Towards Finding Relational Redescriptions

Esther Galbrun & Angelika Kimmig

ILP 2012

[paper to appear at Discovery Science 2012]

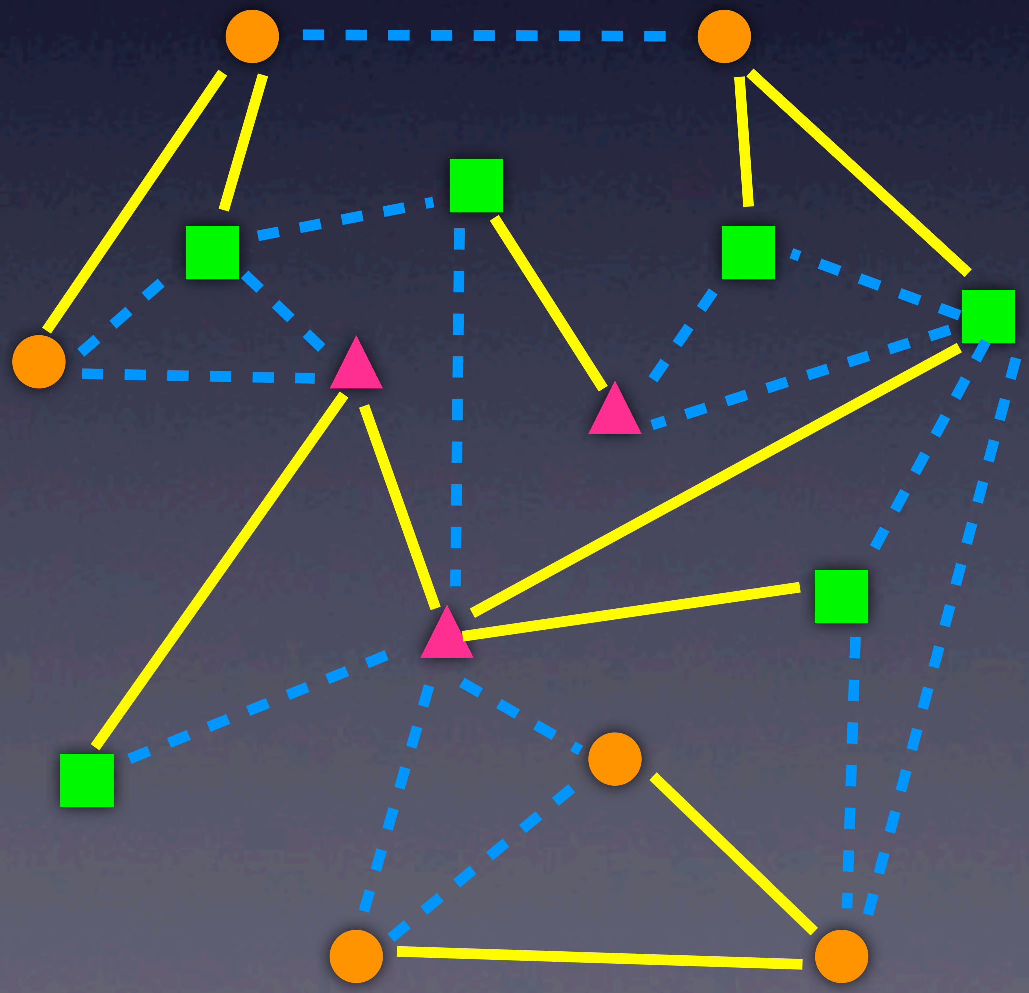
UNIVERSITY OF HELSINKI

KATHOLIEKE UNIVERSITEIT LEUVEN

UNIVERSITY OF MARYLAND

Relational Redescription Mining (RRM)

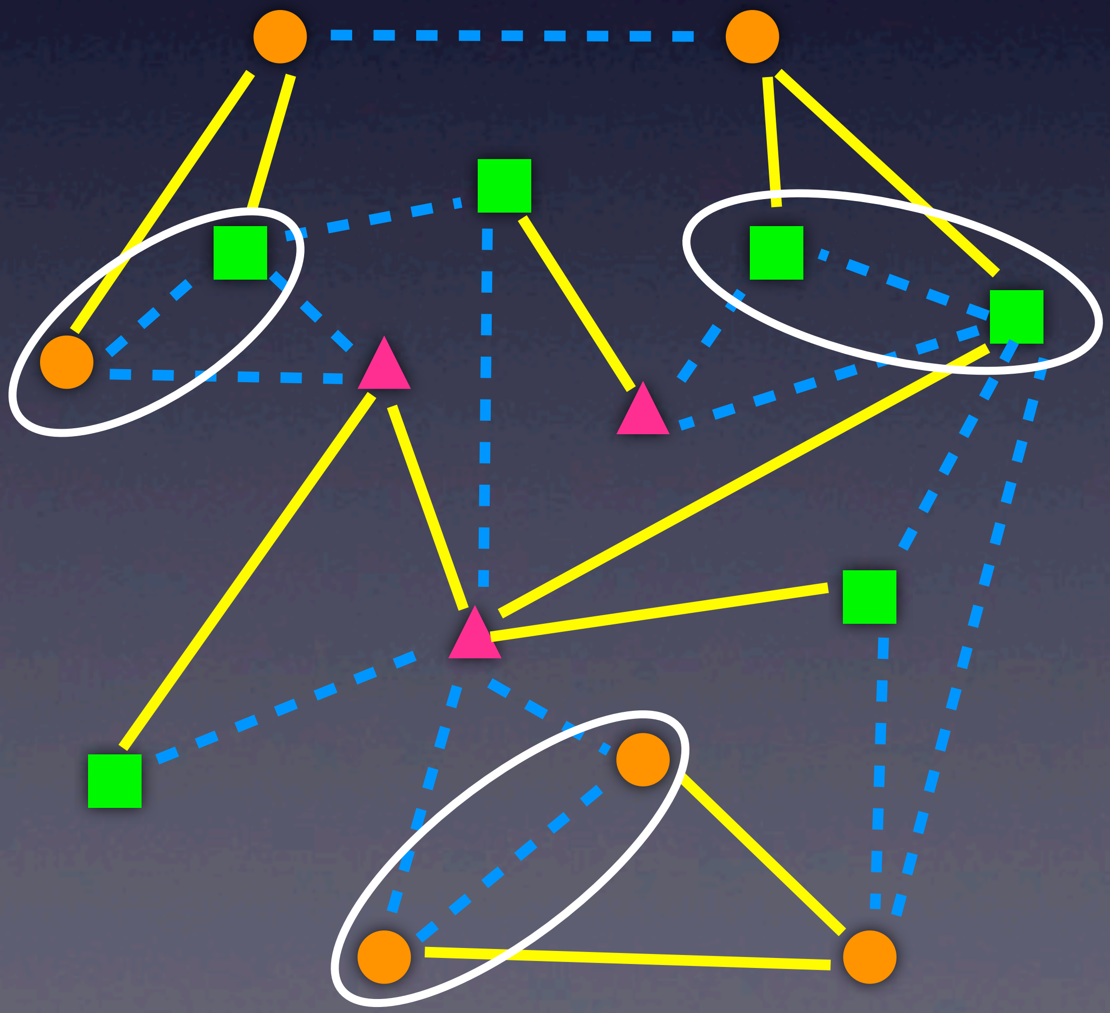
Given: data network

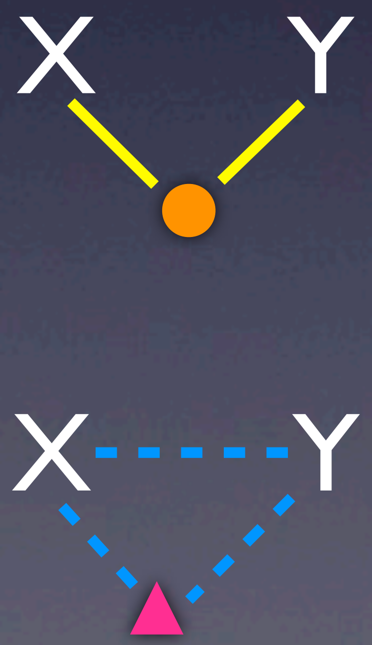


Relational Redescription Mining (RRM)

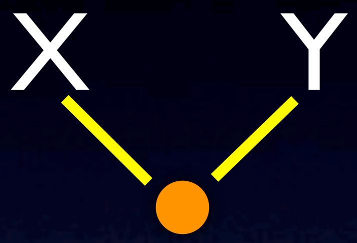
Given: data network

Find: structurally different patterns covering (nearly) same set of node tuples



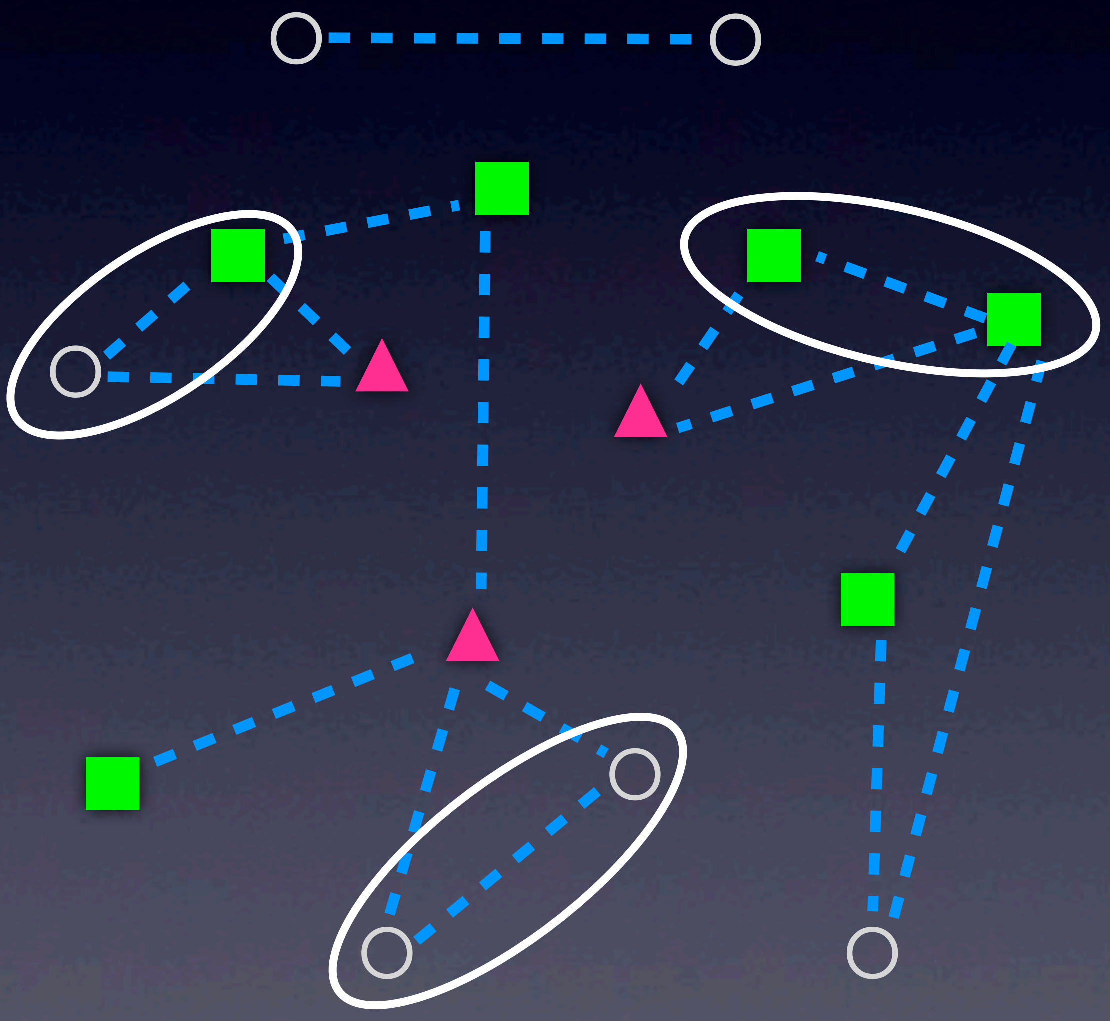


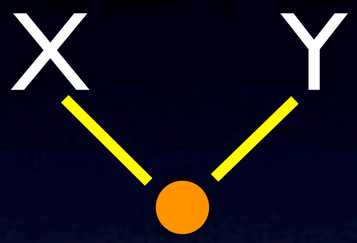
RRM Alternating Scheme



1. Fix first pattern to create examples

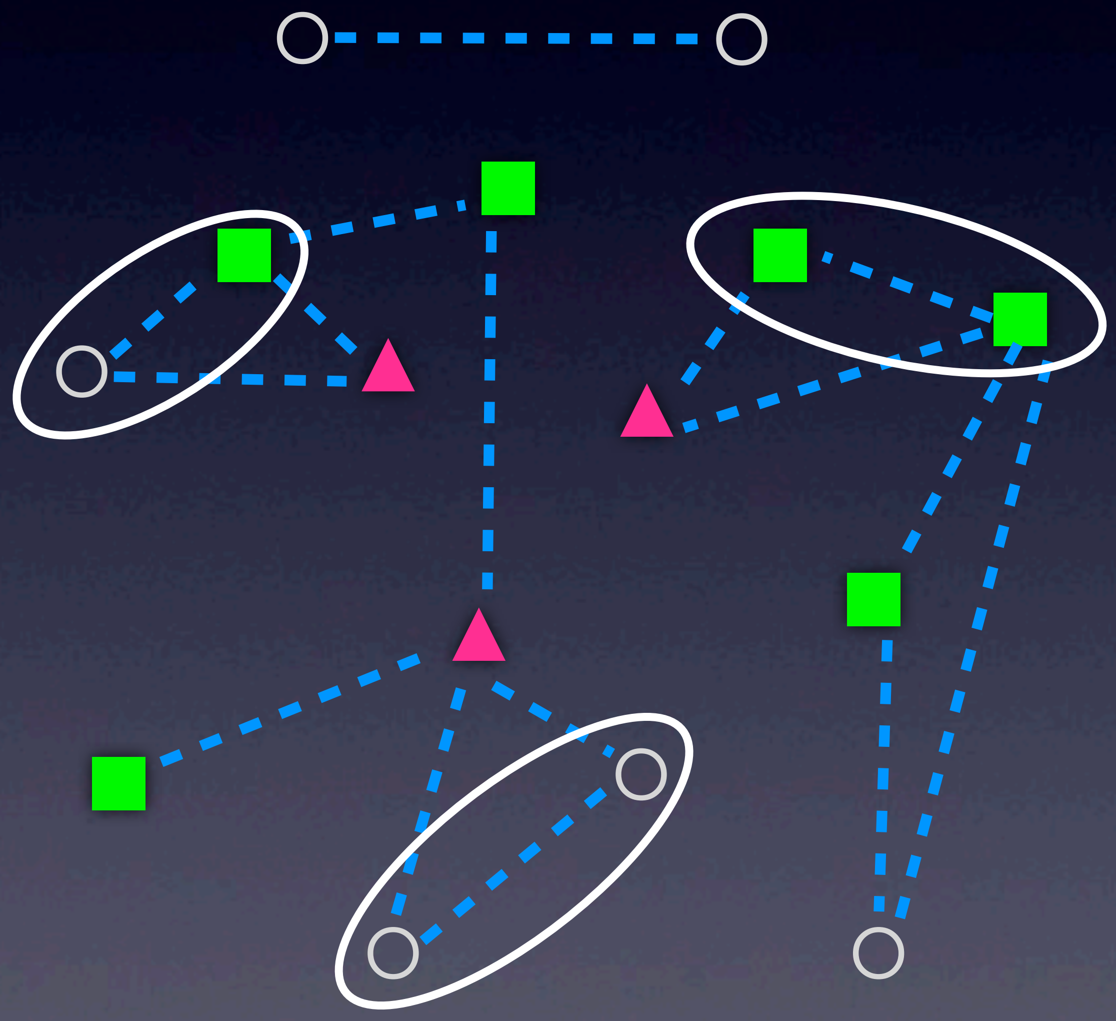
RRM Alternating Scheme

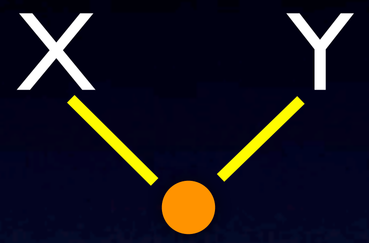




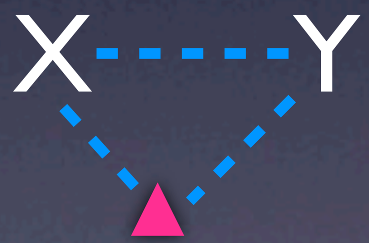
1. Fix first pattern to create examples
2. Consider predicates not used by first pattern

RRM Alternating Scheme

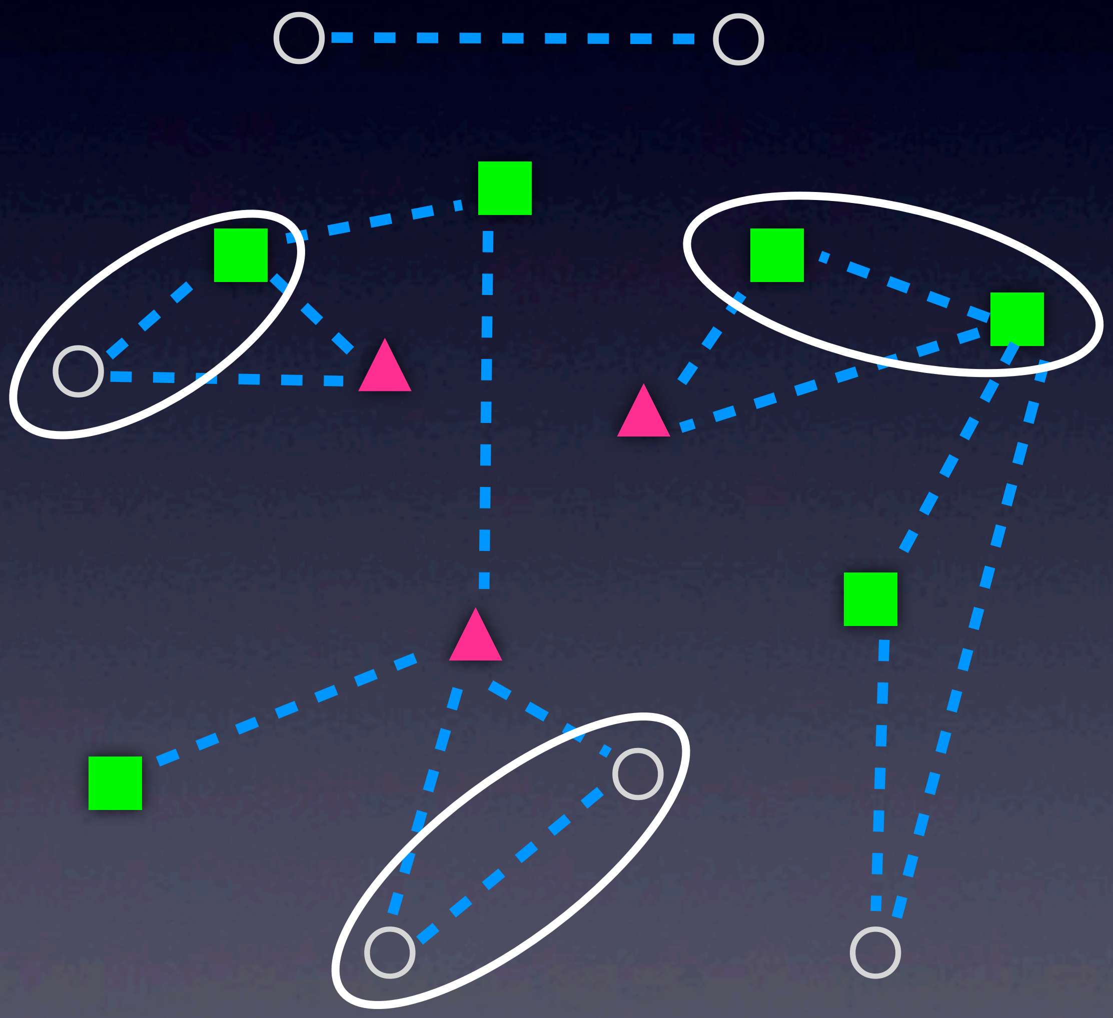


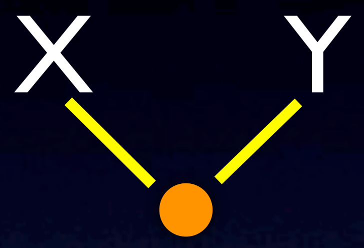


1. Fix first pattern to create examples
2. Consider predicates not used by first pattern
3. Find second pattern

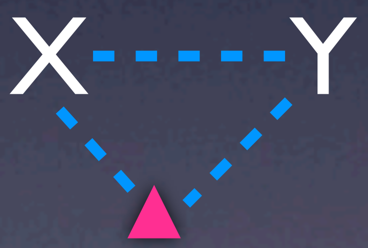


RRM Alternating Scheme



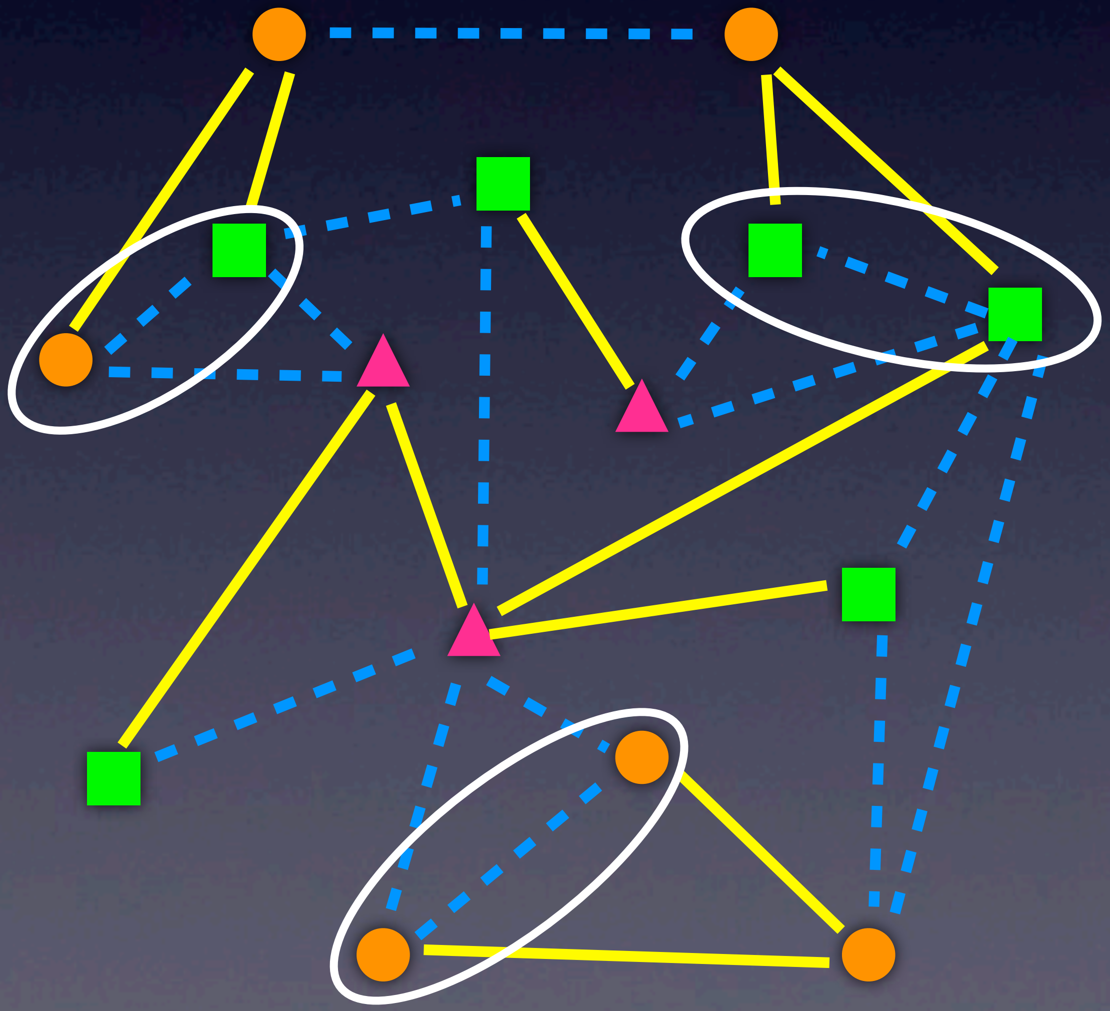


1. Fix first pattern to create examples
2. Consider predicates not used by first pattern
3. Find second pattern
4. Swap roles and iterate



First Step Towards RRM

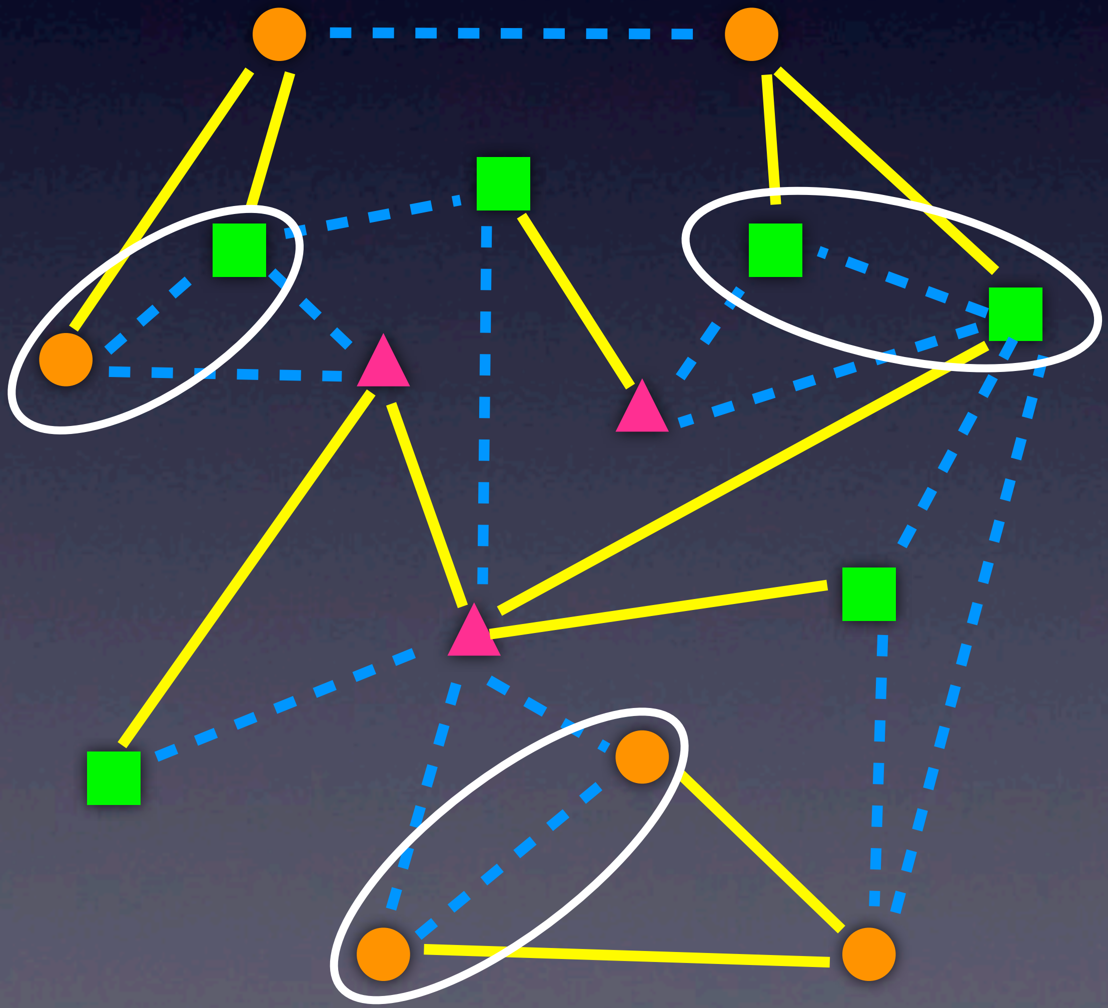
Given: data network &
set of node pairs

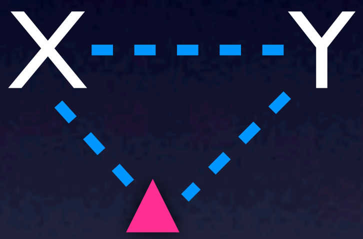


First Step Towards RRM

Given: data network &
set of node pairs

Find: set of patterns
connecting examples

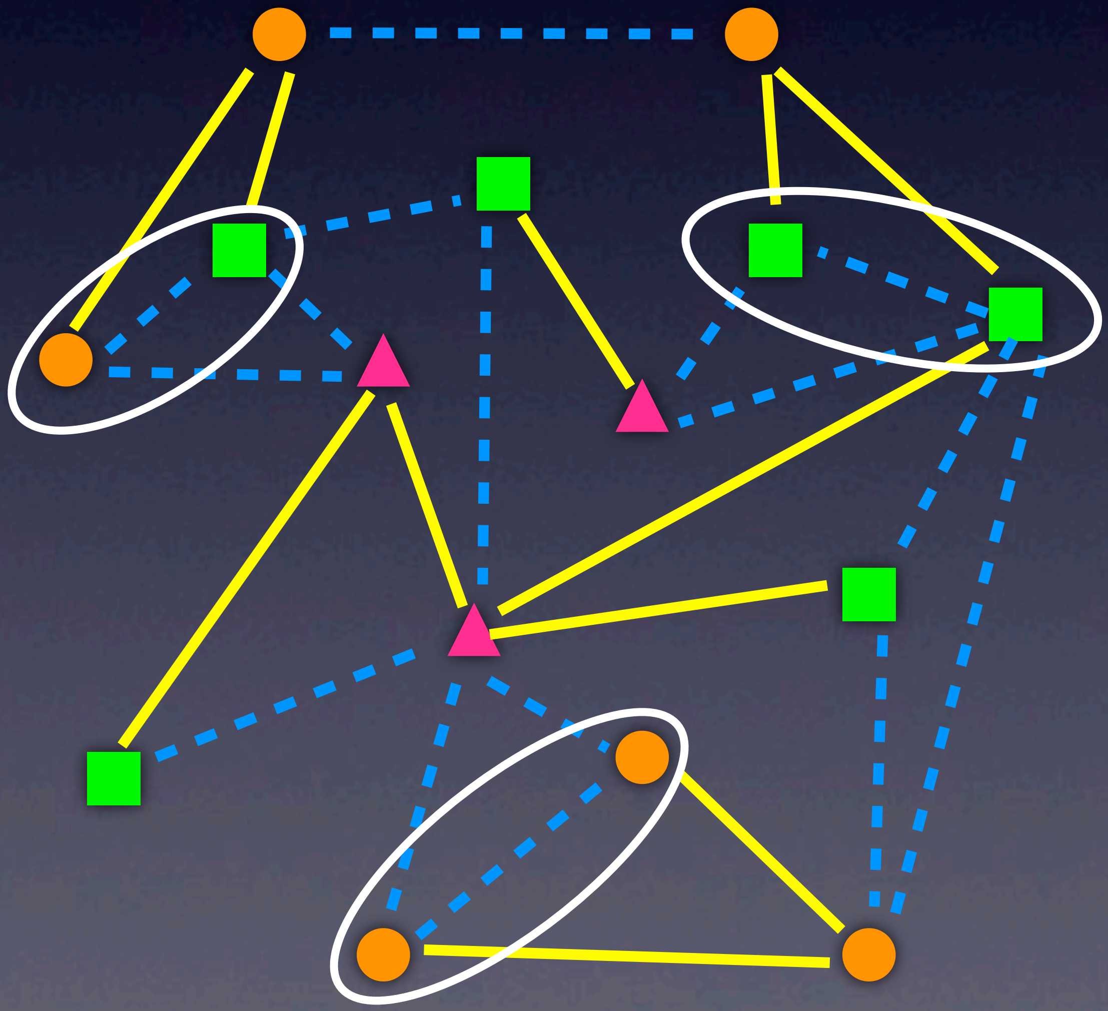


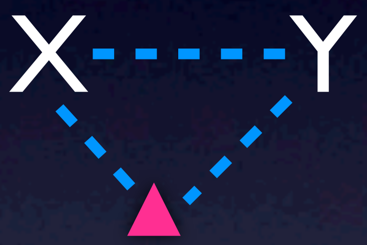


First Step Towards RRM

Given: data network &
set of node pairs

Find: set of patterns
connecting examples





Challenges:

- connected patterns
- complex patterns
- fast mining

Our Approach

1. Extract path descriptions
2. Mine frequent path patterns
3. Build graph patterns from path patterns
4. Select set of graph patterns

Our Approach

1. Extract path descriptions
2. Mine frequent path patterns
3. Build graph patterns from path patterns
4. Select set of graph patterns

constraint-based pattern mining with FIM_CP

Our Approach

1. Extract path descriptions

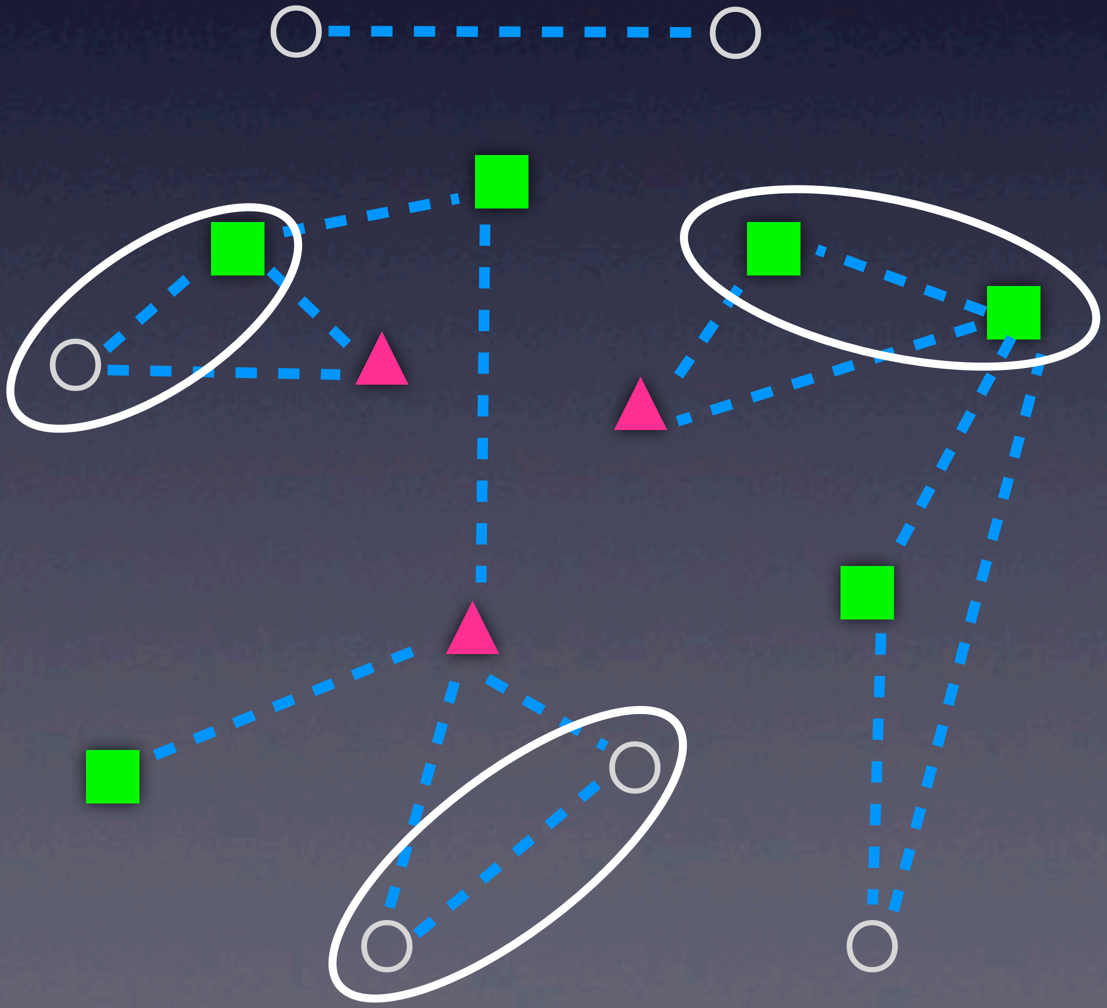
constraint-based pattern mining with FIM_CP

2. Mine frequent path patterns
3. Build graph patterns from path patterns
4. Select set of graph patterns

interleaved

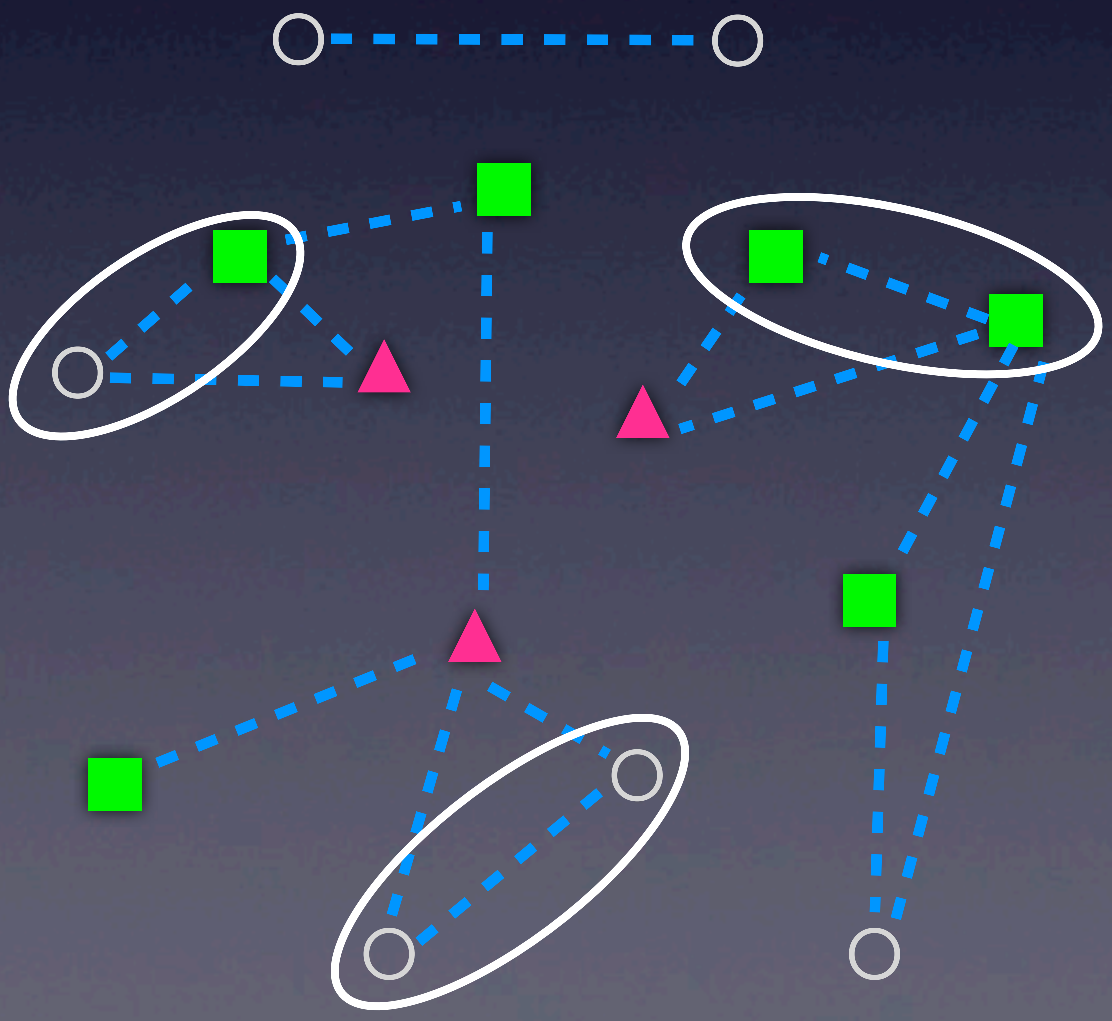
1. Extract path descriptions

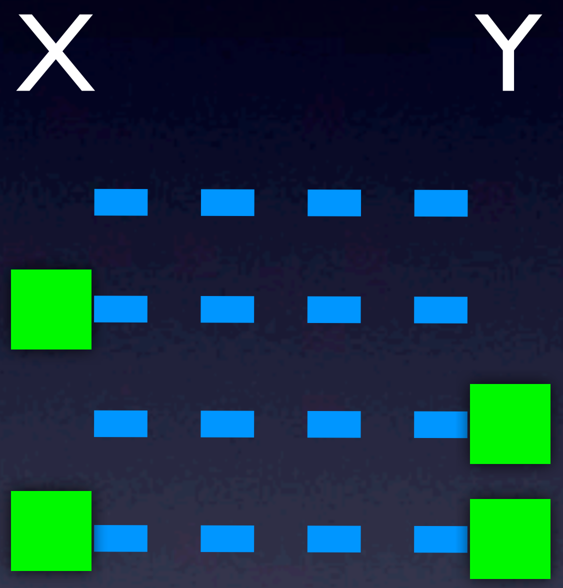
- edge types
- node attributes
- node attribute comparisons

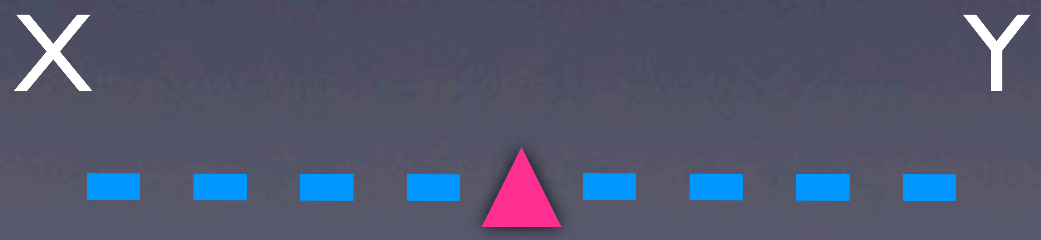


1. Extract path descriptions

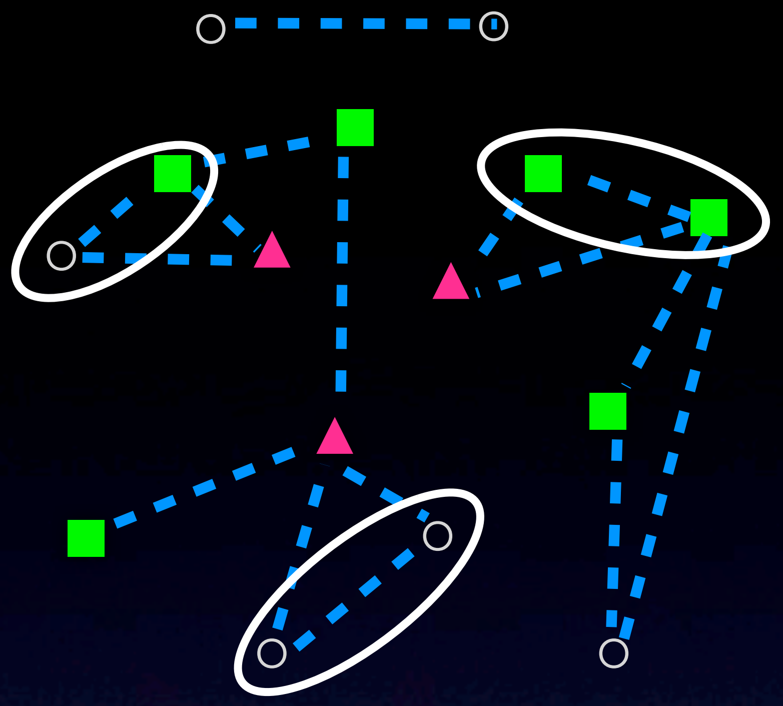
- edge types
- node attributes
- node attribute comparisons

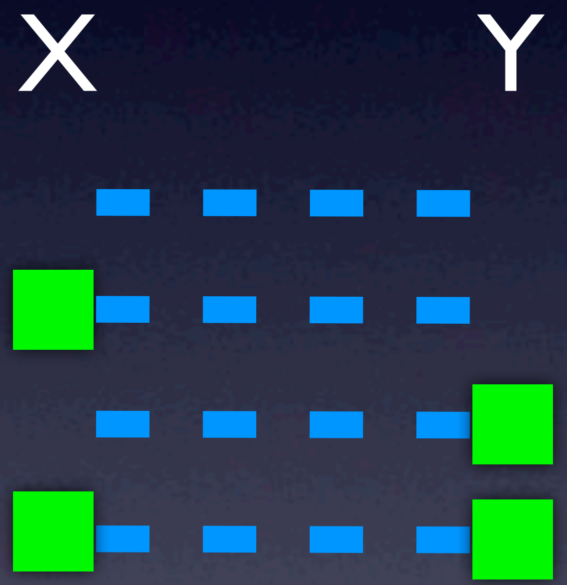


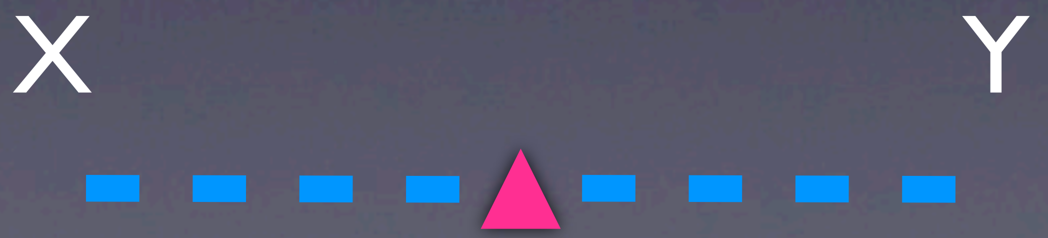




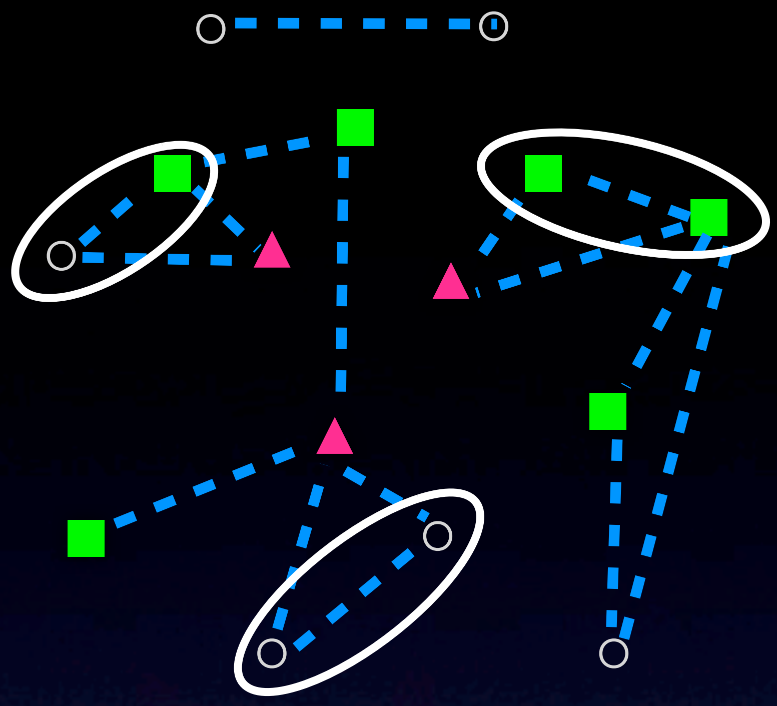
2. Mine frequent paths

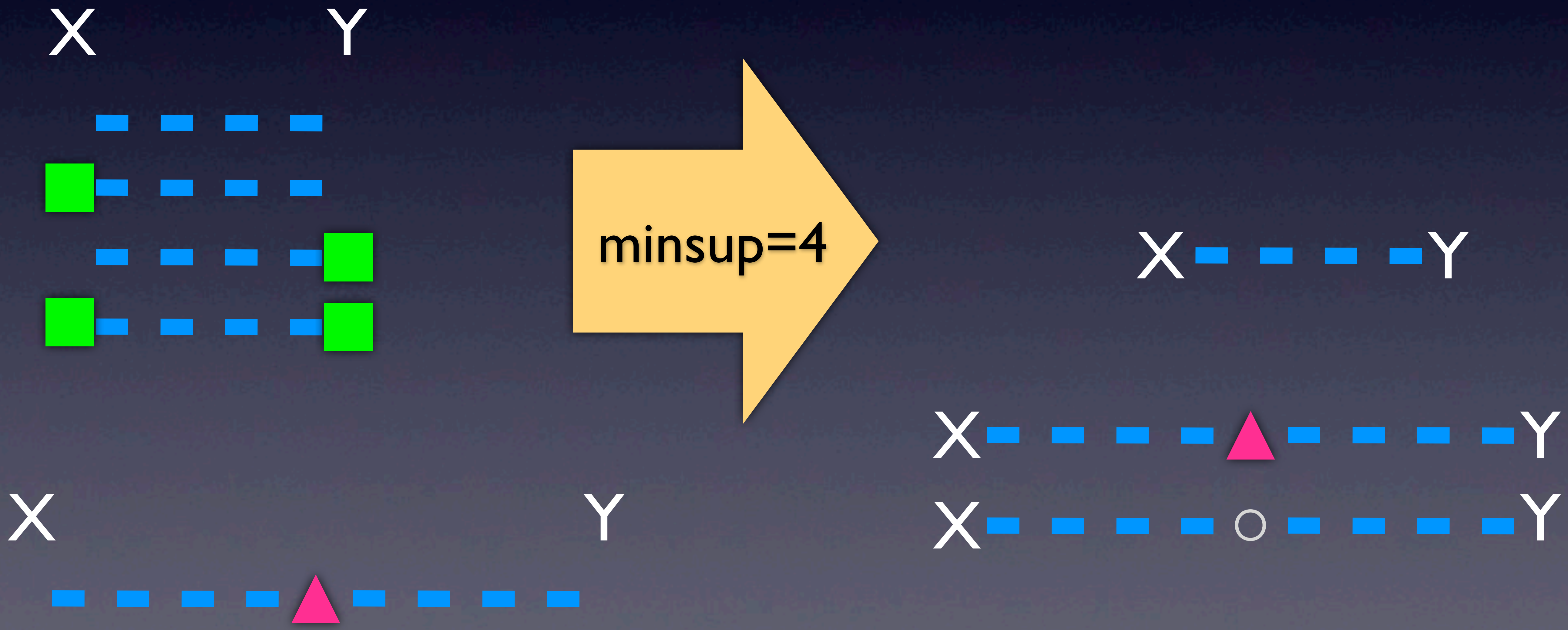






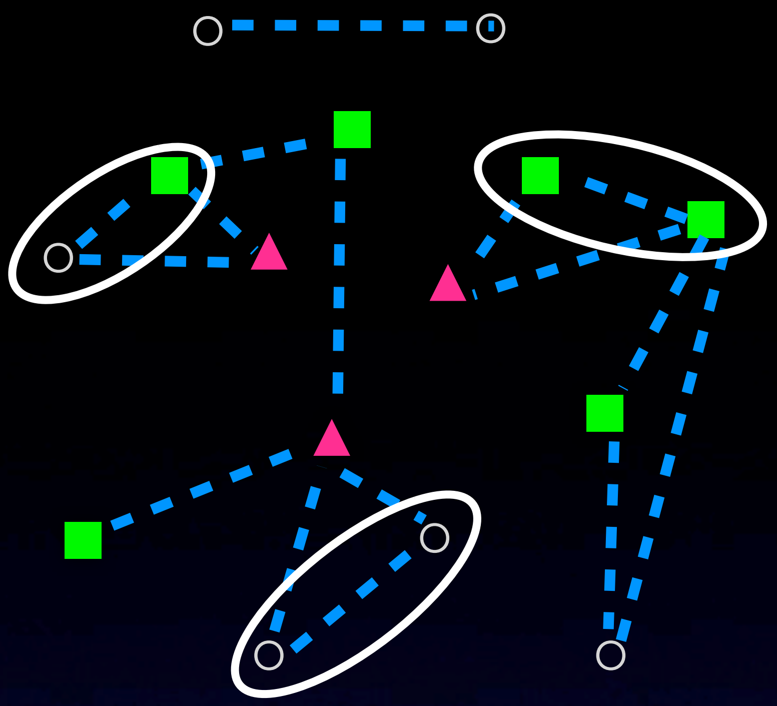
2. Mine frequent paths

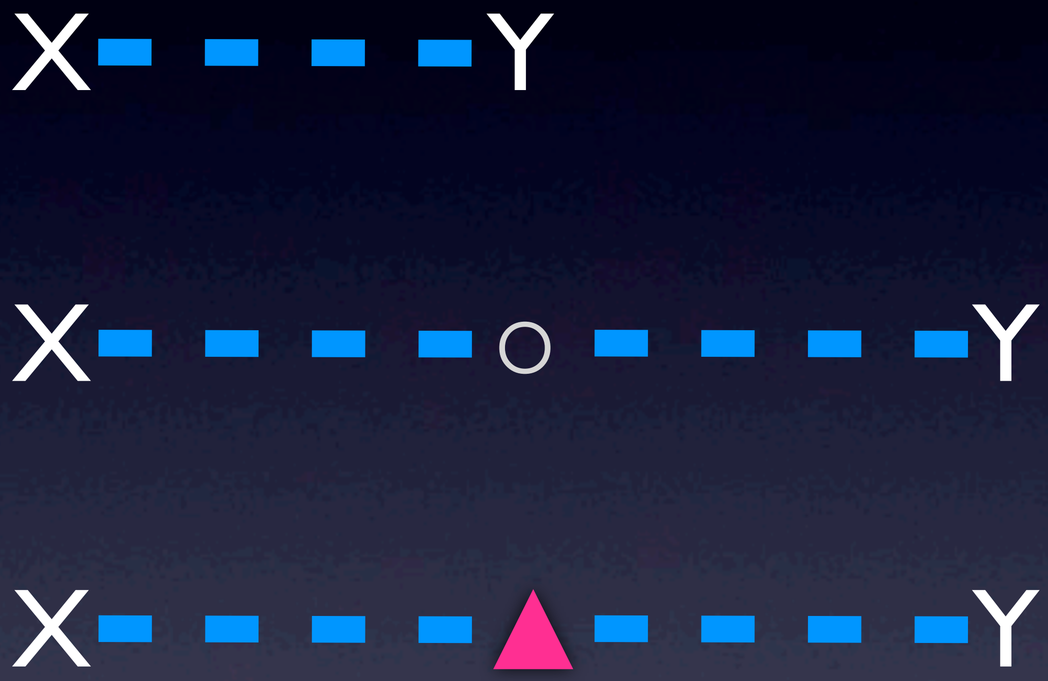




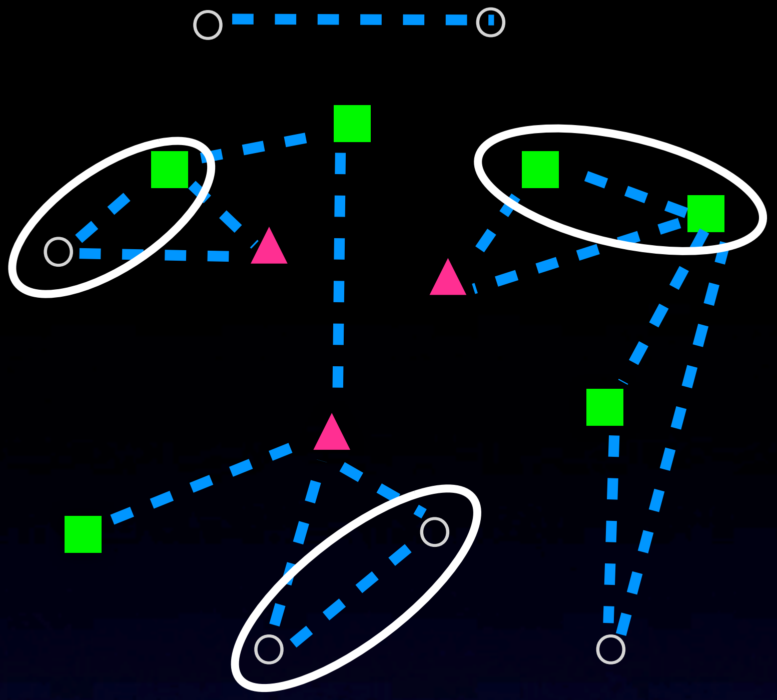
minsup=4

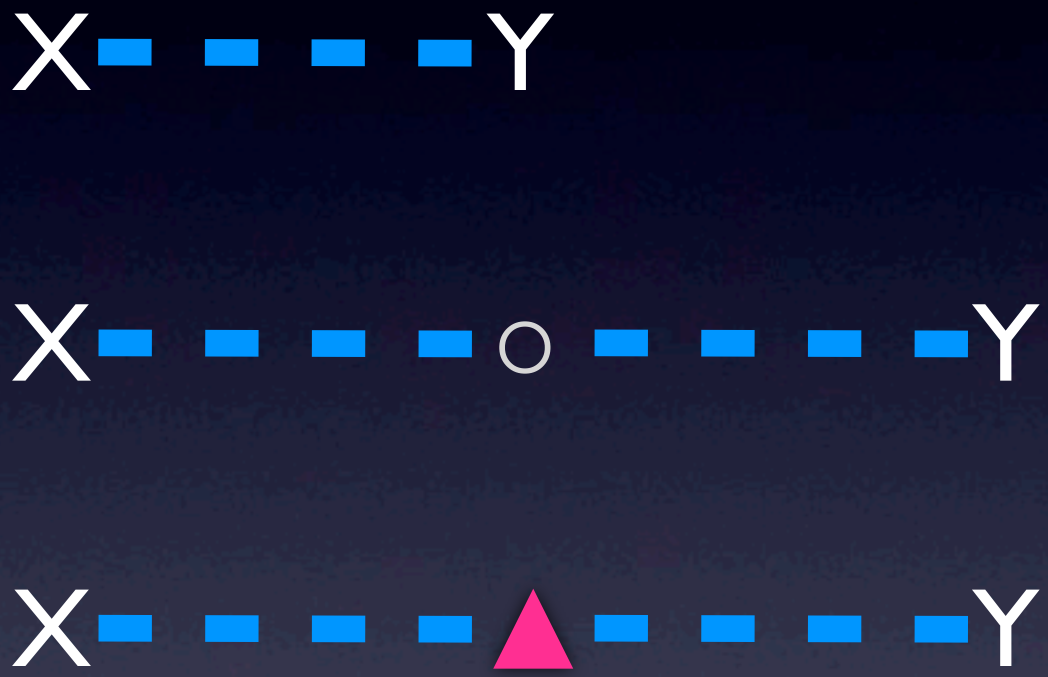
3. Build graphs

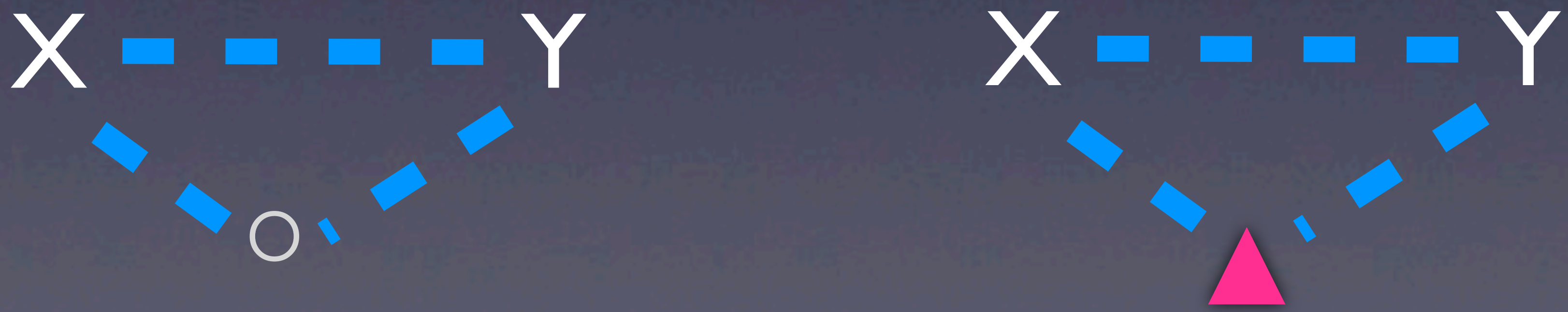




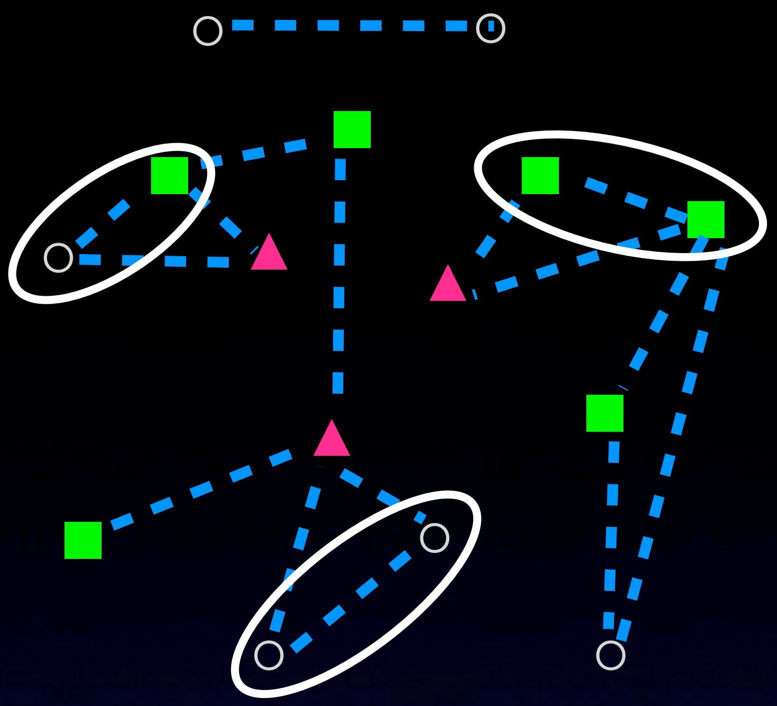
3. Build graphs

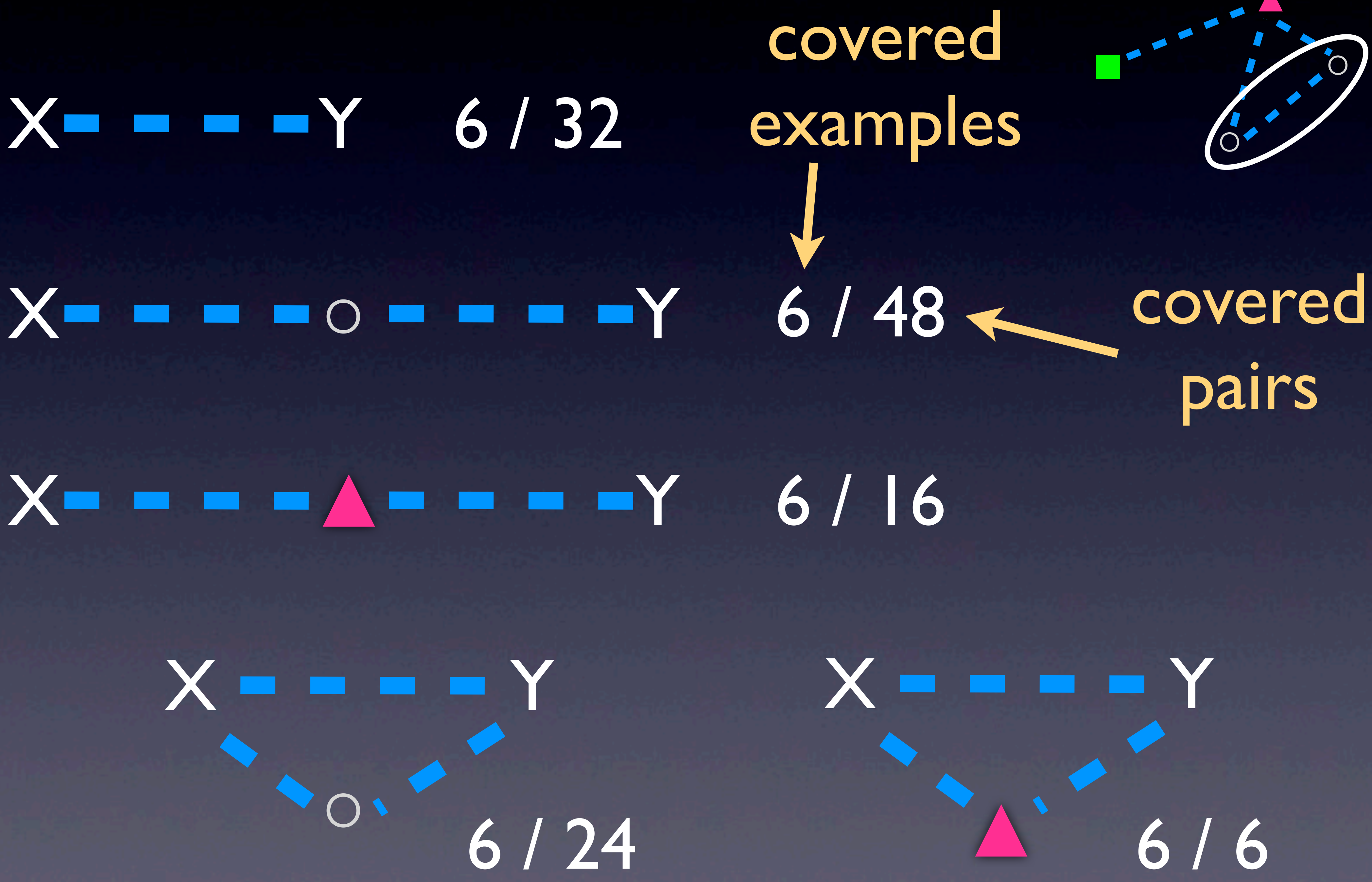






4. Select graphs





6 / 32

covered examples

6 / 48

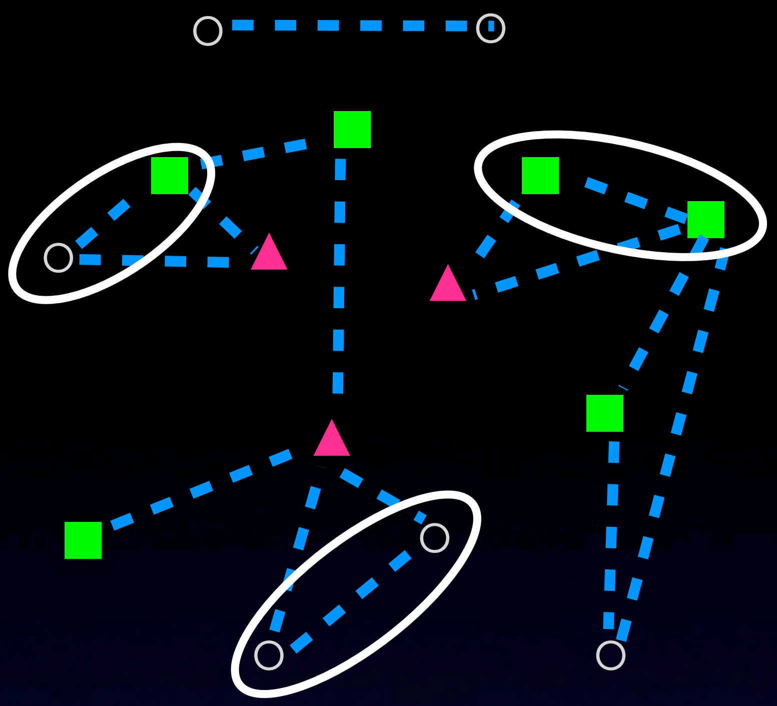
covered pairs

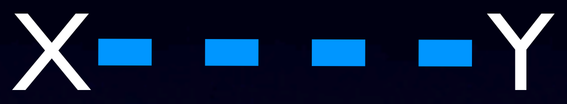
6 / 16

6 / 24

6 / 6

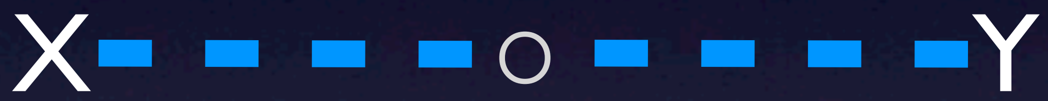
4. Select graphs





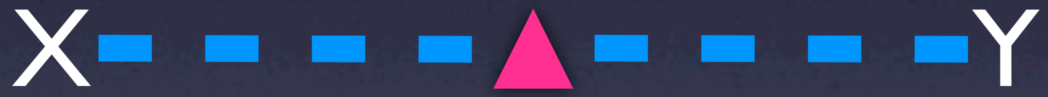
6 / 32

covered examples

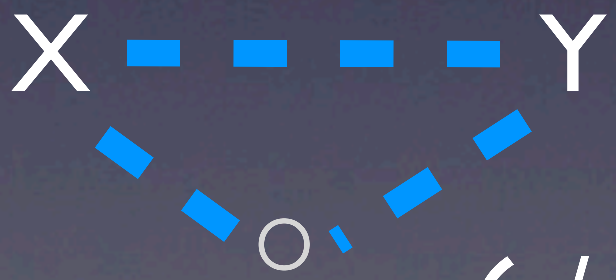


6 / 48

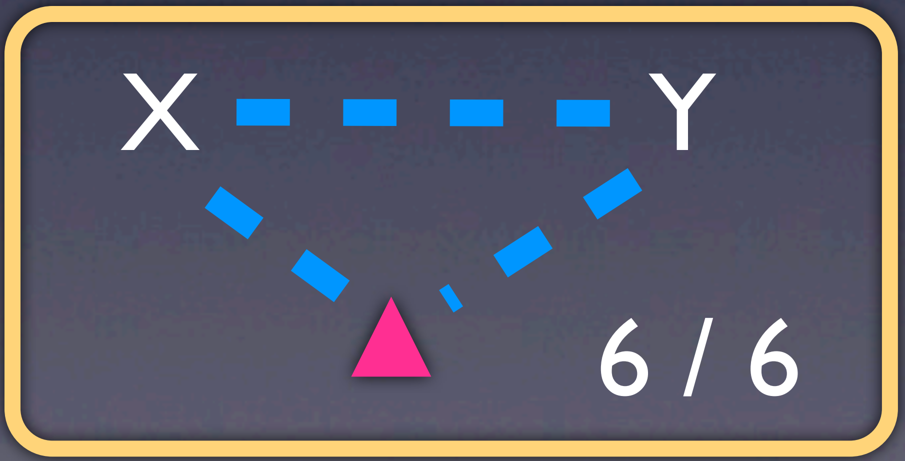
covered pairs



6 / 16

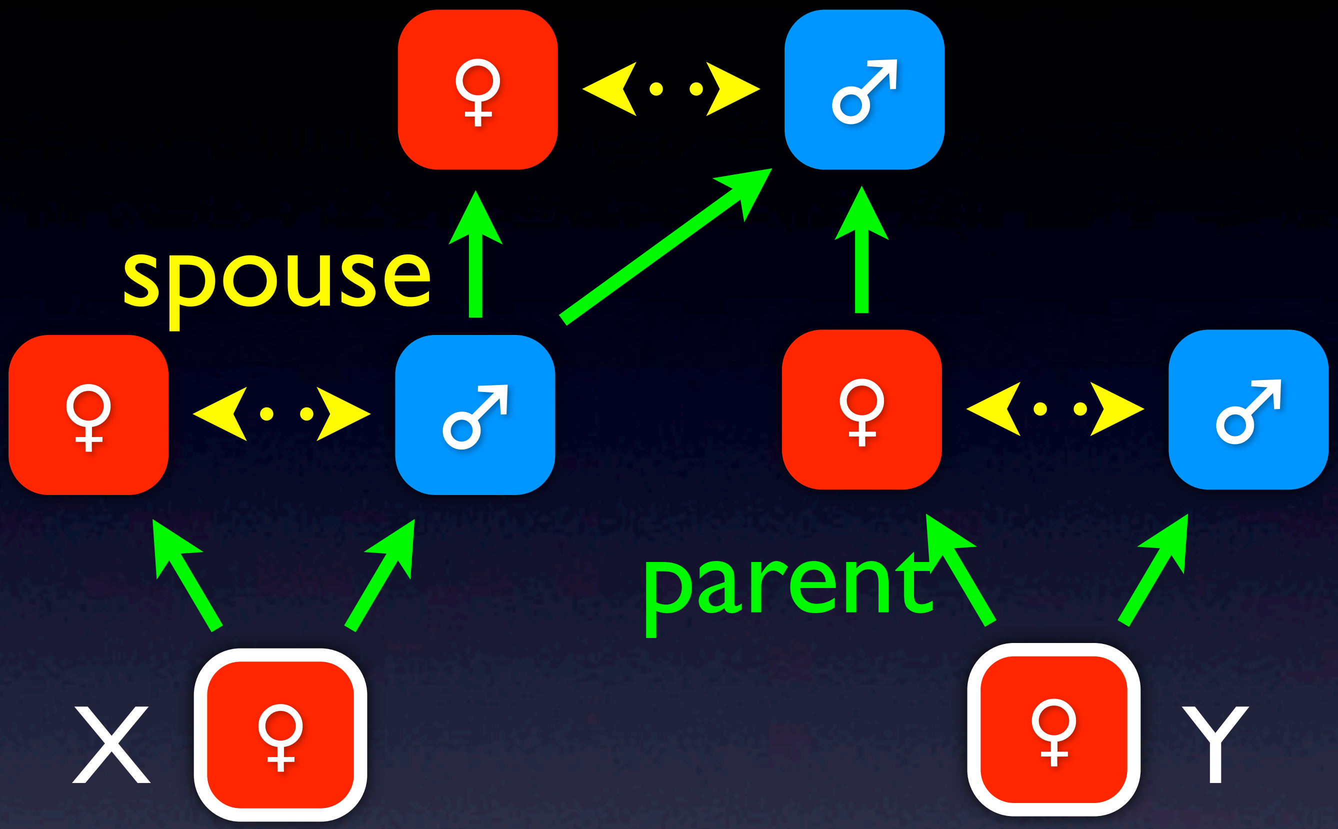


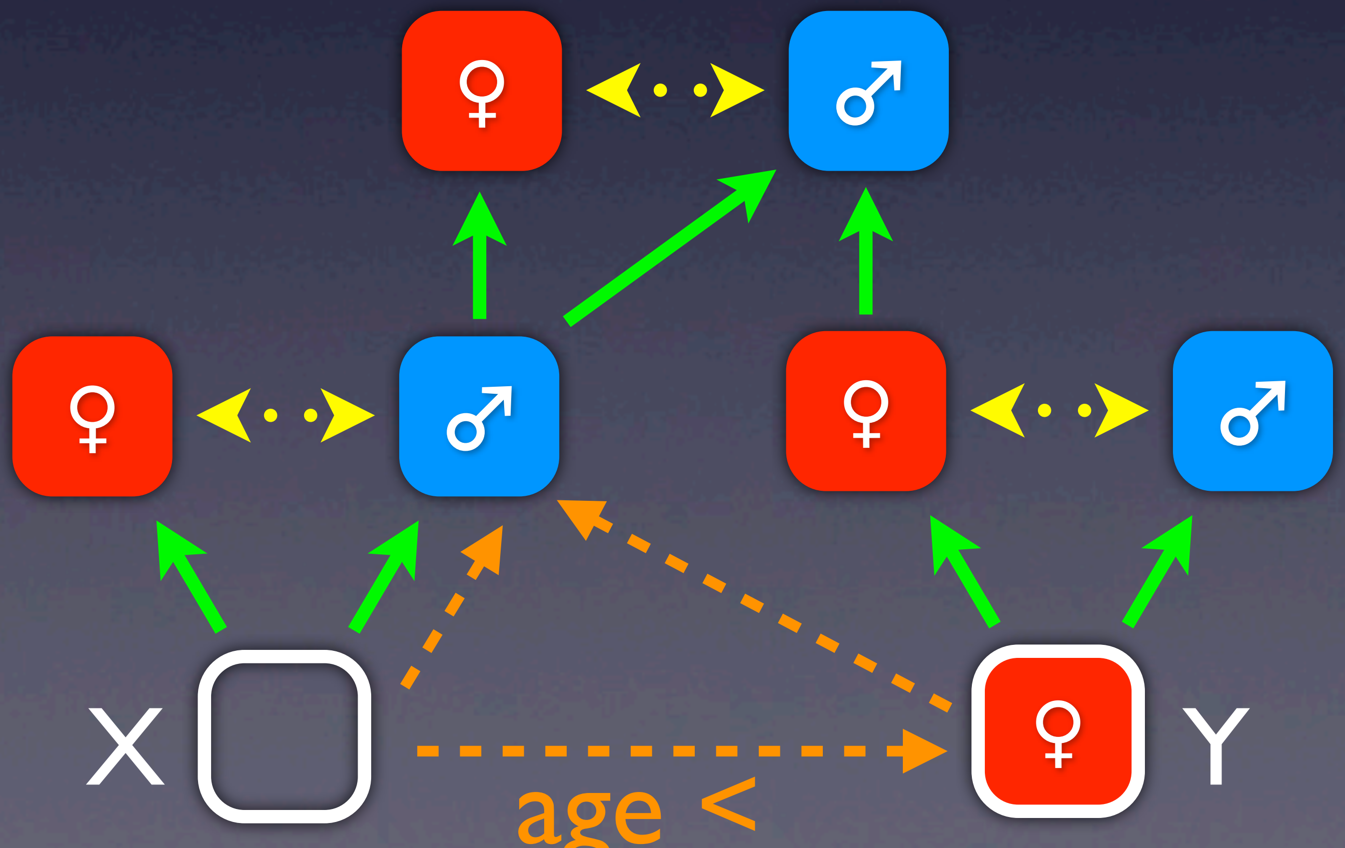
6 / 24



6 / 6

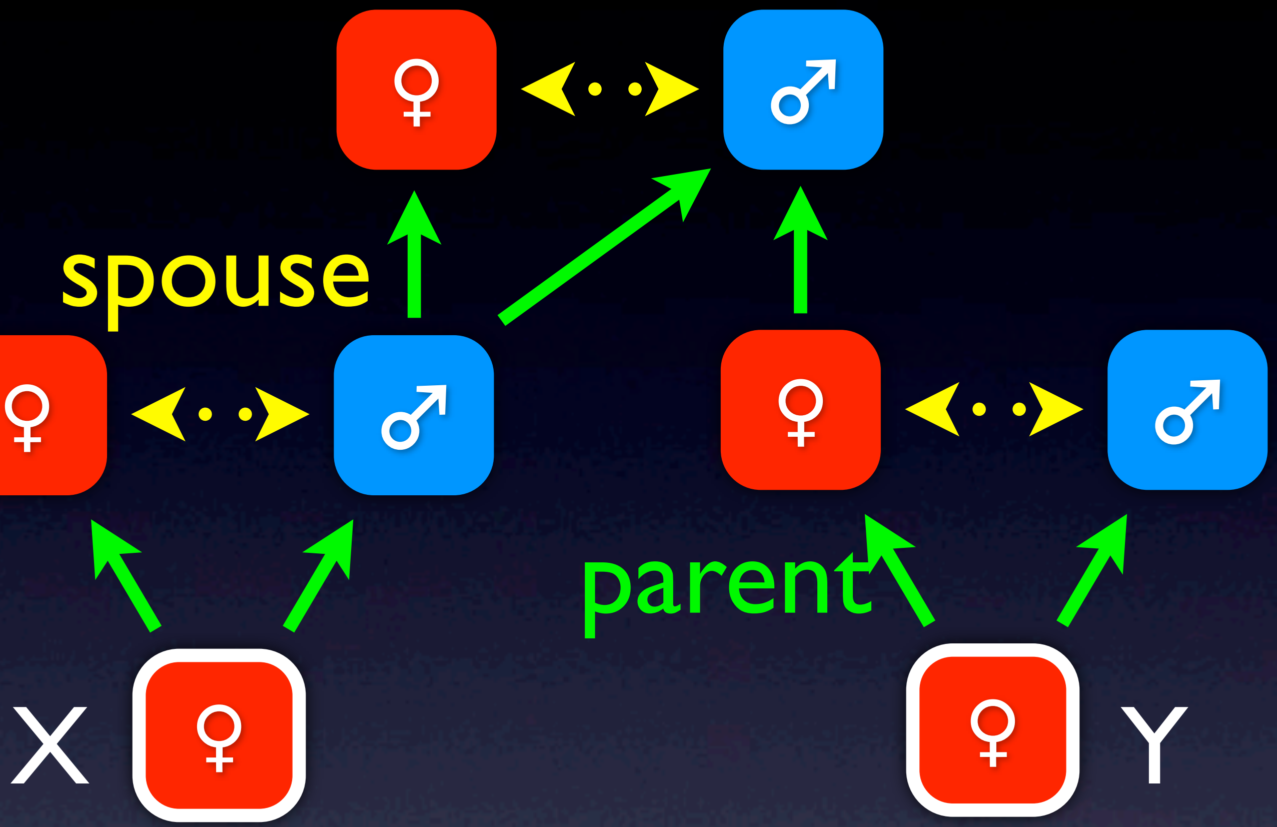
Experiment: Kinship Terms

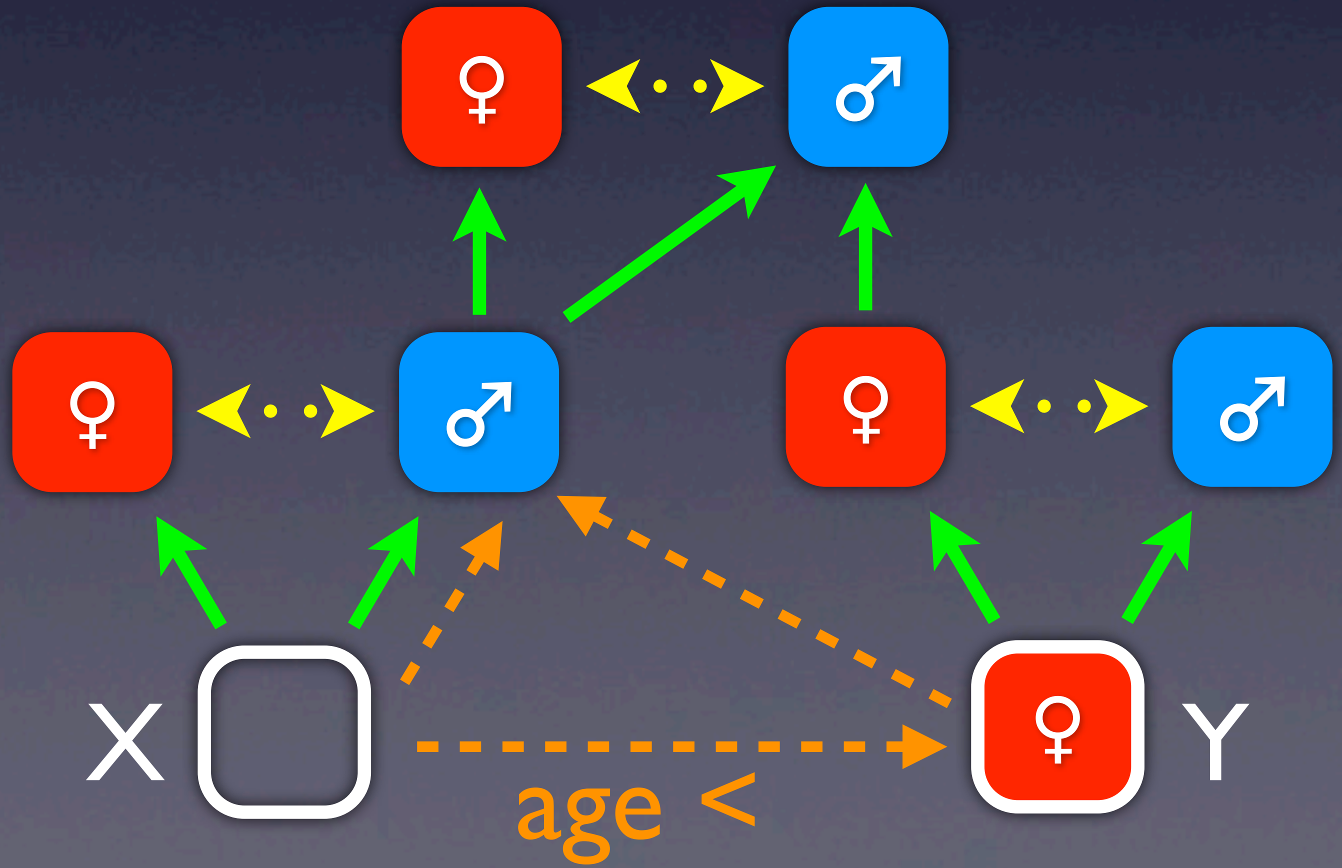




Experiment: Kinship Terms

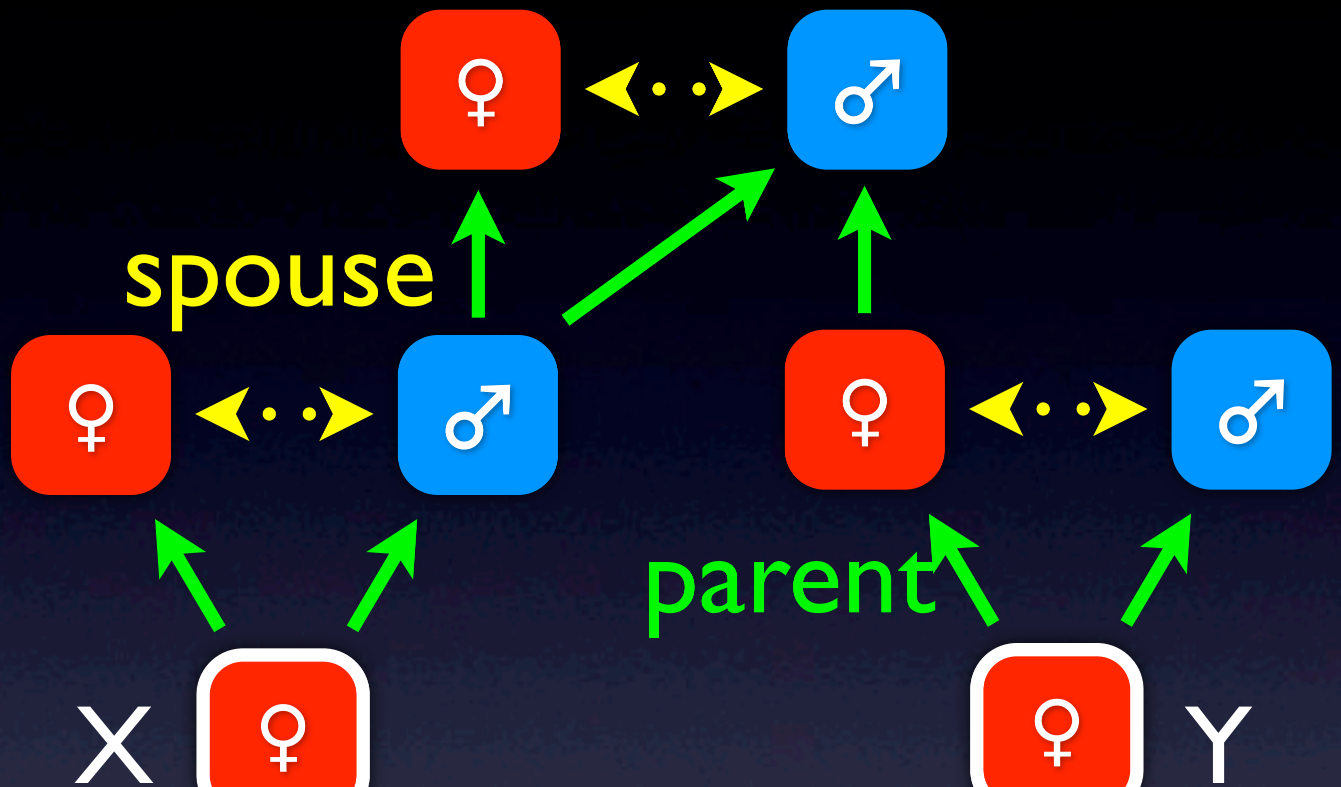
Algyeliya = “my father’s sister’s daughter” or “my mother’s brother’s daughter”

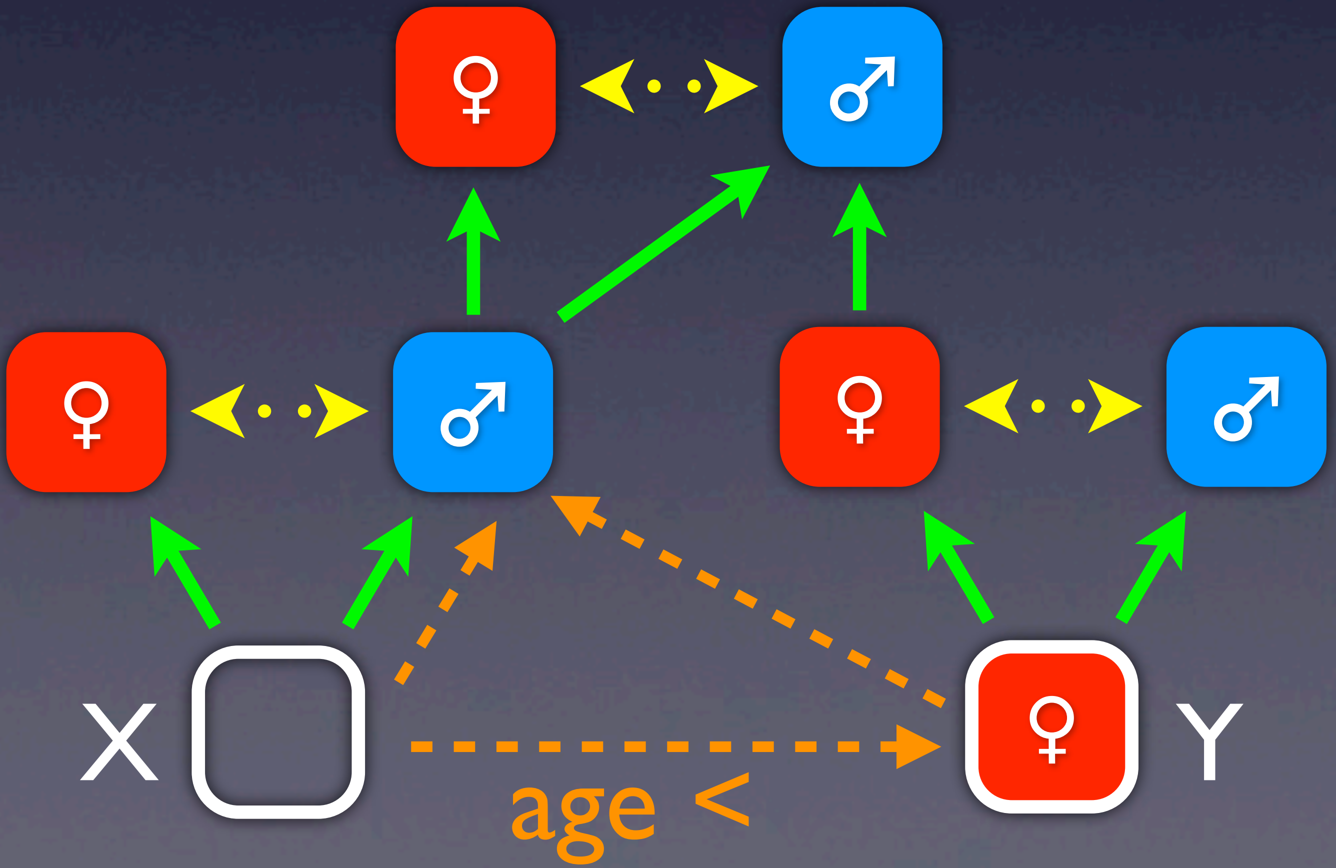




Experiment: Kinship Terms

Algyeliya = “my father’s sister’s daughter” or “my mother’s brother’s daughter”





Challenges:

- connected patterns
- complex patterns
- fast mining

Conclusions

- **Relational redescriptions**: structurally different relational patterns covering same tuples
- **First step**: find disjunctive patterns covering given example pairs
- **Our approach**: constructing patterns from frequent paths
- **Experiments**: approach finds more complex patterns faster than direct relational approach

Thank you!