(Geometric) Formal Concept Analysis

Literature

1982 Rudolf Wille

Restructuring lattice theory: an approach based on hierarchies of concepts

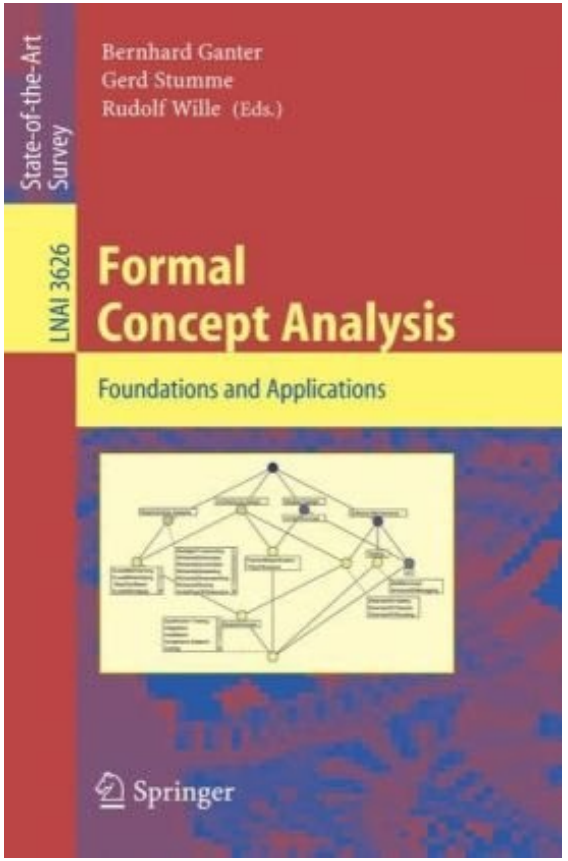
2005 Ganter, Stumme, and Wille

Formal Concept Analysis

2016 Ganter and Obiedkov

Conceptual Exploration

+ much more



Raw data, empirical observations

	rational	emotional	honest	dishonest	optimistic	pessimistic	interested	uninterested	flexible	inflexible	materialistic	idealistic	not fashionable	fashionable	fond of life	depressive	purposeful	unsteady	unconstrained	constrained
Myself	X		X			X		X		X		X	X			X		X		X
Ideal	X		X		X		X		X			X	X		X		X		X	
Father	X		X		X		X		X			X	X		X		X		X	
Mother		X	X			X	X		X		X			X		X		X	X	
Sister		X	X		X		X		X			X	X		X			X	X	
Brother	X		X		X		X		X		X			X	X		X		X	
Otto	X		X			X	X			X	X			X	X			X		X
Anne		X	X			X		X		X	X			X	X			X		X
Eva	X		X		X			X	X			X	X			X		X		X
Elke		X		X	X		X		X		X			X	X		X			X
Ina		X	X		X		X		X			X		X	X		X		X	

Concepts

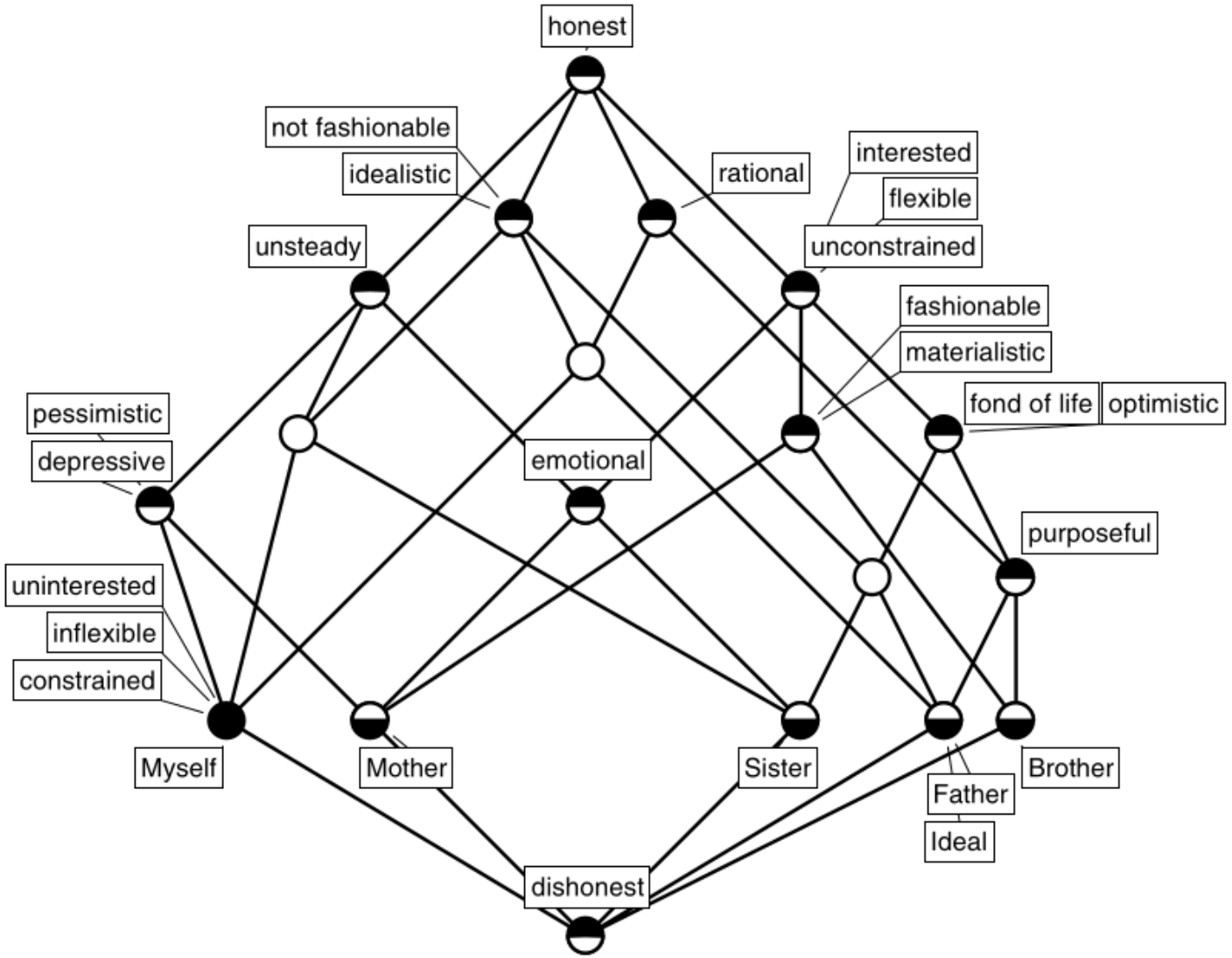
The *extension/extent* of a set of attributes is all the objects having them.

The *intension/intent* of a set of objects is all the attributes they have.

A *concept* is a "perfect" set of objects O and attributes A:

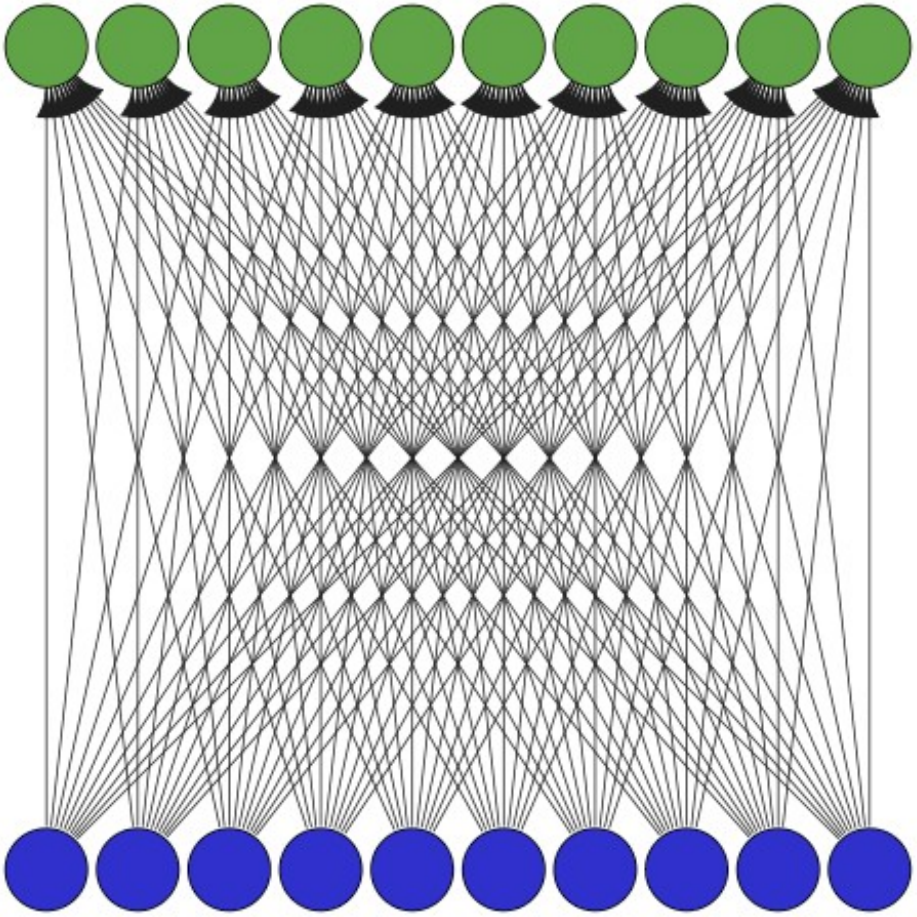
- O is the extent of A
- A is the intent of O

Concept lattice



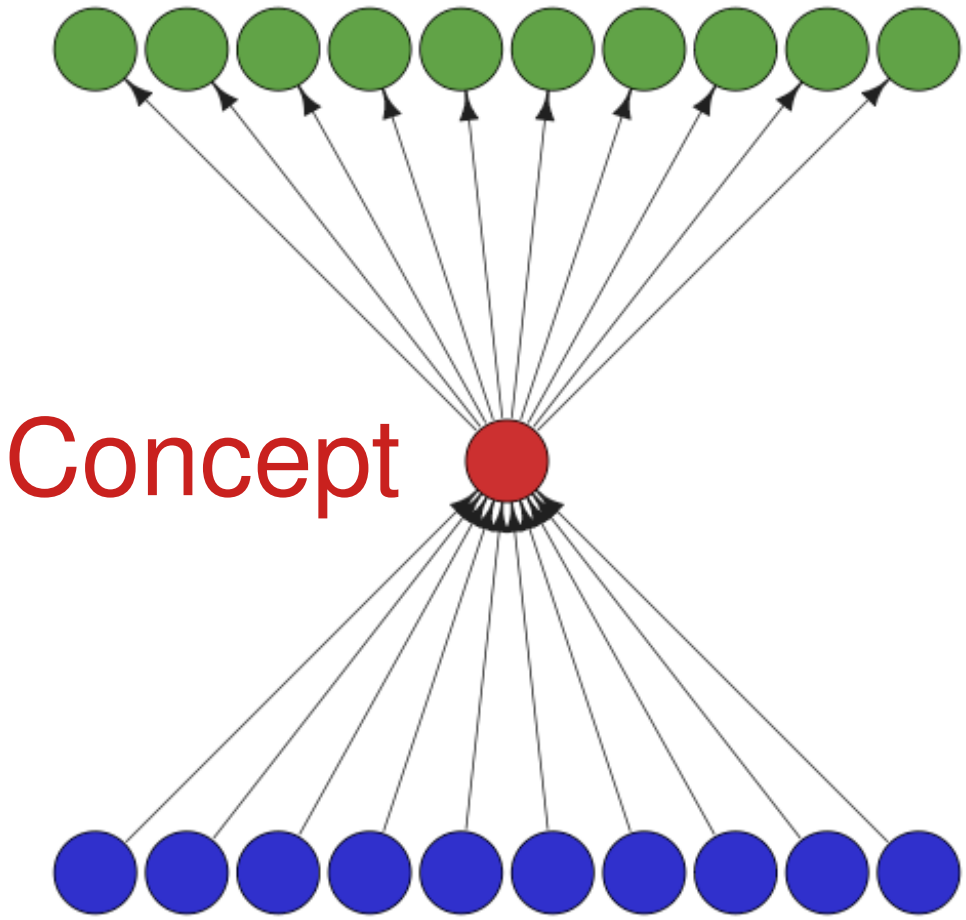
Raw data, empirical observations

Attributes

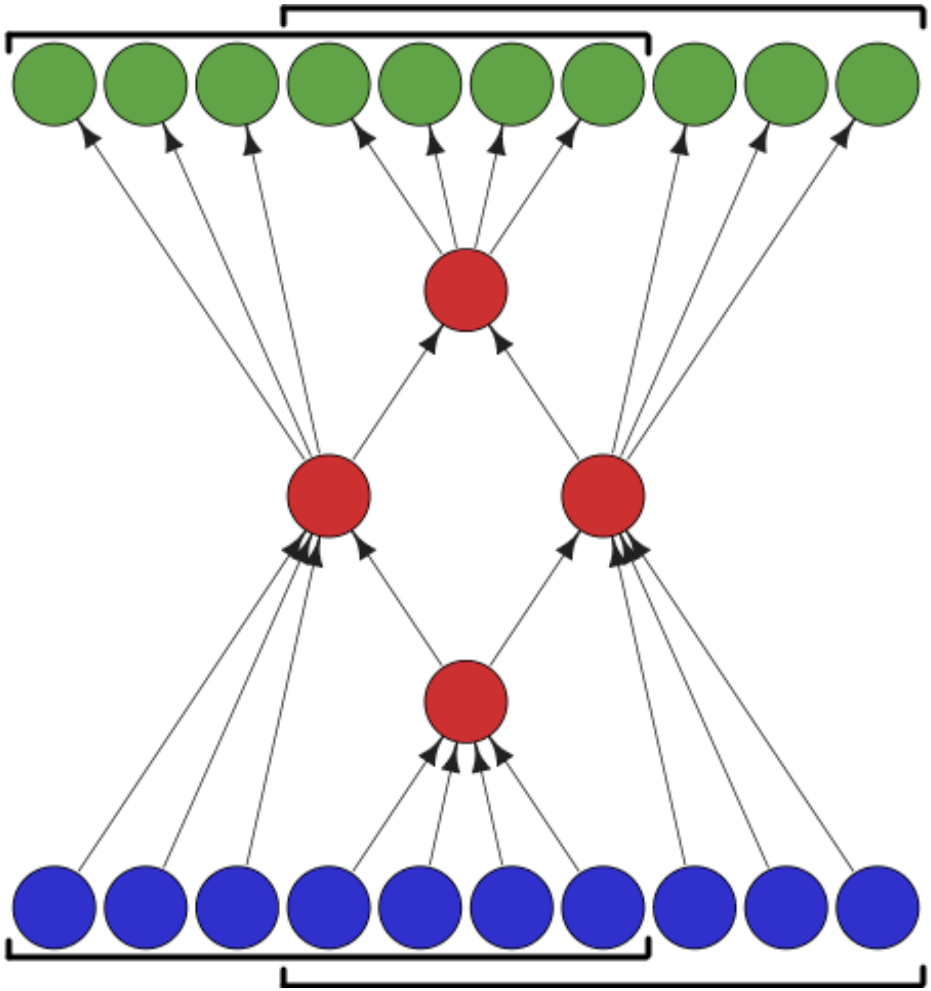


Objects

Concept binding



Concept lattice, simultaneous binding



Theory and algorithms

***Theorem* (Wille)**

The set of concepts forms a complete lattice, i.e. it contains all infimima and suprema.

(interpret concept lattice as closure system)

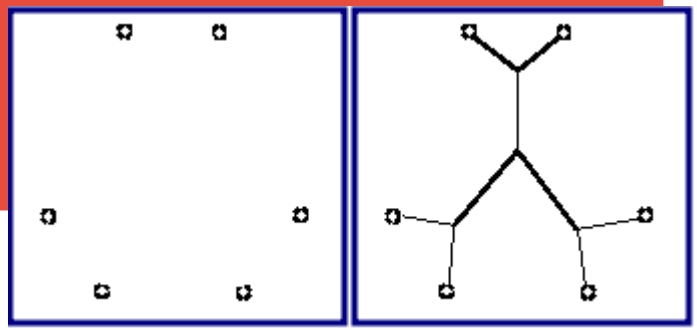
***Algorithm 1* (e.g. Ganter-Obiedkov 2016)**

Find all closed sets by recursive pairwise union.

***Algorithm 2* (e.g. Ganter-Obiedkov 2016)**

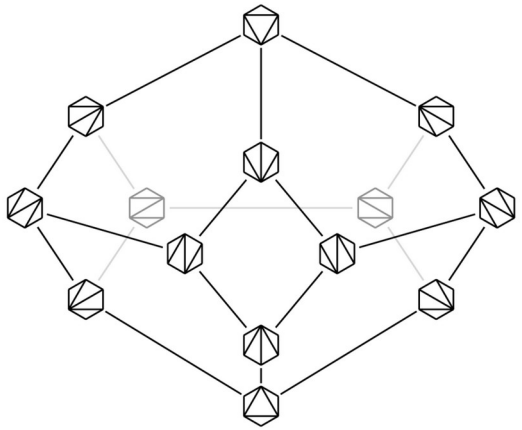
Find all closed sets given attribute ordering, in “lectic order”.

New directions



Theory

- Geometric Steiner problem on data cube
- Recursive hierarchy (concepts as objects and as attributes)
- Adjoint functor formulation
- Algebraic topology of description diagrams. Descriptohedron?



Practice

- Weak/heuristic closure with relaxation parameter
- Weakly supervised version using outcome-based attribute ranking, consider only first lectic concepts

Categorical Description Clustering

Sample data | Data from file

Sample data

pam50_tcga_rand.csv

Dichotomization

- (•) Median
- ( ) 2-population GMM

Relaxation α value

0 — 0.95

Number of sets per level

1 — 400 — 1,000

Number of levels

1 — 4 — 6

Number of feature groups

0 — 200 — 400

Start calculation