FACULTY OF ENGINEERING AND ARCHITECTURE

DEPARTMENT OF ELECTRONICS AND INFORMATION SYSTEMS
INTERNET TECHNOLOGY AND DATA SCIENCE LAB (IDLAB)

PERSONALISED PATTERN MINING

Jefrey Lijffijt
FWO [Pegasus]² Marie Skłodowska-Curie Fellow @ Ghent University

GHENT UNIVERSITY

fwo Research Foundation Flanders Opening new horizons

erc European Research Council

OUTLINE

Formalizing pattern mining

Case studies
- Graph patterns
- Subgroups
- Visualizations

Open challenges

Summary

GHENT UNIVERSITY

A FORMALISATION OF PATTERN MINING

WHAT IS A PATTERN?

GHENT UNIVERSITY

WHAT IS A PATTERN?

Itemset*:* co-occurring attributes in binary data

GHENT UNIVERSITY

WHAT IS A PATTERN?

Itemset*:* co-occurring attributes in binary data

Association rule*:* locally ‘predict’ values of other attributes

GHENT UNIVERSITY

WHAT IS A PATTERN?

Itemset*:* co-occurring attributes in binary data

Association rule*:* locally ‘predict’ values of other attributes

Subgroup: rule predicting specific target variable

GHENT UNIVERSITY

WHAT IS A PATTERN?

Itemset*:* co-occurring attributes in binary data

Association rule*:* locally ‘predict’ values of other attributes

Subgroup: rule predicting specific target variable

Community: densely connected subgraph

WHAT IS A PATTERN?

Itemset*:* co-occurring attributes in binary data

Association rule*:* locally ‘predict’ values of other attributes

Subgroup: rule predicting specific target variable

Community: densely connected subgraph

…

WHAT IS A PATTERN?

Itemset*:* co-occurring attributes in binary data

Association rule*:* locally ‘predict’ values of other attributes

Subgroup: rule predicting specific target variable

Community: densely connected subgraph

…

Clustering? Regression? Probabilistic model?

GHENT UNIVERSITY

WHAT IS A PATTERN?

Itemset*:* co-occurring attributes in binary data

Association rule*:* locally 'predict' values of other attributes

Subgroup: rule predicting specific target variable

Community: densely connected subgraph

…

Clustering? Regression? Probabilistic model?

Patterns specify some aspect of the data

GHENT UNIVERSITY

UNIFIED VIEW OF PATTERNS

Lijffijt et al. 2010, De Bie 2011

Data $\hat{X}$

UNIFIED VIEW OF PATTERNS

Lijffijt et al. 2010, De Bie 2011

Data $\hat{X}$

Data space Ω

- Every $X \in \Omega$ is a value combination what could be $\hat{X}$
- $\hat{X}$ is constant, but we do not *know* the values

GHENT UNIVERSITY

UNIFIED VIEW OF PATTERNS

Lijffijt et al. 2010, De Bie 2011

Data $\hat{X}$

Data space Ω

– Every $X \in \Omega$ is a value combination what could be $\hat{X}$

– $\hat{X}$ is constant, but we do not *know* the values

Pattern P is a set $P \subseteq \Omega$ s.t. $\hat{X} \in P$

– That is, it may limit the possibilities for $\hat{X}$

GHENT UNIVERSITY

UNIFIED VIEW OF PATTERNS

Lijffijt et al. 2010, De Bie 2011

Data $\hat{X}$ ← G = (V, E)

Data space Ω

– Every $X \in \Omega$ is a value combination what could be $\hat{X}$

– $\hat{X}$ is constant, but we do not *know* the values

Pattern P is a set $P \subseteq \Omega$ s.t. $\hat{X} \in P$

– That is, it may limit the possibilities for $\hat{X}$

GHENT UNIVERSITY

UNIFIED VIEW OF PATTERNS

Lijffijt et al. 2010, De Bie 2011

Data $\hat{X}$ ← G = (V, E)

Data space Ω ← All possible E for this set V

– Every $X \in \Omega$ is a value combination what could be $\hat{X}$

– $\hat{X}$ is constant, but we do not *know* the values

Pattern P is a set $P \subseteq \Omega$ s.t. $\hat{X} \in P$

– That is, it may limit the possibilities for $\hat{X}$

GHENT UNIVERSITY

UNIFIED VIEW OF PATTERNS

Lijffijt et al. 2010, De Bie 2011

G = (V, E)

In case V is known

Data $\hat{X}$

All possible E for this set V

Data space Ω

– Every $X \in \Omega$ is a value combination what could be $\hat{X}$

– $\hat{X}$ is constant, but we do not *know* the values

Pattern P is a set $P \subseteq \Omega$ s.t. $\hat{X} \in P$

– That is, it may limit the possibilities for $\hat{X}$

UNIFIED VIEW OF PATTERNS

Lijffijt et al. 2010, De Bie 2011

Data $\hat{X}$

G = (V, E)

In case V is known

Data space Ω

All possible E for this set V

– Every $X \in \Omega$ is a value combination what could be $\hat{X}$

– $\hat{X}$ is constant, but we do not *know* the values

Pattern P is a set $P \subseteq \Omega$ s.t. $\hat{X} \in P$

– That is, it may limit the possibilities for $\hat{X}$

Certain edges that do or do not exist

GHENT UNIVERSITY

UNIFIED VIEW OF PATTERNS

Lijffijt et al. 2010, De Bie 2011

Data $\hat{X}$

Data space Ω

- Every $X \in \Omega$ is a value combination what could be $\hat{X}$
- $\hat{X}$ is constant, but we do not *know* the values

Pattern P is a set $P \subseteq \Omega$ s.t. $\hat{X} \in P$

- That is, it may limit the possibilities for $\hat{X}$

GHENT UNIVERSITY

UNIFIED VIEW OF PATTERNS

Lijffijt et al. 2010, De Bie 2011

Data $\hat{X}$

Data matrix

Data space Ω

– Every $X \in \Omega$ is a value combination what could be $\hat{X}$

– $\hat{X}$ is constant, but we do not *know* the values

Pattern P is a set $P \subseteq \Omega$ s.t. $\hat{X} \in P$

– That is, it may limit the possibilities for $\hat{X}$

GHENT UNIVERSITY

UNIFIED VIEW OF PATTERNS

Lijffijt et al. 2010, De Bie 2011

Data $\hat{X}$ Data matrix

Data space Ω All possible value combinations

- Every $X \in \Omega$ is a value combination what could be $\hat{X}$
- $\hat{X}$ is constant, but we do not *know* the values

Pattern P is a set $P \subseteq \Omega$ s.t. $\hat{X} \in P$

- That is, it may limit the possibilities for $\hat{X}$

GHENT UNIVERSITY

UNIFIED VIEW OF PATTERNS

Lijffijt et al. 2010, De Bie 2011

Data $\hat{X}$

Data matrix

Data space Ω

All possible value combinations

– Every $X \in \Omega$ is a value combination what could be $\hat{X}$

– $\hat{X}$ is constant, but we do not *know* the values

Pattern P is a set $P \subseteq \Omega$ s.t. $\hat{X} \in P$

– That is, it may limit the possibilities for $\hat{X}$

Range, moment, correlation, clustering, …

GHENT UNIVERSITY

INTERESTINGNESS OF PATTERNS

GHENT UNIVERSITY

THE AIM OF PATTERN MINING

Suppose: aim is to **inform** the user about the data

THE AIM OF PATTERN MINING

Suppose: aim is to **inform** the user about the data

Form of communication

– Computer $\xrightarrow{P}$ user

GHENT UNIVERSITY

THE AIM OF PATTERN MINING

Suppose: aim is to **inform** the user about the data

Form of communication

– Computer $\xrightarrow{P}$ user

Central premise of pattern mining:
Patterns can provide *information*
in a compressed form

GHENT
UNIVERSITY

THE AIM OF PATTERN MINING

Suppose: aim is to **inform** the user about the data

Form of communication

– Computer $\xrightarrow{P}$ user

Central premise of pattern mining:
Patterns can provide *information*
in a compressed form

How to optimize this communication?

≈ Maximal structure in minimal time

GHENT UNIVERSITY

INFORMATION THEORETIC APPROACH

De Bie 2011, 2013

Pattern contains *information*

- What we really learn about data

Pattern has a *descriptional complexity*

- How time consuming is it to internalize

INFORMATION THEORETIC APPROACH

De Bie 2011, 2013

Learning is being surprised

– What we learn depends on what we know/expect

In particular, need to define $\Pr(P)$

– Typically from mass or density function over Ω

Used a lot in data mining, but often implicitly

– Models for random graphs, sequences, matrices, …

GHENT UNIVERSITY

BACKGROUND MODEL

$\Pr(P)$ is called the background model (or distribution)

Many options

- Maximum Entropy model Tatti 2008, De Bie 2011
- Randomization Gionis et al. 2006, Lijffijt et al. 2014
- Other probabilistic models

BACKGROUND MODEL

$\Pr(P)$ is called the background model (or distribution)

Explicitly neutral model given certain expectations

Many options

- Maximum Entropy model Tatti 2008, De Bie 2011
- Randomization Gionis et al. 2006, Lijffijt et al. 2014
- Other probabilistic models

GHENT UNIVERSITY

BACKGROUND MODEL: EXAMPLE

$\Pr(P)$ is called the background model (or distribution)

Given a graph G = (V, E), and Ω all possible E

- Without knowledge: MaxEnt dist. is Uniform over Ω
- Given degree for every vertex: find MaxEnt dist. through optimization

GHENT UNIVERSITY

BACKGROUND MODEL: EXAMPLE

$\Pr(P)$ is called the background model (or distribution)

Given a graph G = (V, E), and Ω all possible E

- Without knowledge: MaxEnt dist. is Uniform over Ω
- Given degree for every vertex: find MaxEnt dist. through optimization

Cannot do arbitrary constraints, but many types of expectations lead to tractable parameter inference

GHENT UNIVERSITY

THE FORMALIZATION

De Bie 2011, 2013

Recall: pattern P is a set $\hat{X} \in P \subseteq \Omega$

Information Content (IC) of P is $-log(\Pr(P))$

Description Length (DL) of P depends on syntax

Subjective Interestingness (SI) of P is $\dfrac{\mathrm{IC}(P)}{\mathrm{DL}(P)}$

GHENT UNIVERSITY

THE FORMALIZATION

De Bie 2011, 2013

Recall: pattern P is a set $\hat{X} \in P \subseteq \Omega$

Derived from information gain

Information Content (IC) of P is $-log(\Pr(P))$

Description Length (DL) of P depends on syntax

Subjective Interestingness (SI) of P is $\dfrac{\mathrm{IC}(P)}{\mathrm{DL}(P)}$

GHENT UNIVERSITY

THE FORMALIZATION

De Bie 2011, 2013

Recall: pattern P is a set $\hat{X} \in P \subseteq \Omega$

Derived from information gain

Information Content (IC) of P is $-log(\Pr(P))$

Description Length (DL) of P depends on syntax

Subjective Interestingness (SI) of P is $\frac{\mathrm{IC}(P)}{\mathrm{DL}(P)}$

Rate at which we gain information

GHENT UNIVERSITY

THE FORSIED PROCESS

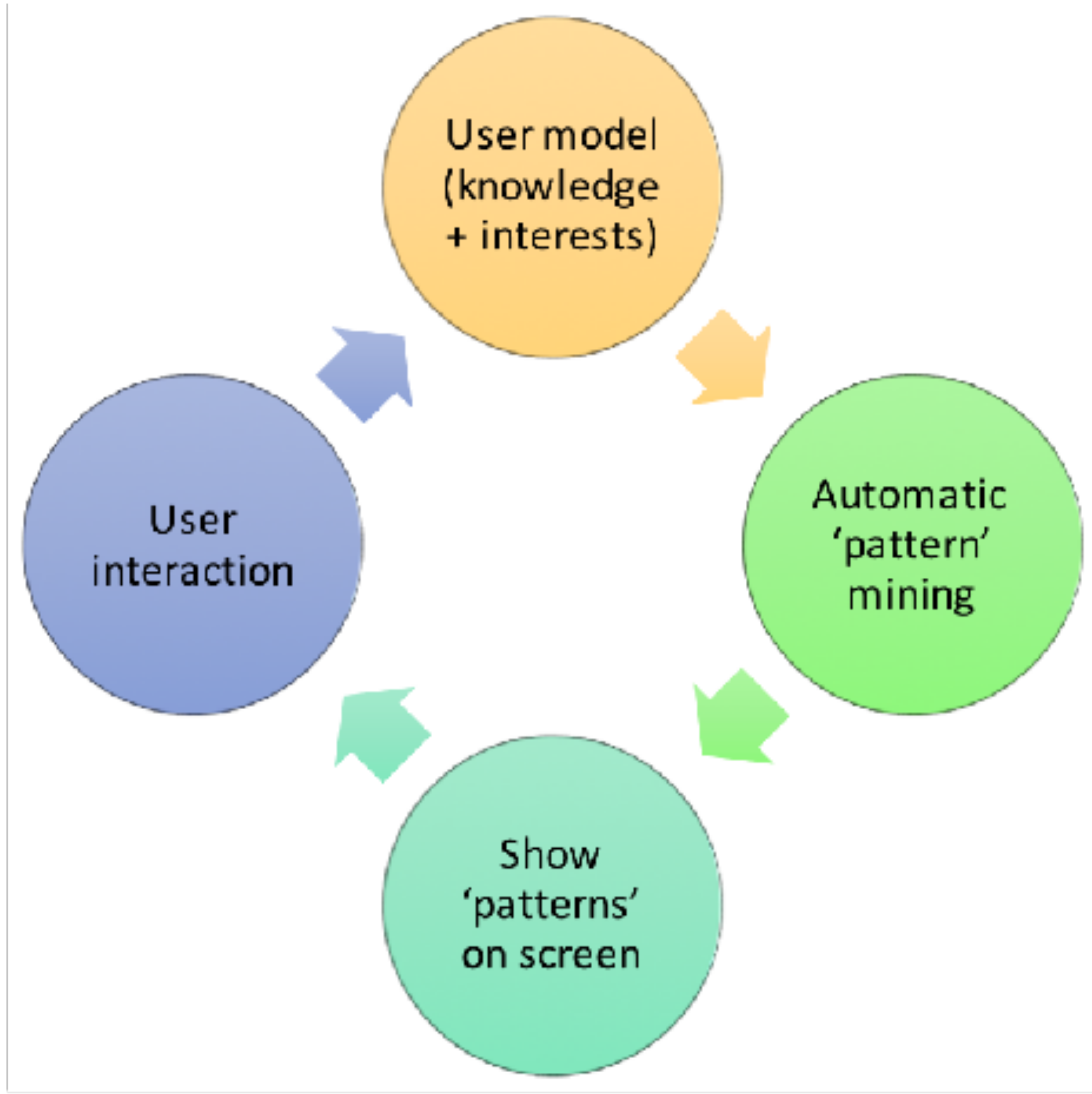
1. Background model
2. Pattern syntax
   – IC & DL of patterns
3. Mining
4. Update background model
5. Iterative mining

GHENT UNIVERSITY

THE FORSIED PROCESS

1. Background model
2. Pattern syntax
   – IC & DL of patterns
3. Mining
4. Update background model
5. Iterative mining

Personalised Pattern Mining



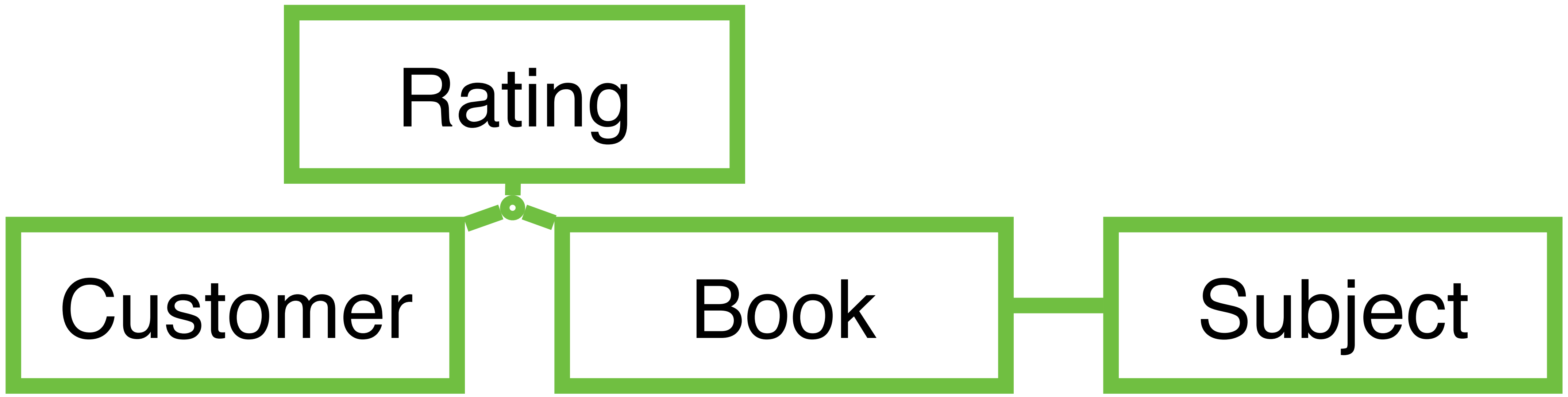
GHENT UNIVERSITY

CONNECTIONS IN RELATIONAL DATA

GHENT UNIVERSITY

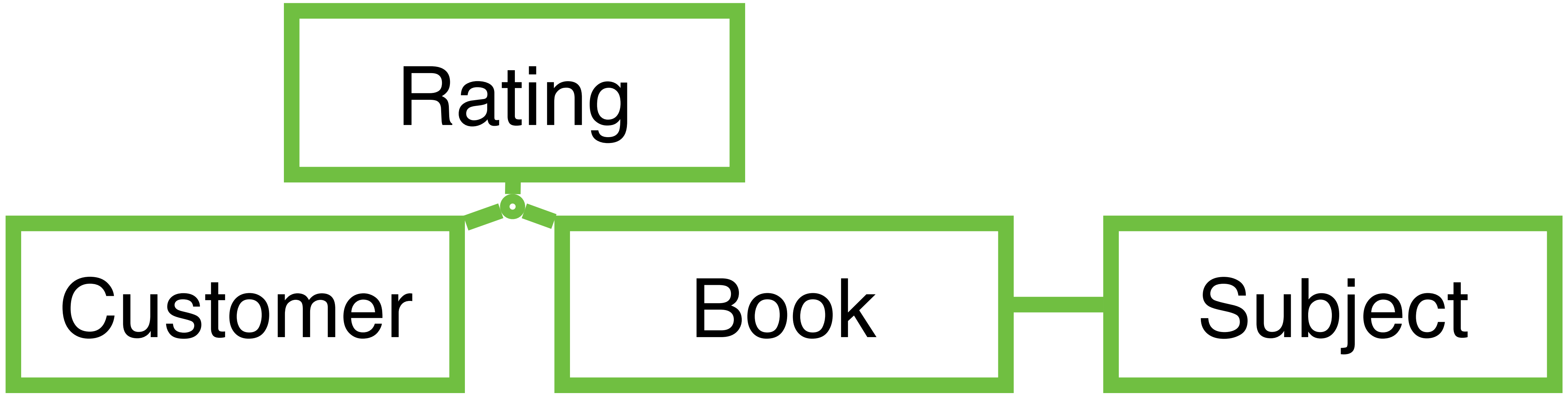
RELATIONAL DATA

Example: Amazon ratings



RELATIONAL DATA

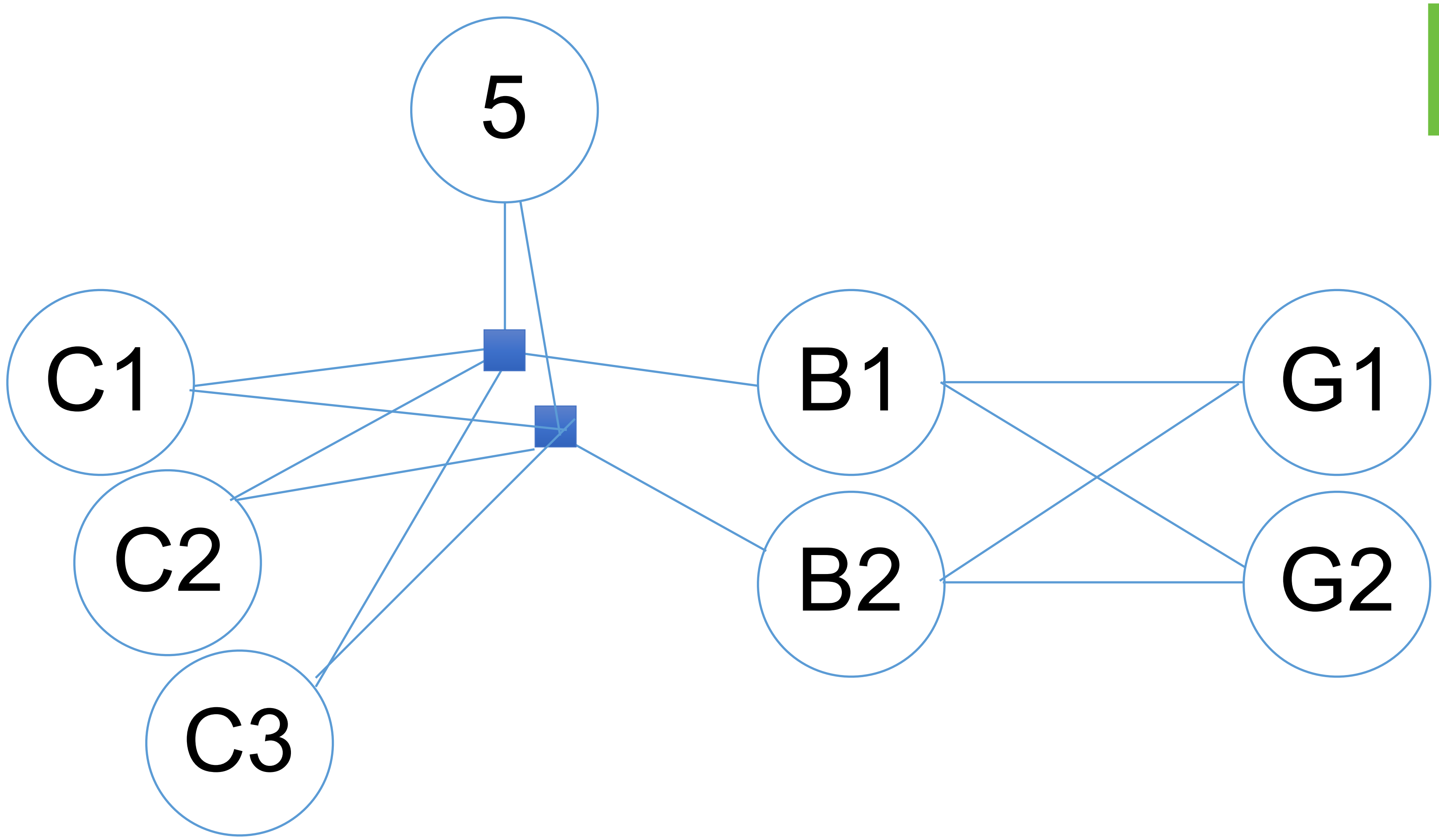
Example: Amazon ratings

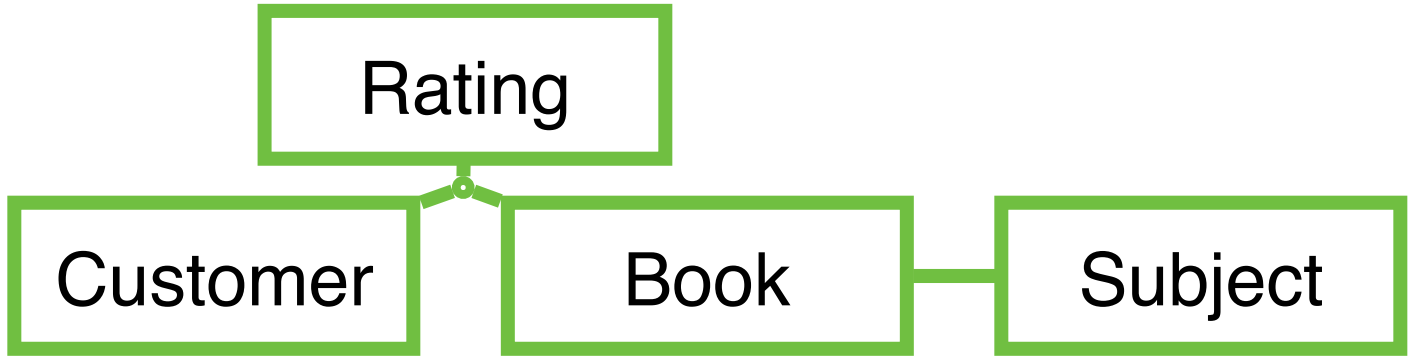


Patterns like “Many customers rate all seven volumes of Harry Potter highly, which all belong to Fantasy and Fiction”

PATTERN SYNTAX

Fully connected set of entities





INFORMATION CONTENT

‘Compression’ due to fully connectedness

– Only have to communicate vertices: $P \subseteq V$

If background model factorizes over edges:

$$\mathrm{IC}(P) = \sum_{(v,w) \in P \times P} \Pr((v,w))$$

GHENT UNIVERSITY

DESCRIPTION LENGTH

Proportional to number of vertices:

$$\mathrm{DL}(P) = \gamma|P| + \eta$$

NB. Absolute values for SI are irrelevant, only ranking matters. Hence either parameter can be constant

MINING

RMiner: exhaustive candidate enumeration Spyropoulou et al. 2014

Constraint programming version: branch and bound Guns et al. 2016

CP works well, but scalability remains an issue

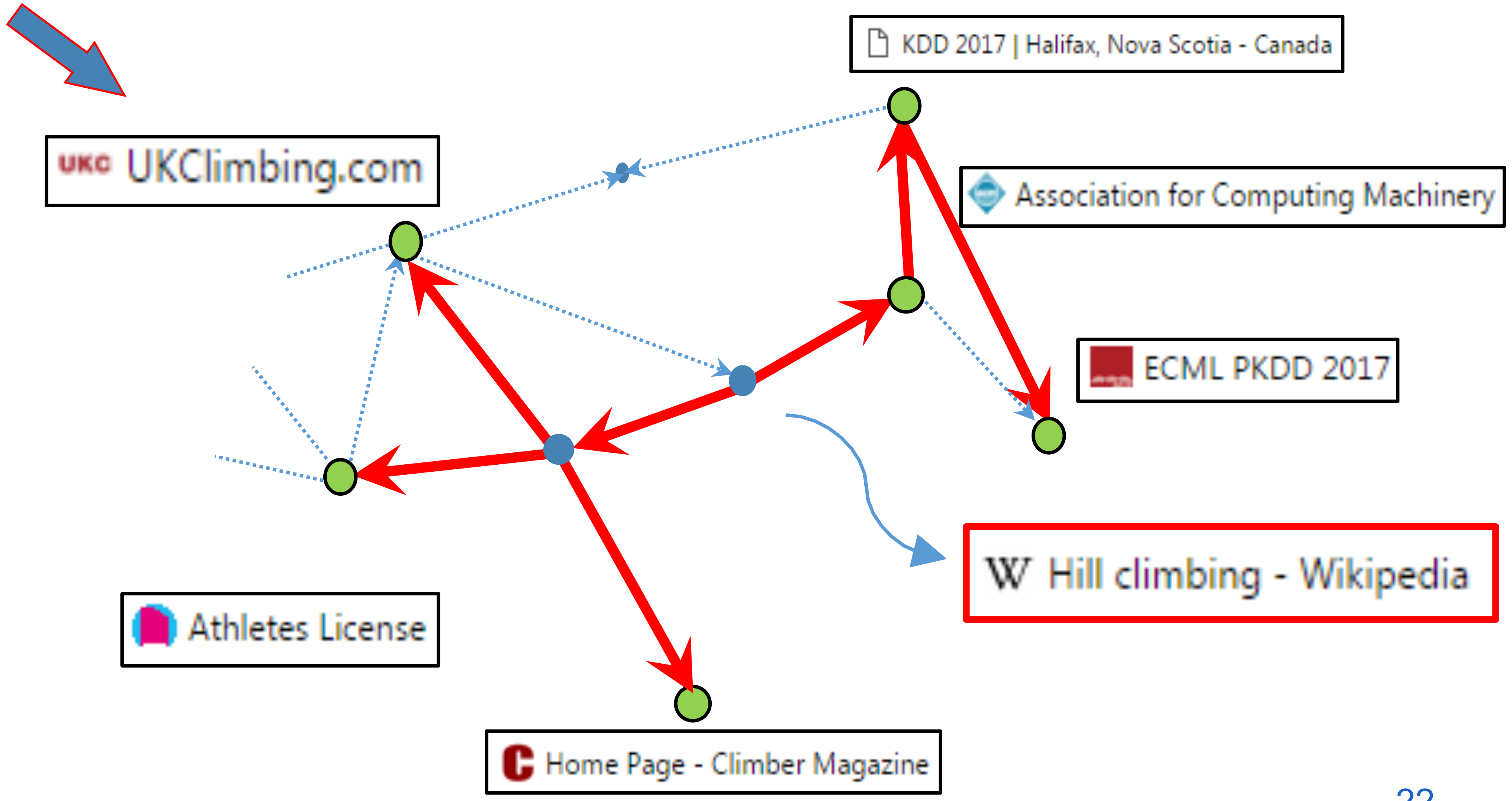
No heuristic algorithms developed yet

CONNECTING TREES IN GRAPHS

GHENT UNIVERSITY

EXAMPLE

Adriaens et al. 2017



GHENT UNIVERSITY

PATTERN SYNTAX

Adriaens et al. 2017

Given a query Q, integer k

Pattern: arborescence of height $\leq k$ that connects Q

- All leafs must be query nodes
- May not exist

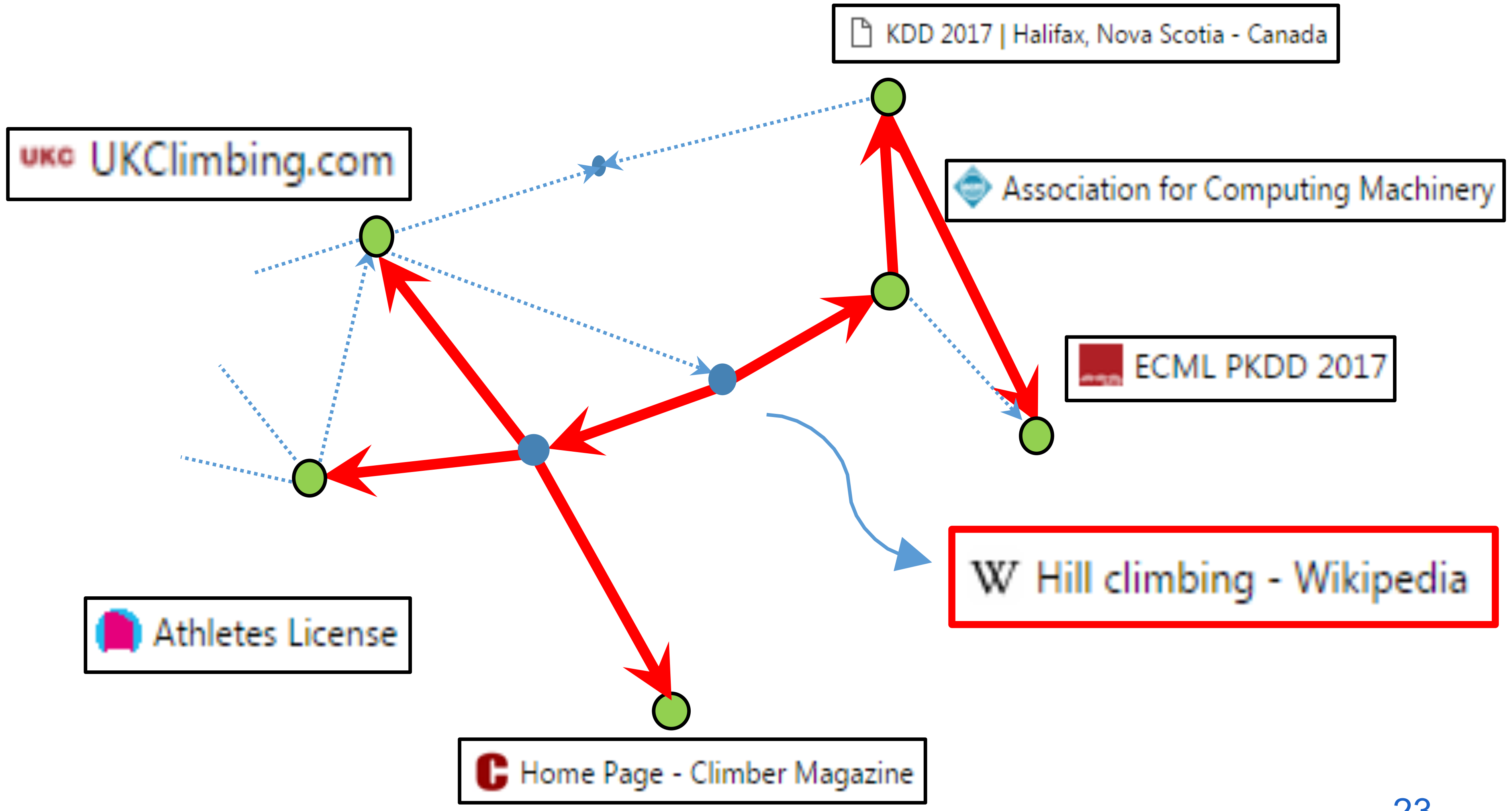
PATTERN SYNTAX

Adriaens et al. 2017

Given a query Q, integer k

Pattern: arborescence of height $\leq k$ that connects Q

- All leafs must be query nodes
- May not exist



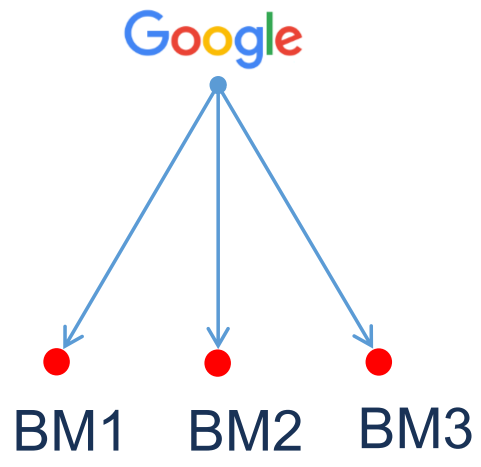
GHENT UNIVERSITY

QUANTIFICATION

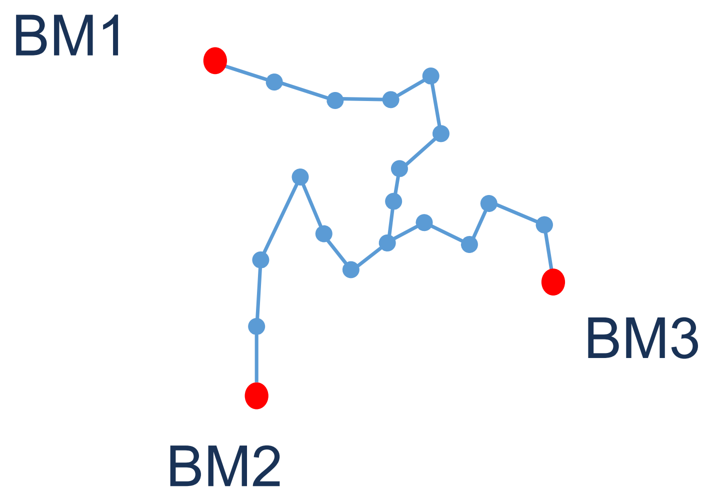
Adriaens et al. 2017

Idea: tree is informative if

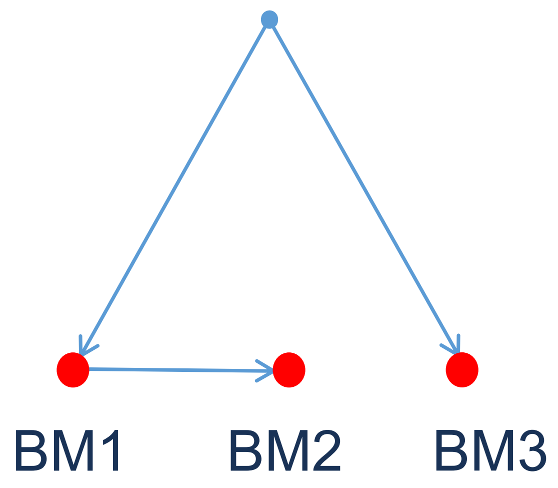
- (links to) vertices are unexpected
- vertices share parents



Non-informative



Non-minimal



Minimal and informative

GHENT UNIVERSITY

QUANTIFICATION

Information Content like before

$$\mathrm{IC}(P) = \sum_{(v,w) \in P \times P} \Pr((v,w))$$

Communicate (new) vertices + parent for each vertex

$$\mathrm{DL}(T) =$$
$$(|V_T| - |Q| + 1)\log(|V| - |Q| + 1) + |V_T|\log(|V_T| + 1)$$

GHENT UNIVERSITY

MINING

Adriaens et al. 2017

Heuristic construction of trees

- Not so straightforward to greedily construct a tree

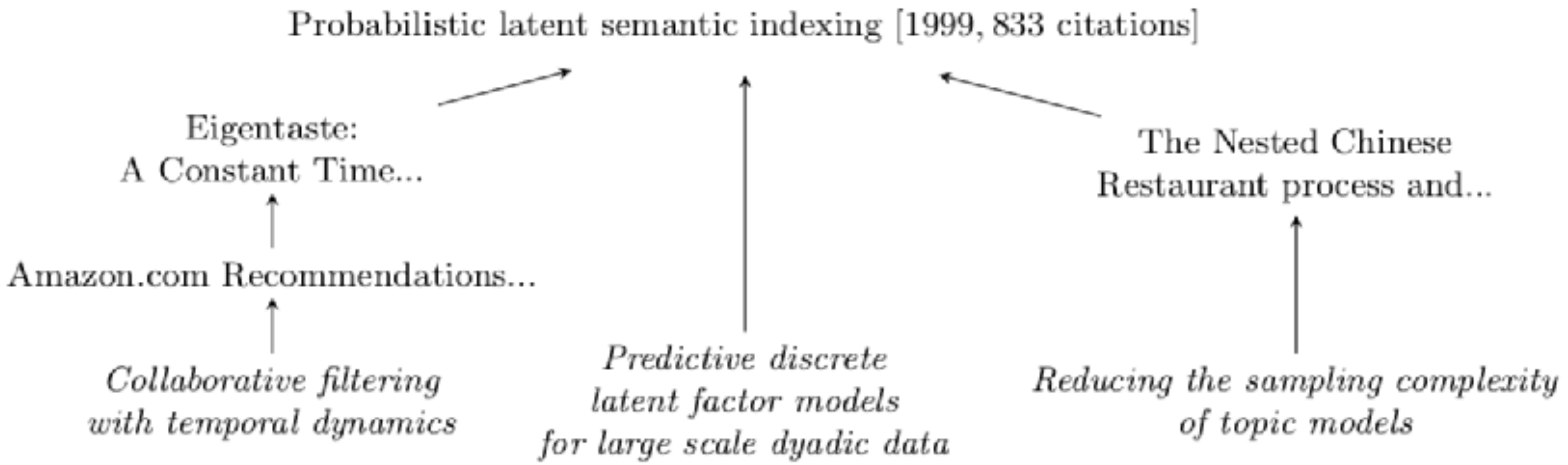
Background distribution for growing graph:

- Vertices can only have links to older vertices
- Parameter inference through intelligent grouping

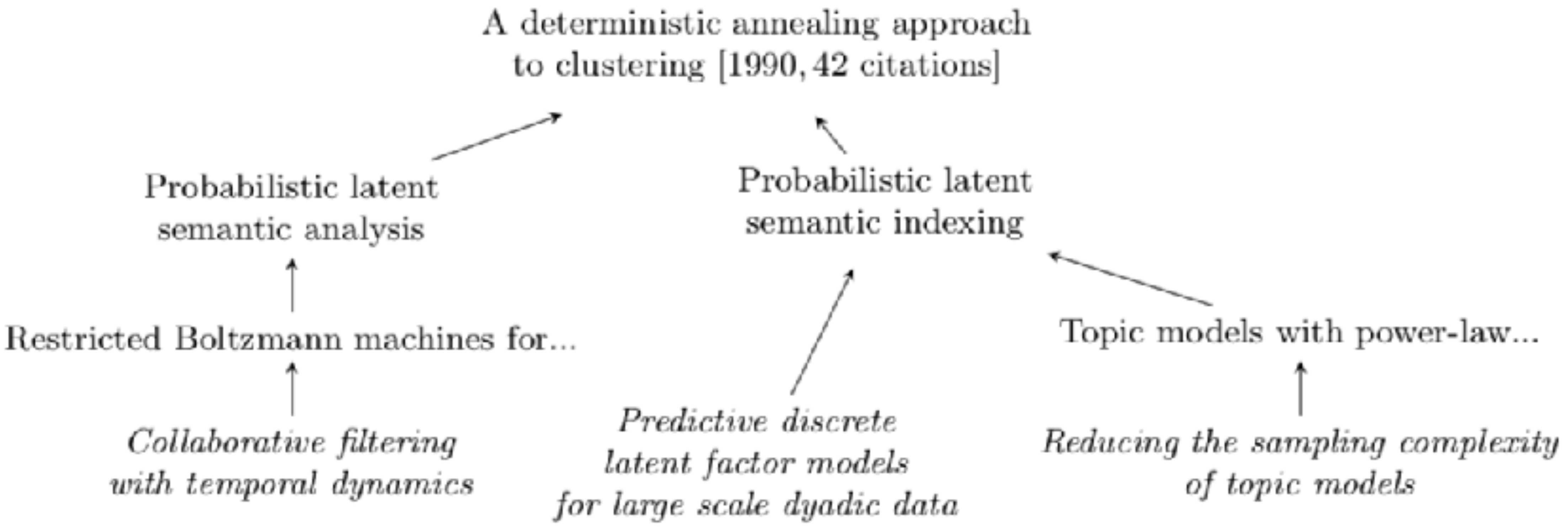
GHENT UNIVERSITY

EXAMPLE RESULT: KDD BEST PAPERS

Without prior:



With degree and time:

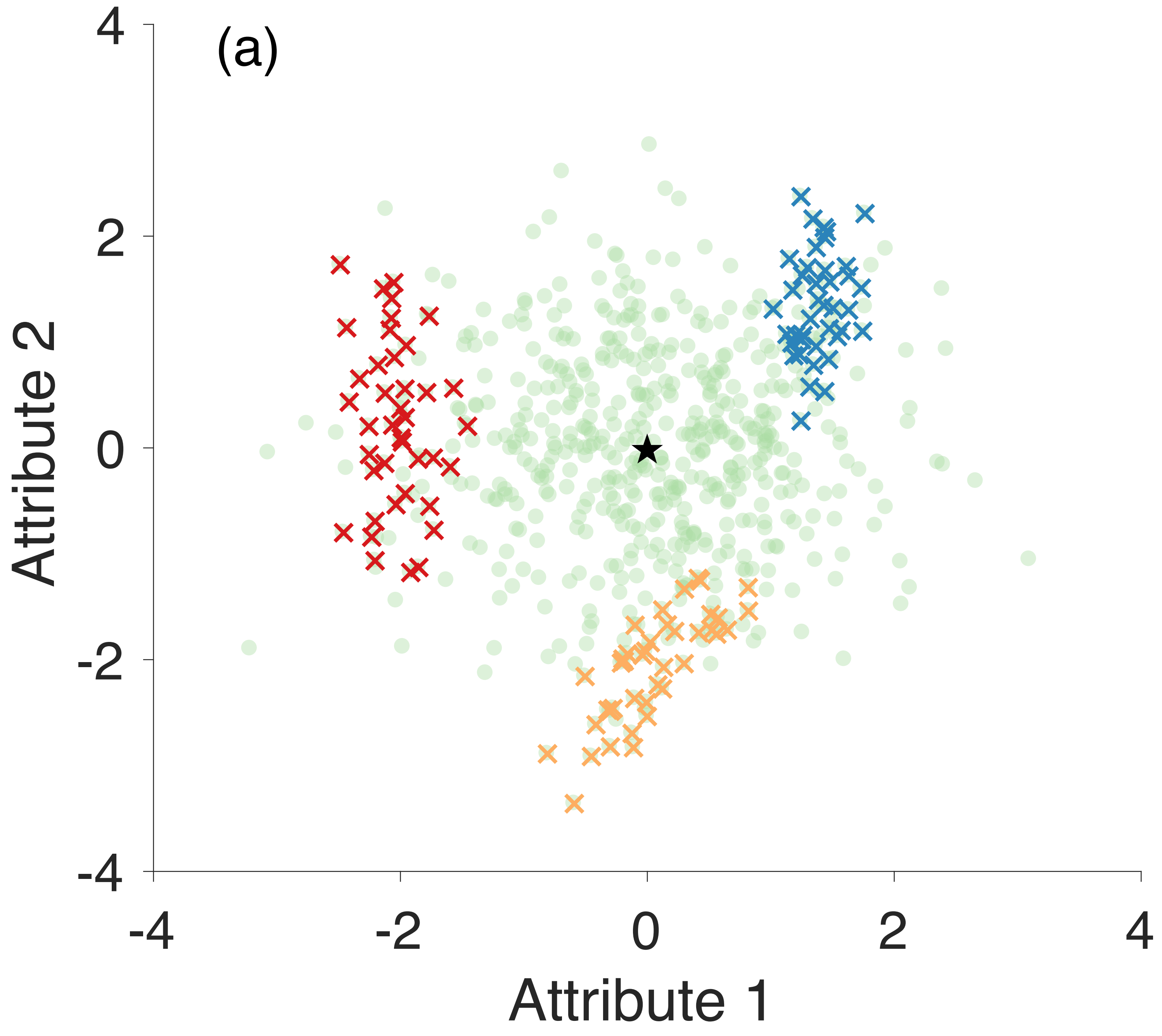


GHENT UNIVERSITY

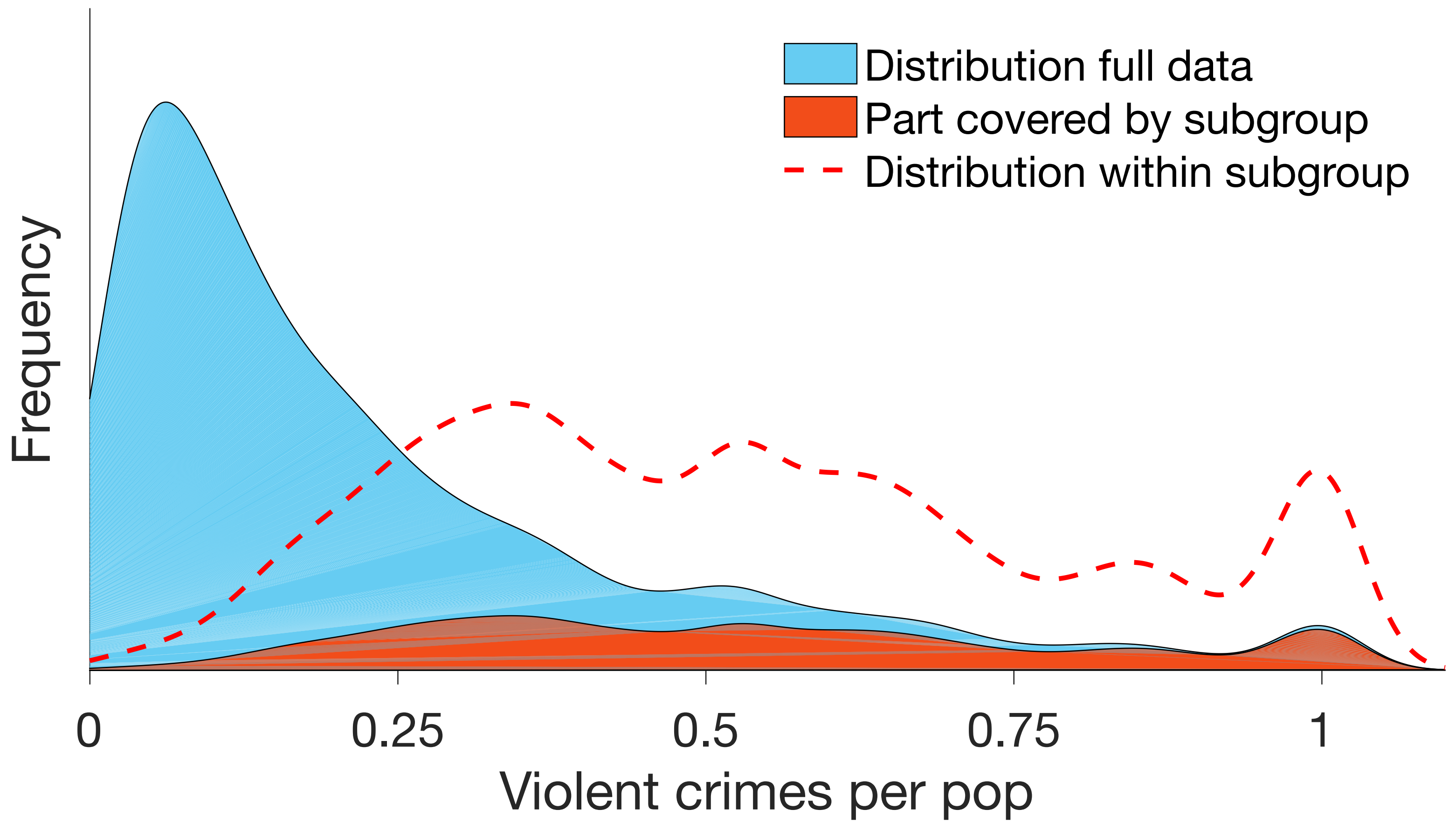
SUBGROUP DISCOVERY IN NUMERIC DATA

EXAMPLE

Lijffijt et al. *under review*



Condition Z → High crime rate



GHENT UNIVERSITY

PATTERN SYNTAX

Lijffijt et al. *under review*

Condition $Z \rightarrow$ means are $f(\hat{X}_Z)$

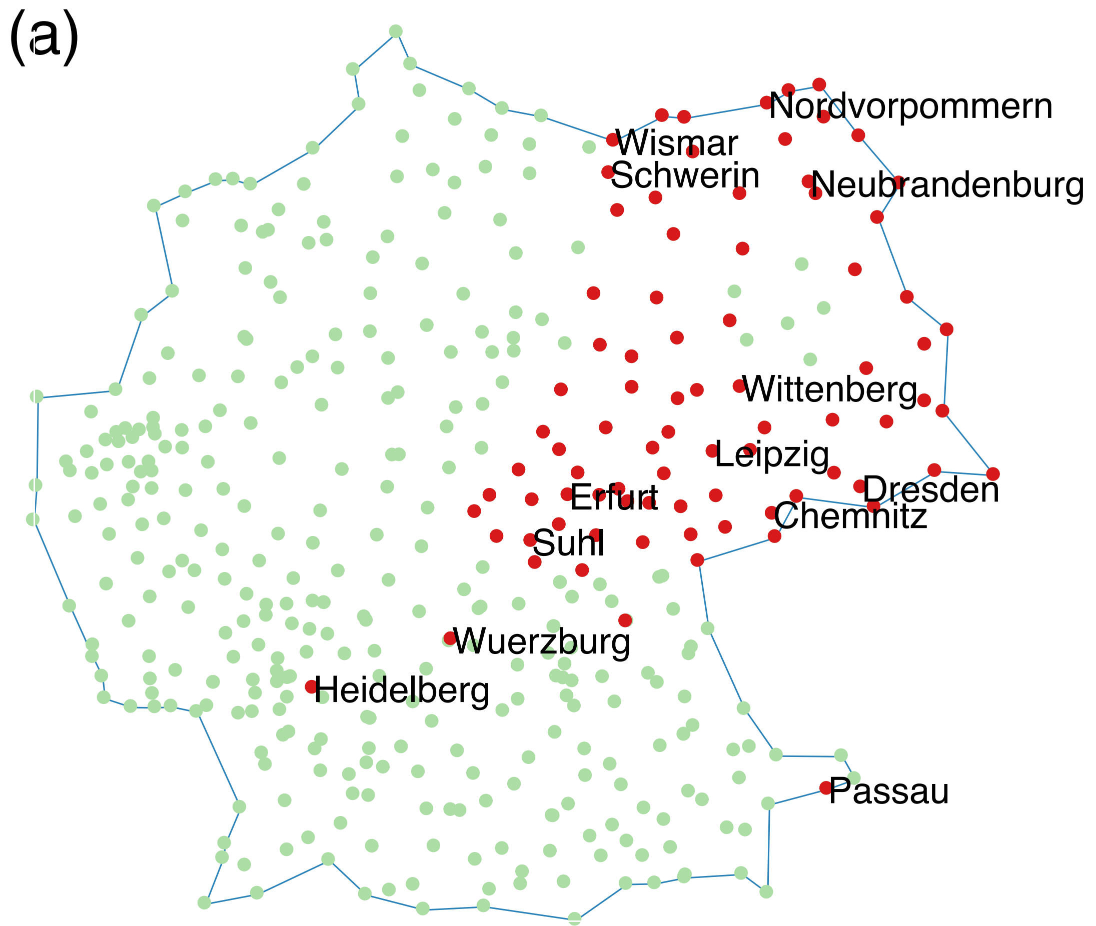
Condition $Z \rightarrow$ variance in direction w is $g(\hat{X}_Z, w)$

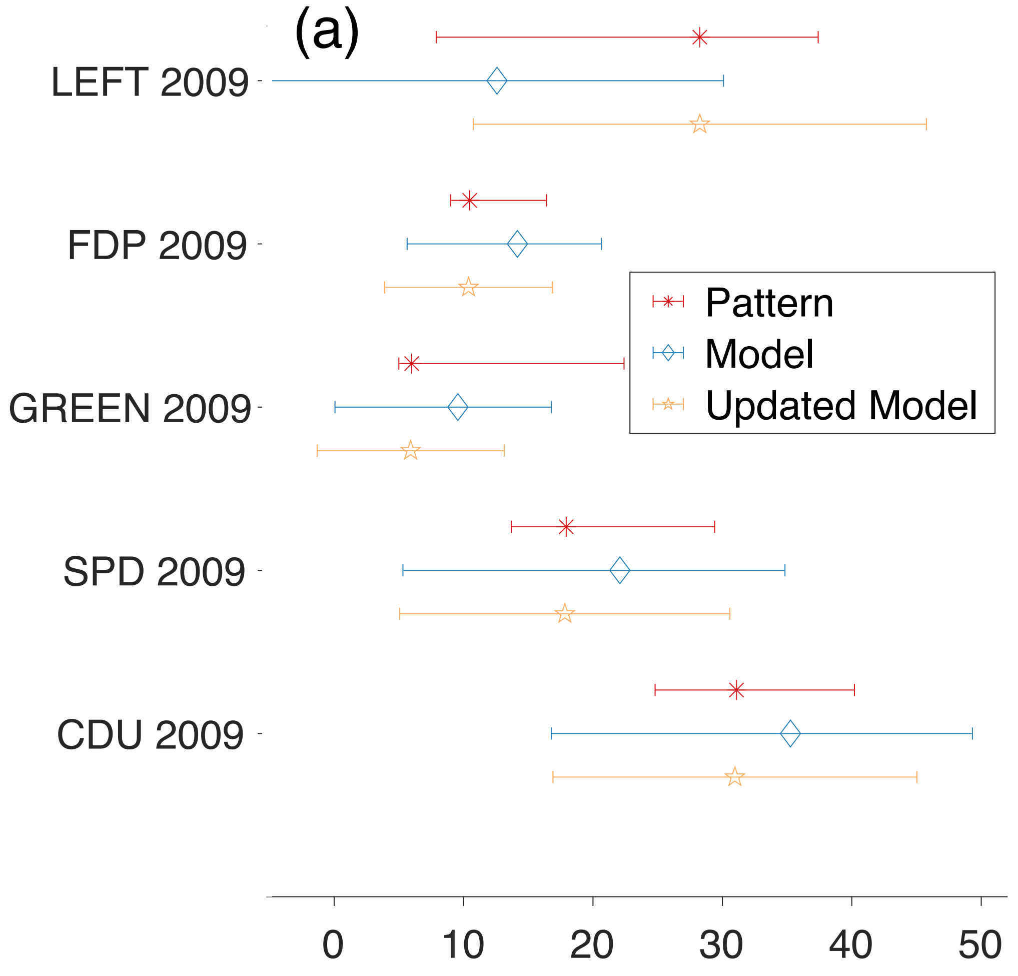
PATTERN SYNTAX

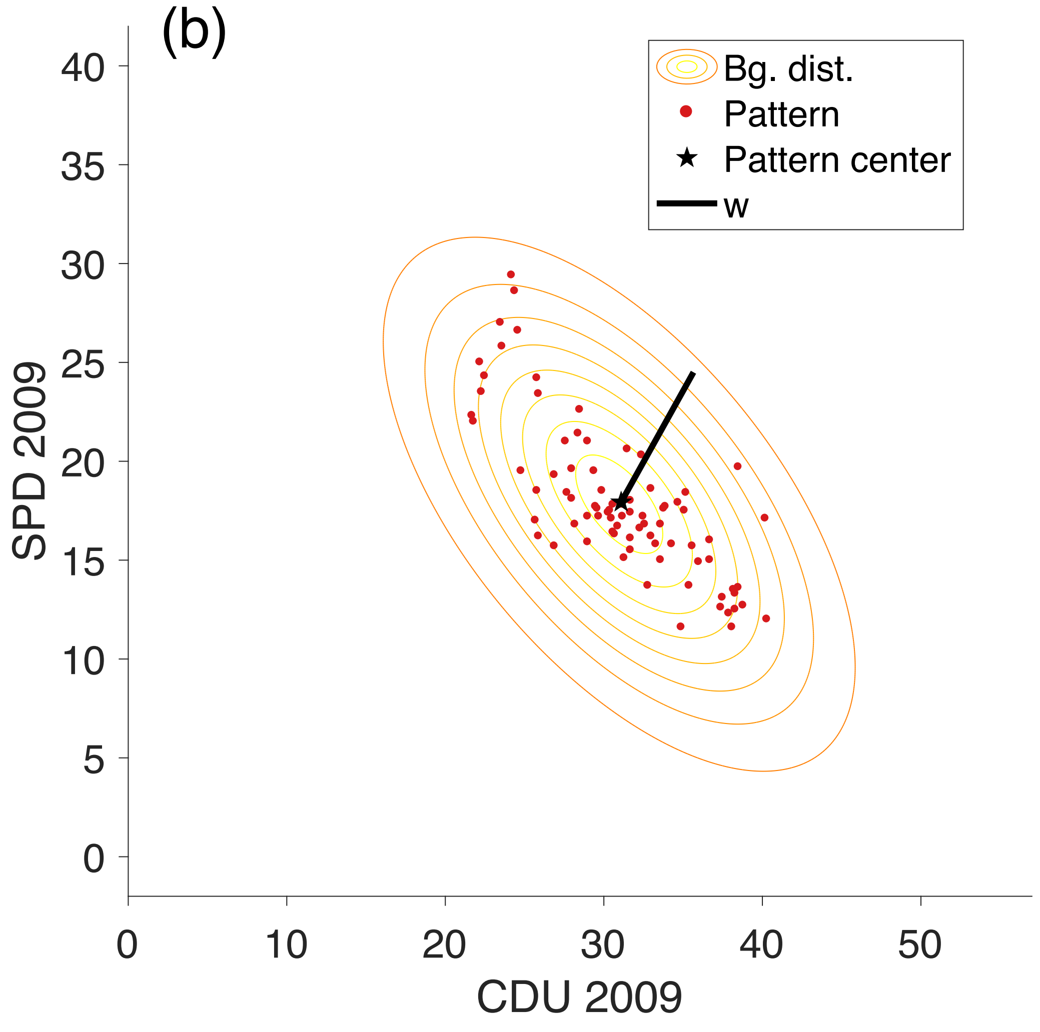
Lijffijt et al. *under review*

Condition $Z \rightarrow$ means are $f(\hat{X}_Z)$

Condition $Z \rightarrow$ variance in direction w is $g(\hat{X}_Z, w)$







GHENT UNIVERSITY

Condition: areas with few children

QUANTIFICATION

Surprisal of the location

$$\mathrm{IC}(f(\hat{X}_Z)) =$$
$$\log\left((2\pi)^d|\Sigma_Z|\right)/2 + \left(f(\hat{X}_Z - \mu_Z\right)^T \Sigma_Z^{-1}\left(f(\hat{X}_Z - \mu_Z\right)/2$$

Surprisal of the spread

$$\mathrm{IC}(g(\hat{X}_Z, w)) \approx \log\left(2^{\frac{m}{2}}\Gamma(m/2)\right)$$
$$+\alpha - (m/2-1)\log\left(\left(g(\hat{X}_Z, w - \beta\right)/\alpha\right) + \left(g(\hat{X}_Z, w - \beta\right)/(2\alpha)$$

Description length

$$\mathrm{DL}(\cdot) = \gamma|\mathrm{Cond}|(+\gamma) + \eta$$

GHENT UNIVERSITY

QUANTIFICATION

Too much algebra to explain

Surprisal of the location

$$\mathrm{IC}(f(\hat{X}_Z)) =$$
$$\log\left((2\pi)^d|\Sigma_Z|\right)/2 + \left(f(\hat{X}_Z - \mu_Z\right)^T \Sigma_Z^{-1}\left(f(\hat{X}_Z - \mu_Z\right)/2$$

Surprisal of the spread

$$\mathrm{IC}(g(\hat{X}_Z, w)) \approx \log\left(2^{\frac{m}{2}}\Gamma(m/2)\right)$$
$$+\,\alpha - (m/2-1)\log\left(\left(g(\hat{X}_Z, w - \beta\right)/\alpha\right) + \left(g(\hat{X}_Z, w - \beta\right)/(2\alpha)$$

Description length

$$\mathrm{DL}(\cdot) = \gamma|\mathrm{Cond}|(+\gamma) + \eta$$

GHENT UNIVERSITY

QUANTIFICATION

Too much algebra to explain

Surprisal of the location

$$\mathrm{IC}(f(\hat{X}_Z)) =$$
$$\log\left((2\pi)^d|\Sigma_Z|\right)/2 + \left(f(\hat{X}_Z - \mu_Z\right)^T \Sigma_Z^{-1}\left(f(\hat{X}_Z - \mu_Z\right)/2$$

Surprisal of the spread

$$\mathrm{IC}(g(\hat{X}_Z, w)) \approx \log\left(2^{\frac{m}{2}}\Gamma(m/2)\right)$$
$$+\,\alpha - (m/2-1)\log\left(\left(g(\hat{X}_Z, w-\beta\right)/\alpha\right) + \left(g(\hat{X}_Z, w-\beta\right)/(2\alpha)$$

Description length

$$\mathrm{DL}(\cdot) = \gamma|\mathrm{Cond}|(+\gamma) + \eta$$

Weigth times number of conditions + constant
(+spread has one term more than location)

GHENT UNIVERSITY

MINING

Lijffijt et al. *under review*

Finding subgroups:

Beam-search through off-the-shelf toolbox (Cortana)

Finding weight vector for spread pattern:

Manifold learning toolbox (ManOpt)

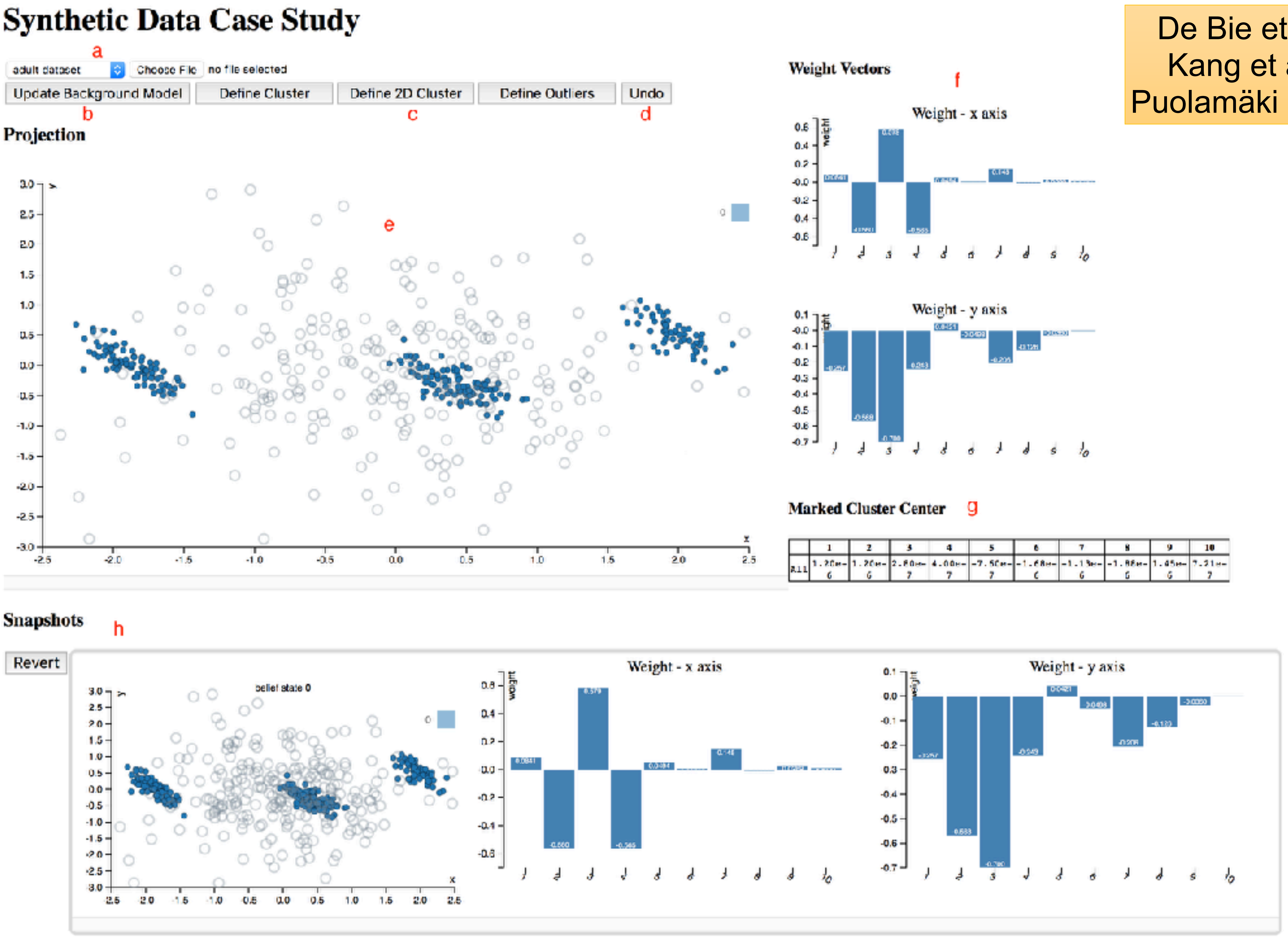
GHENT UNIVERSITY

VISUALIZATIONS AS PATTERNS

GHENT UNIVERSITY

EXAMPLE

De Bie et al. 2016
Kang et al. 2016
Puolamäki et al. 2016

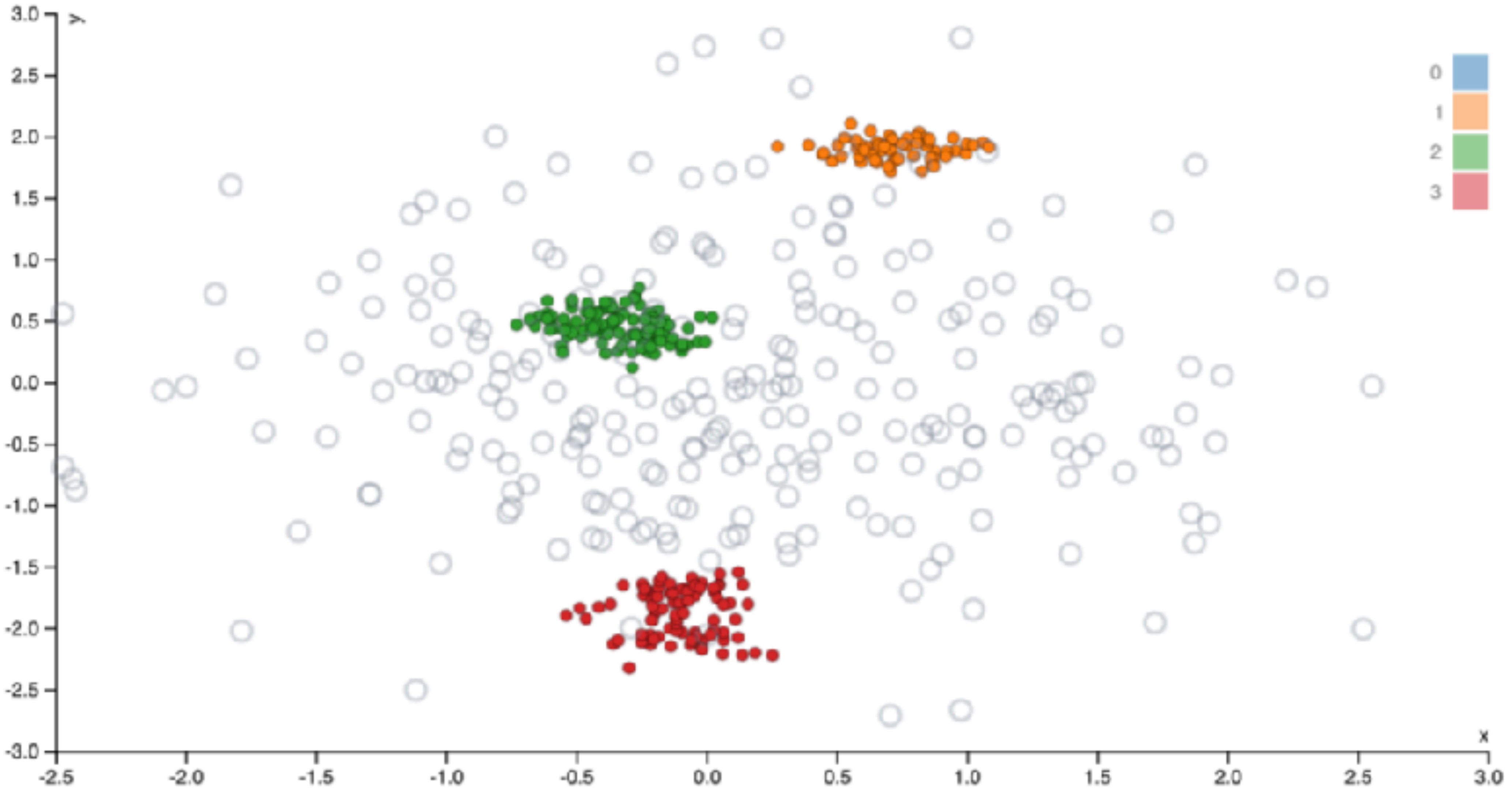


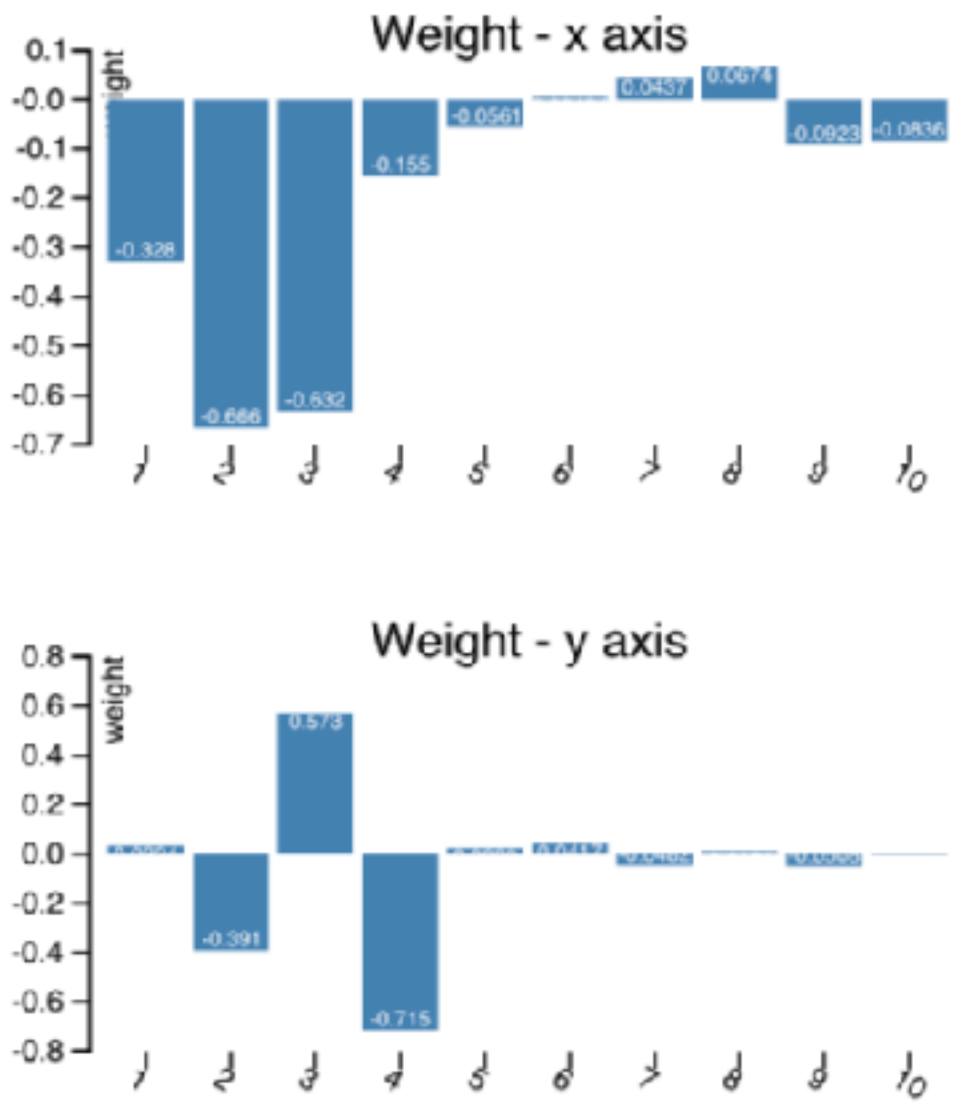
GHENT UNIVERSITY

PATTERN SYNTAX

De Bie et al. 2016
Kang et al. 2016
Puolamäki et al. 2016

Weight vector and (approximate) projection

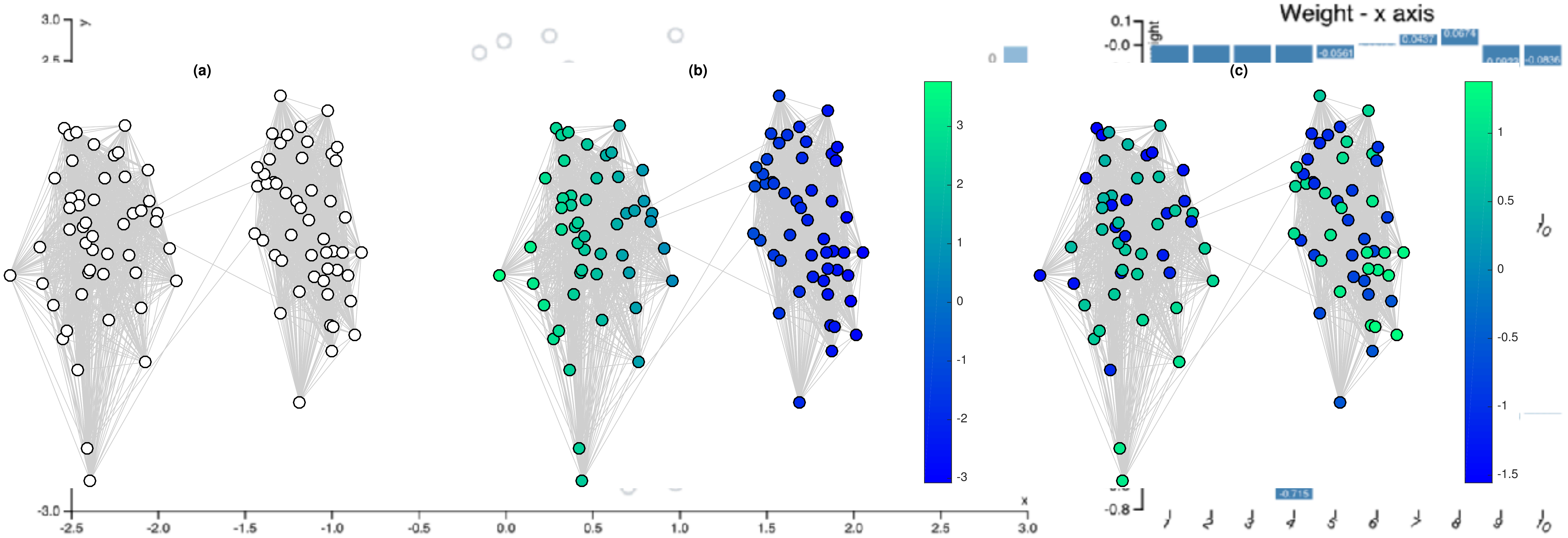




PATTERN SYNTAX

De Bie et al. 2016
Kang et al. 2016
Puolamäki et al. 2016

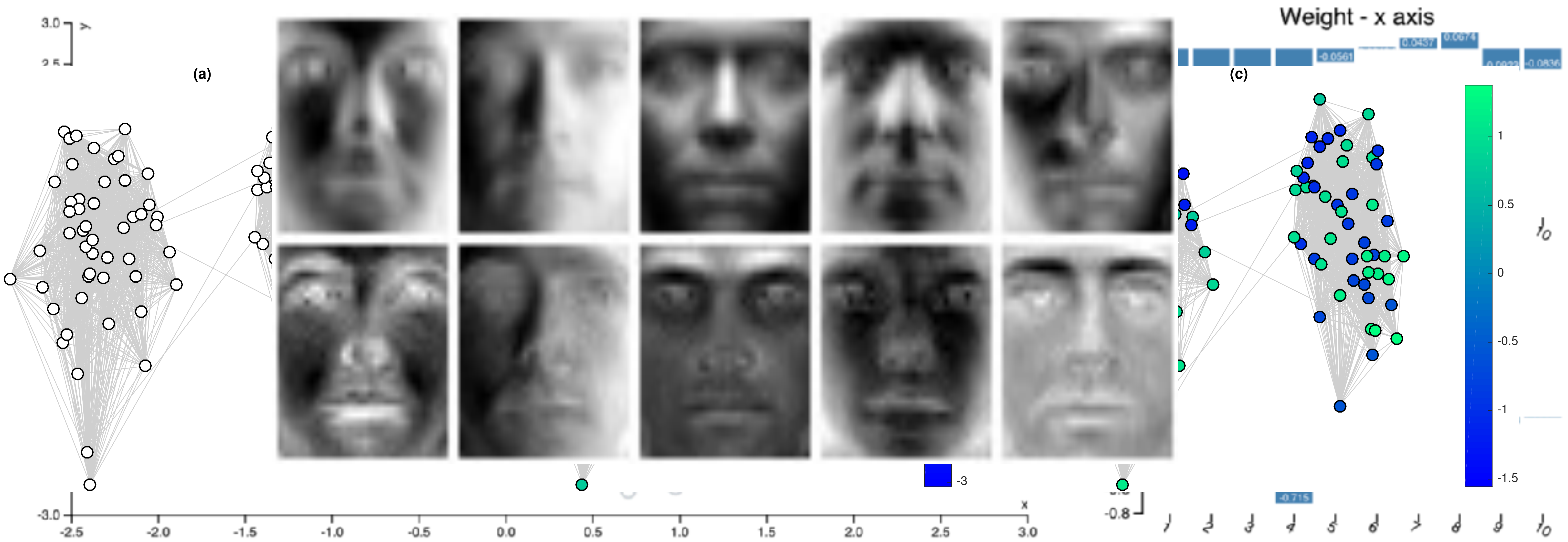
Weight vector and (approximate) projection



PATTERN SYNTAX

De Bie et al. 2016
Kang et al. 2016
Puolamäki et al. 2016

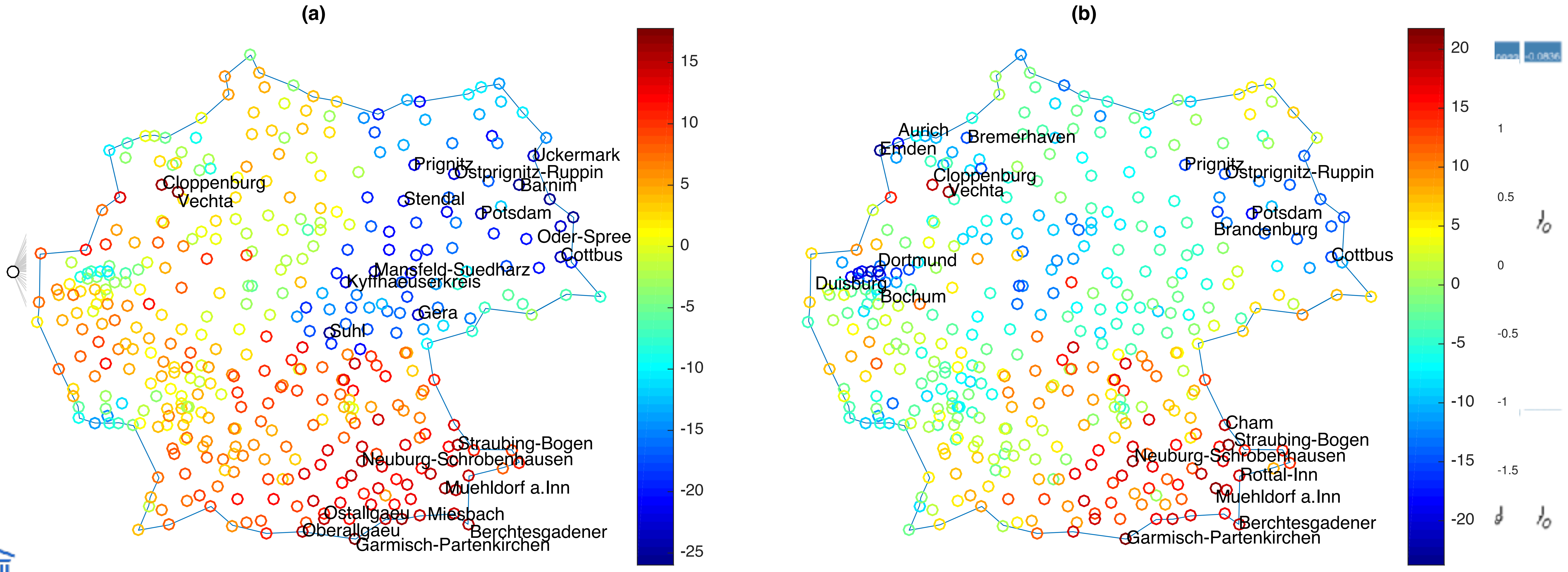
Weight vector and (approximate) projection



PATTERN SYNTAX

De Bie et al. 2016
Kang et al. 2016
Puolamäki et al. 2016

Weight vector and (approximate) projection

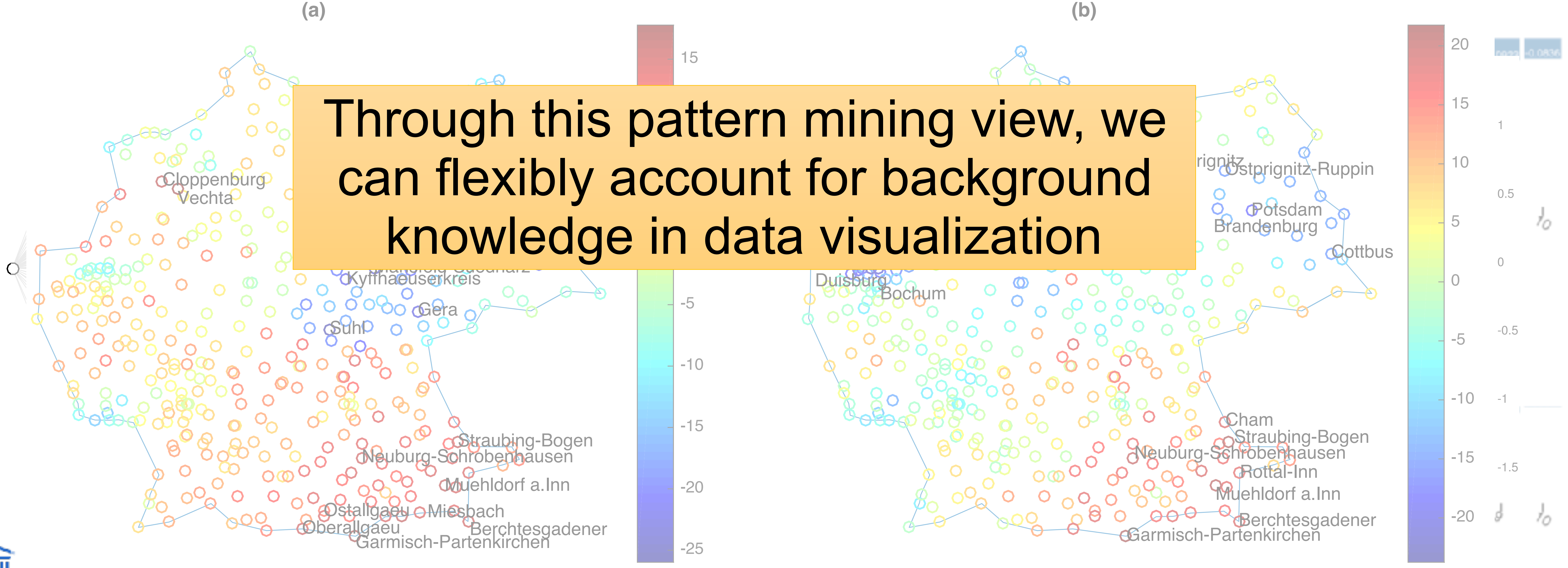


GHENT UNIVERSITY

PATTERN SYNTAX

De Bie et al. 2016
Kang et al. 2016
Puolamäki et al. 2016

Weight vector and (approximate) projection

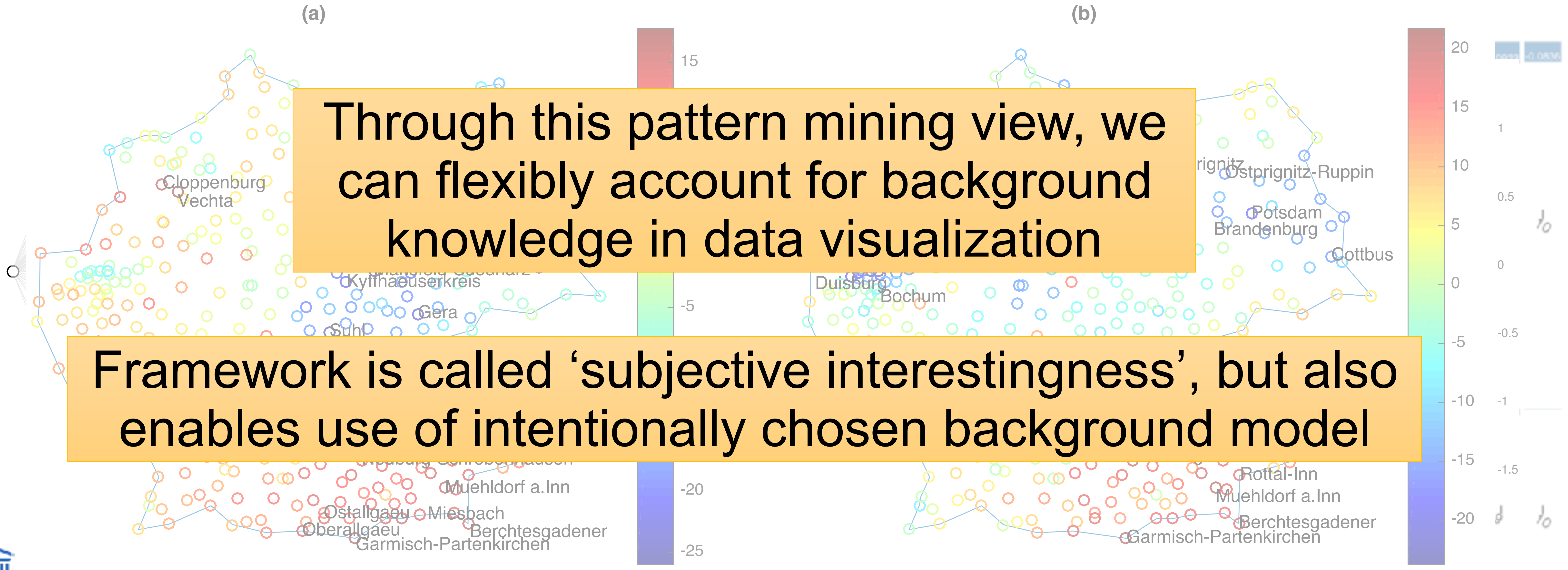


Through this pattern mining view, we can flexibly account for background knowledge in data visualization

PATTERN SYNTAX

De Bie et al. 2016
Kang et al. 2016
Puolamäki et al. 2016

Weight vector and (approximate) projection



Through this pattern mining view, we can flexibly account for background knowledge in data visualization

Framework is called 'subjective interestingness', but also enables use of intentionally chosen background model

GHENT UNIVERSITY

MINING

De Bie et al. 2016
Kang et al. 2016
Puolamäki et al. 2016

Background model inference with gradient descent

Mining patterns is a manifold learning problem

– Optimize through toolbox (ManOpt)

GHENT UNIVERSITY

OPEN CHALLENGES

GHENT UNIVERSITY

THE FORSIED PROCESS

1. Background model
2. Pattern syntax
   – IC & DL of patterns
3. Mining
4. Update background model
5. Iterative mining

BACKGROUND MODEL

How to provide specification as end-user

- Objective
- Subjective

Parameter inference (of MaxEnt model)

GHENT UNIVERSITY

PATTERN SYNTAX

Should correspond to the task

- Probably constructed by researchers
- Users need to be able to select appropriate syntax

Not yet investigated: syntaxes with mixed relevance

- E.g., for graphs with various vertex types

IC & DL OF PATTERNS

Requires extensive knowledge of maths

Not related to cognition yet in any way

MINE GOOD/BEST PATTERNS

All the usual algorithmic challenges

- Typically NP-hard
  - Depends on b.g. model and pattern syntax
- Search strategies: greedy, beam search, branch & bound, SGD, black-box optimization
  - Consider bounds / approximability

UPDATE BACKGROUND MODEL

Insert what the user has learned

How about forgetting?

What kind of inference capabilities do humans have?

– Track remaining degree of vertices in graph, degrees of freedom for projections?

GHENT UNIVERSITY

ITERATIVE MINING

May be possible to re-use state of previous mining step

- E.g., in branch-and-bound
- Non-overlapping patterns are generally unaffected

Iterative scheme has an 1-1/e bound for greedy

- For the budget used so far

SUMMARY

SUMMARY

A generic approach to define patterns

Information Theory perspective on interestingness

Examples of mining interesting/surprising

- Dense and connecting vertex sets in graphs
- Subgroups in numeric data
- Projections for visualization

Many open challenges remaining

FACULTY OF ENGINEERING
AND ARCHITECTURE

Jefrey Lijffijt, Dr. Sc. (Tech.)

FWO [Pegasus]² MSC Fellow

Internet Technology and Data Science Lab
Dept. of Electronics and Information Systems

E jefrey.lijffijt@ugent.be
M +32 474 53 94 64

users.ugent.be/~jlijffij/

jlijffijt
@jlijffijt
jefrey

GHENT
UNIVERSITY

FACULTY OF ENGINEERING
AND ARCHITECTURE

Jefrey Lijffijt, Dr. Sc. (Tech.)

FWO [Pegasus]² MSC Fellow

Internet Technology and Data Science Lab
Dept. of Electronics and Information Systems

E jefrey.lijffijt@ugent.be
M +32 474 53 94 64

users.ugent.be/~jlijffij/

jlijffijt
@jlijffijt
jefrey

Interested? Visit us!
And we are hiring :-)

GHENT
UNIVERSITY