大学共同利用機関法人 情報・システム研究機構
統計数理研究所
The Institute of Statistical Mathematics

Graph theoretic methods for Data Visualization:

2. Navigating high-dimensional spaces and the `RnavGraph R` package

Wayne Oldford

based on joint work with Adrian Waddell and Catherine Hurley

WATERLOO | MATHEMATICS
STATISTICS AND ACTUARIAL SCIENCE

2011 年度の統計数理研究所共同研究集会
「データ解析環境Rの整備と利用」

Tutorial B2

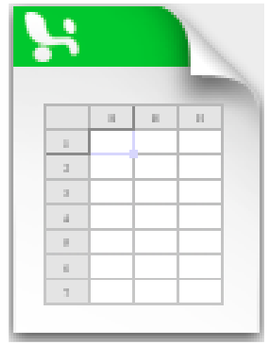
WATERLOO

Challenge

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

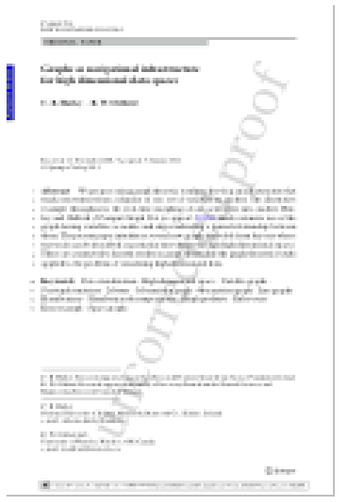
Challenge

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



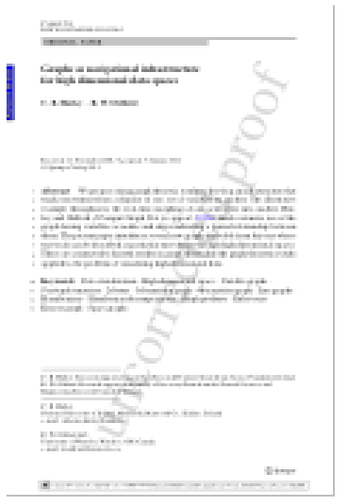
Challenge

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



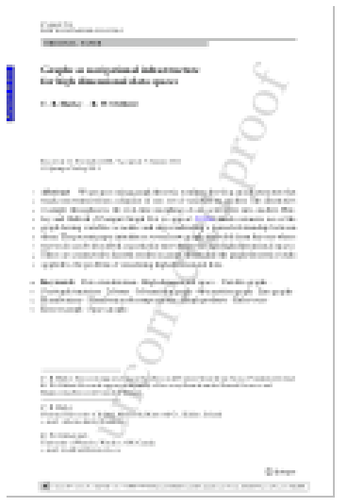
Challenge

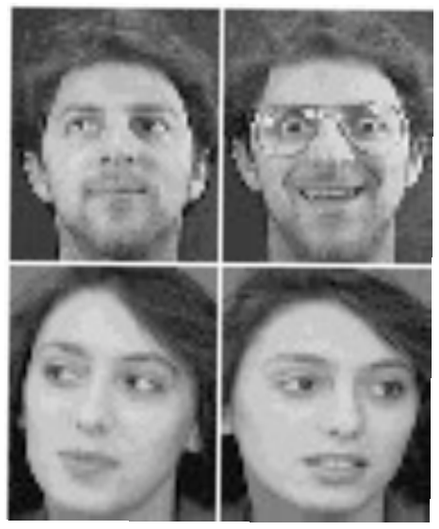
- p values on each of n individuals
- modern data: n, or p, or both, can be very large



Challenge

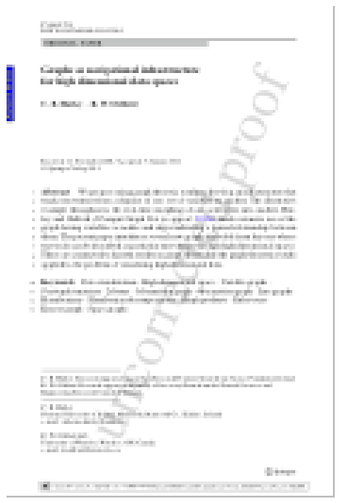
- p values on each of n individuals
- modern data: n, or p, or both, can be very large

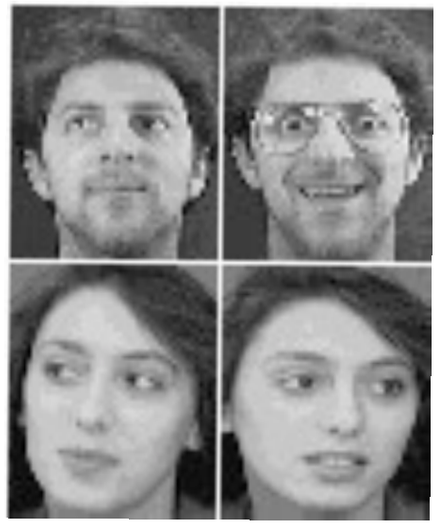




Challenge

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



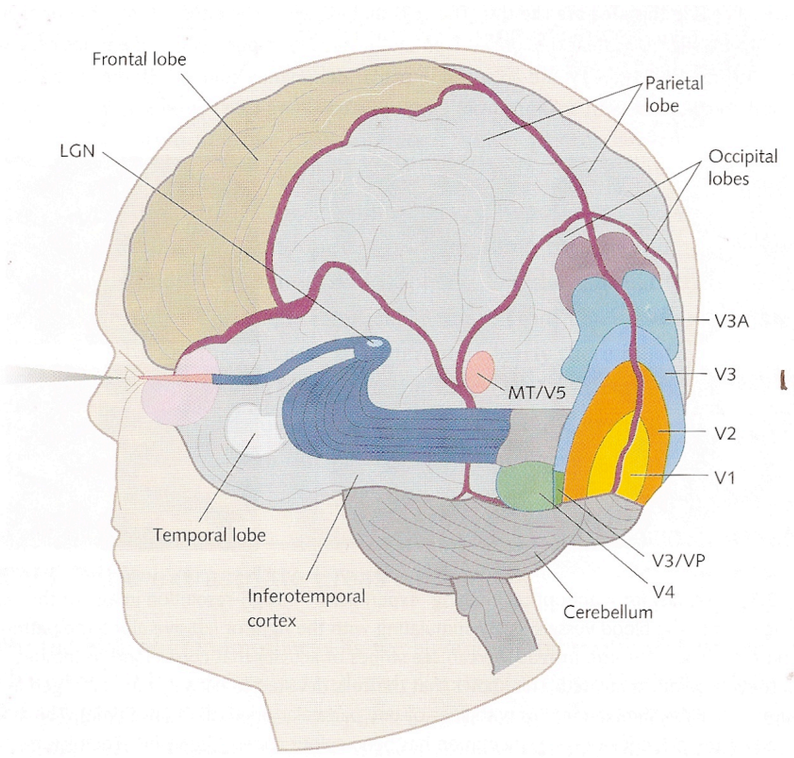


- can have non-obvious variables, complex, unanticipated structure, …

Visualization

powerful human visual system

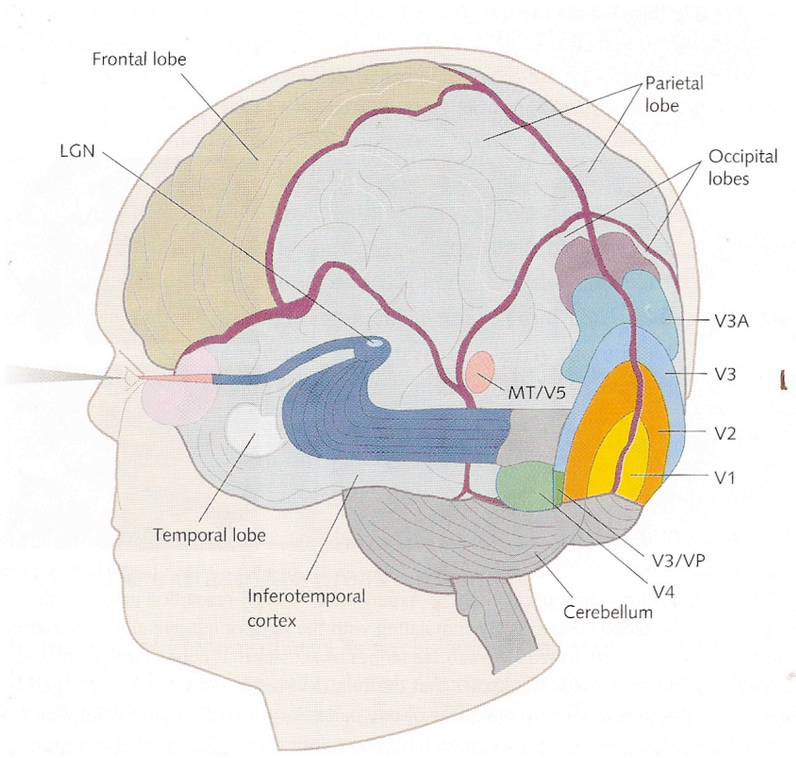
- use a variety of cues:
  - proximity, movement, shape, colour, texture, ...
- patterns, relations, like and unlike, ...
- recognition and discovery
- structure need not be anticipated



Visualization

powerful human visual system

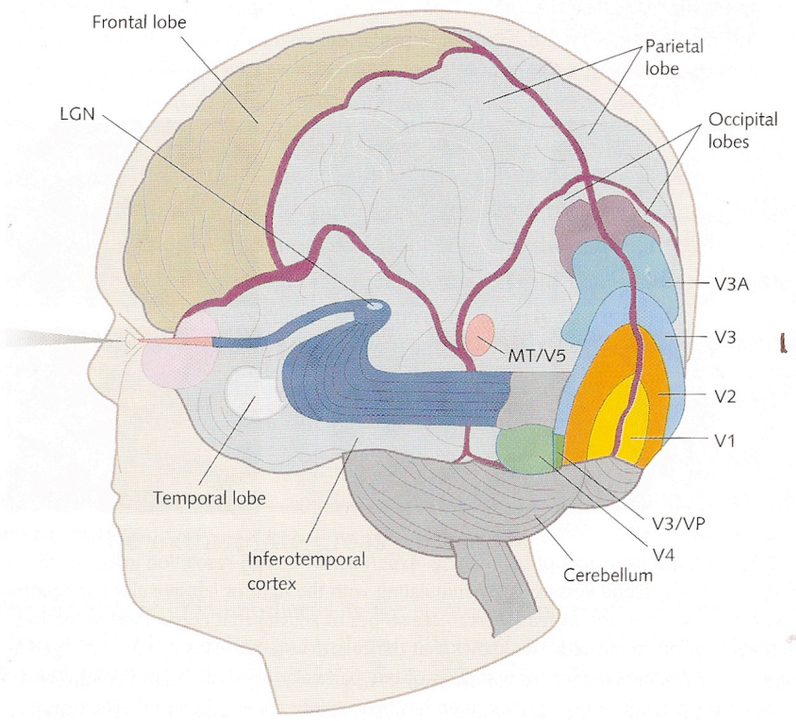
- use a variety of cues:
  - proximity, movement, shape, colour, texture, ...
- patterns, relations, like and unlike, ...
- recognition and discovery
- structure need not be anticipated



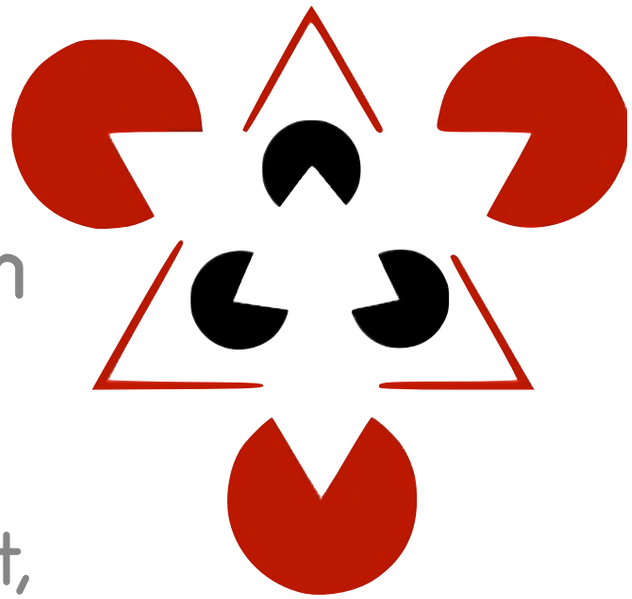
Visualization

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

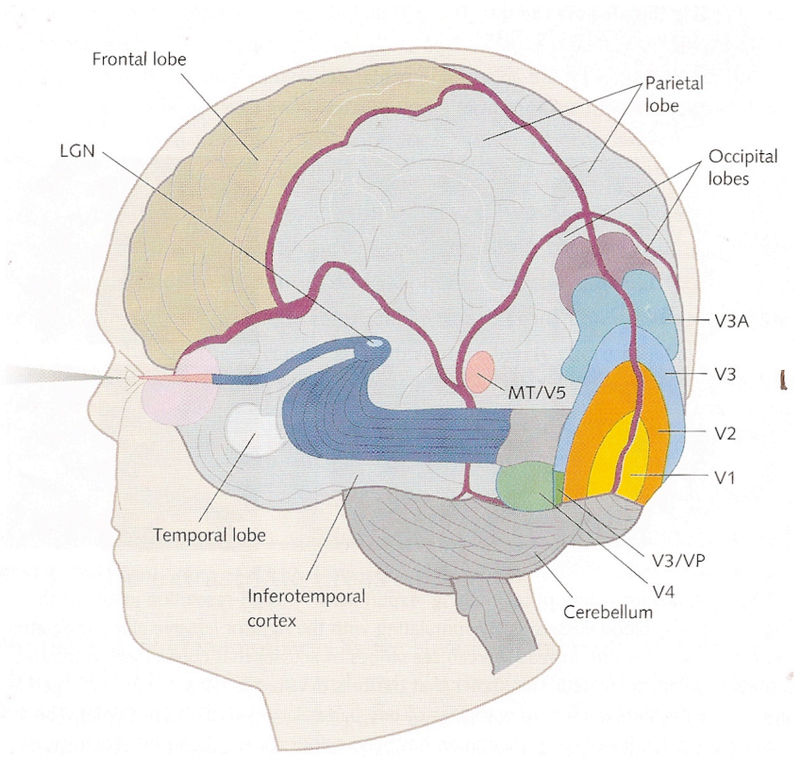


Visualization



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

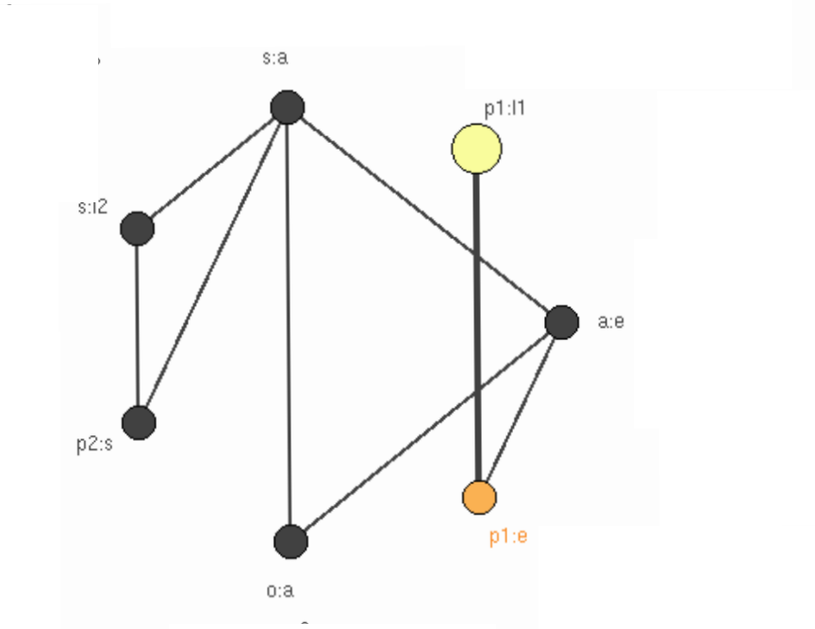
  ... at a time

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

Navigation Graphs

Connecting low-d spaces

- node = variable pair
- edges connect nodes that share a variable
- could display scatterplot at each node

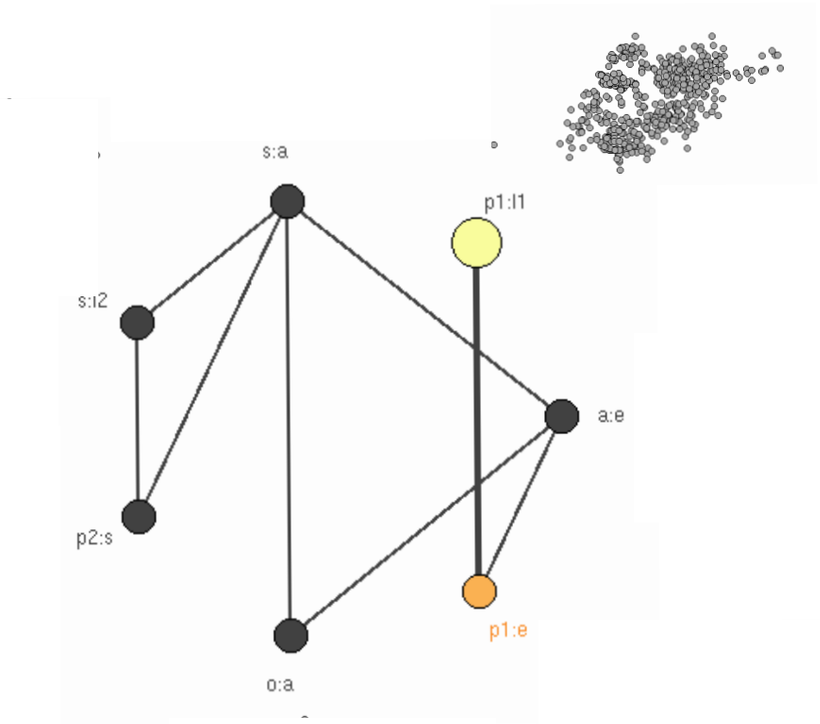


- edges are 3D transitions
- high dimensional space is explored by moving from one 2D space to another through 3D (or 4D) transitions
- track/map exploration
- suggest routes

Navigation Graphs

Connecting low-d spaces

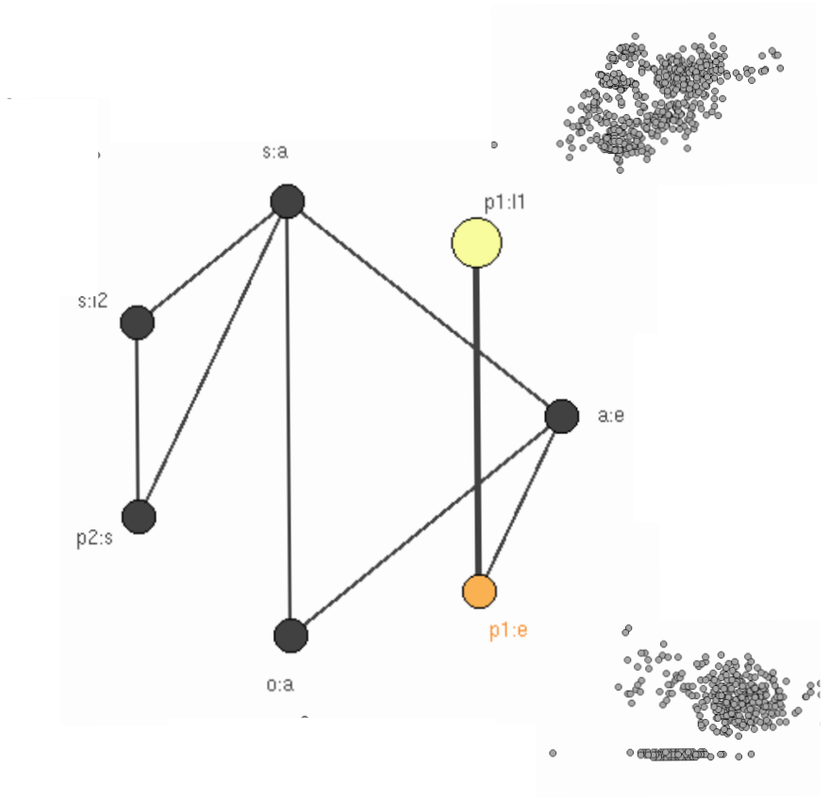
- node = variable pair
- edges connect nodes that share a variable
- could display scatterplot at each node
- edges are 3D transitions
- high dimensional space is explored by moving from one 2D space to another through 3D (or 4D) transitions
- track/map exploration
- suggest routes



Navigation Graphs

Connecting low-d spaces

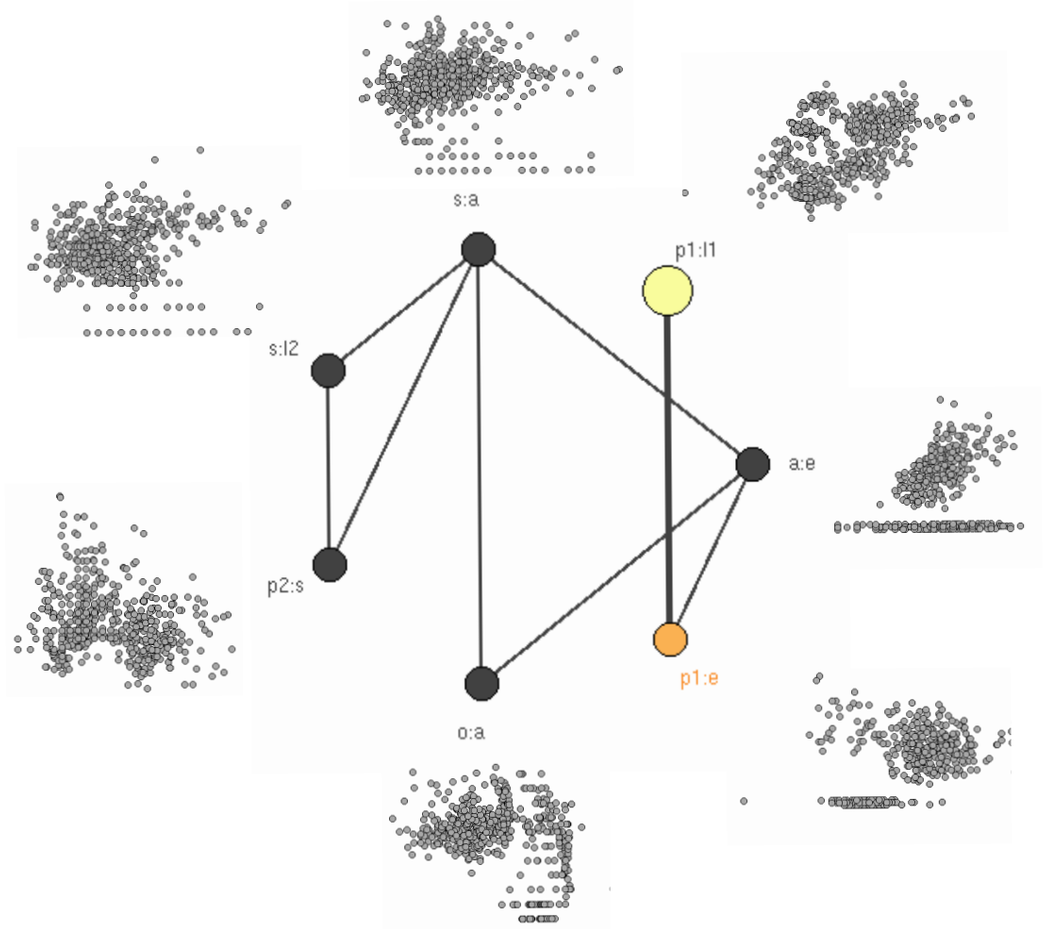
- node = variable pair
- edges connect nodes that share a variable
- could display scatterplot at each node
- edges are 3D transitions
- high dimensional space is explored by moving from one 2D space to another through 3D (or 4D) transitions
- track/map exploration
- suggest routes



Navigation Graphs

Connecting low-d spaces

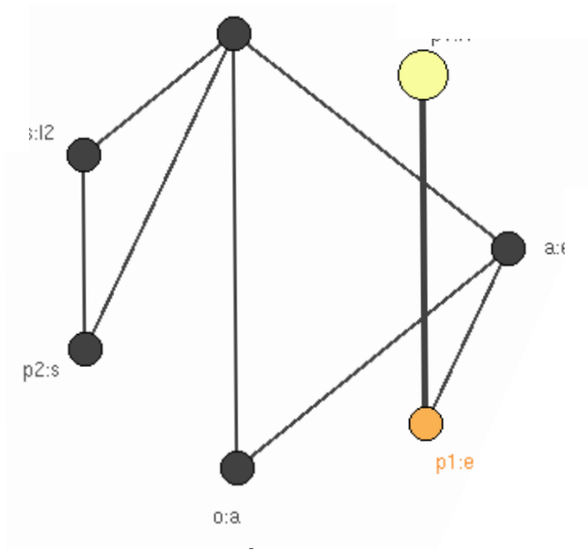
- node = variable pair
- edges connect nodes that share a variable
- could display scatterplot at each node
- edges are 3D transitions
- high dimensional space is explored by moving from one 2D space to another through 3D (or 4D) transitions
- track/map exploration
- suggest routes



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
- high dimensional space is explored by moving from one 2D space to another through 3D (or 4D) transitions
- track/map exploration
- suggest routes

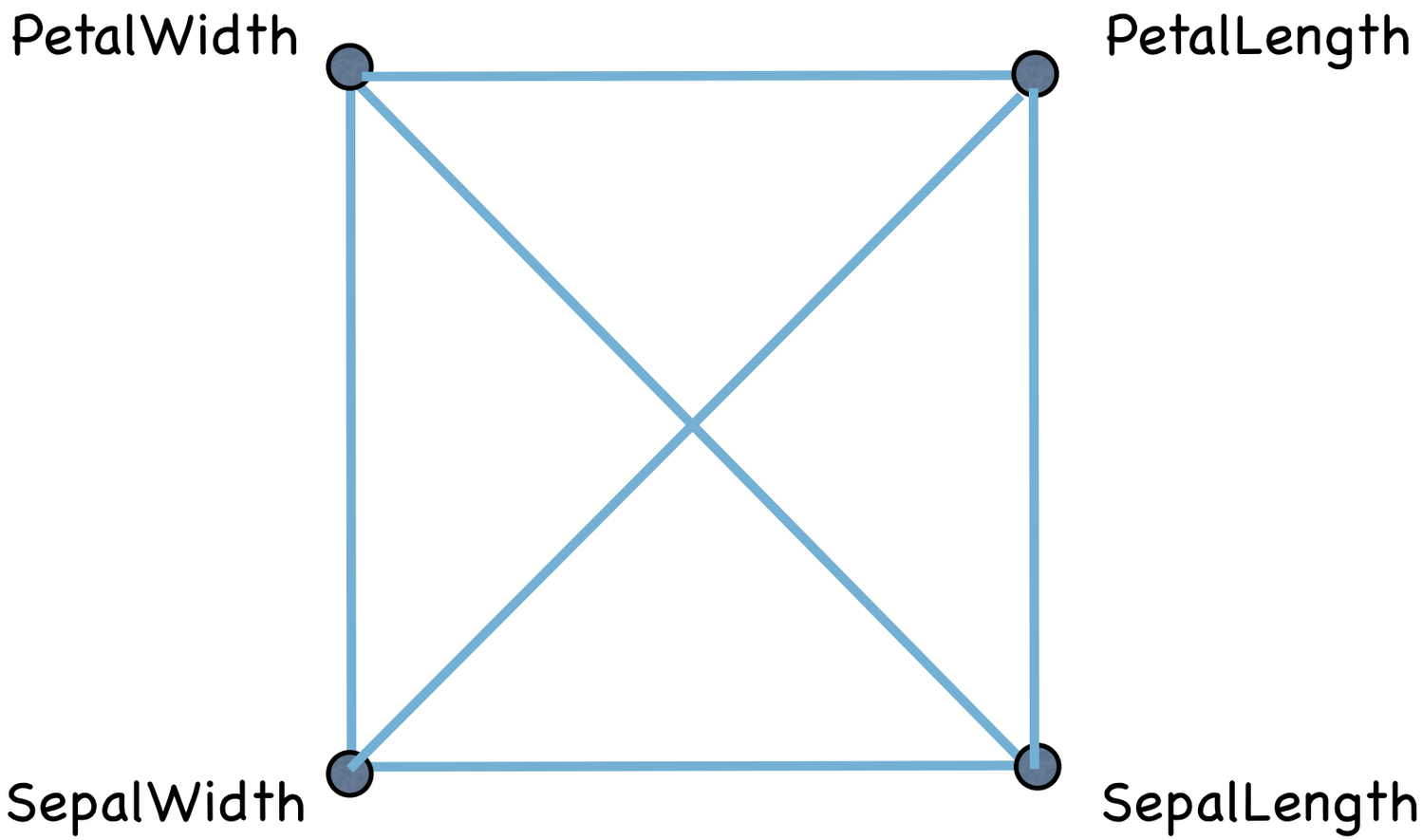
RNavgraph

... R implementation

RnavGraph demo

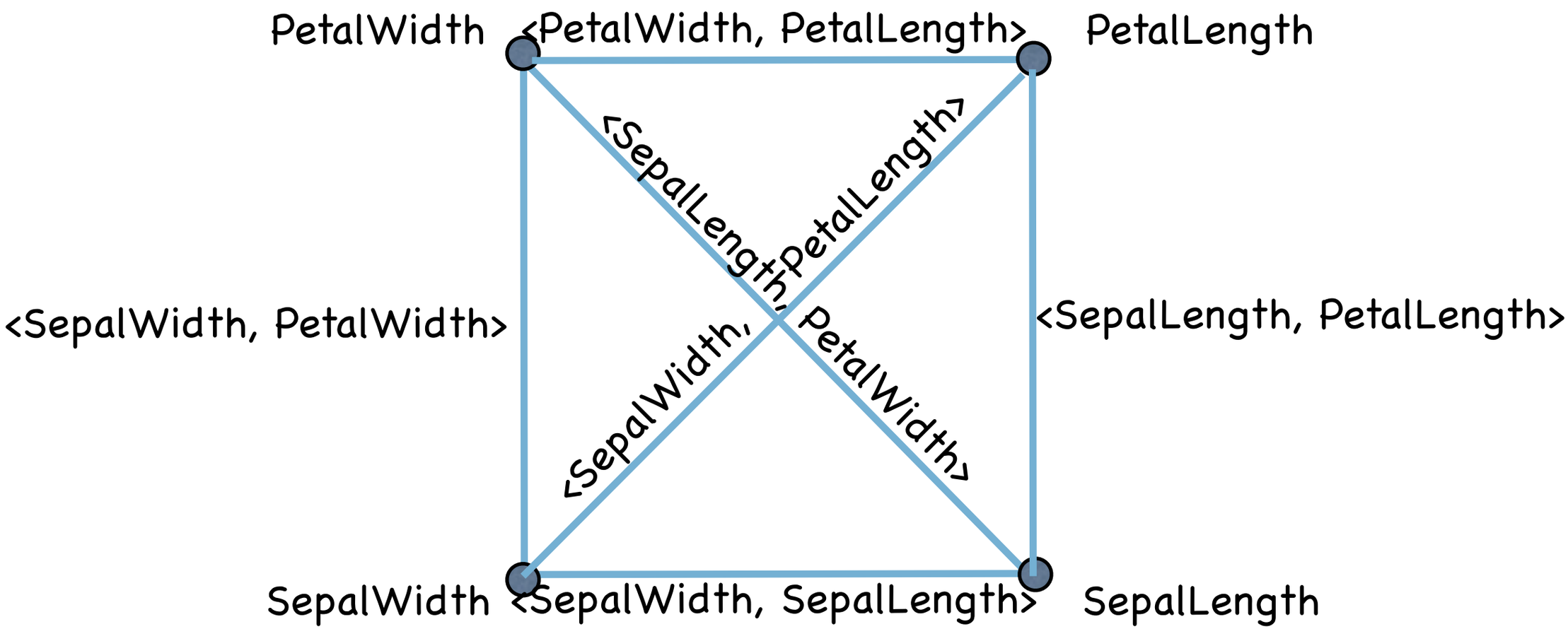
Graph construction

Construction: Line graph of the variable graph



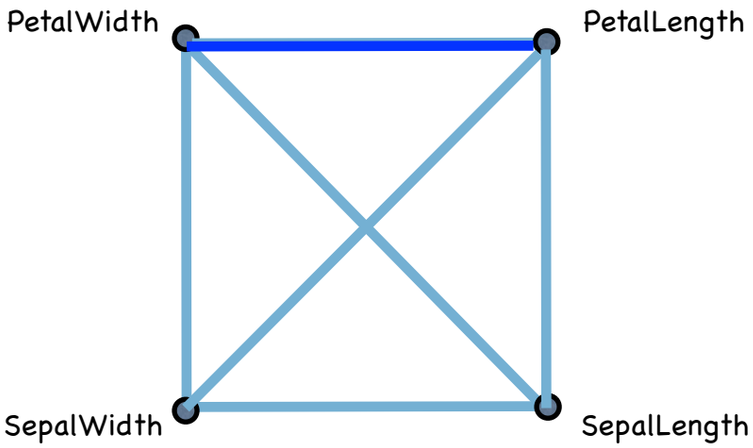
Graph construction

Construction: Line graph of the variable graph



Graph construction

Construction: Line graph of the variable graph



<PetalWidth, PetalLength>

<SepalLength, PetalWidth>

<SepalLength, PetalLength>

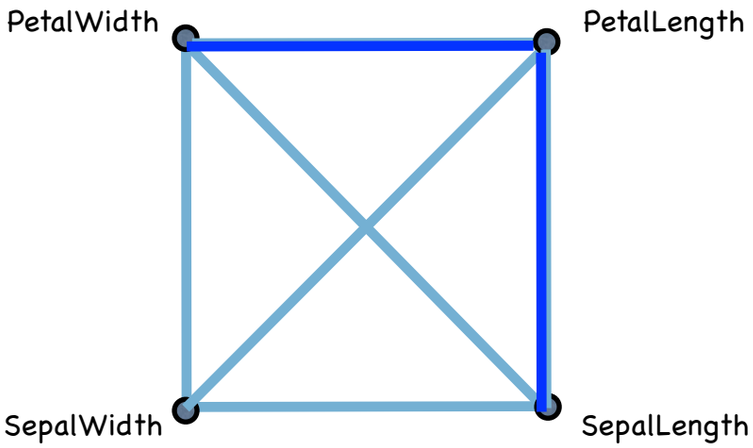
<SepalWidth, PetalWidth>

<SepalWidth, PetalLength>

<SepalWidth, SepalLength>

Graph construction

Construction: Line graph of the variable graph



<PetalWidth, PetalLength>

<SepalLength, PetalWidth>

<SepalLength, PetalLength>

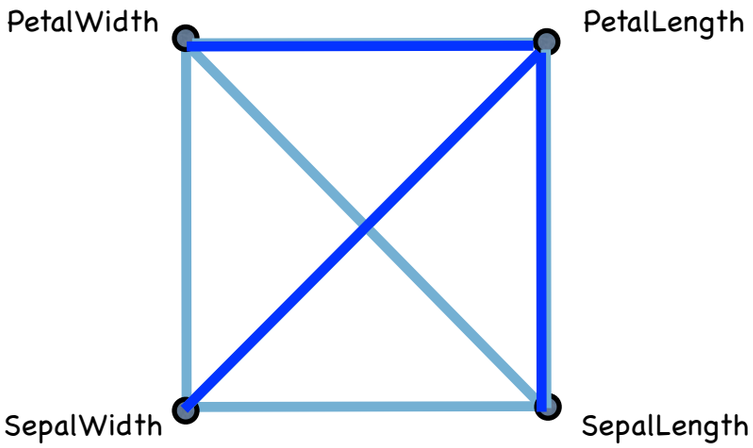
<SepalWidth, PetalWidth>

<SepalWidth, PetalLength>

<SepalWidth, SepalLength>

Graph construction

Construction: Line graph of the variable graph



<PetalWidth, PetalLength>

<SepalLength, PetalWidth>

<SepalLength, PetalLength>

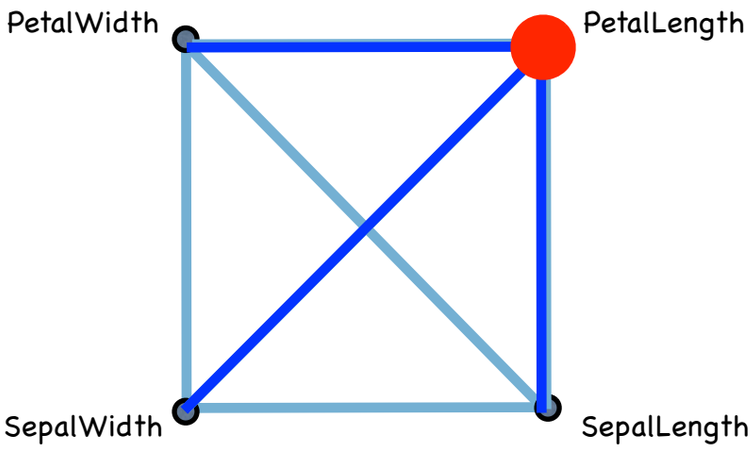
<SepalWidth, PetalWidth>

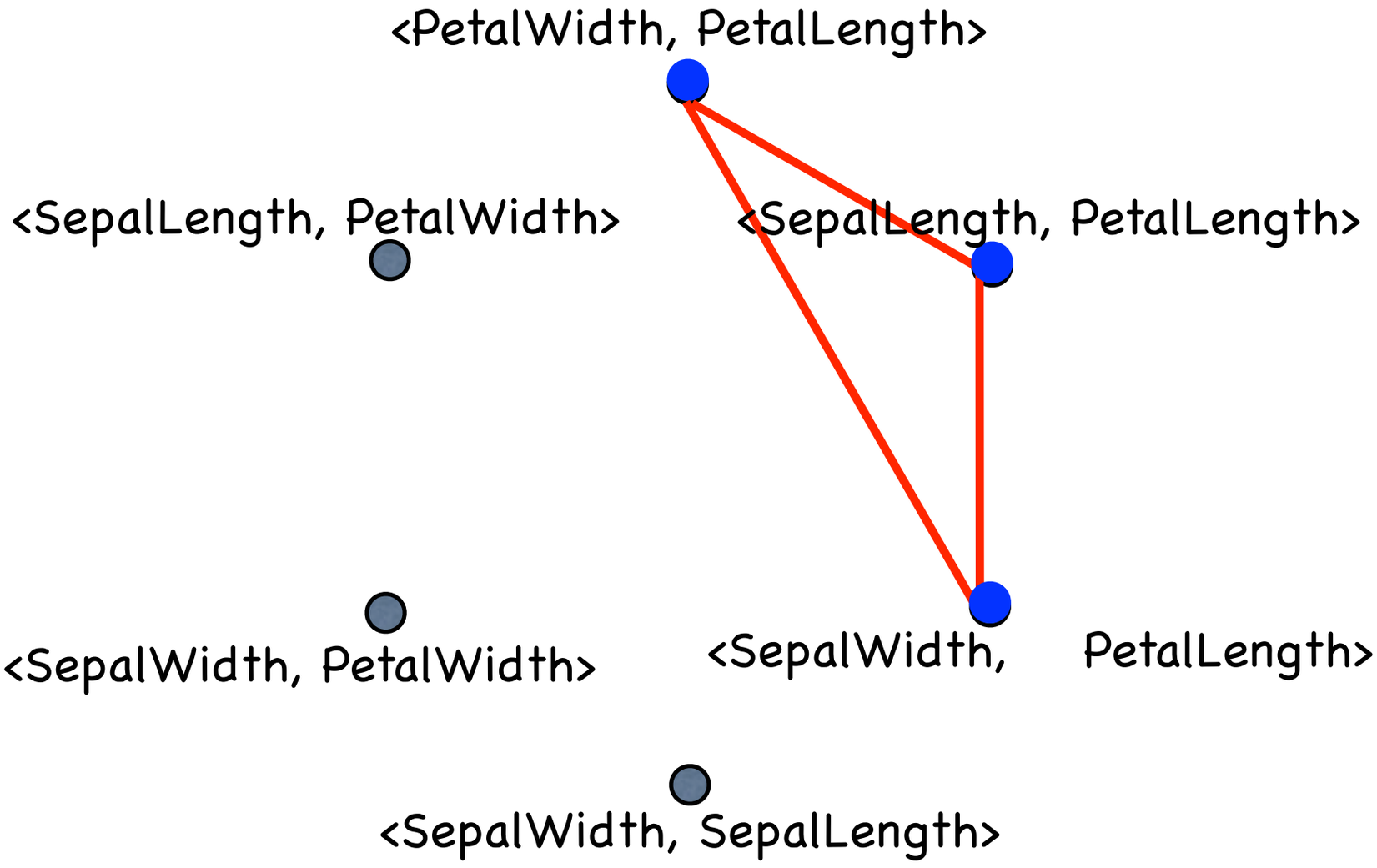
<SepalWidth, PetalLength>

<SepalWidth, SepalLength>

Graph construction

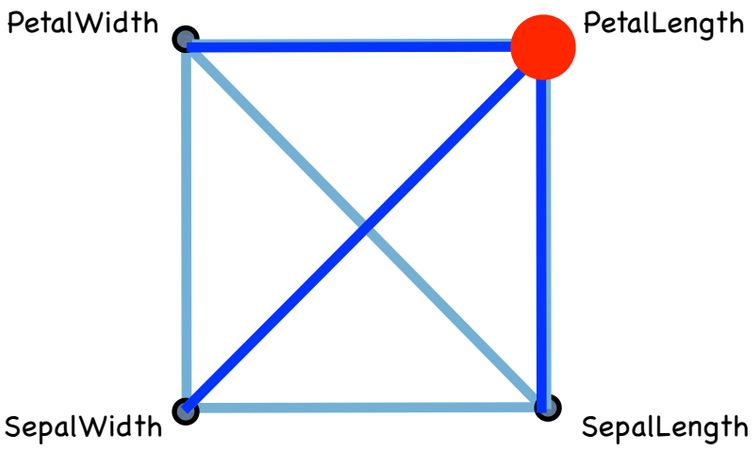
Construction: Line graph of the variable graph

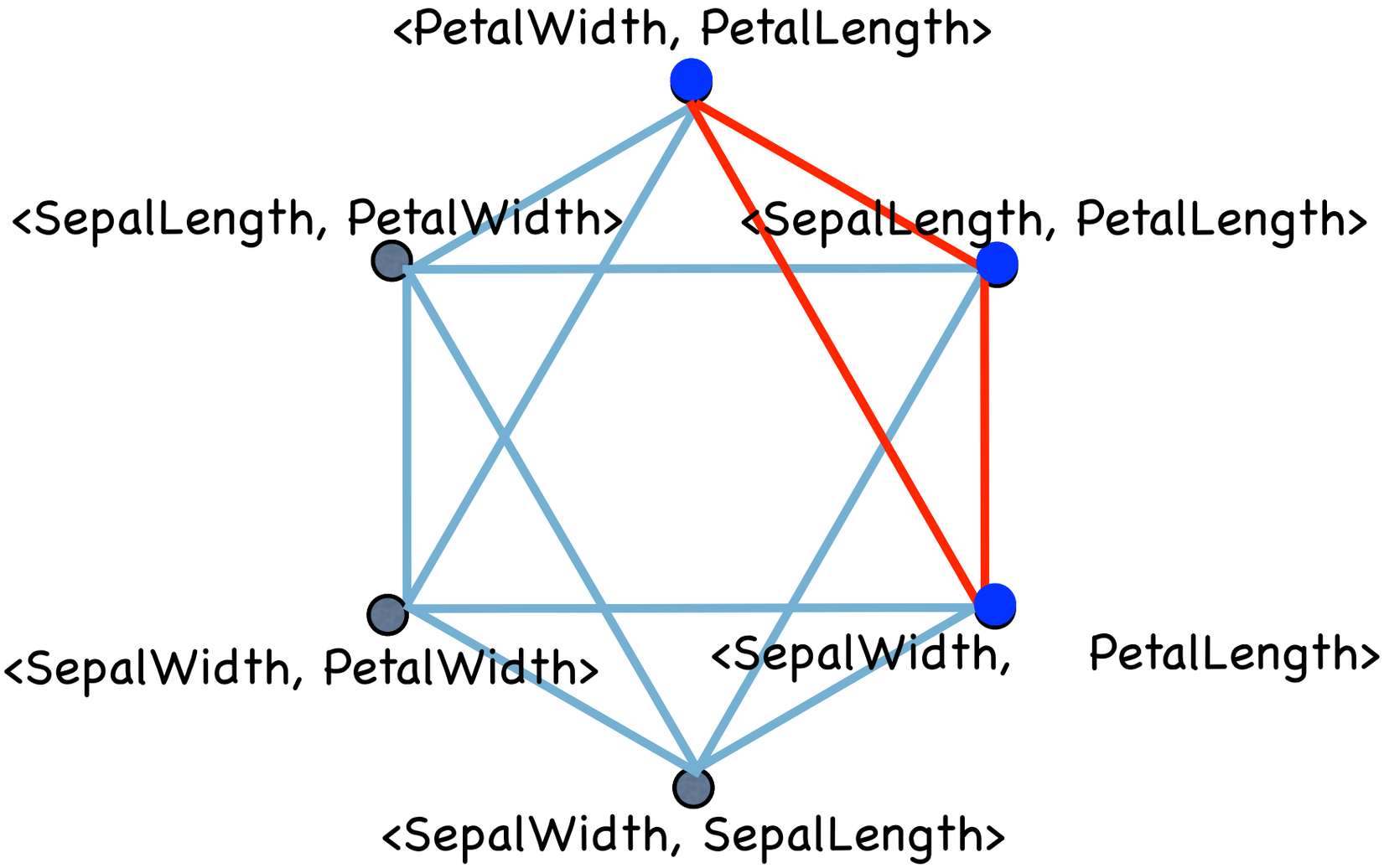




Graph construction

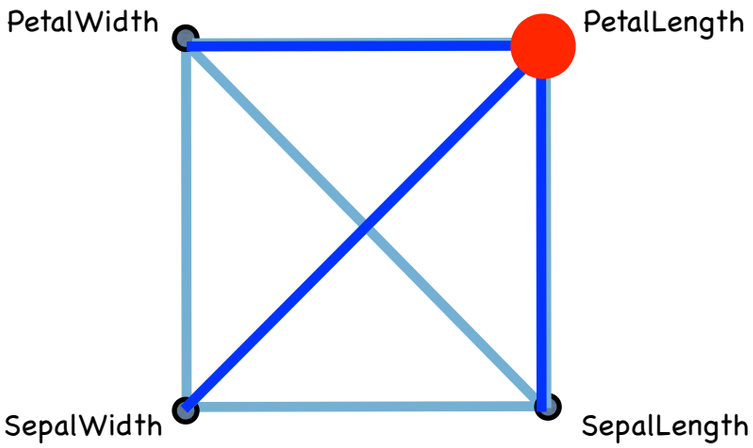
Construction: Line graph of the variable graph



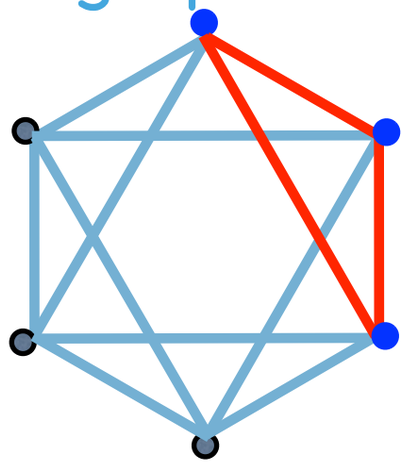


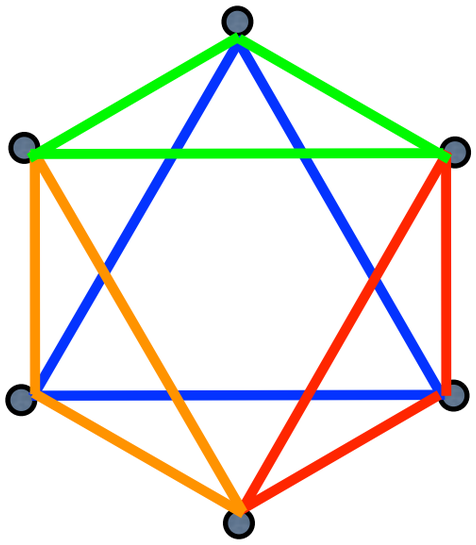
Graph construction

Construction: Line graph of the variable graph



variables complete graph
<--> line graph
<--> 3D transition graph

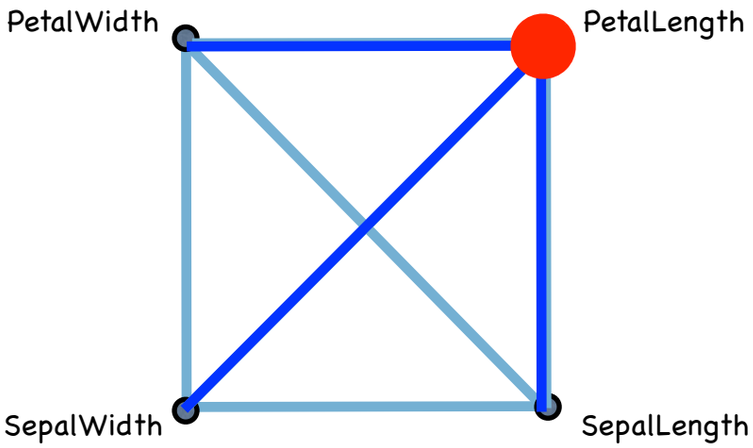




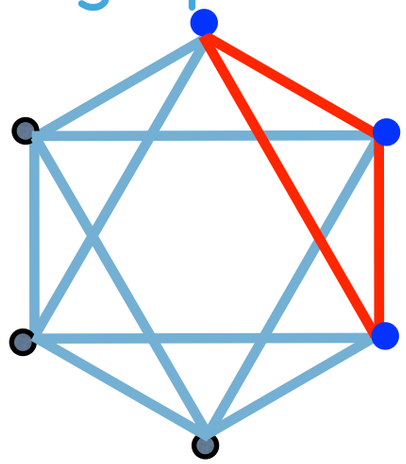
<PetalWidth, PetalLength, SepalLength>
<PetalWidth, PetalLength, SepalWidth>
<SepalLength, SepalWidth, PetalWidth>
<SepalLength, SepalWidth, PetalLength>

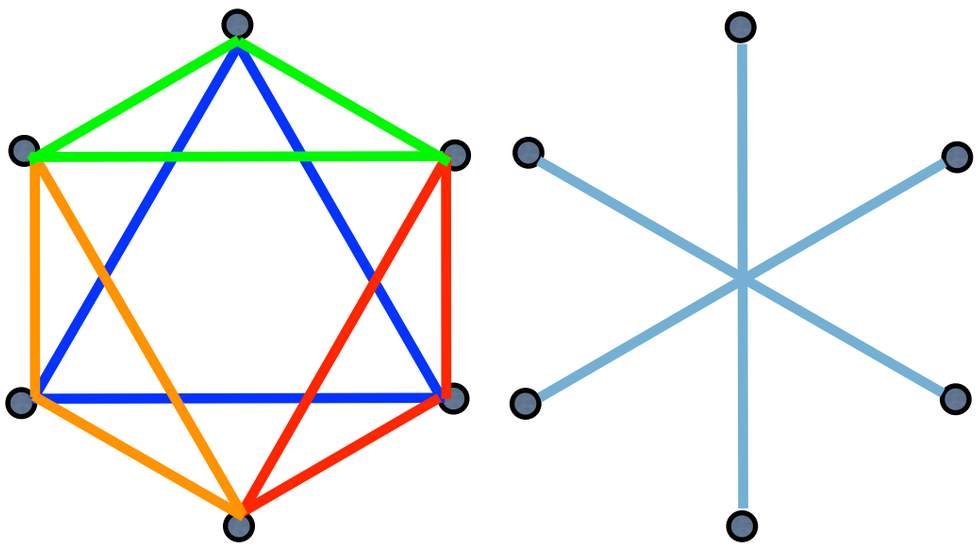
Graph construction

Construction: Line graph of the variable graph



variables complete graph
<--> line graph
<--> 3D transition graph



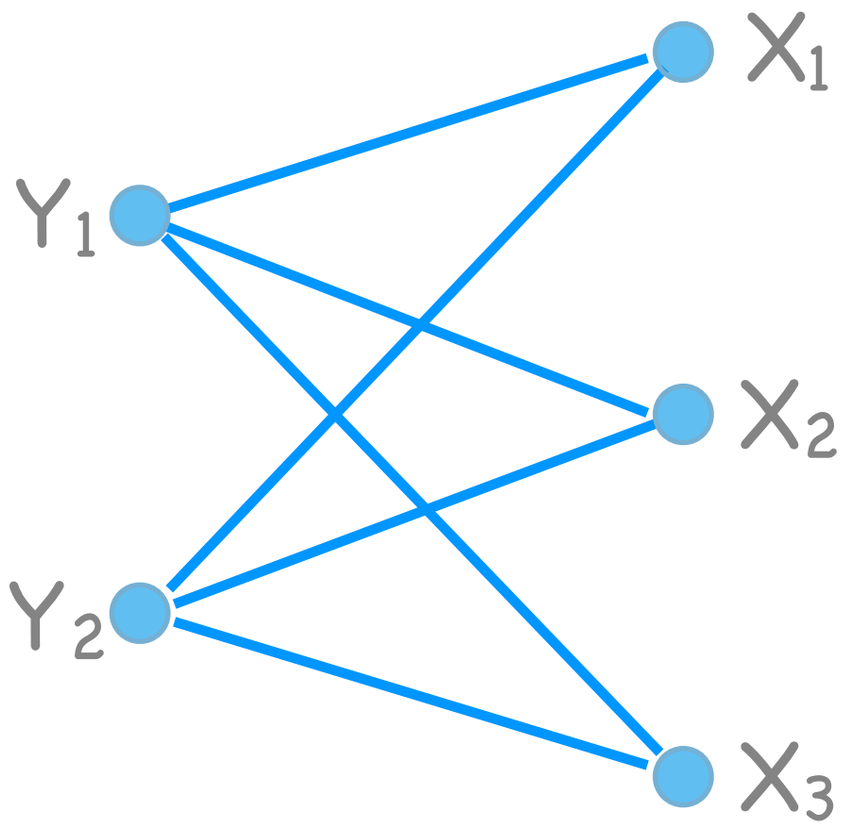


Complement(Line graph) = 4D transition graph

<PetalWidth, PetalLength, SepalLength>
<PetalWidth, PetalLength, SepalWidth>
<SepalLength, SepalWidth, PetalWidth>
<SepalLength, SepalWidth, PetalLength>

Graph construction

- From any variable graph G=(V,E)

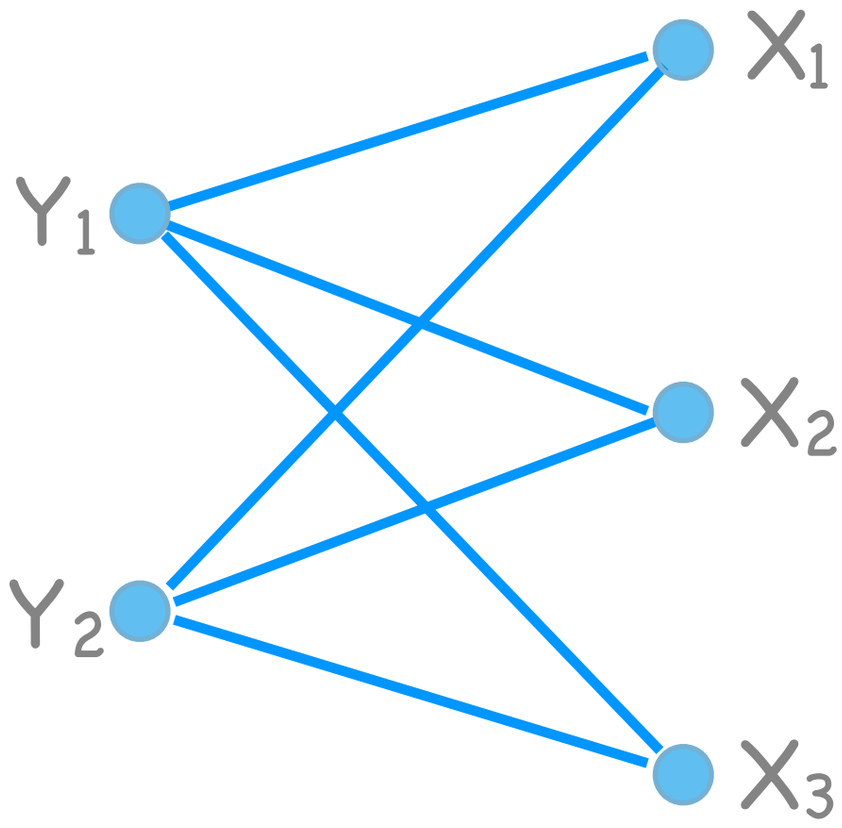


Variable graph

G

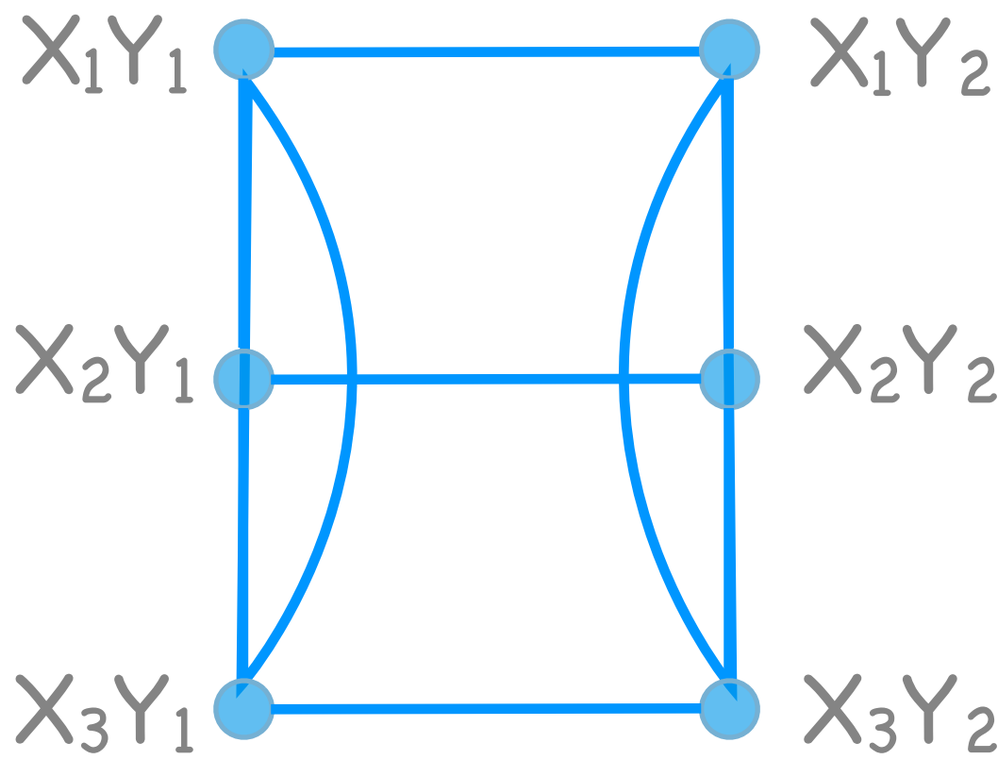
Graph construction

- From any variable graph G=(V,E)



Variable graph

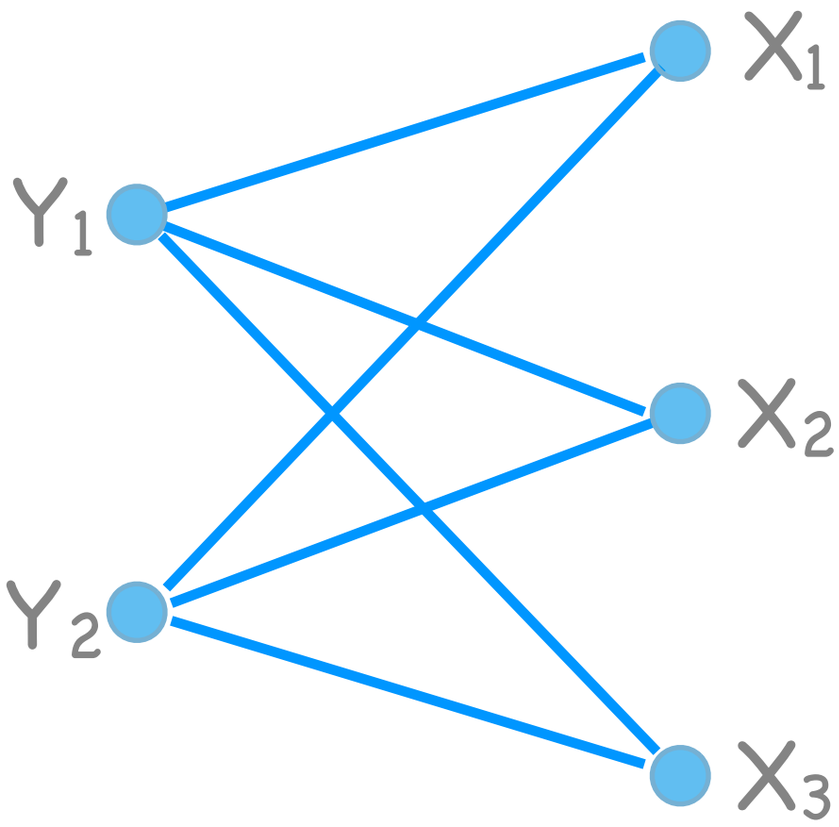
G



3d transition graph $L(G)$

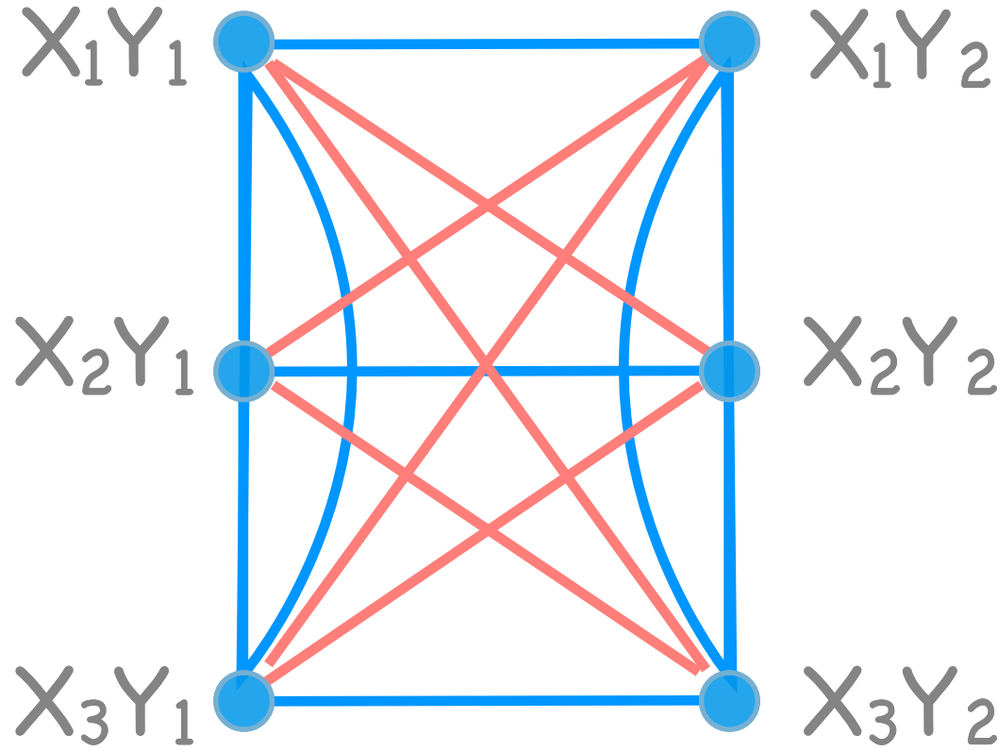
Graph construction

- From any variable graph G=(V,E)



Variable graph

G

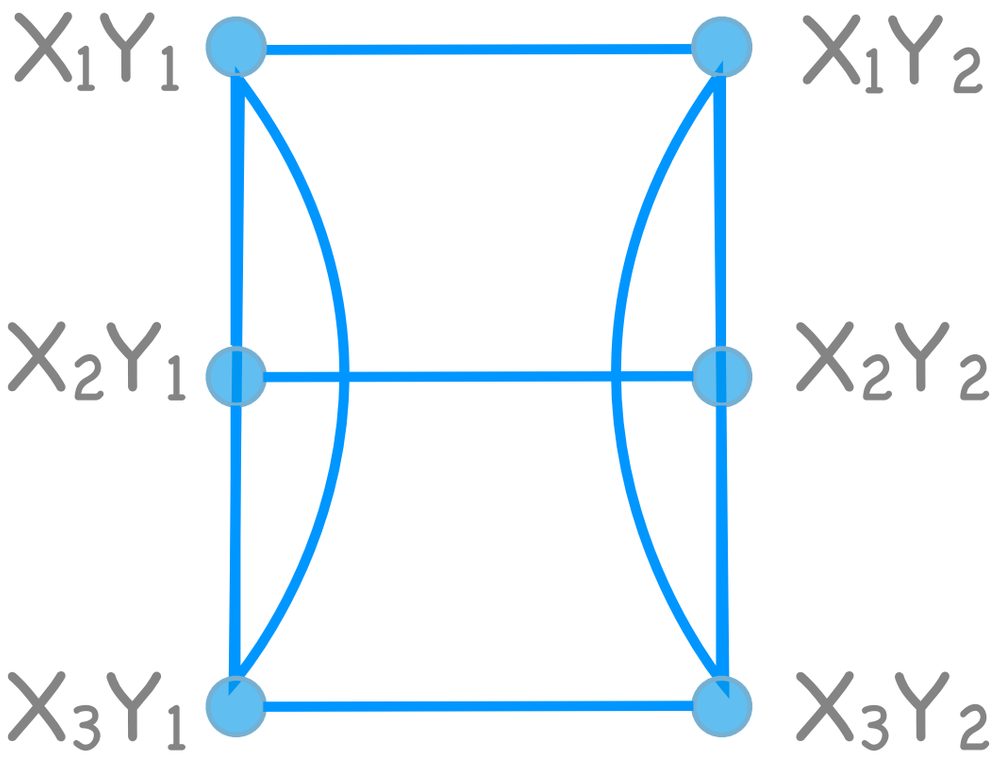


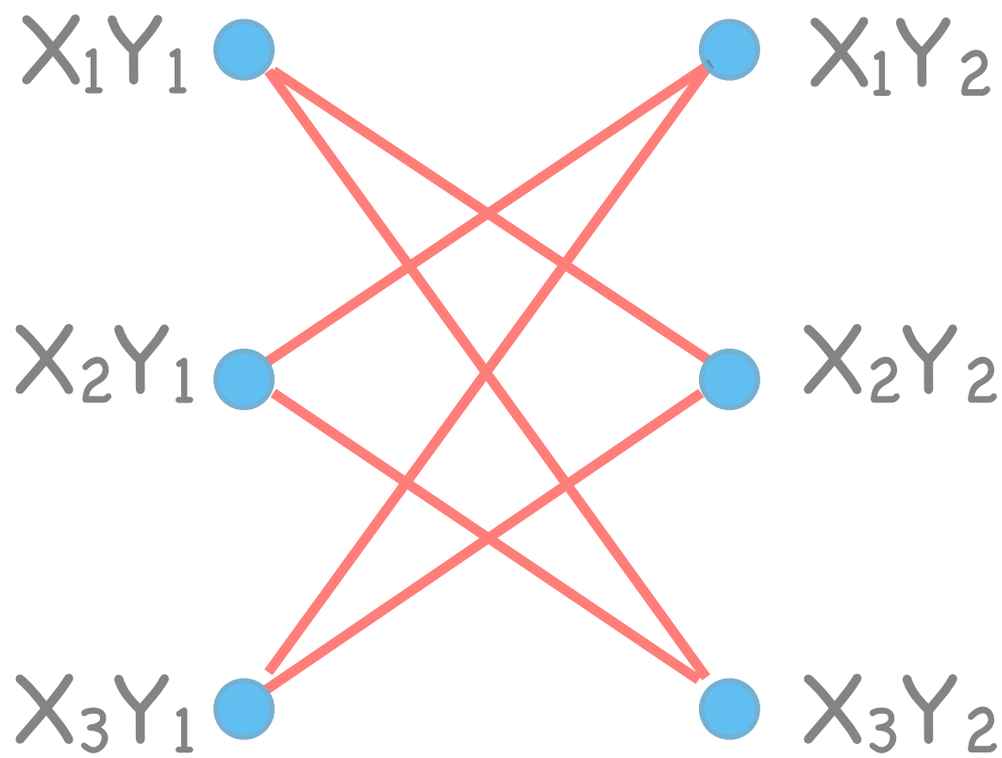
3d transition graph $L(G)$

4d transition graph $\overline{L(G)}$

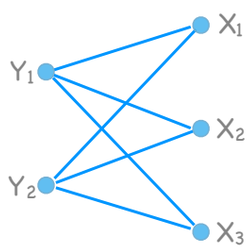
Graph construction

- From any variable graph G=(V,E)





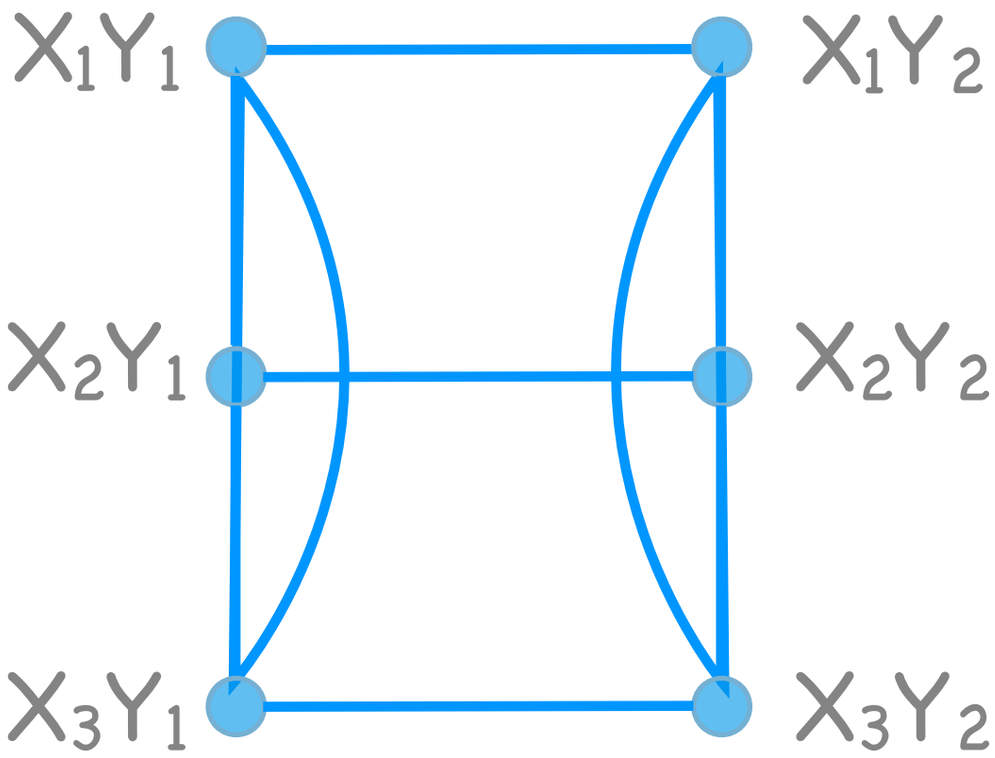
3d transition graph $L(G)$

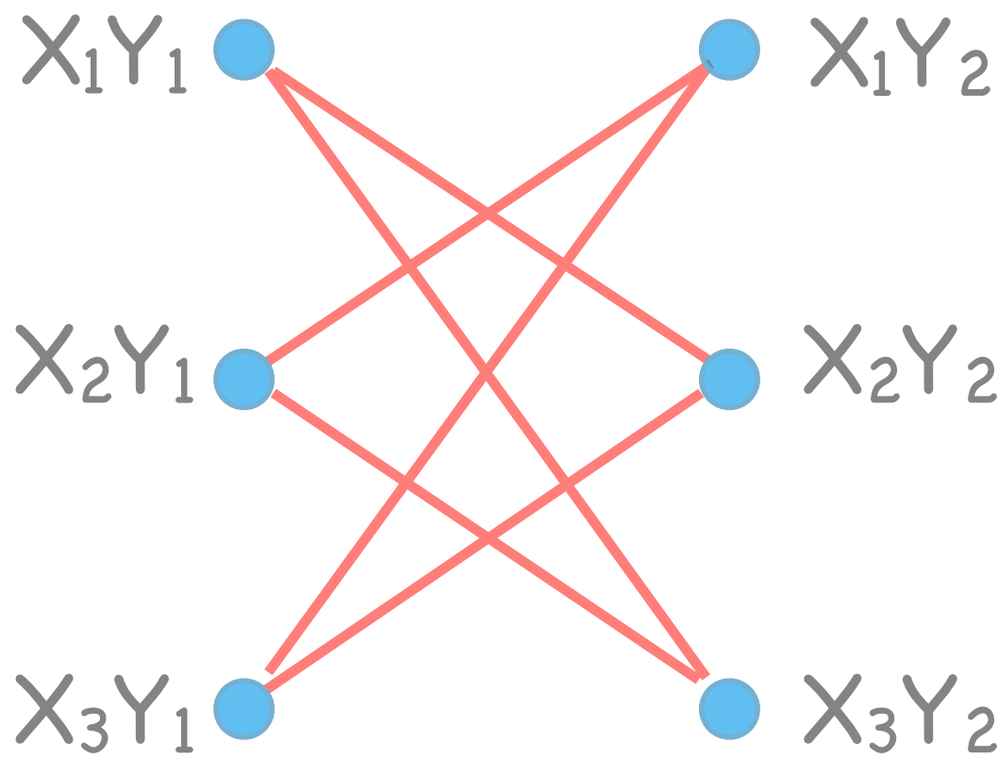


4d transition graph $\overline{L(G)}$

Graph construction

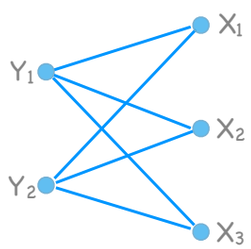
- From any variable graph G=(V,E)





3d transition graph $L(G)$

4d transition graph $\overline{L(G)}$



<- Put edges ONLY on interesting pairs

Graph structures

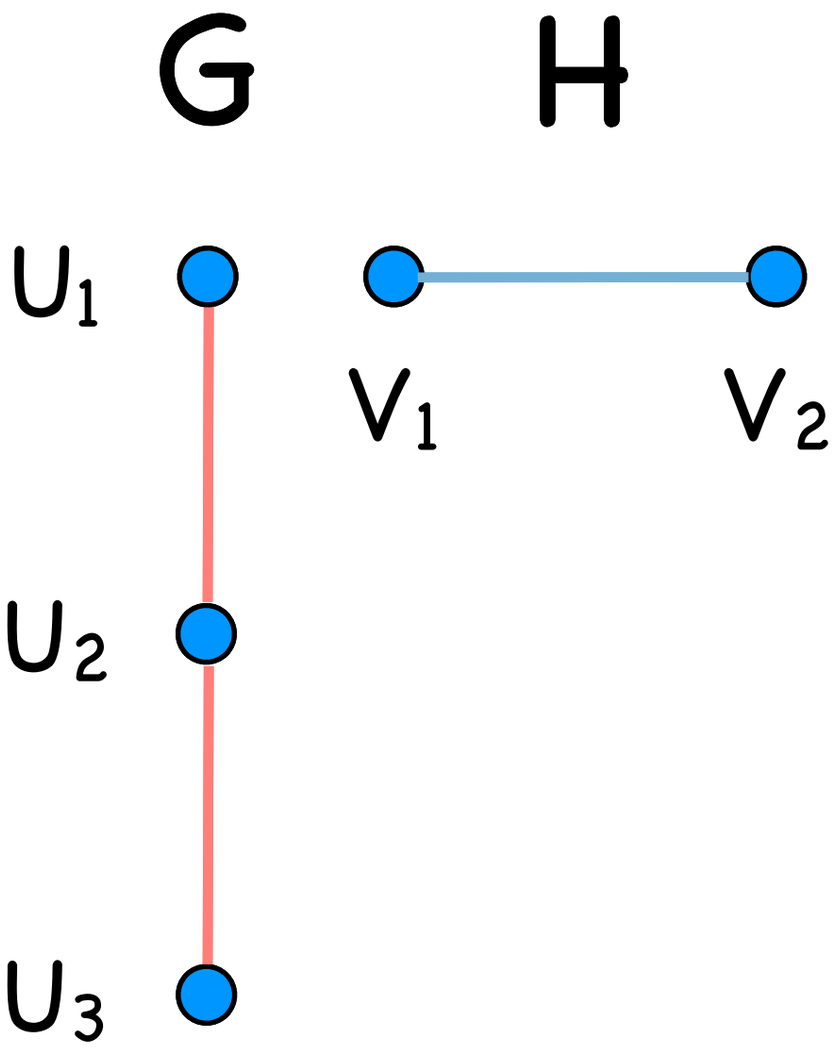
- Graph construction is actually general
  - start with **any** graph G on the variables
  - its line graph $L(G)$ **will** be a 3D-transition graph
  - the complement of the line graph $\overline{L(G)}$ **will** be a 4D-transition graph

- Might start with a variable graph that connects variables only if they have and interesting relation.

Graph Products

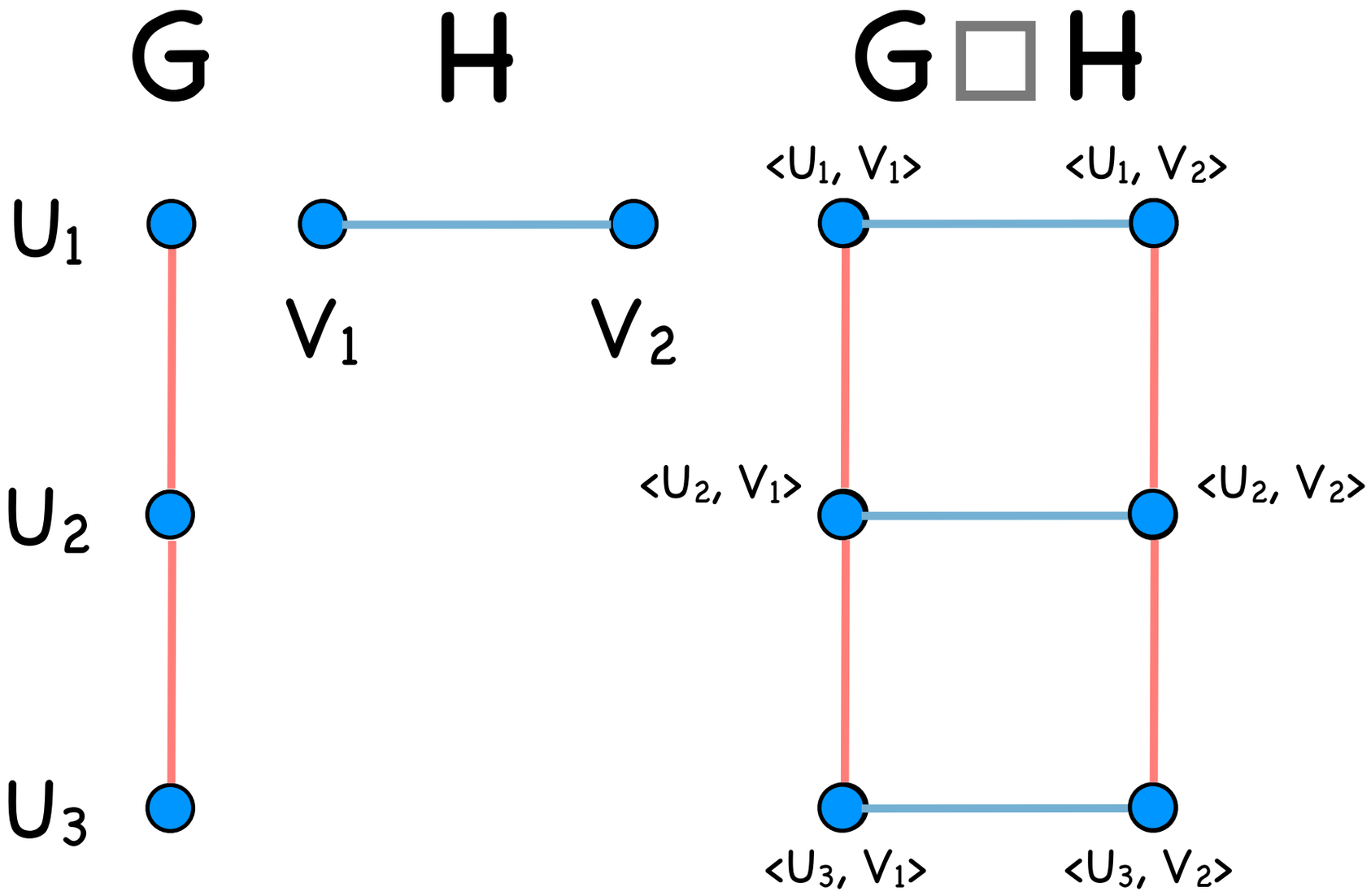
- Another general construction: graph products

Graph Products



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

Graph Products

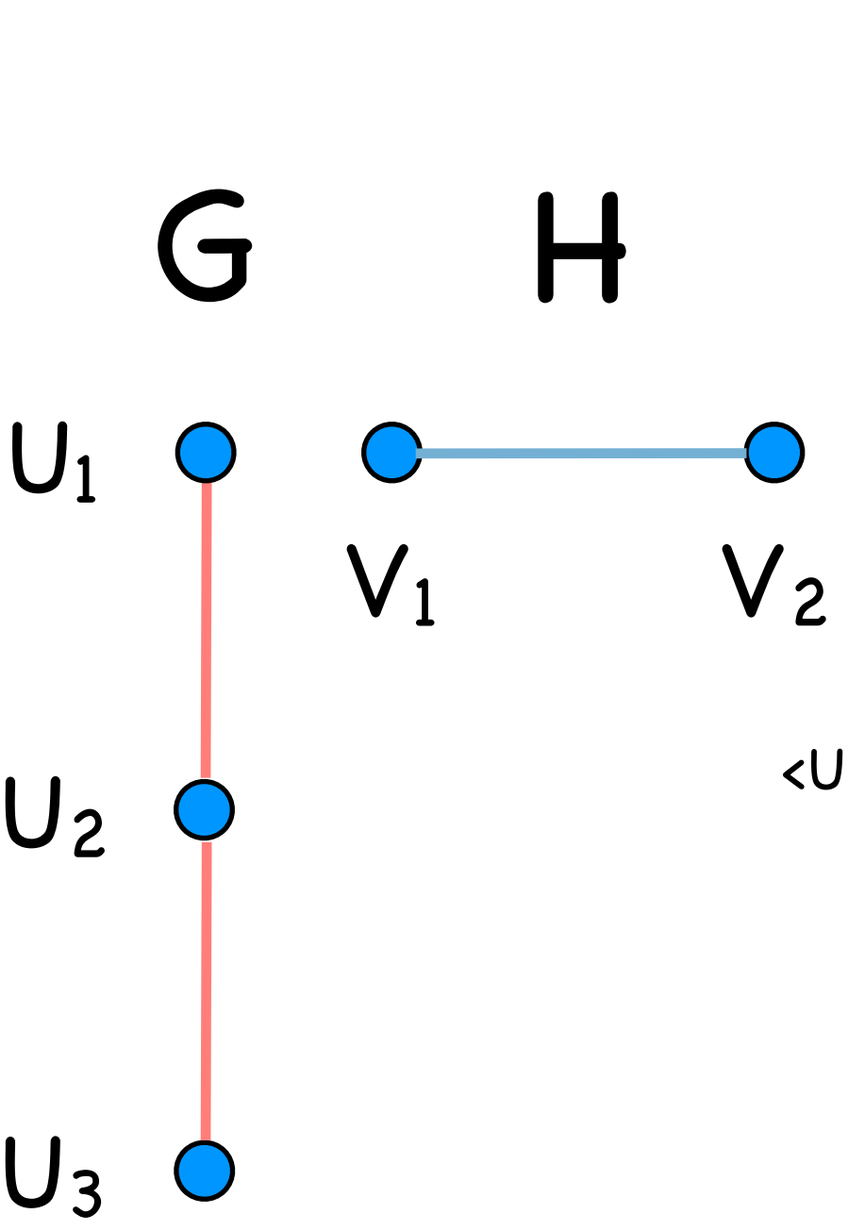


E.g.
explanatory U (or Xs),
responses V (or Ys)

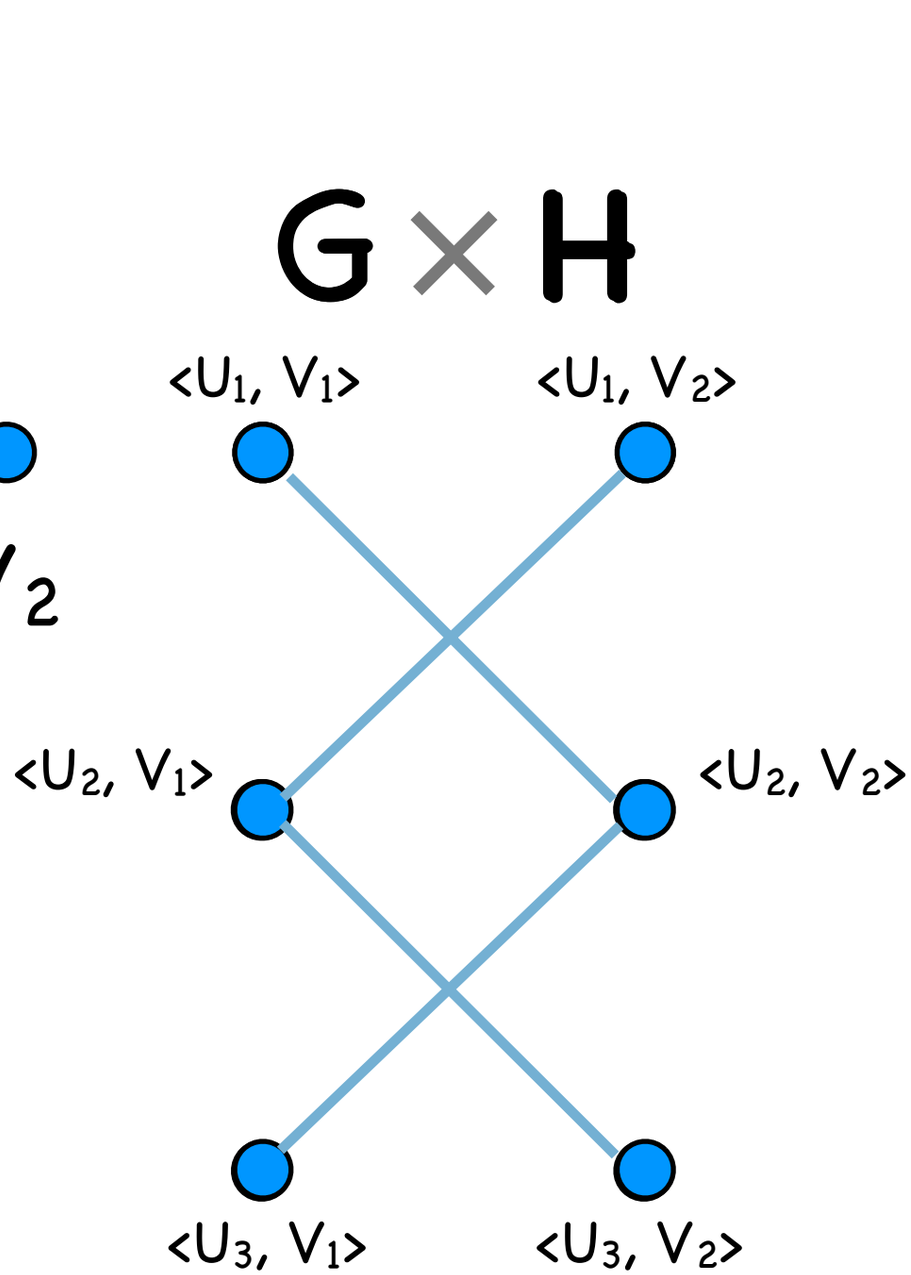
Cartesian product

3D transition graph

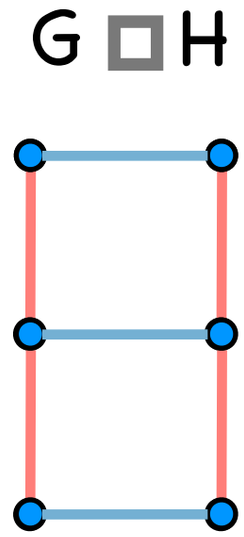
Graph Products



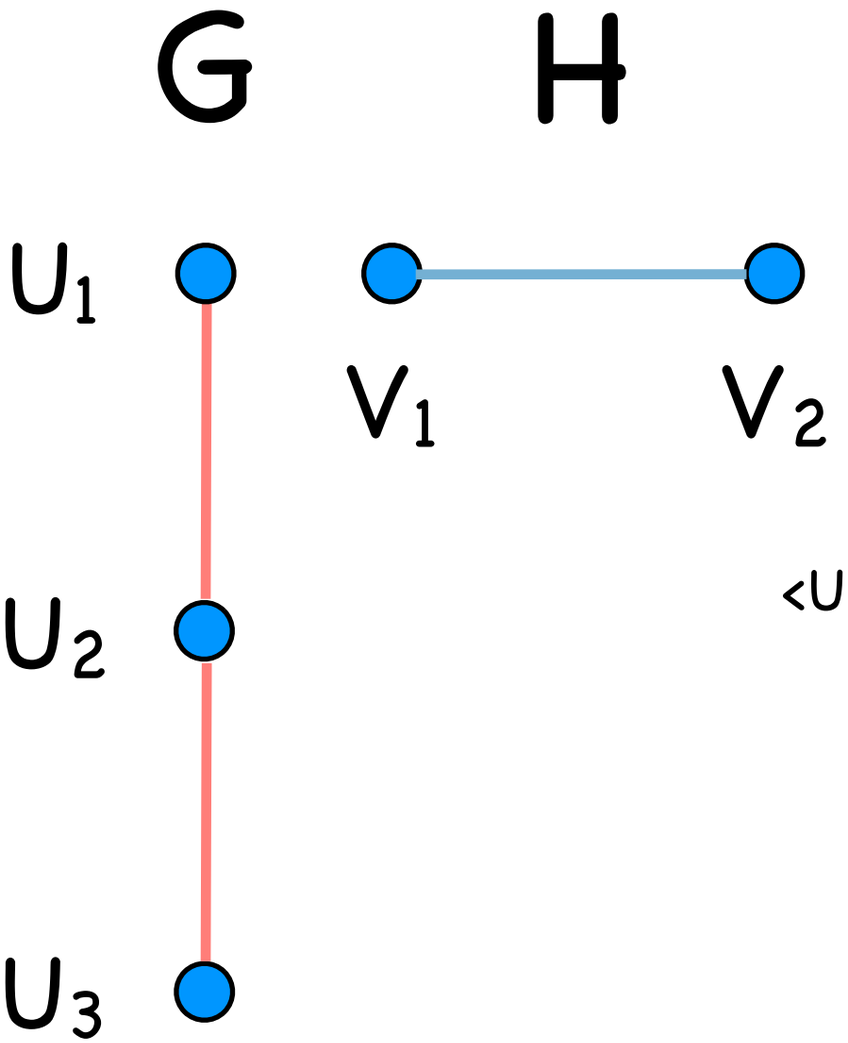
E.g.
explanatory U (or Xs),
responses V (or Ys)

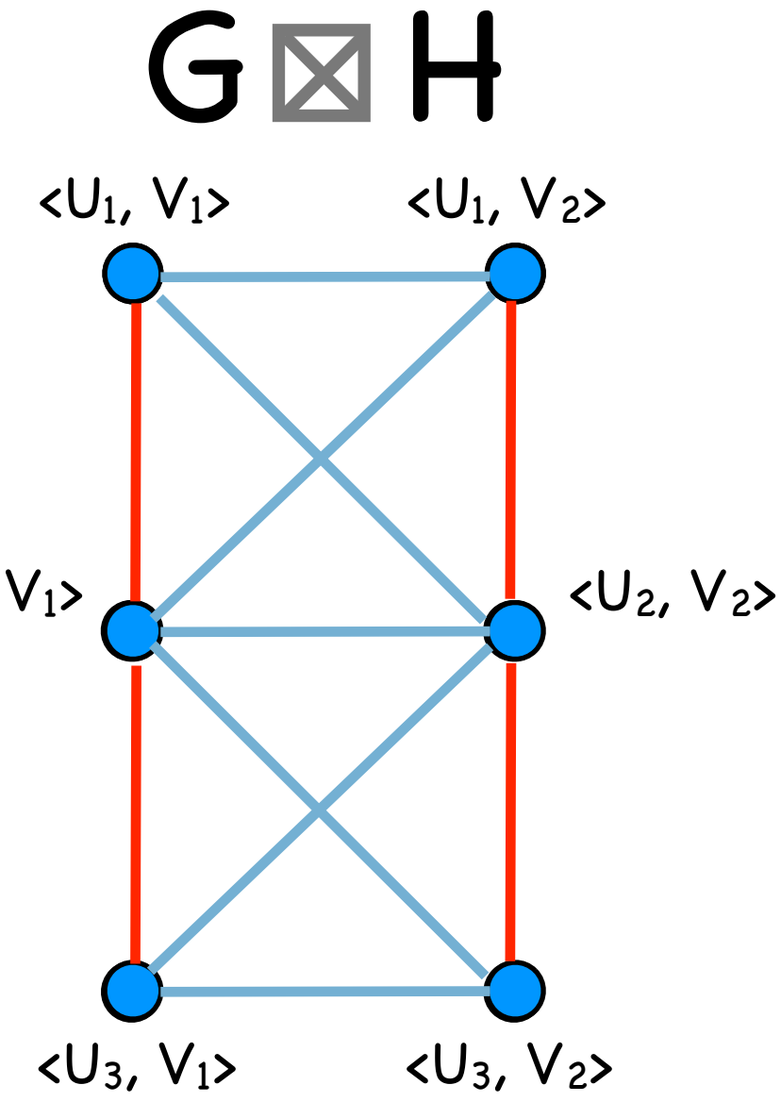


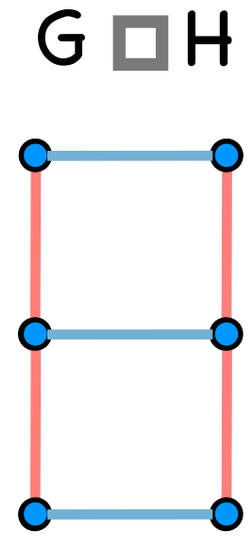
Tensor product
4D transition graph

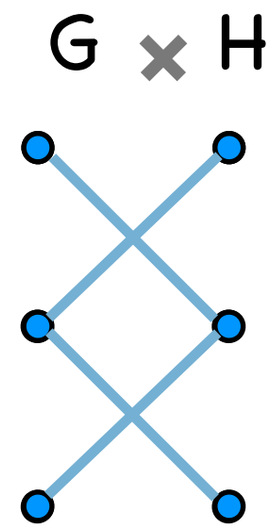


Graph Products









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

Strong product

All are symmetric

Conclusion

- Graphs provide some navigational infrastructure
  - Formal structure of graphs → new viz tools, maps, routes
  - Special structures exist and can be exploited (3D transitions, 4D transitions, Hamiltonians, Eulerians, graph products, ...)
    - graph algorithms, ...
- Need new measures of interest ...evaluate routes
- New applications: images, documents, etc
- Low d spaces are natural and informative.

Thank you

Thank you

御清聴ありがとうございました

Questions?

質問はありますか？

Papers

Hurley & Oldford:

- Graphs as navigational infrastructure for high dimensional data spaces (**Comp Stats 2011**).

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