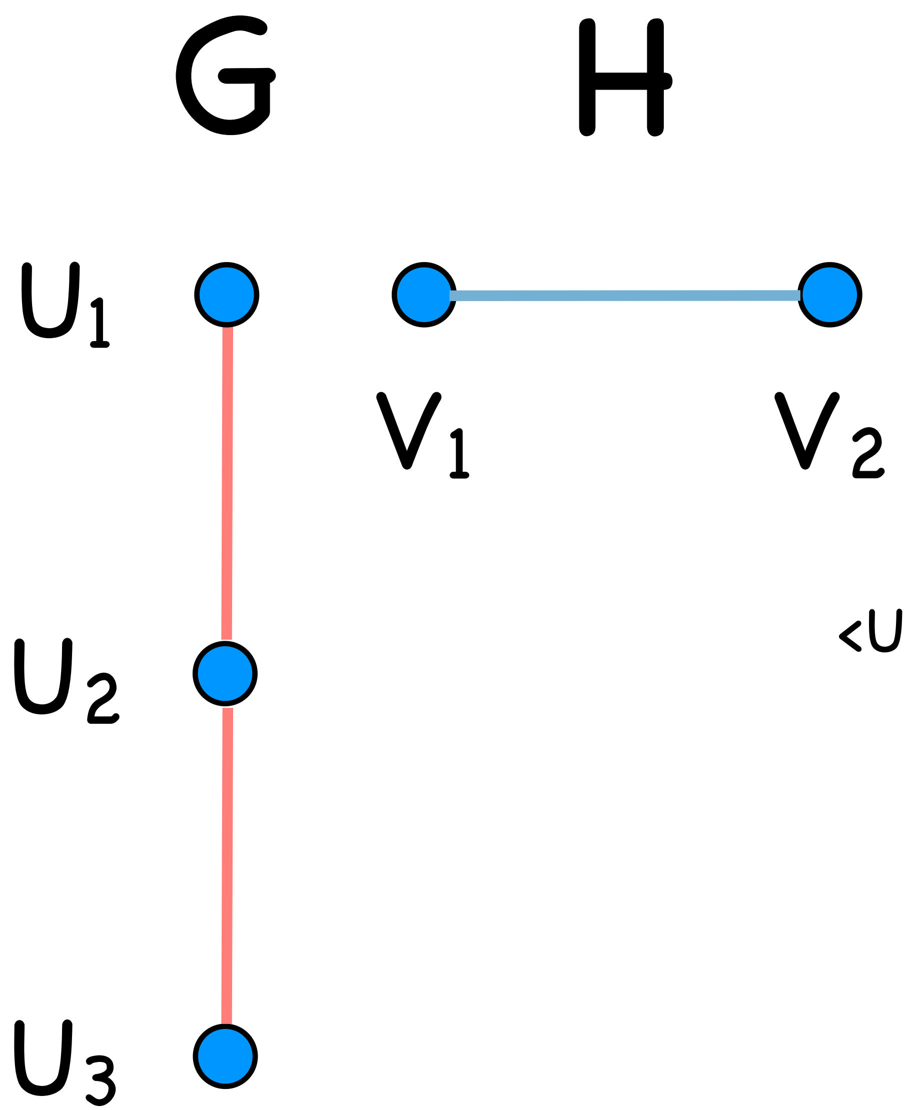
Graph Products



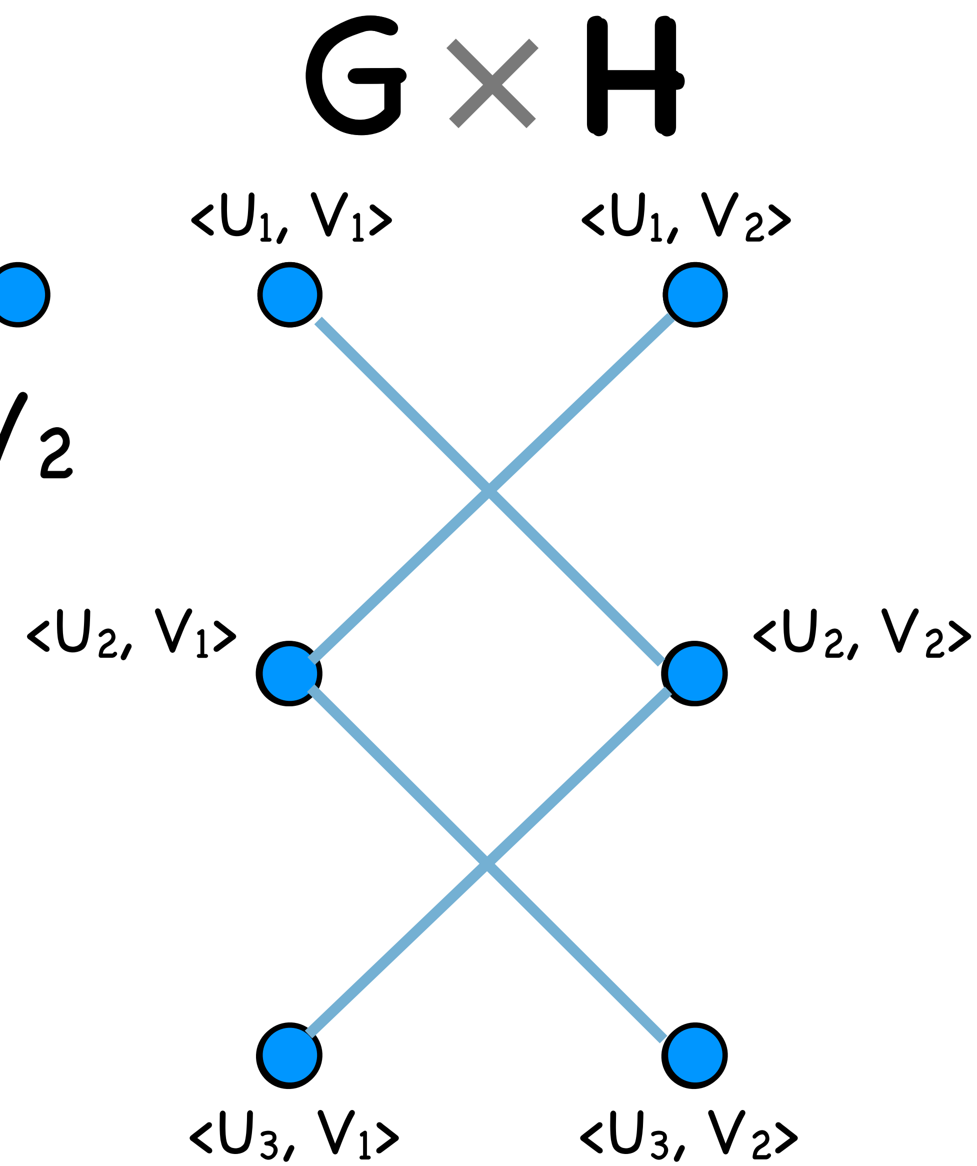
E.g.
explanatory U (or Xs),
responses V (or Ys)

Cartesian product
3D transition graph

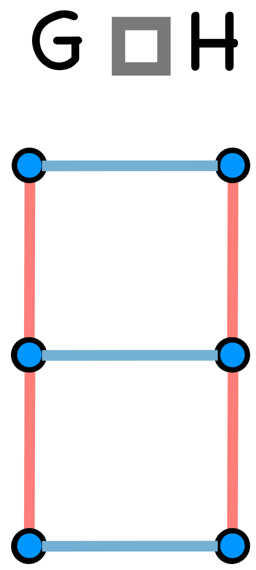
Graph Products



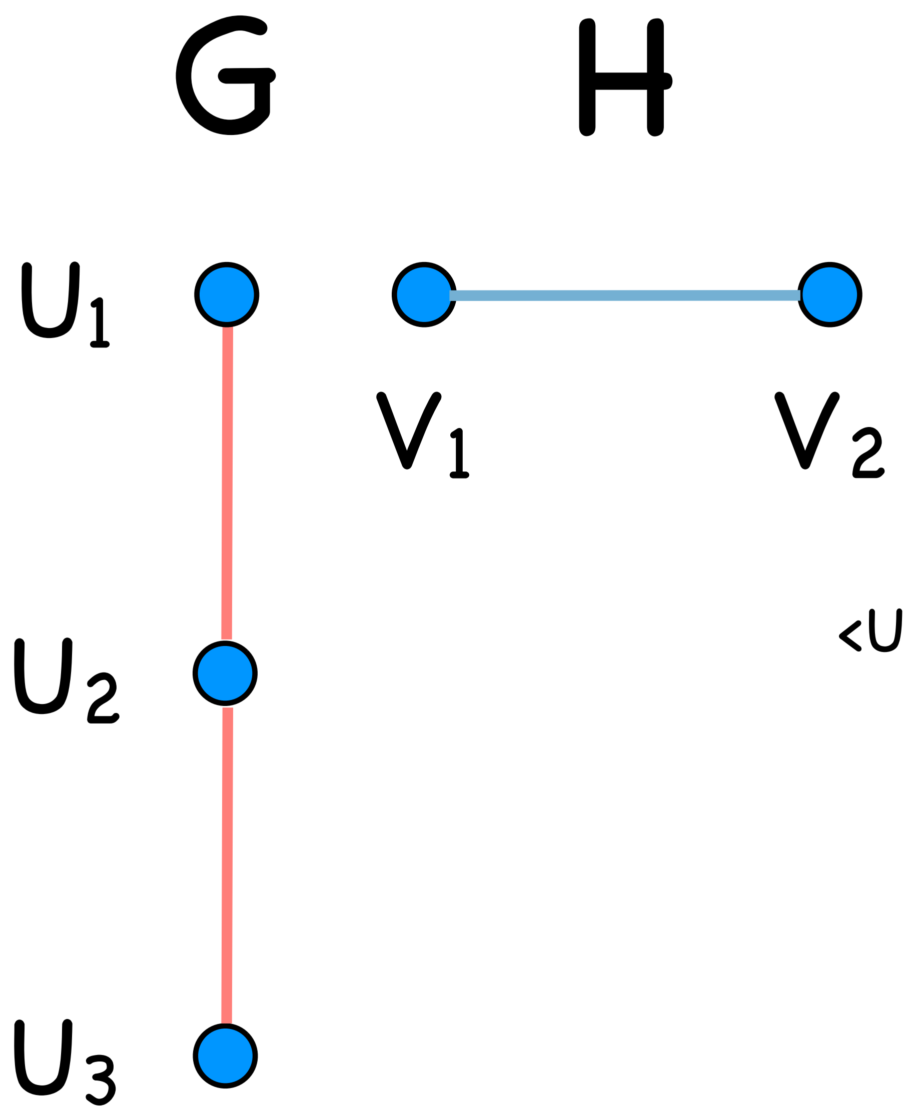
E.g.
explanatory U (or Xs),
responses V (or Ys)

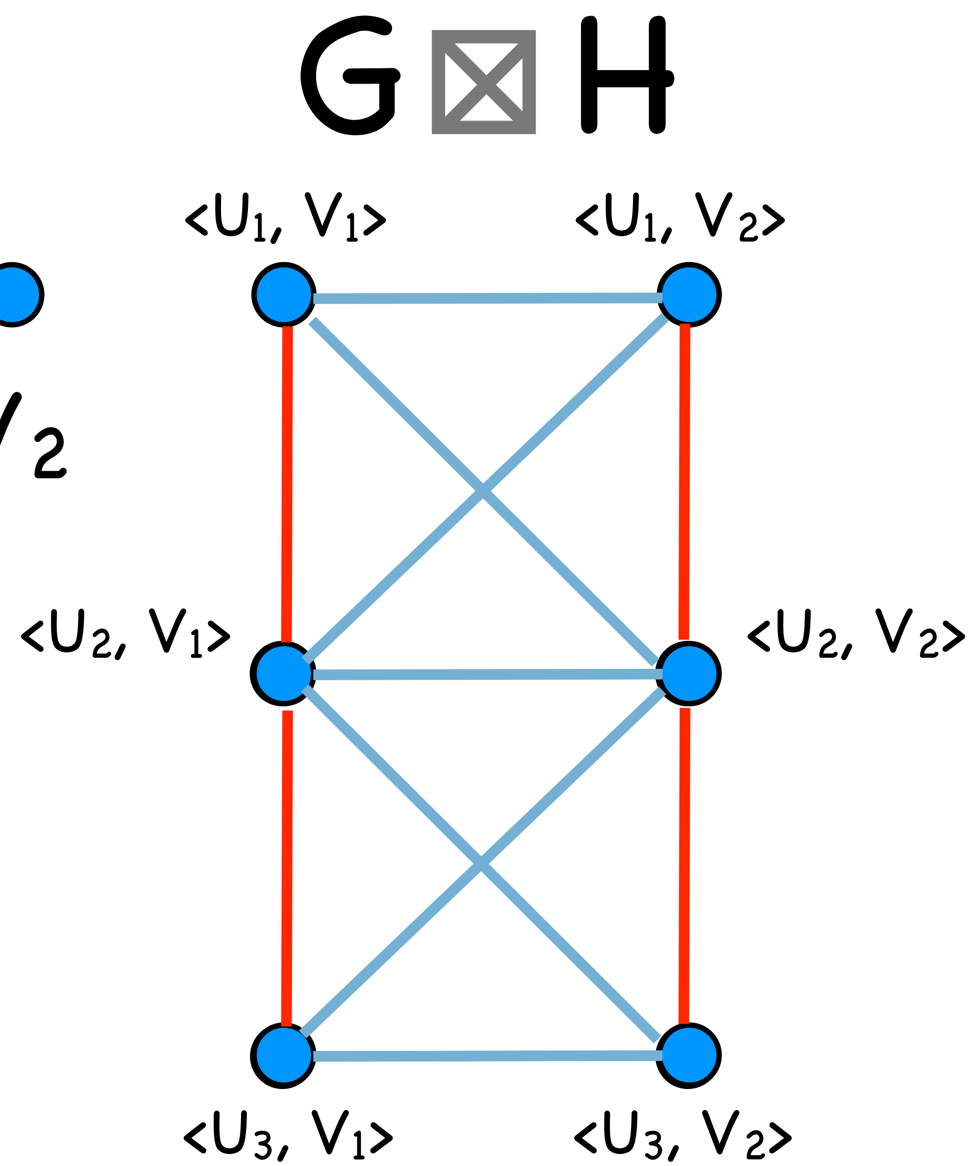


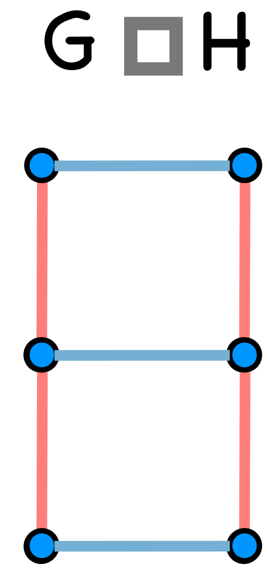
Tensor product
4D transition graph

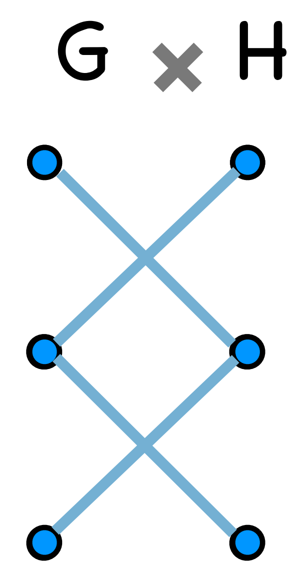


Graph Products









E.g.
explanatory U (or Xs),
responses V (or Ys)

Strong product

All are symmetric

Challenge

Large p => large graphs

- scagnostics work well
- sometimes context suggests small graphs (e.g. via products)
- but when p is very large, so is $\binom{p}{2}$
- dimensionality reduction methods could be employed.

Example: images

Frey: 1,965 movie frames

Example: images

Frey: 1,965 movie frames

28 x 20 array

Example: images

Frey: 1,965 movie frames

28 x 20 array

560 dimensions

Example: images

Frey: 1,965 movie frames

28 x 20 array

560 dimensions

explore via low dimensional spaces

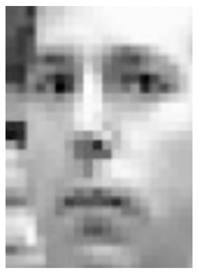
Example: images

Frey: 1,965 movie frames

560 dimensions

Example: images

Frey: 1,965 movie frames



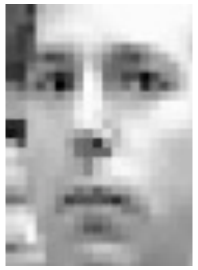
560 dimensions

Using LLE: local linear embedding

k=12 neighbours

Example: images

Frey: 1,965 movie frames



560 dimensions

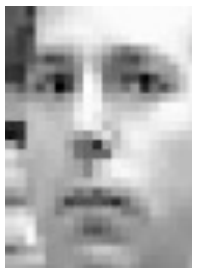
Using LLE: local linear embedding

k=12 neighbours

reduce to 5

Example: images

Frey: 1,965 movie frames



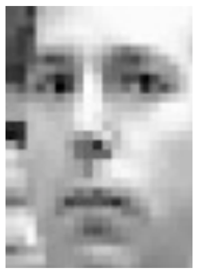

560 dimensions

reduce to 5

interactive low-d view

Example: images

Frey: 1,965 movie frames



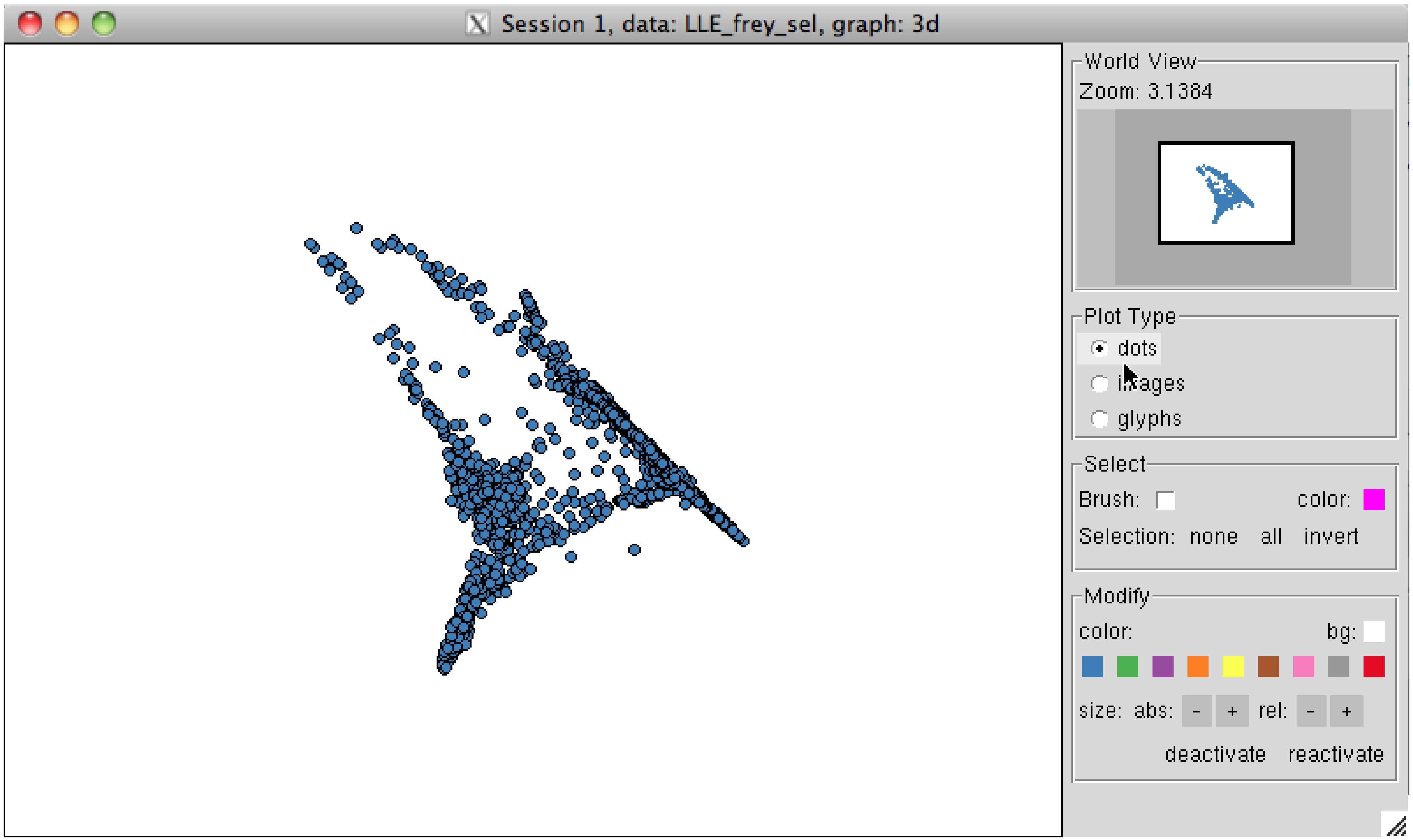
560 dimensions

reduce to 5

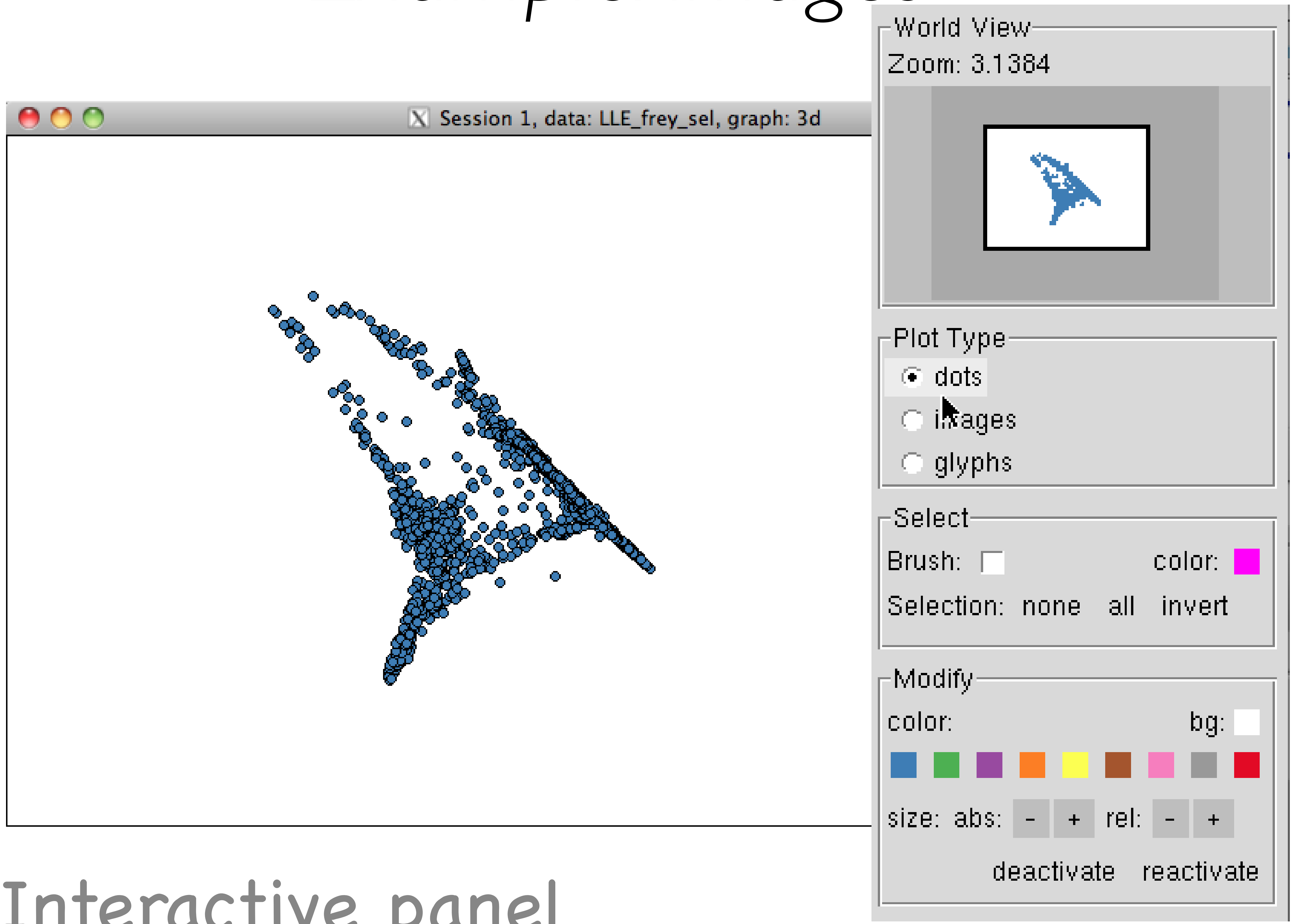
interactive low-d view

connect low-d views

Example: images

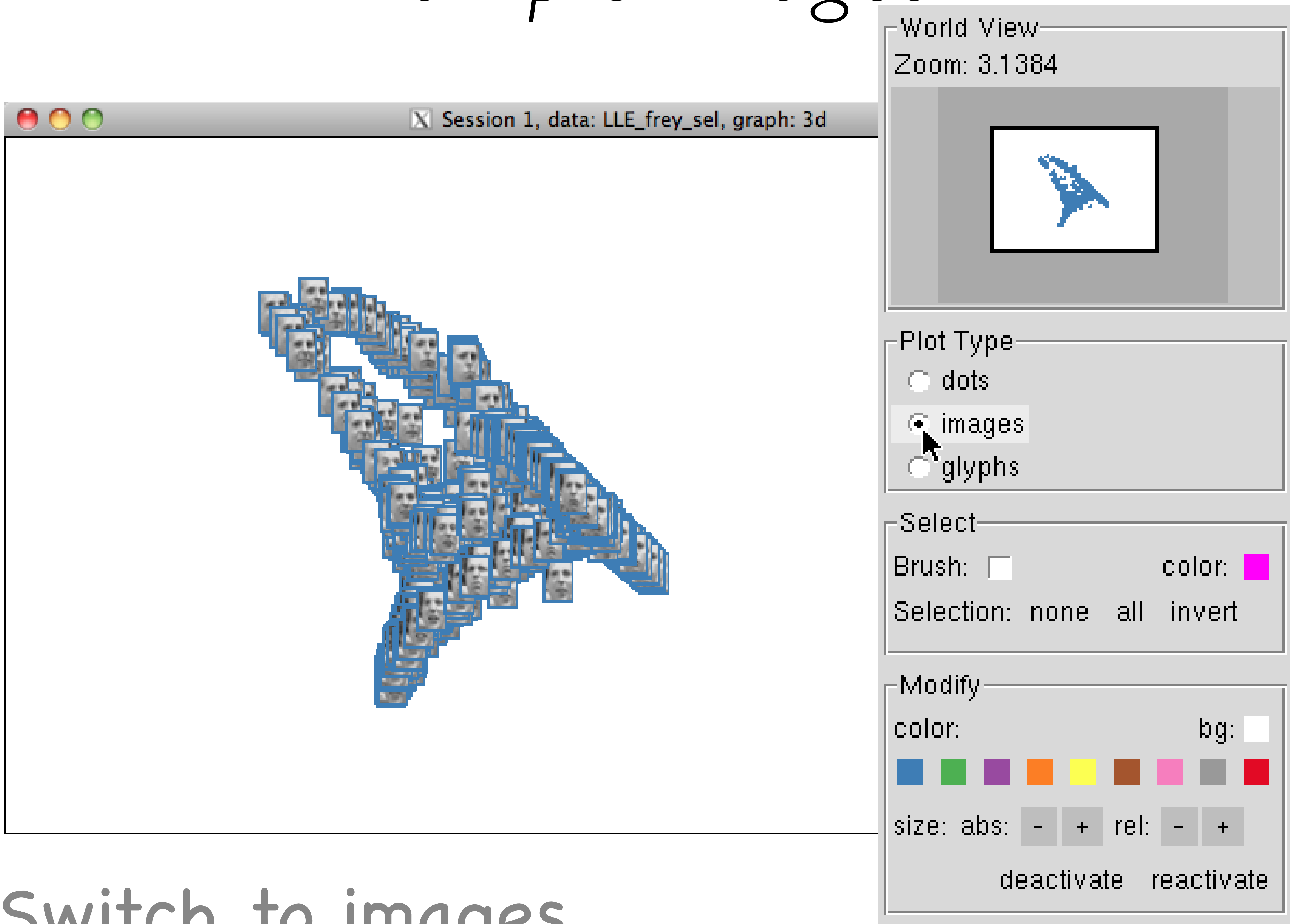


Example: images



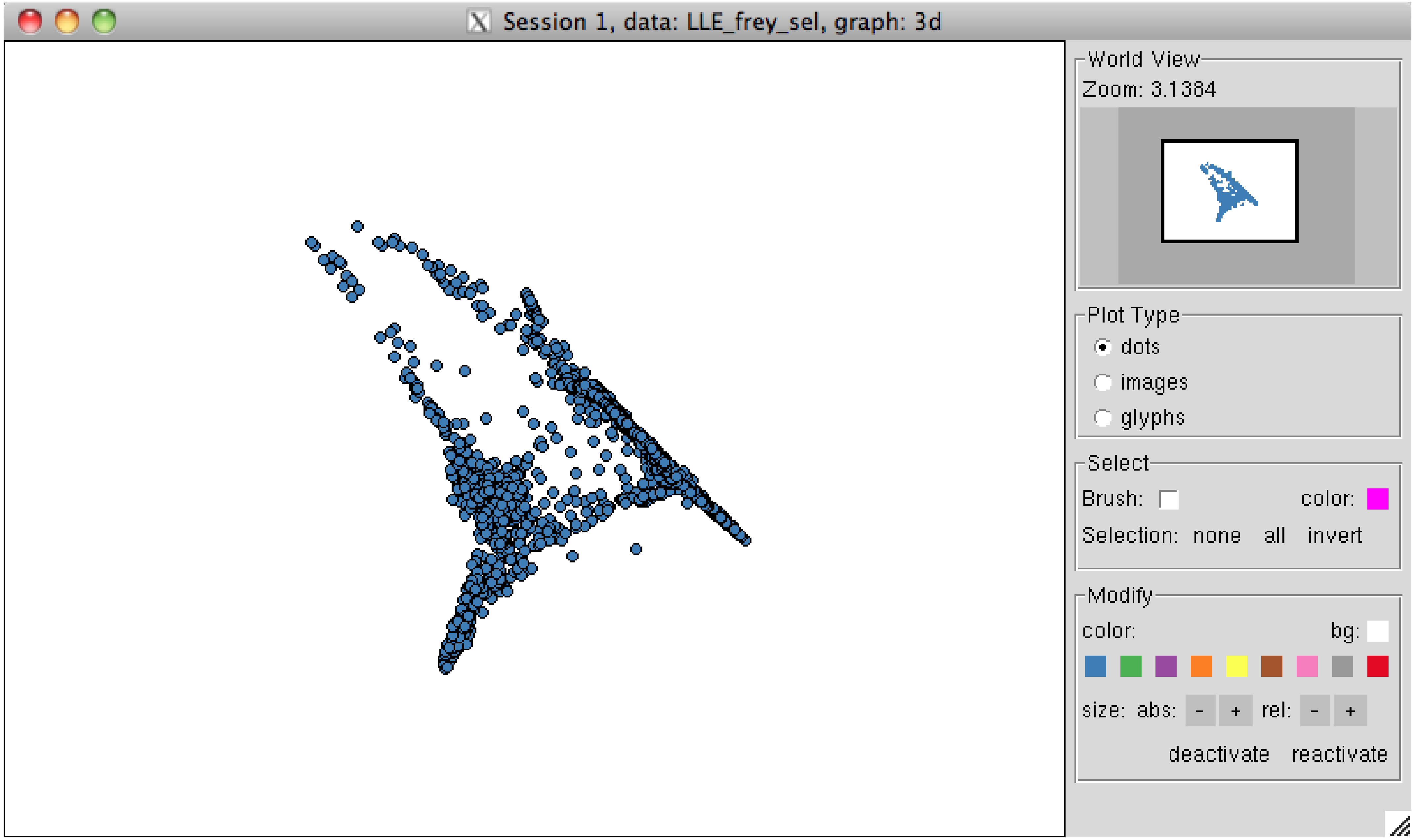
Interactive panel

Example: images



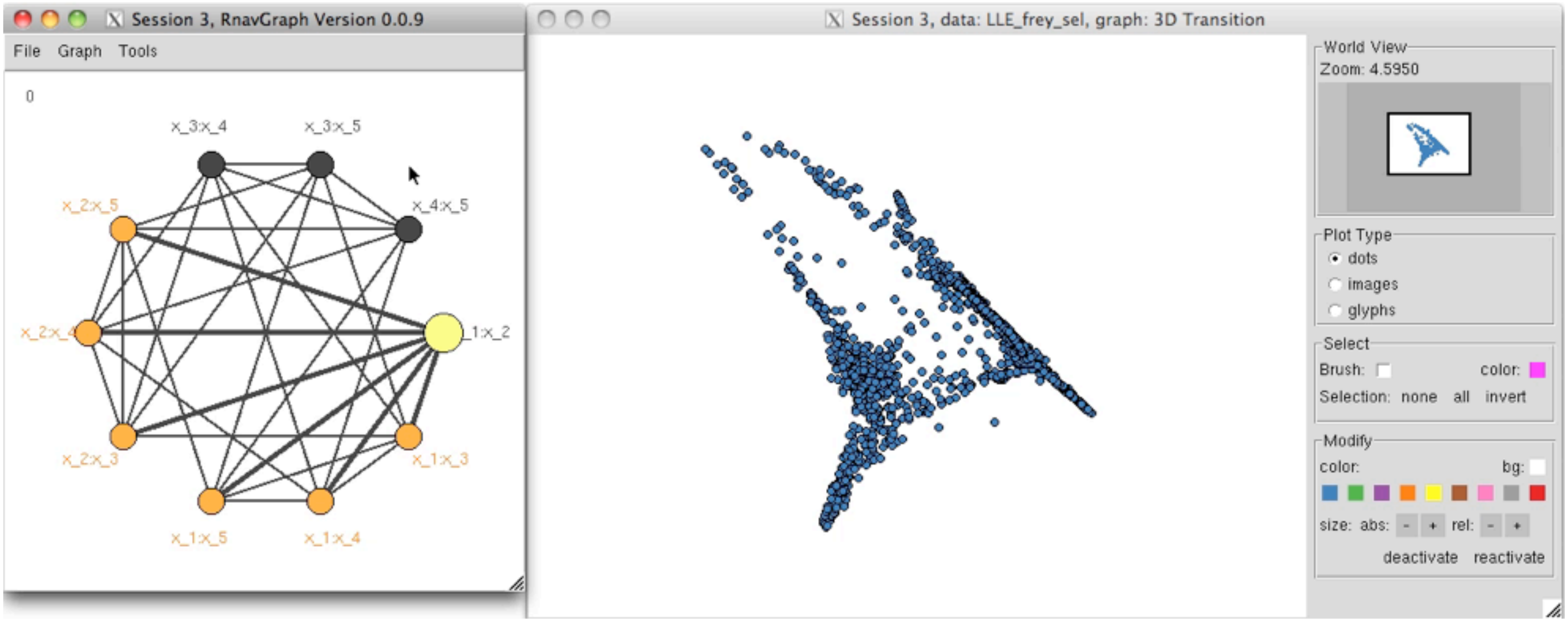
Switch to images

Example: images



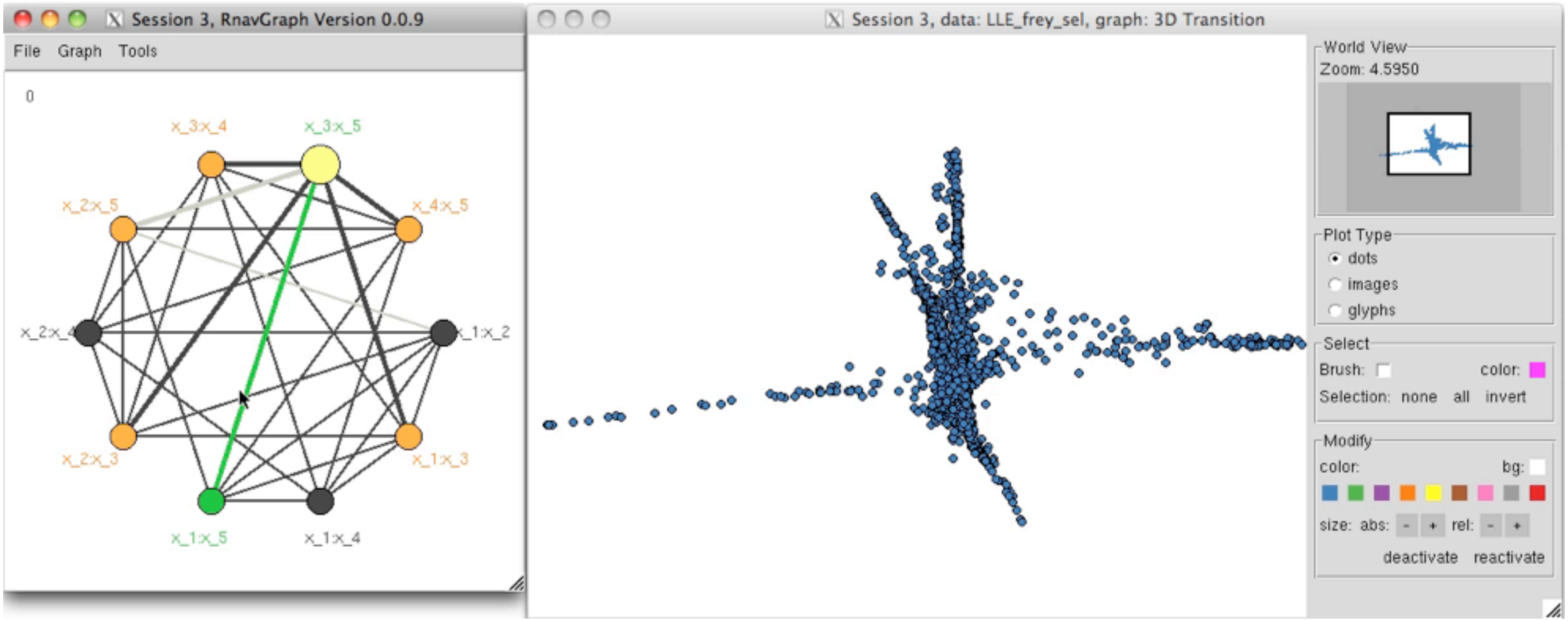
Back to dots

Example: images



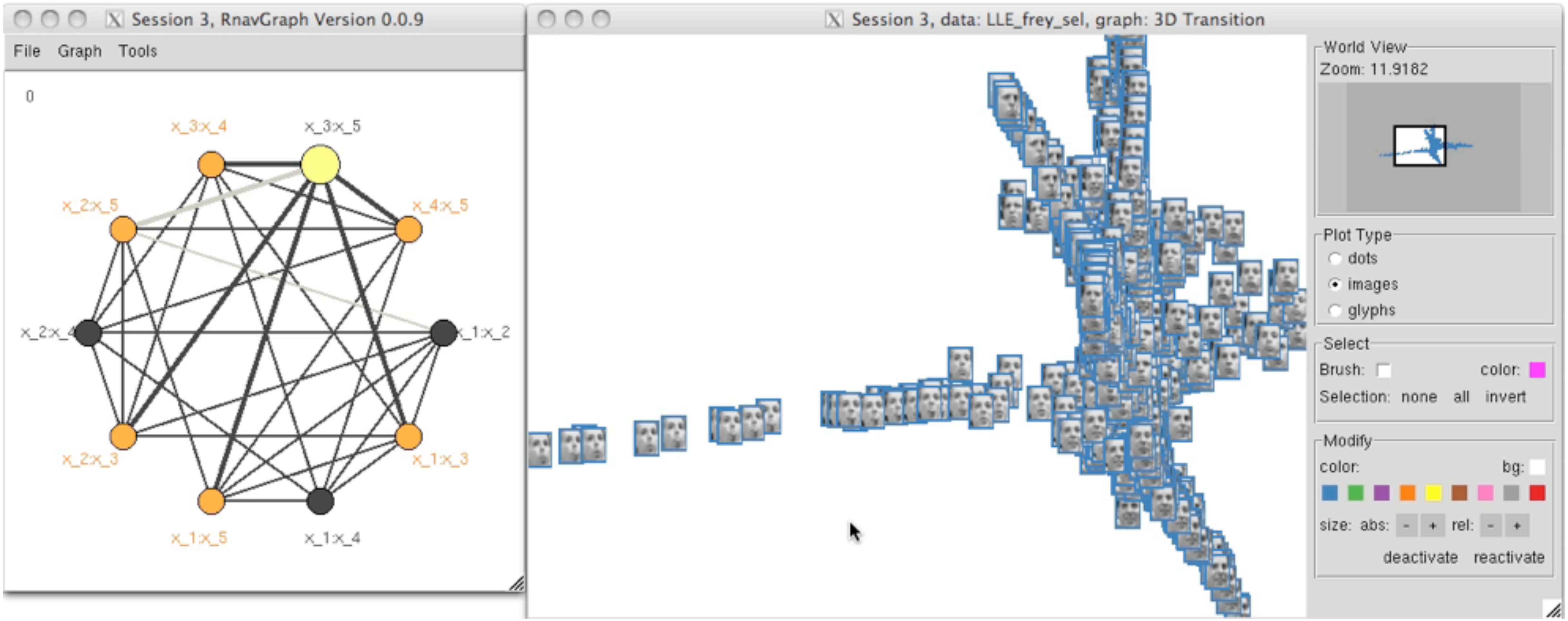
Lots of structure ... explored in 5d

Example: images



Lots of structure ... explored in 5d

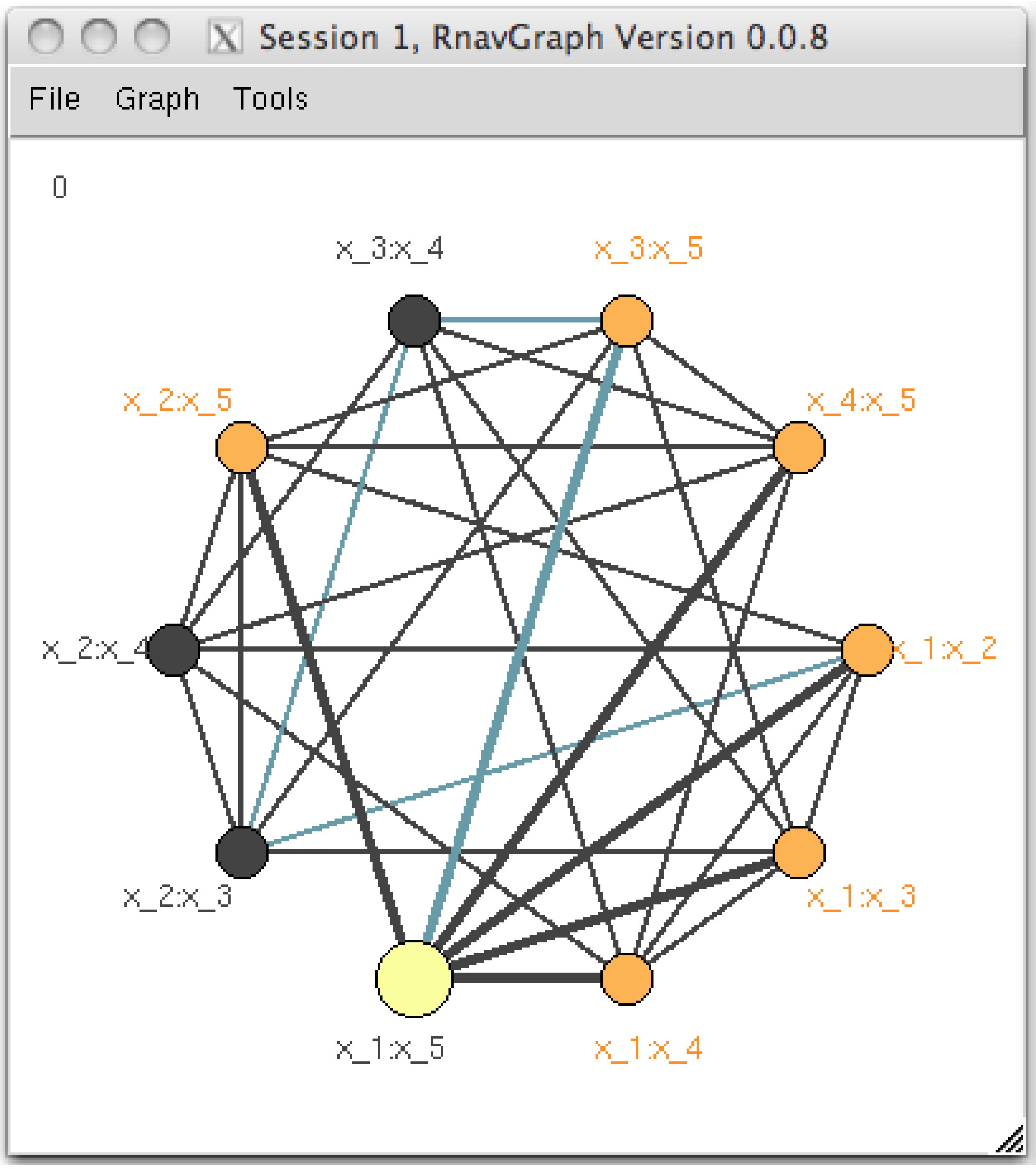
Example: images



Lots of structure ... explored in 5d

Aside: 4d transitions

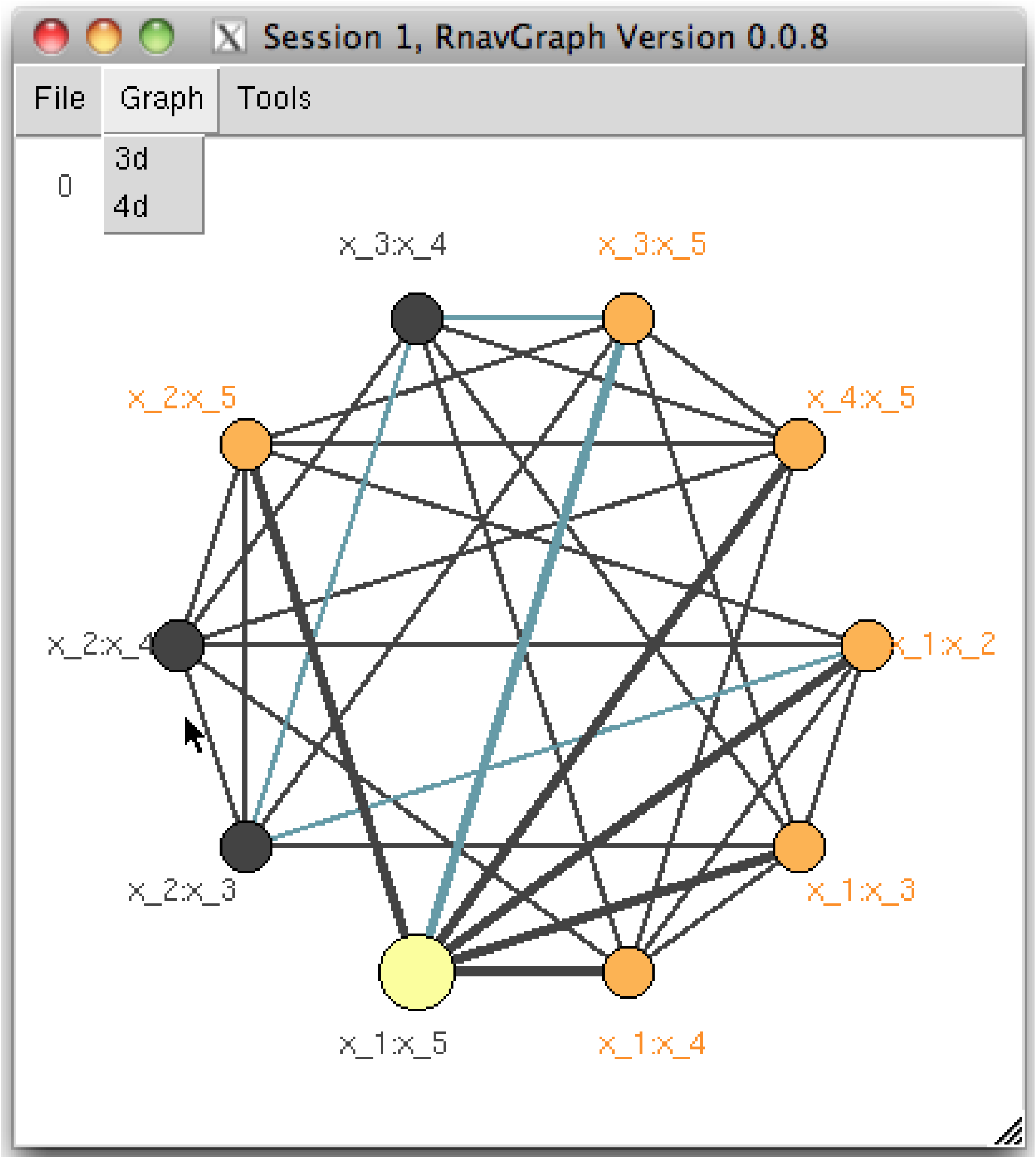
3d and 4d transition graphs



3d transition graph

Aside: 4d transitions

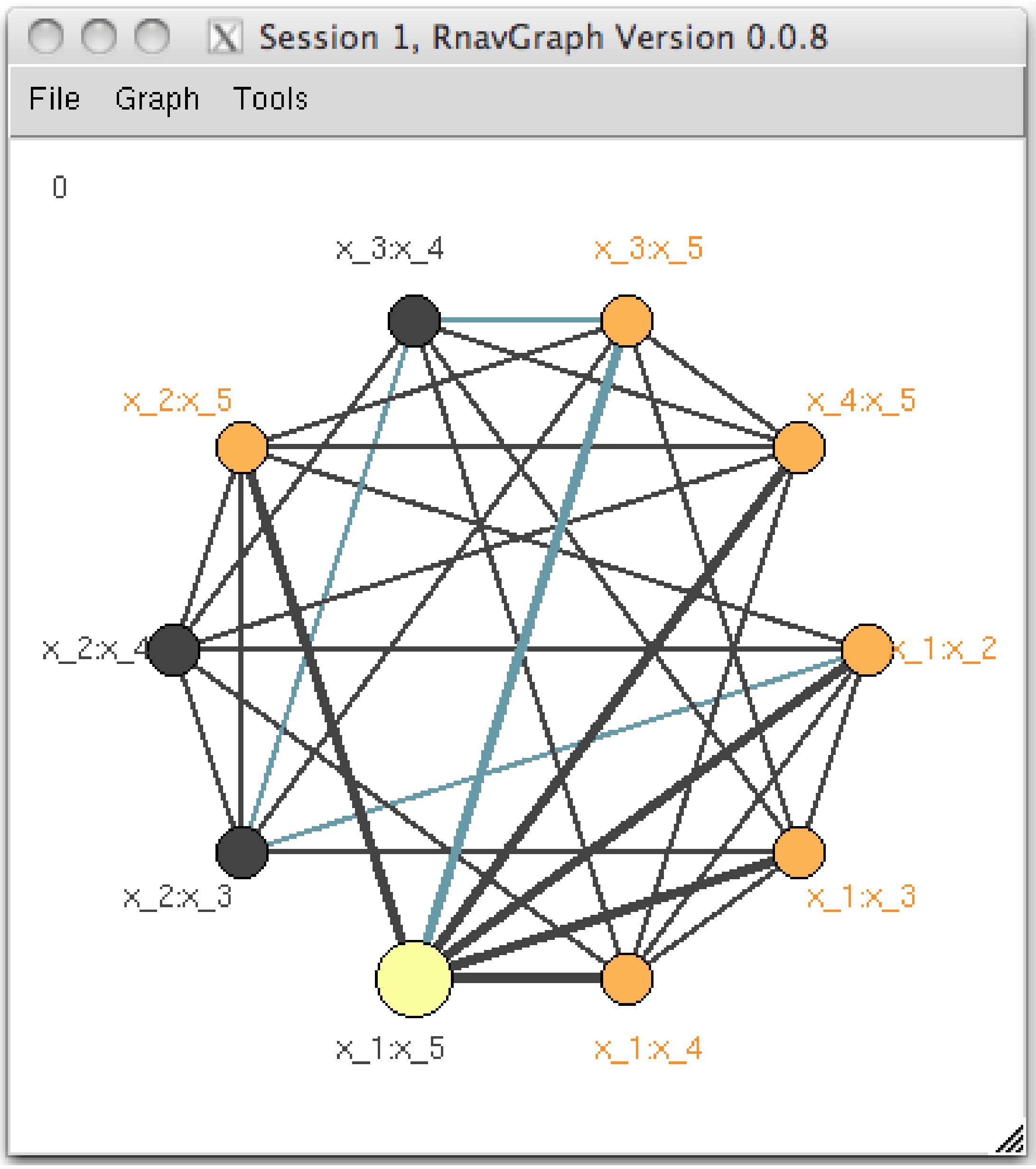
3d and 4d transition graphs



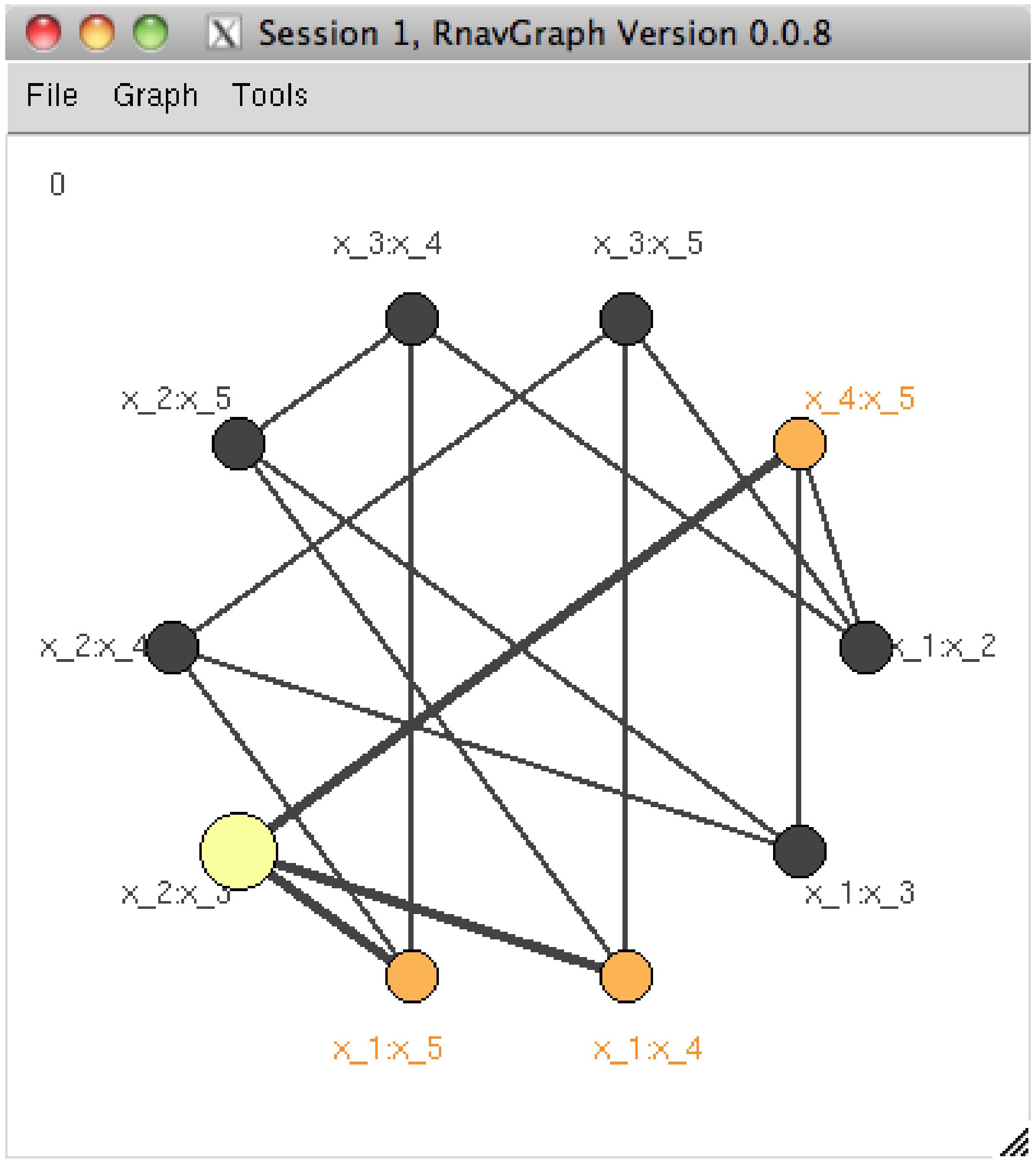
3d transition graph

Aside: 4d transitions

3d and 4d transition graphs



3d transition graph

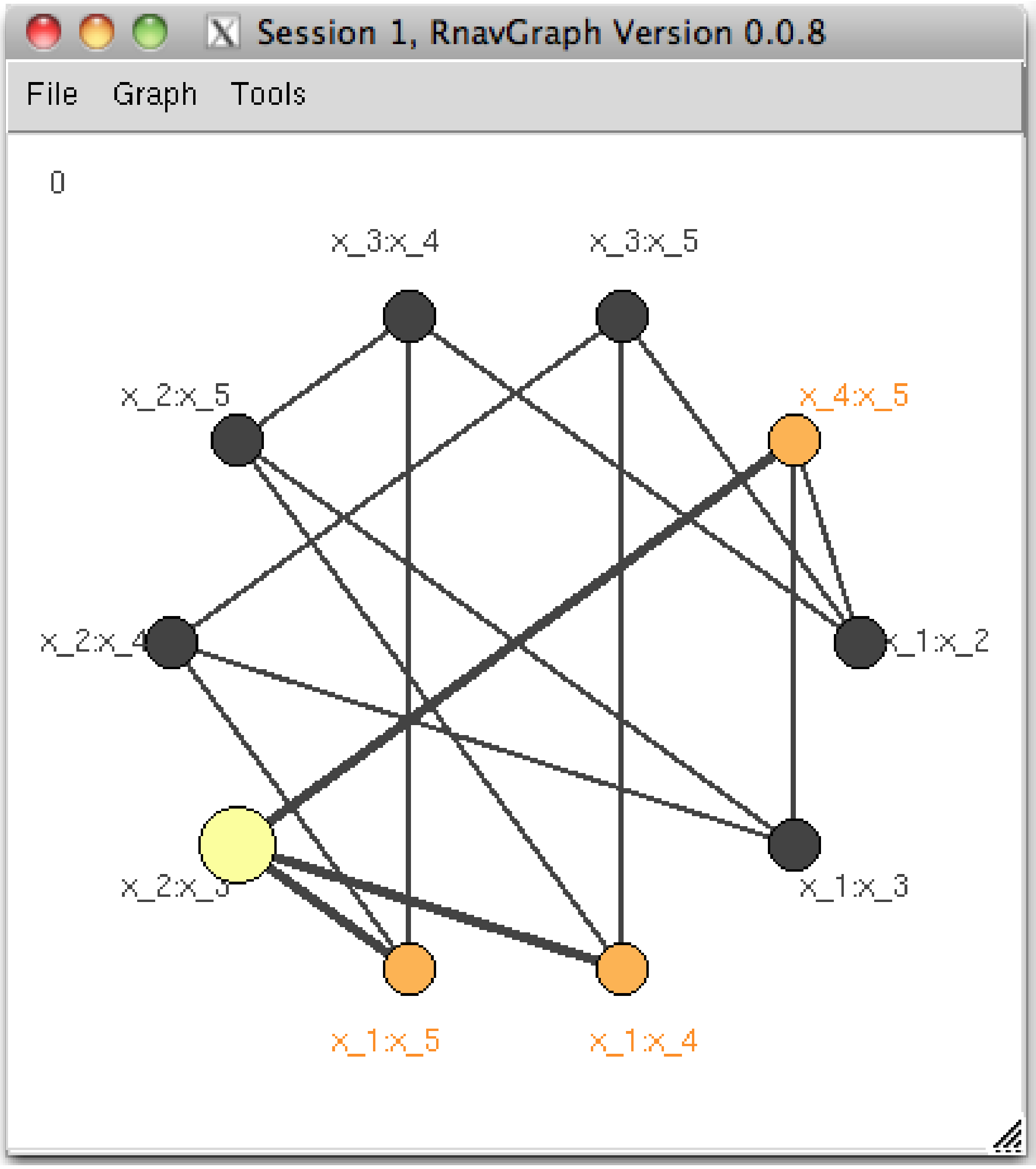


its complement

a 4d transition graph

Aside: 4d transitions

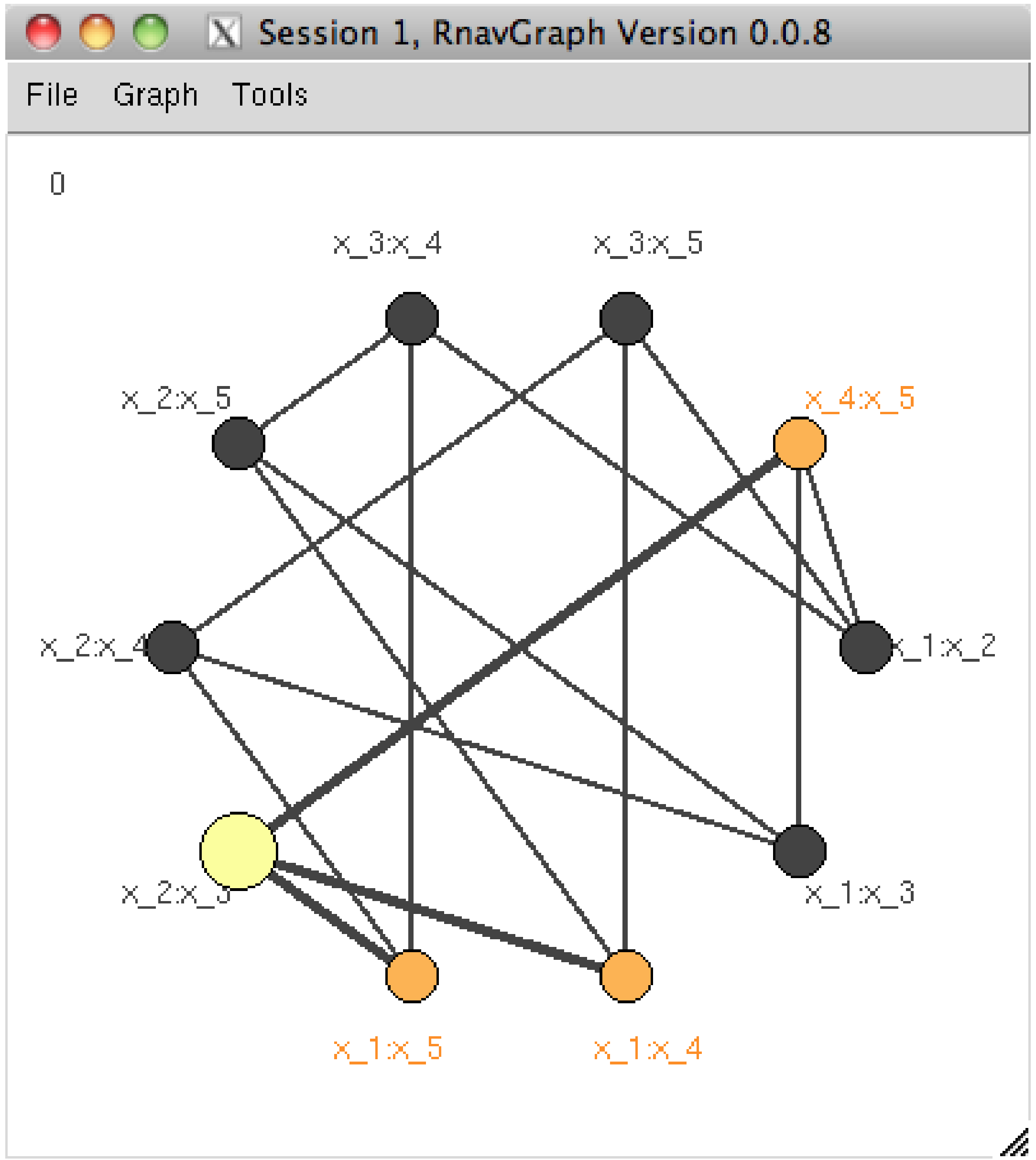
3d and 4d transition graphs



a 4d transition graph

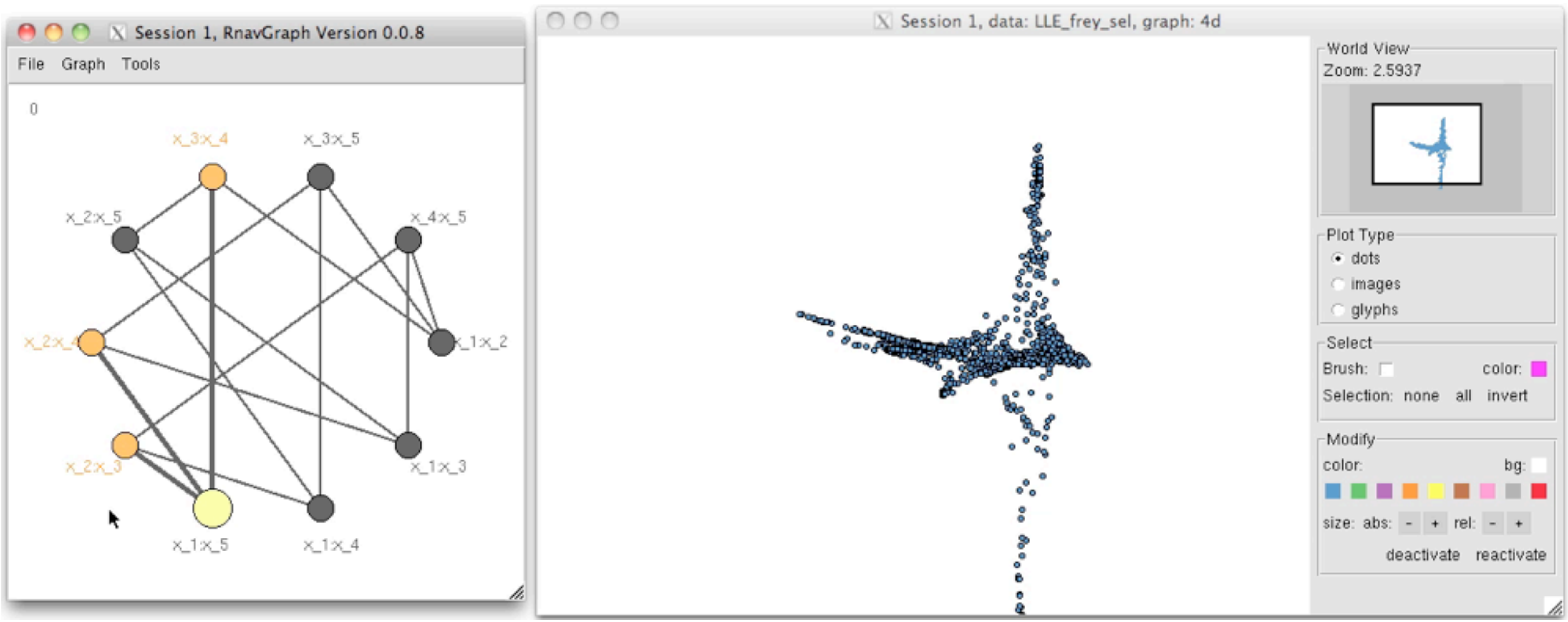
Aside: 4d transitions

3d and 4d transition graphs



Aside: 4d transitions

3d and 4d transition graphs



4d navGraph

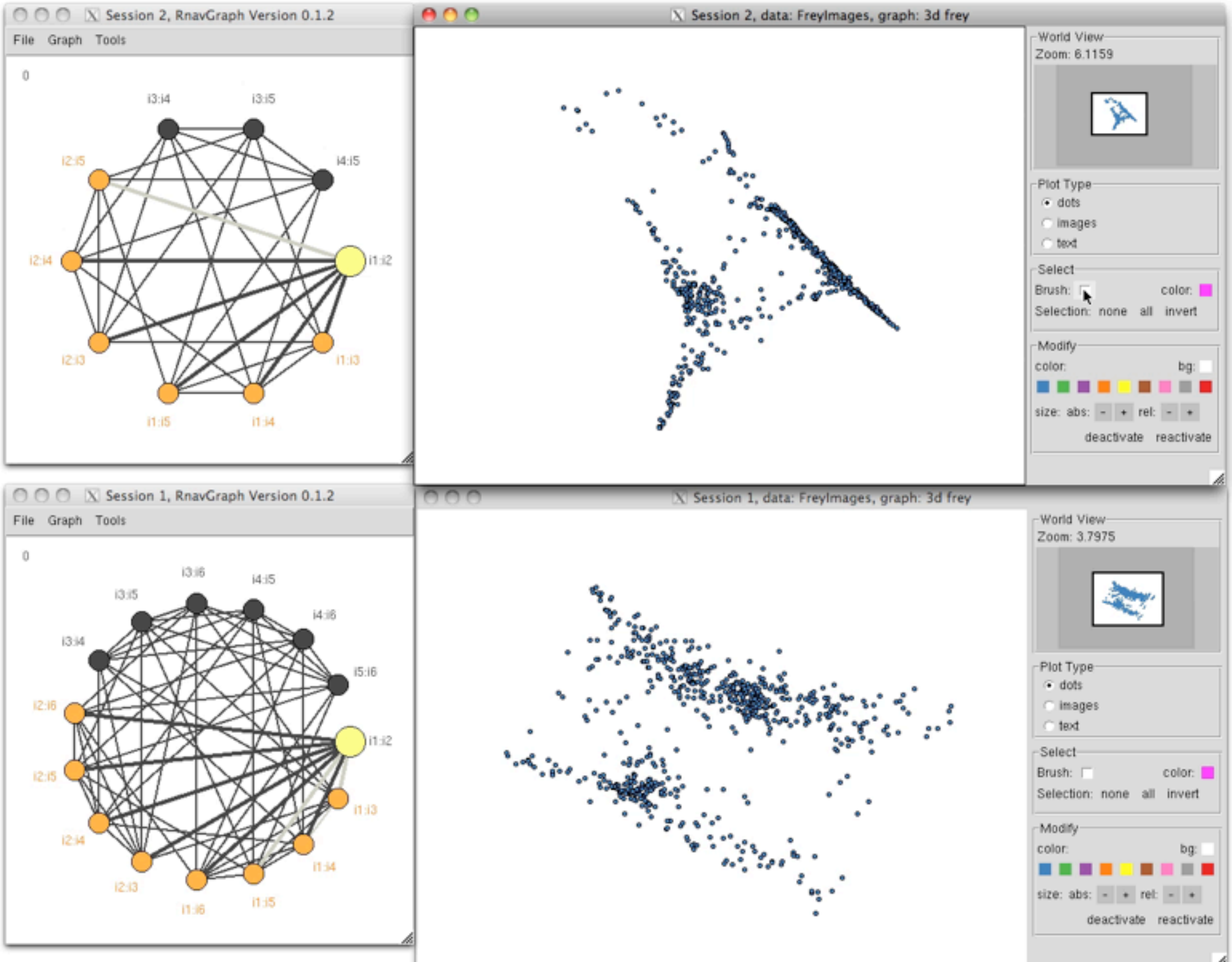
Observe the 4d transition

NOT a rigid rotation

Can link across NavGraph Sessions

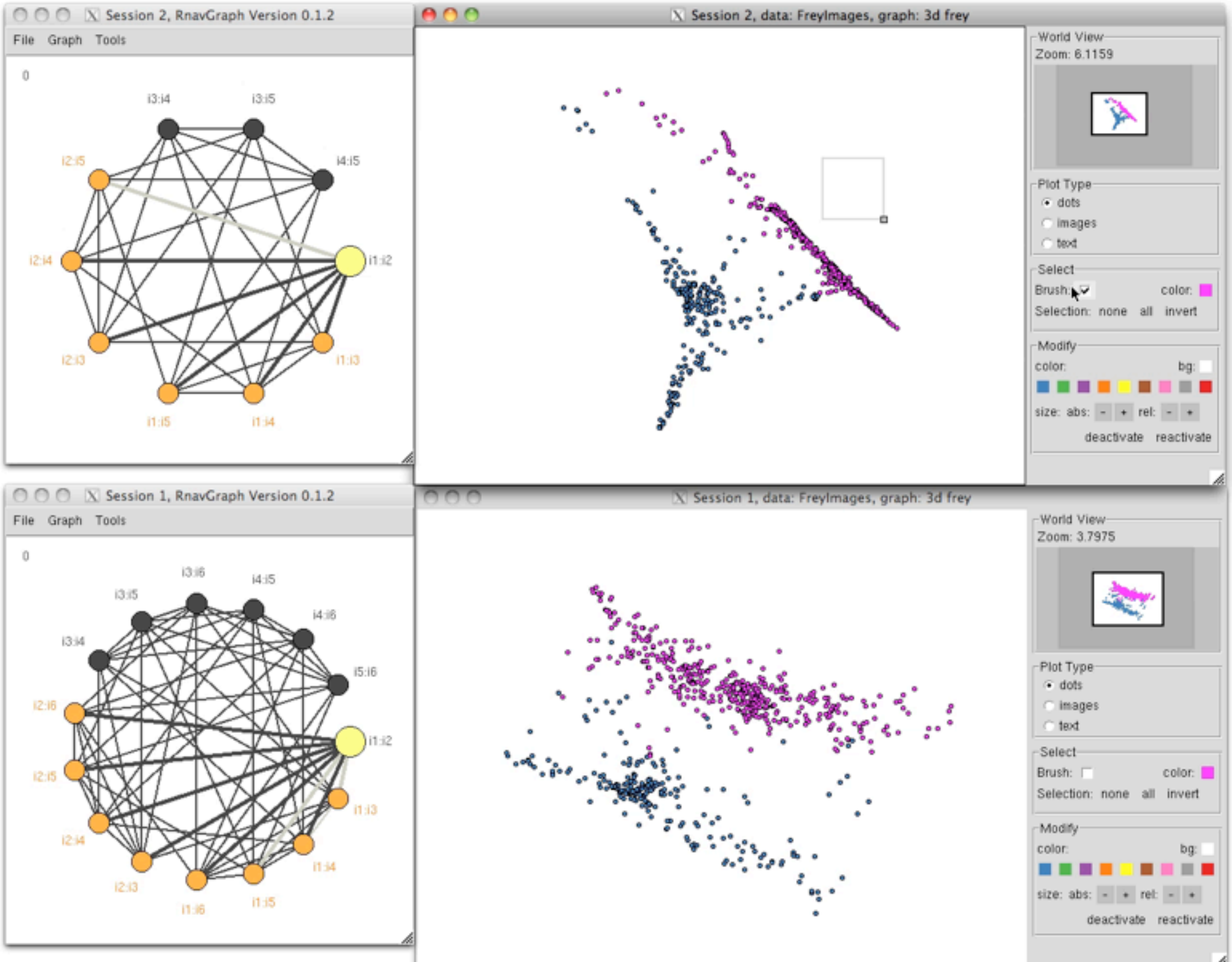
Here LLE and ISOMAP embeddings

Can link across NavGraph Sessions



Here LLE and ISOMAP embeddings

Can link across NavGraph Sessions

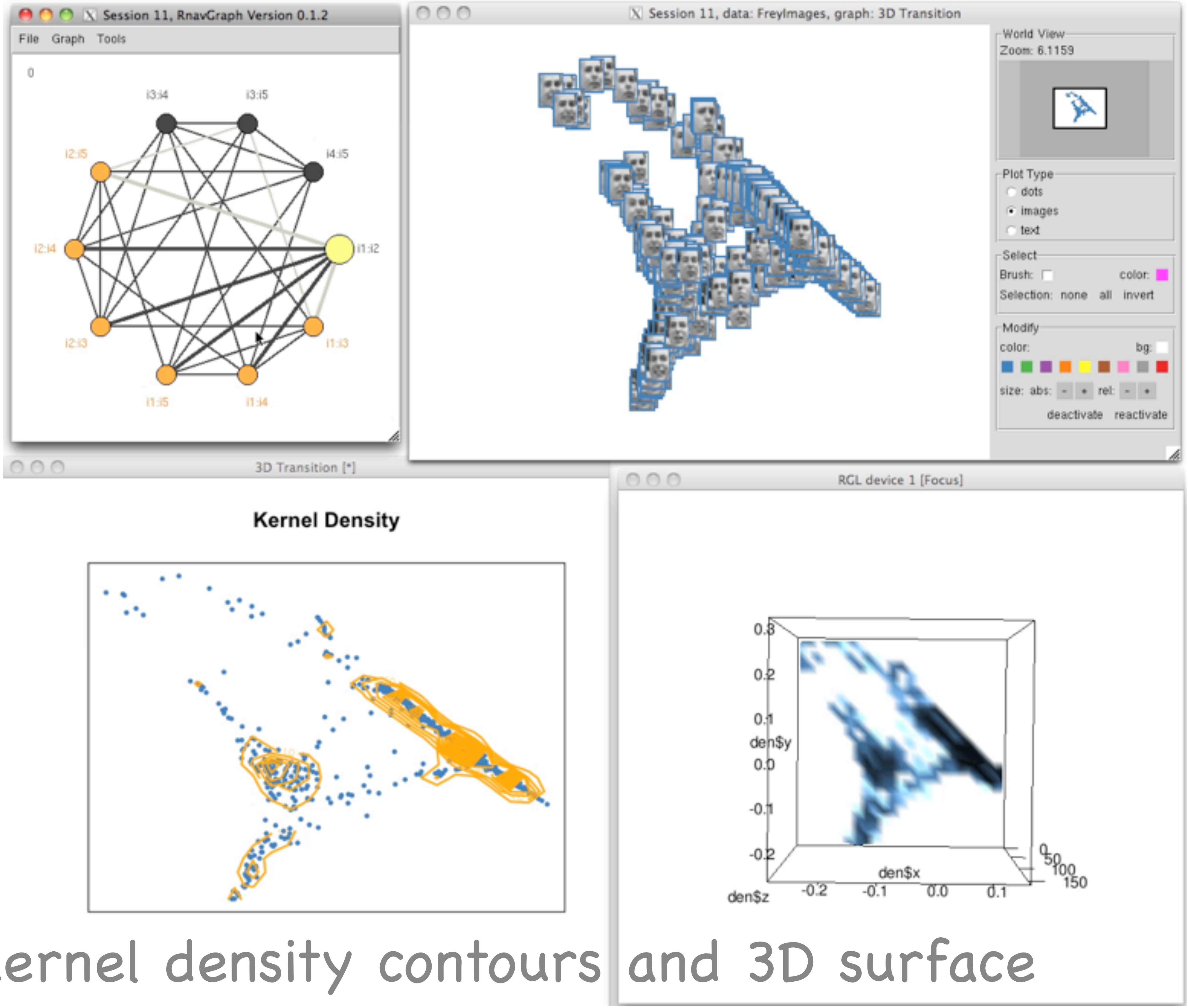


Here LLE and ISOMAP embeddings

Multiple visualizations

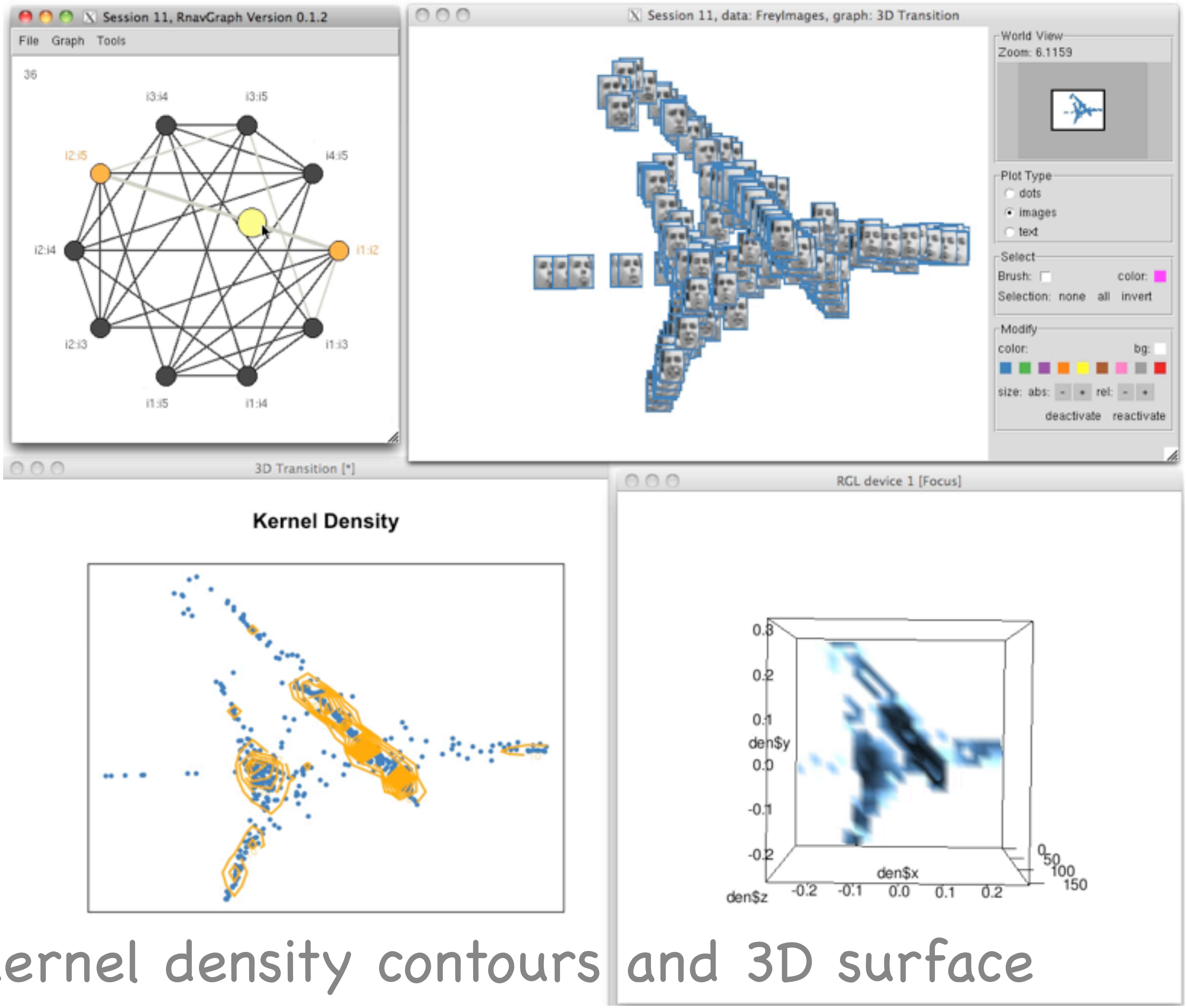
Kernel density contours and 3D surface

Multiple visualizations



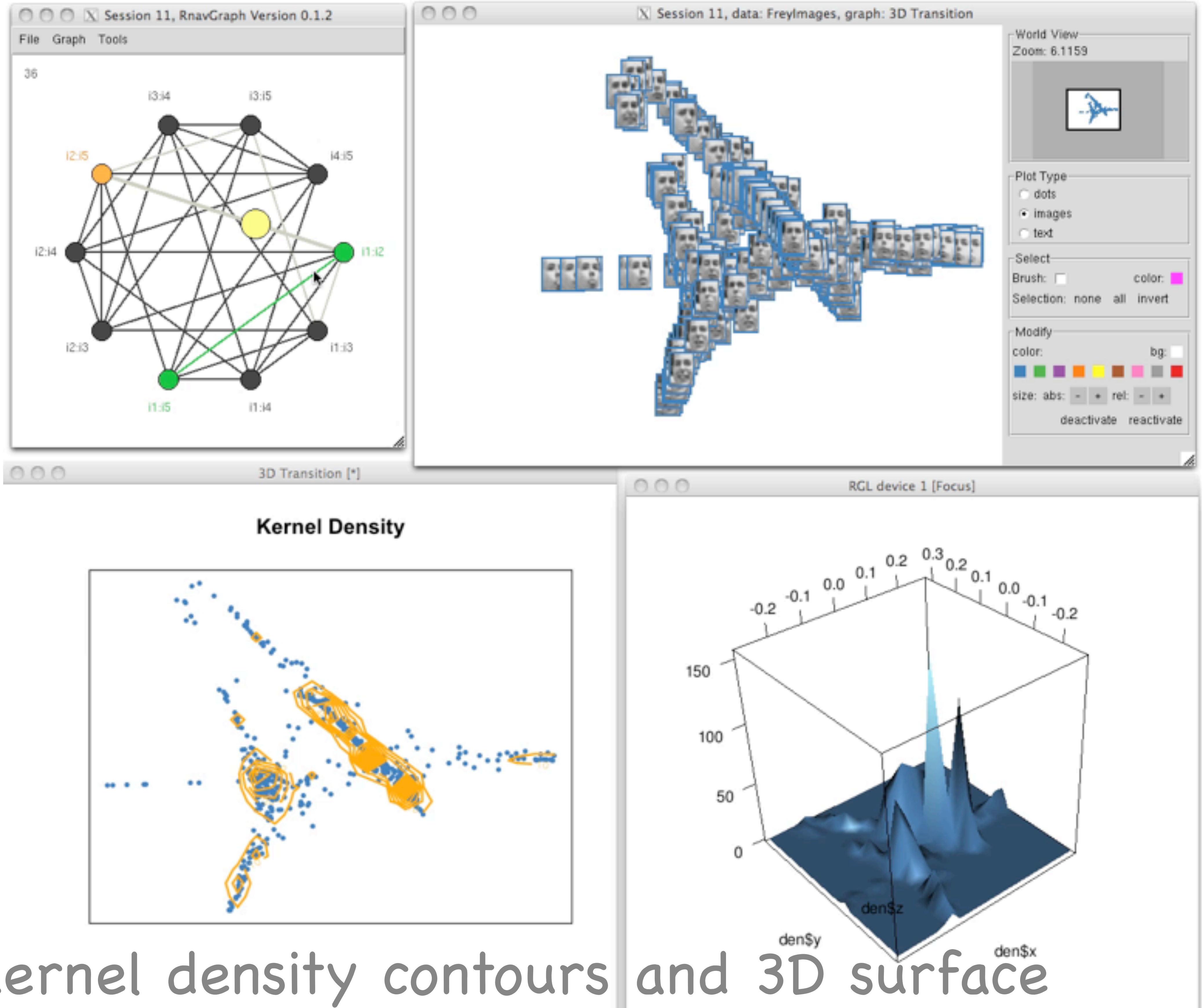
Kernel density contours and 3D surface

Multiple visualizations



Kernel density contours and 3D surface

Multiple visualizations



Kernel density contours and 3D surface

Axis systems

- Choice of coordinate axis layout
  - Orthogonal (`RnavGraph` `R` package)
  - Radial (`PairViz` `R` package)
  - Parallel (`PairViz` `R` package)
- Find a good order of axes
  - Complete graphs on variables only
  - Hamiltonian paths, Eulerian tours, Hamiltonian decompositions
  - greedy methods, TSPs

Summary

Graph theory structure

- graphs as maps to navigate high-dimensional space
- graph walks as low dimensional trajectories
- focus on interesting walks
- needs interactive data visualization
- capitalize on visual ability

Summary

Graph theory structure

- organizes order of axes (e.g. radial, parallel, orthog.)
- use interesting orders (correlations, scagnostics, etc.)
- organizes ANY display order (e.g. multiple comparisons)

Summary

Try it yourself

- R packages (available on CRAN):
  - `PairViz` Hurley & Oldford
  - `RnavGraph` Waddell & Oldford

Thank you

Thank you

Questions?

有問題嗎？

有问题吗？

質問はありますか？

질문이 있으십니까?

Papers

Hurley & Oldford:

- Graphs as navigational infrastructure for high dimensional data spaces (**Comp Stats 2011**)
- Pairwise display of high dimensional information via Eulerian tours and Hamiltonian decompositions (**JCGS, 2010**)
    - Eulerian tour algorithms for data visualization and the `PairViz` package (**Comp Stats 2011**)
    - `PairViz` R package ... available on CRAN.

Oldford & Waddell:

- Visual clustering of high-dimensional data by navigating low-dimensional spaces (**ISI Dublin, 2011**)
- RnavGraph: A visualization tool for navigating through high dimensional data (**ISI Dublin, 2011**)
- `RnavGraph` R package ... available on CRAN

Oldford & Zhou:

- Tree Ensemble Reduced Clustering via a Graph Algebraic Framework. submitted