Generalizing rings of sets in a star-spangled manner: Star-shaped subgroup discovery

Johan Christian Clausen Dahl: Blick auf Dresden bei Vollmondschein, Bearbeitung GS

Part I: Formal concept analysis (FCA)

RESTRUCTURING LATTICE THEORY:
AN APPROACH BASED ON HIERARCHIES OF CONCEPTS

Rudolf Wille
Fachbereich Mathematik
Technische Hochschule Darmstadt
6100 Darmstadt
Federal Republic of Germany

ABSTRACT

Lattice theory today reflects the general status of current mathematics: there is a rich production of theoretical concepts, results, and developments, many of which are reached by elaborate mental gymnastics; on the other hand, the connections of the theory to its surroundings are getting weaker and weaker, with the result that the theory and even many of its parts become more isolated. Restructuring lattice theory is an attempt to reinvigorate connections with our general culture by interpreting the theory as concretely as possible, and in this way to promote better communication between lattice theorists and potential users of lattice theory.

Formal concept analysis (FCA)

Given: formal context $\mathbb{K} := (G, M, I)$ where

- G is a set of objects,
- M is a set of attributes,
- $I \subseteq G \times M$ is a binary relation with the interpretation $(g, m) \in I$ iff object g has attribute m.
- Aim: Describe I with the help of so-called formal concepts.

Definition (formal concept)

Let $\mathbb{K} := (G, M, I)$ be a formal context. A pair (A, B) where $A \subseteq G$ is a set of objects and $B \subseteq M$ is a set of attributes is called a formal concept if

1. All objects in A have all attributes in B.
2. The set A is maximal w.r.t. property 1.
3. The set B is maximal w.r.t. property 1.

In such a case, we call A the extent and B the intent of the formal concept (A, B).

Remark

If we define for $A \subseteq G$ and $B \subseteq M$ the associated sets

$$A' = \{m \in M \mid \forall g \in A : (g, m) \in I\}$$
$$B' = \{g \in G \mid \forall m \in B : (g, m) \in I\},$$

then a pair (A, B) with $A \subseteq G$ and $B \subseteq M$ is a formal concept iff $B = A'$ and $A = B'$.

Example

	verletzlich	verschlossen	selbstsicher	pflichtbewusst	herzlich	schwierig	aufmerksam	leicht beleidigt	nicht jähzornig	ängstlich	geschwätzig	oberflächlich	sensibel	ehrgeizig
Ich selbst	×	×	×		×	×	×		×	×			×	×
Mein Ideal	×		×	×	×		×		×				×	×
Vater	×	×		×	×	×	×	×	×	×		×	×	×
Mutter	×	×		×	×	×		×	×	×		×	×	×
Schwester	×	×		×	×	×	×		×	×			×	×
Schwager			×	×	×		×				×	×		×

Interviewdaten aus einer Therapie von Anorexia nervosa

[Ganter, 2013, p. 62]

Example

	verletzlich	verschlossen	selbstsicher	pflichtbewusst	herzlich	schwierig	aufmerksam	leicht beleidigt	nicht jähzornig	ängstlich	geschwätzig	oberflächlich	sensibel	ehrgeizig
Ich selbst	×	×	×		×	×	×		×	×			×	×
Mein Ideal	×		×	×	×		×		×				×	×
Vater	×	×		×	×	×	×	×	×	×		×	×	×
Mutter	×	×		×	×	×		×	×	×		×	×	×
Schwester	×	×		×	×	×	×		×	×			×	×
Schwager			×	×	×		×				×	×		×

Interviewdaten aus einer Therapie von Anorexia nervosa

[Ganter, 2013, p. 62]

Example

	verletzlich	verschlossen	selbstsicher	pflichtbewusst	herzlich	schwierig	aufmerksam	leicht beleidigt	nicht jähzornig	ängstlich	geschwätzig	oberflächlich	sensibel	ehrgeizig
Ich selbst	×	×	×		×	×	×		×	×			×	×
Mein Ideal	×		×	×	×		×		×				×	×
Vater	×	×		×	×	×	×	×	×	×		×	×	×
Mutter	×	×		×	×	×		×	×	×		×	×	×
Schwester	×	×		×	×	×	×		×	×			×	×
Schwager			×	×	×		×				×	×		×

Interviewdaten aus einer Therapie von Anorexia nervosa

[Ganter, 2013, p. 62]

The complete lattice of formal concepts

Theorem (Basic theorem on concept lattices)

Let $\mathbb{K} := (G, M, I)$ be a formal context. The set of all formal concepts, together with the subconcept-relation

$$(A, B) \leq (C, D) :\iff A \subseteq C \ \& \ B \supseteq D$$

builds a complete lattice where the infimum and the supremum of an arbitrary family $(A_i, B_i)_{i \in I}$ of formal concepts are given by

$$\bigwedge_{i \in I} (A_i, B_i) = \left(\bigcap_{i \in I} A_i, \left(\bigcap_{i \in I} A_i \right)'' \right)$$

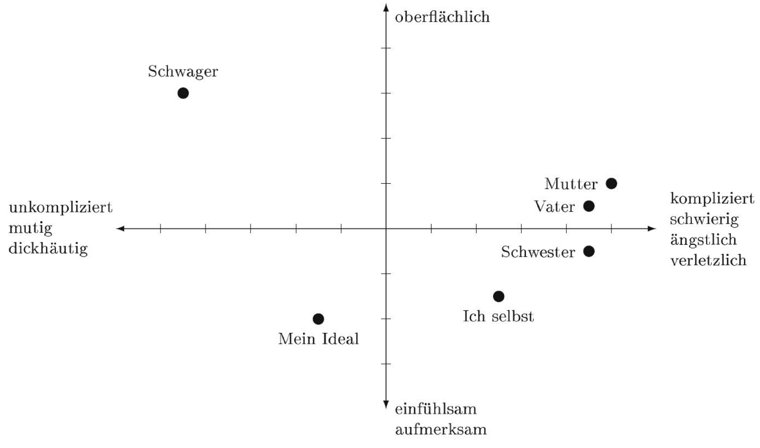
$$\bigvee_{i \in I} (A_i, B_i) = \left(\left(\bigcap_{i \in I} B_i \right)'', \bigcap_{i \in I} B_i \right).$$

The complete lattice of all formal concepts of a formal context $\mathbb{K}$ is denoted with $\mathfrak{B}(\mathbb{K})$.

Remark

From the basic theorem it follows that the family of all concept extents is closed under arbitrary intersections. Furthermore, the whole set G is always a concept extent. Thus, the family of all concept extents is a closure system. (A closure system on a basic set V is a family $\mathcal{S} \subseteq 2^V$ which contains V and that is closed under arbitrary intersections, relevant later...)

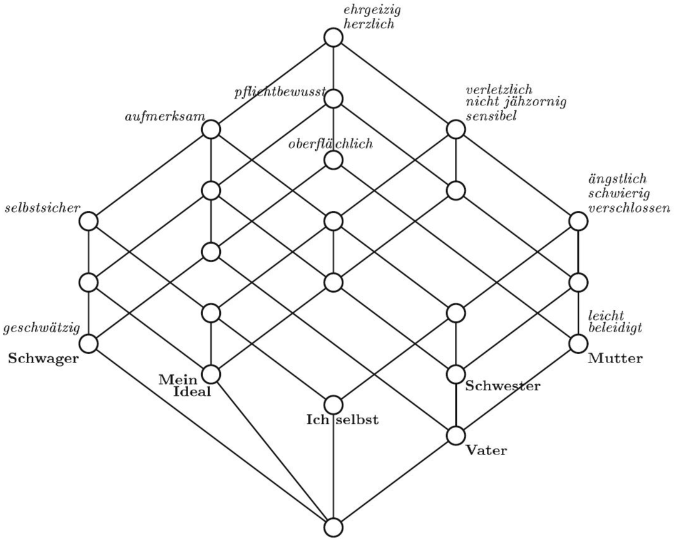
Example



Ein Biplot von Interviewdaten

[Ganter, 2013, p. 62]

Example



Der Begriffsverband der Interviewdaten

[Ganter, 2013, p. 63]

Part II: Subgroup discovery (SD)

Subgroup discovery (SD)

Definition (Subgroup discovery)

"In subgroup discovery, we assume we are given a so-called population of individuals (objects, customer,...) and a property of those individuals we are interested in. The task of subgroup discovery is then to discover the subgroups of the population that are statistically "most interesting" i.e. are as large as possible and have the most unusual statistical (distributional) characteristics with respect to the property of interest." [Wrobel, 2001]

Subgroup discovery in the language of formal concept analysis

Problem statement

Given a formal context $\mathbb{K} = (G, M, I)$, a target variable $t : G \longrightarrow \mathcal{A}$ and a quality function $q_t : 2^G \longrightarrow \mathbb{R}$, which "measures" the interestingness of subsets of objects w.r.t. the target t, the task of subgroup discovery is to find that k formal concepts (A, B), for which the quality-values

$$q_t(A)$$

are the largest. In the sequel: $k = 1$, so (ignoring non-uniqueness issues) we want to compute

$$D^+ := \max_{(A,B)\in\mathfrak{B}(\mathbb{K})} q_t(A)$$

(and a corresponding argmax).

Example (of classical subgroup discovery: Dimensions of justice and election intentions in the ALLBUS 2014)

- $G = \{g_1, \ldots, g_{1412}\}$ set of 1412 respondents as objects.
- $M = \{m_1, \ldots, m_8\}$ set of 8 statements about social justice that intend to measure 4 dimensions of social justice.
- $I \subseteq G \times M$ binary relation with interpretation $(g, m) \in I$ if respondent g agrees to statement m.
- target $t : G \longrightarrow \{0, 1\}$:

$$g \mapsto \begin{cases} 1 & \text{if respondent } g \text{ intents to elect e.g., CDU-CSU} \\ 0 & \text{else} \end{cases}$$

- quality function $q_t : 2^G \longrightarrow \mathbb{R} : A \mapsto n^{\alpha} \cdot (p - p_0)$, where

$$\alpha \in [0, 1];\ n = |A|;\ p_0 = \frac{\sum_{g \in G} t(g)}{|G|} \text{ and } p = \frac{\sum_{g \in A} t(g)}{|A|}.$$

Dimensionen sozialer Gerechtigkeit (cf., [Liebig and May, 2009])

Leistungsprinzip (principle of achievement):

Es ist gerecht, wenn Personen, die im Beruf viel leisten, mehr verdienen als andere.

Gerecht ist, wenn jede Person nur das bekommt, was sie sich durch eigene Anstrengungen erarbeitet hat.

(It is fair if people who perform well in their jobs earn more than others.

It is fair when each person receives only what he or she has earned through his or her own efforts.)

Gleichheitsprinzip (principle of equality):

Gerecht ist, wenn alle die gleichen Lebensbedingungen haben.

Es ist gerecht, wenn Einkommen und Vermögen in unserer Gesellschaft an alle Personen gleich verteilt sind.

(It is fair when everyone has the same living conditions.

It is fair when income and wealth in our society are distributed equally to all persons.)

Dimensionen sozialer Gerechtigkeit

Anrechtsprinzip (principle of entitlement):

Es ist gerecht, wenn Personen, die aus angesehenen Familien stammen, dadurch Vorteile im Leben haben.

Es ist gerecht, wenn diejenigen, die in einer Gesellschaft oben stehen, bessere Lebensbedingungen haben als diejenigen, die unten stehen. (It is just when people who come from respected families have advantages in life as a result. It is just for those at the top of a society to have better living conditions than those at the bottom.)

Bedarfsprinzip (principle of demand):

Eine Gesellschaft ist gerecht, wenn sie sich um die Schwachen und Hilfsbedürftigen kümmert.

Es ist gerecht, wenn Personen, die Kinder oder pflegebedürftige Angehörige zu versorgen haben, besondere Unterstützung und Vergünstigungen erhalten. (A society is just when it cares for the weak and those in need of help. It is just when people who have children or dependents in need of care receive special support and benefits.)

response scale (details not important):

0 Keine Teilnahme an Split B (Code 1 in V4)

5 stimme voll zu

4 stimme etwas zu

3 weder noch

2 lehne etwas ab

1 lehne ganz ab

9 Keine Angabe

response scale target attribute t (details not important):

0 Nicht wahlberechtigt, da keine deutsche Staatsbürgerschaft

1 CDU bzw. CSU

2 SPD

3 FDP

4 Bündnis 90/Die Grünen

6 Die Linke

20 NPD

41 Piratenpartei

42 AfD (Alternative für Deutschland)

90 Andere Partei, und zwar:...

91 Würde nicht wählen

97 Angabe verweigert

98 Weiß nicht

99 Keine Angabe

Results (details not important, its only an illustration):

	CDU-CSU	SPD	FDP	GRUENE	LINKE	NPD	PIRATEN	AFD	A.P.	N.W.
p_0	0.311	0.227	0.042	0.128	0.093	0.012	0.023	0.055	0.015	0.093
p	0.385	0.316	0.130	0.213	0.131	0.030	0.058	0.082	0.040	0.157
n	771	320	193	488	716	364	363	680	272	338
q_t	57.29	28.48	16.80	41.45	27.07	6.62	12.52	18.44	6.96	21.64
$p - p_0$	0.074	0.089	0.087	0.085	0.038	0.018	0.034	0.027	0.026	0.064
$\frac{p-p_0}{p_0}$	0.239	0.393	2.048	0.663	0.404	1.510	1.475	0.491	1.719	0.690

CDU:	"UM.SCHWAECHERE.KUEMMERN"	≥ 2
	"BEKOMMEN..WAS.ERARBEITET.WURDE"	≥ 2
	"UNTERSTUETZUNG.VON.PFLEGENDEN"	≥ 2
	"EINKOMMEN.GLEICH.VERTEILT"	≤ 2
SPD:	"MEHR.LEISTUNG..MEHR.VERDIENST"	≥ 3
	"GLEICHE.LEBENSBEDINGUNGEN"	≥ 2
	"UM.SCHWAECHERE.KUEMMERN"	≥ 3
	"BEKOMMEN..WAS.ERARBEITET.WURDE"	≥ 3
	"UNTERSTUETZUNG.VON.PFLEGENDEN"	≥ 3
	"EINKOMMEN.GLEICH.VERTEILT"	≥ 3
	"WENN.OBENSTEHENDE.BESSER.LEBEN"	≥ 2

FDP:	"MEHR.LEISTUNG..MEHR.VERDIENST"	≥ 5
	"UM.SCHWAECHERE.KUEMMERN"	≥ 3
	"UNTERSTUETZUNG.VON.PFLEGENDEN"	≥ 4
	"EINKOMMEN.GLEICH.VERTEILT"	≤ 1
	"EINKOMMEN.GLEICH.VERTEILT"	≤ 3
	"EINKOMMEN.GLEICH.VERTEILT"	≤ 4
	"WENN.OBENSTEHENDE.BESSER.LEBEN"	≥ 3
GRUENE:	"MEHR.LEISTUNG..MEHR.VERDIENST"	≥ 2
	"VORTEILE.DURCH.HERKUNFT"	≤ 3
	"BEKOMMEN..WAS.ERARBEITET.WURDE"	≤ 4
	"UNTERSTUETZUNG.VON.PFLEGENDEN"	≥ 4
	"WENN.OBENSTEHENDE.BESSER.LEBEN"	≤ 4

Computing $D^+ := \max_{(A,B)\in\mathfrak{B}(\mathbb{K})} q_t(A)$

One approach in subgroup discovery

defines a search-space of all subgroups described by subgroup descriptions (i.e., value-specifications of certain attributes) and then scans through all subgroup descriptions from less specific to more specific ones. Then, the search space can be pruned by so-called optimistic estimates that bound the quality-value that is maximally obtainable by a further refinement of an envisaged subgroup description.

Another approach related to formal concept analysis

additionally uses the fact that it is enough to look not at all arbitrary subset descriptions, but only on the set of all formal concept intents (or extents).

Computing $D^+ := \max_{(A,B)\in\mathfrak{B}(\mathbb{K})} q_t(A)$

(Our) MILP-apprach

uses a mixed integer linear program formulation and the joint modeling of concept extents and concept intents as decision variables and implements formal implications between objects and attributes as inequality constraints.

Computing $D^+ := \max_{(A,B)\in\mathfrak{B}(\mathbb{K})} q_t(A)$

(Our) MILP-apprach

uses a mixed integer linear program formulation and the joint modeling of concept extents and concept intents as decision variables and implements formal implications between objects and attributes as inequality constraints.

Note

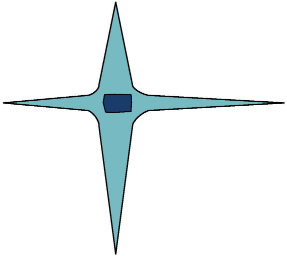
- In every case, computing D^+ is generally a very difficult task.
- Therefore, we want to make the problem simpler.
- Instead of the family of all concept extents we now introduce the family of all star-shaped sets (w.r.t. a so-called betweenness relation).
- This is usually a larger family then the family of all extents, but it is easier to handle.

Part III: Local rings of sets

Motivation: A geometrical inspiration

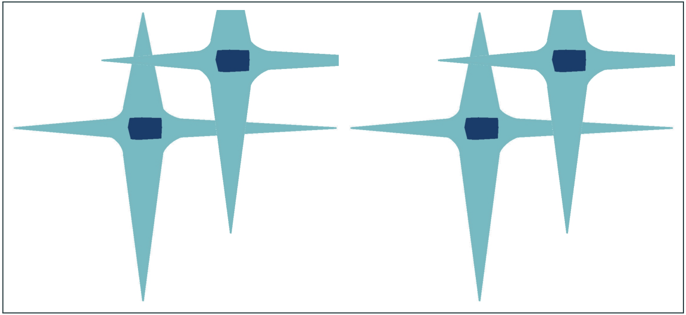
Definition (Star-shaped set)

A set $S \subseteq \mathbb{R}^d$ in d-dimensional Euclidean space $\mathbb{R}^d$ is called star-shaped if there exists a point $r \in S$ such that every other point $p \in S$ is visible from r, i.e., the whole line $\overline{rp}$ lies in S. In this case, any such point r is called a reference point or a center point of S. The set of all reference points of a star-shaped set S, denoted by $\ker(S)$, is called the kernel of S.



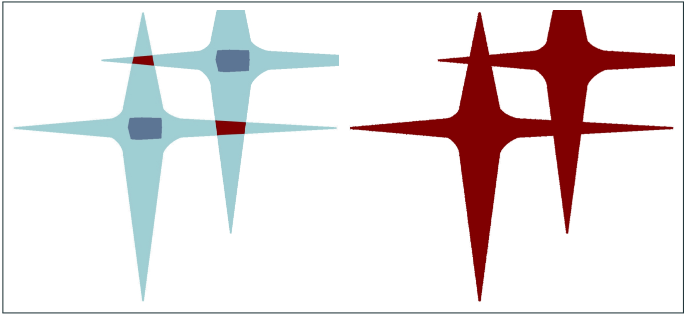
Observations

- For $d \geq 2$ the family of all star-shaped sets of $\mathbb{R}^d$ is generally neither closed under intersection nor closed under union.



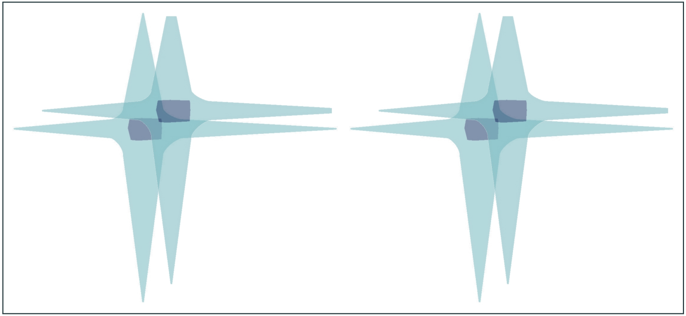
Observations

- For $d \geq 2$ the family of all star-shaped sets of $\mathbb{R}^d$ is generally neither closed under intersection nor closed under union.



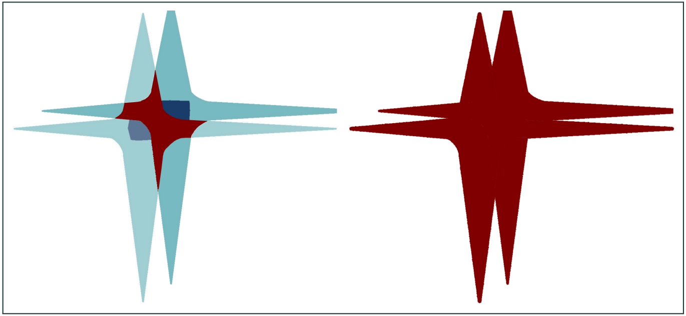
Observations

- But an arbitrary intersection/union of star-shaped sets with overlapping kernels is again a star-shaped set.



Observations

- But an arbitrary intersection/union of star-shaped sets with overlapping kernels is again a star-shaped set.



Definition (Closure system, Ring of sets)

Let $\mathcal{S} \subseteq 2^V$ be a family of subsets of a basic set V. Then $\mathcal{S}$ is called

- closure system, if it contains V and is additionally closed under arbitrary intersections.
- ring of sets, if it is closed under arbitrary intersections and under arbitrary unions.

Families of sets ii

Definition (local ring of sets, cored ring-bundle)

We shall call a family $\mathcal{S} \subseteq 2^V$ of subsets of V a local ring of sets if there exists a kernel-map

$$\ker : \mathcal{S} \longrightarrow 2^V$$

with $\emptyset \neq \ker(S) \subseteq S$ for all $S \in \mathcal{S} \backslash \{\emptyset\}$, such that for every arbitrary family $(S_i)_{i \in I}$ of sets $S_i \in \mathcal{S}$ with overlapping kernels (i.e.: $\bigcap_{i \in I} \ker(S_i) \neq \emptyset$) both the intersection $\bigcap_{i \in I} S_i$ as well as the union $\bigcup_{i \in I} S_i$ are again in the family $\mathcal{S}$.

Furthermore, we may call a pair $\mathcal{R} = (\mathcal{S}, ker)$ a cored ring-bundle if $\mathcal{S}$ is a local ring of sets and *ker* is an associated kernel-map with the above property.

Observations in the nomenclature from above

The family of all star-shaped subsets of $\mathbb{R}^d$ (for $d \geq 2$) is neither a closure system, nor a ring of sets, but a local ring of sets.

Remark

Not every arbitrary family of sets is a local ring of sets.

Example

$V = \{a, b, c, d\}$ and $\mathcal{S} = \binom{V}{2}$.

Then, there does not exist any kernel-map *ker*, because

$$ker(\{a,b\}) \ni a \Longrightarrow \begin{array}{c} ker(\{a,c\}) = \{c\} \\ ker(\{a,d\}) = \{d\} \end{array} \Longrightarrow \begin{array}{c} c \notin ker(\{c,d\}) \\ d \notin \mathrm{ker}(\{c,d\}) \end{array}$$

and

$$ker(\{a,b\}) \ni b \Longrightarrow \begin{array}{c} ker(\{b,c\}) = \{c\} \\ ker(\{b,d\}) = \{d\} \end{array} \Longrightarrow \begin{array}{c} c \notin ker(\{c,d\}) \\ d \notin \mathrm{ker}(\{c,d\}) \end{array}$$

Examples of closures systems and rings of sets

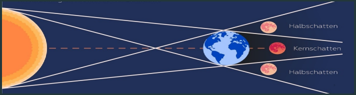
Definition

Let $\mathbb{V} = (V, \leq)$ be a partially ordered set (poset). For $x, y \in V$ define

- $\uparrow x := \{y \in V \mid y \geq x\}$
- $\downarrow x := \{y \in V \mid y \leq x\}$
- $[x, y] := \uparrow x \cap \downarrow y$.

- If $\mathbb{V}$ builds a complete lattice, then the set $\mathcal{I}(\mathbb{V}) := \{[x, y] \mid x, y \in V\}$ of all intervals of $\mathbb{V}$ is a closure system.
- The set $\mathcal{U}(\mathbb{V}) := \{U \subseteq V \mid x \in U \implies \uparrow x \subseteq U\}$ of all upsets of $\mathbb{V}$ is a ring of sets.

Intermezzo: Betweenness-relations



Let $Z \subseteq V^3$ be a ternary relation on V. (Think of a betweenness-relation where $(x, y, z) \in Z$ is interpreted as "y lies between x and z" or as "z is reachable from x via y".)

Definition (betweenness-relation, cf., [Birkhoff, 1940, p. 2, Ex.4])

For a ternary relation $Z \subseteq V^3$ and $x, z \in V$ define $\overline{xz}^Z := \{y \in V \mid (x, y, z) \in Z\}$.Then, one may call a ternary relation $Z \subseteq V^3$ a betweenness-relation if it satisfies the following properties:

B0: $\overline{xz}^Z = \overline{zx}^Z$ (outer symmetry)

B1: $y \in \overline{xz}^Z \ \& \ z \in \overline{xy}^Z \implies y = z$ (conditional antisymmetry)

B2: $y \in \overline{xz}^Z \implies \overline{xy}^Z \subseteq \overline{xz}^Z$ (inner transitivity)

B3: $y \in \overline{xz}^Z \ \& \ z \in \overline{yw}^Z \ \& \ y \neq z \implies y \in \overline{xw}^Z$ (outer transitivity)

B4: $y \in \overline{xz}^Z \ \& \ z \in \overline{xw}^Z \implies z \in \overline{yw}^Z$ (compositionality)

Examples of betweenness-relations and non-betweenness relations i

1. The case of a linear space $\mathbb{V} = (V, +, \cdot)$:

$$(x, y, z) \in Z \text{ iff } y \text{ lies on the line } \overline{xz}$$
$$\iff \exists \lambda \in [0, 1] : y = \lambda \cdot x + (1 - \lambda) \cdot z.$$

2. The case of a metric space (M, d):

$$(x, y, z) \in Z \text{ iff the triangle inequality is fulfilled with equality}$$
$$\iff d(x, z) = d(x, y) + d(y, z)$$

3. The case of a metric space (M, d):

$$(x, y, z) \in Z \text{ iff the triangle inequality is almost fulfilled with equality}$$
$$\iff d(x, y) + d(y, z) \leq c \cdot d(x, z) \text{ with } c > 1$$

Examples of betweenness-relations and non-betweenness relations ii

4. The case of a metric space (M, d):

$$
\begin{aligned}
(x, y, z) \in Z \text{ iff } y \text{ is closer to } x \text{ than } z \\
\iff d(y, x) \leq d(z, x)
\end{aligned}
$$

5. The case of a poset $\mathbb{V} = (V, \leq)$:

$$(x, y, z) \in Z :\iff x \leq y \leq z \text{ or } z \leq y \leq x.$$

6. The case of a product of losets $\mathbb{V} = \prod_{i=1}^{p} \mathbb{L}_i$ (with $\mathbb{L}_i = (L_i, \leq_i)$ linearly ordered and with the interordinal betweenness-relation):

$$
\begin{aligned}
(x, y, z) \in Z \text{ iff } y \text{ lies between } x \text{ and } z \text{ w.r.t. every dimension } i = 1, \ldots, p \\
\iff \forall i \in \{1, \ldots, p\} : x_i \leq_i y_i \leq_i z_i \text{ or } z_i \leq_i y_i \leq_i x_i.
\end{aligned}
$$

Two canonical betweenness relations in FCA: I: An asymmetric notion

- Given formal context $\mathbb{K} = (G, M, I)$.
- $\boxed{\text{Define } Z \subseteq G^3 \text{ via: } (x, y, z) \in Z :\iff \{x\}' \cap \{z\}' \subseteq \{y\}'.}$

Two canonical betweenness relations in FCA: I: An asymmetric notion

- Given formal context $\mathbb{K} = (G, M, I)$.
- $\boxed{\text{Define } Z \subseteq G^3 \text{ via: } (x, y, z) \in Z :\iff \{x\}' \cap \{z\}' \subseteq \{y\}'.}$
- In terms of formal implications this reads as: $(x, y, z) \in Z \iff \{x, z\} \longrightarrow \{y\}$.
- Or in words: y lies between x and z if y has all the attributes that x and z have in common.
- Or in FCA language: $(x, y, z) \in Z \iff$ object y shares the common crosses of x and z.
- **Note**: This relation satisfies B0 and B2, but not B1, B3 and B4.

Two canonical betweenness relations in FCA: II: A symmetrized notion

- Given formal context $\mathbb{K} = (G, M, I)$.
- Define $Z \subseteq G^3$ via:

$$(x, y, z) \in Z :\Longleftrightarrow \{x\}' \cap \{z\}' \subseteq \{y\}' \quad \& \\ (\{x\}')^c \cap (\{z\}')^c \subseteq (\{y\}')^c.$$

Two canonical betweenness relations in FCA: II: A symmetrized notion

- Given formal context $\mathbb{K} = (G, M, I)$.
- Define $Z \subseteq G^3$ via:

$$(x, y, z) \in Z :\iff \{x\}' \cap \{z\}' \subseteq \{y\}' \quad \& \\ (\{x\}')^c \cap (\{z\}')^c \subseteq (\{y\}')^c.$$

- Or in words: y lies between x and z if y has all the attributes that x and z have in common **and** y has no attribute that both x and z do not have.
- Or in FCA language: $(x, y, z) \in Z \iff$ object y shares the common crosses of x and z **and** the common non-crosses of x and z.
- **Note**: This relation satisfies B0, B1 and B2 (but not B3 - B4).

Examples of local rings of sets i

Definition (Star-shaped set w.r.t. a ternary relation Z)

Let V be a basic set and let $Z \subseteq V^3$ be a ternary relation on V. A subset $A \subseteq V$ for which there exists a center element $r \in A$ with the property

$$p \in A \Longrightarrow \overline{rp}^Z \subseteq A$$

may be called a star-shaped set w.r.t. the relation Z, because every point $p \in A$ is visible from r in the sense that every further point y between r and p is in A.

Examples of local rings of sets ii

Observation

The set

$$\mathcal{ST}(Z) := \{A \subseteq V \mid \exists r \in A : p \in A \Longrightarrow \overline{rp}^{z} \subseteq A\}$$

of all star-shaped sets of $\mathbb{V}$ (w.r.t. Z) is a local ring of sets.

Examples of local rings of sets iii

Remark

- By switching from star-shaped sets in $\mathbb{R}^d$ to $\mathcal{ST}(Z)$, we essentially did only modify the basic set and the notion of the term "between".
- The proof is completely analogous to that for conventional star-shaped sets in $\mathbb{R}^d$.
- Note that for the proof, no structural properties of the betweenness-relation of $\mathbb{R}^d$ are needed.

The simple proof (durchmeditieren)

Let $\ker : \mathcal{ST}(Z) \longrightarrow 2^V$ map every star-shaped set A to the set of all its reference points. Let furthermore $(A_i)_{i\in I}$ be a family of star-shaped sets with a non-empty overlap (i.e.: $\bigcap\limits_{i\in I} \ker(A_i) \neq \emptyset$).

Let $B := \bigcap\limits_{i\in I} A_i$ and choose some arbitrary common reference point $r \in \bigcap\limits_{i\in I} \ker(A_i)$. Consider now an arbitrary point $p \in B$. For a further point $y \in V$ with $(r, y, p) \in Z$ we have $y \in A_i$ for all $i \in I$, because all A_i are star-shaped sets with reference point r and $p \in A_i$ for all $i \in I$. Thus $y \in B$ and the intersection B of the star-shaped family $(A_i)_{i\in I}$ is a star-shaped set.

Let furthermore $C := \bigcup\limits_{i\in I} A_i$. Consider now an arbitrary $p \in C$ and an arbitrary $y \in V$ with $(r, y, p) \in Z$. Then $p \in A_i$ for some $i \in I$ and because A_i was star-shaped with reference point r, we have $y \in A_i$ and thus $y \in C$, which shows that the union C of the family $(A_i)_{i\in I}$ is a star-shaped set.

A representation theorem for cored ring-bundles

Theorem

i) For every ternary relation Z on a basic set V the associated family of star-shaped sets (w.r.t. Z)

$$\mathcal{ST}(Z) := \{A \subseteq V \mid \exists r \in A : p \in A \Longrightarrow \overline{rp}^z \subseteq A\},$$

together with the kernel-map

$$ker_Z : \mathcal{ST}(Z) \longrightarrow 2^V : \quad A \mapsto \{r \in A \mid p \in A \Longrightarrow \overline{rp}^z \subseteq A\}$$

builds a cored ring-bundle.

ii) Furthermore, every finite cored ring-bundle $\mathcal{R} = (\mathcal{S}, ker)$ (only meaning that $\mathcal{S}$ is finite) on a basic set V can be obtained as the family of star-shaped sets w.r.t. the ternary relation

$$Z_{\mathcal{R}} = \{(r, y, p) \in V^3 \mid \forall S \in \mathcal{S} : r \in \mathrm{ker}(S) \;\&\; p \in S \Longrightarrow y \in S\}.$$

Note

- If we use the asymmetric canonical FCA betweenness relation Z_{asym}, then all concept extents are special star-shaped sets.
- Thus, the family of all star-shaped sets is a superfamily of all concept extents.
- The same holds for the symmetrized canonical FCA betweenness relation Z_{sym}.
- Since $Z_{sym} \subseteq Z_{asym}$ we have $\mathcal{ST}(Z_{sym}) \supseteq \mathcal{ST}(Z_{asym})$.

Part IV: Linear programming on families of sets

Linear programming on families of sets

Let $V = \{v_1, \ldots, v_k\}$ be a finite basic set and let $\mathcal{S} \subseteq 2^V$ be a family of subsets of V. Identify every arbitrary set $S \subseteq V$ with its indicator function written as a vector $s = (s_1, \ldots, s_k) \in \{0, 1\}^k$ with $s_i = 1 \iff v_i \in S$.

We now consider the problem of computing the supremum statistic

$$D^+ := \max_{S \in \mathcal{S}} \langle w, s \rangle$$

for some given fixed objective vector $w \in \mathbb{R}^k$.

Computing D^+

- To compute
$$D^+ = \max_{S \in \mathcal{S}} \langle w, s \rangle$$
it would be very helpful to efficiently describe the family $\mathcal{S}$.
- The structure of $\mathcal{S}$ actually helps!
- Of course, the fact that the expression $\langle w, s \rangle$ is linear in s also helps.

Describing closure systems via contextual logic

Any closure system $\mathcal{S} \subseteq 2^V$ can be described with the help of formal implications:

Definition (Formal implication)

A formal implication is a pair (A, B) of subsets of V, denoted by

$$A \longrightarrow B.$$

We say that a formal implication $A \longrightarrow B$ is valid in a family $\mathcal{S} \subseteq 2^V$ if every set $S \in \mathcal{S}$ that contains every element of the premise A always also contains every element of the conclusion B. In this case we also say that $\mathcal{S}$ respects the implication $A \longrightarrow B$. Additionally, we say that a set $C \subseteq V$ respects the implication $A \longrightarrow B$ if in the case that it contains all elements of A, it also contains every element of B.

Theorem

Any finite closure system $\mathcal{S}$ is uniquely characterized by the set $\mathfrak{I}(\mathcal{S})$ of all its valid formal implications via

$$\mathcal{S} = \{C \subseteq V \mid C \text{ respects all implications of } \mathfrak{I}(\mathcal{S})\}.$$

Computing D^+

$$D^+ = \max_{S \in \mathcal{S}} \langle w, s \rangle$$

If $\mathcal{S}$ is a closure system: Use description of $\mathcal{S}$ via the set $\mathfrak{I}(\mathcal{S})$ of all valid formal implications (or a generating subset $\mathfrak{J} \subseteq \mathfrak{I}(\mathcal{S})$):

The demand $S \in \mathcal{S}$ can be implemented via inequality constraints: For example the implication $\{v_i\} \longrightarrow \{v_j\}$ can be implemented by the inequality constrain $s_i \leq s_j$.

General implications $A \longrightarrow B$ can be implemented by the inequality constraint

$$\sum_{i:v_i \in A} s_i - |A| + 1 \leq \frac{1}{|B|} \sum_{i:v_i \in B} s_i.$$

Computing D^+ as an integer linear program (ILP)

One can compute $D^+ = \max_{S \in \mathcal{S}} \langle w, s \rangle$ by solving the integer linear program

$$\boxed{\begin{aligned} \langle w, s \rangle &\longrightarrow \max \\ w.r.t. \\ s &\in \{0,1\}^k \\ \forall (A,B) \in \mathfrak{J} : \sum_{i : v_i \in A} s_i - |A| + 1 &\leq \frac{1}{|B|} \sum_{i : v_i \in B} s_i \end{aligned}}$$

Computing D^+ as an integer linear program (ILP)

- However, computing the set $\mathfrak{I}(\mathcal{S})$ (or a base $\mathfrak{I}$) of implications is often computationally intractable.
- For the case that the family $\mathcal{S}$ is the family of all extents of a formal context $\mathbb{K} = (G, M, I)$, one can model jointly both the extents and the intents with some vector $(e_1, \ldots, e_m, i_1, \ldots, i_n) \in \{0, 1\}^{m+n}$ where $m = |G|$ and $n = |M|$.
- Then, one can implement the relationships between the extent and the intent with inequalities.
- For example consider that object e_i does not have attribute i_j. Then, if attribute i_j is in the intent (i.e.: $i_j = 1$), the object e_i cannot be in the extent (i.e.: $e_i = 0$) and vice versa. This can be implemented with the inequality $e_i + i_j \leq 1$

Computing D^+ for formal contexts as a mixed integer linear program (MILP)

Result

One can compute the statistic

$$\sup_{e \in \text{extents}(\mathbb{K})} \langle w, e \rangle$$

by solving an integer linear program with $m + n$ decision variables and $\mathcal{O}(m + n)$ constraints.

Observation

One can remove the demand that the decision variables $(e_1, \ldots, e_m)$ describing the extent are integer. Thus, one has to solve a mixed integer program with n binary variables and m continuous variables.

Computing D^+ on a ring of sets

If $\mathcal{S}$ is a ring of sets (containing the basic set V and $\emptyset$), a generating set $\mathfrak{I}$ of implications of the form $\{v_i\} \longrightarrow \{v_j\}$ can be found. For computing D^+, we have to solve the ILP

$$\boxed{\begin{array}{r} \langle w, s\rangle \longrightarrow \max \\ w.r.t. \\ s \in \{0,1\}^k \\ \forall(\{v_i\},\{v_j\}) \in \mathfrak{I} : s_i \leq s_j \end{array}}$$

Computing D^+ on a ring of sets

Observation

In this situation, the demand $s \in \{0,1\}^k$ can actually be relaxed to the demand $s \in [0,1]^k$, and the ILP problem reduces to the far more tractable linear program (LP)

$$\boxed{\begin{array}{r} \langle w^x - w^y, s\rangle \longrightarrow \max \\ w.r.t. \\ s \in [0,1]^k \\ \forall(\{v_i\},\{v_j\}) \in \mathfrak{J} : s_i \leq s_j \end{array}}$$

Complexity of LP and (M)ILP i

- Worst case time complexity of simplex method for solving a LP: exponential in input size (e.g., Klee and Minty [1972]), the same for (M)ILP (if simplex method is used for solving the relaxed solutions within a branch and bound/branch and cut approach).
- But e.g. average-case time complexity (cf., Borgwardt [2012]) and smoothed time complexity (cf., Spielman and Teng [2004]) of simplex method only polynomial in input size.
- For (M)ILP analysis far more difficult.
- In our concrete situation (e.g., subgroup discovery): (M)ILP seems to be very data dependent.

Complexity of LP and (M)ILP ii

$\Longrightarrow$ Difference between (M)ILP and LP may be of special importance for simulating D^+ under some null hypothesis! (Experience for (M)ILP: More time needed under H_0 than under actually observed data, seemingly the same for conventional subgroup discovery algorithms)

Computing D^+ on cored ring-bundles

- The complexity reduction by switching from (M)ILP to LP is very alluring. How can we make use of it for computing D^+ on a cored ring-bundle, which is generally not a ring of sets (and not even a closure system)?
- Simply quantify over all possible reference points!

Computing D^+ on cored ring-bundles

- Given cored ring-bundle $\mathcal{R} = (\mathcal{S}, ker)$ over basic set V with associated betweenness-relation $Z_{\mathcal{R}}$.
- To compute
$$\sup_{S \in \mathcal{S}} \langle w, s \rangle,$$
firstly solve for every $r \in V$ the program
$$D_r^+ := \sup_{S \in \mathcal{S} : \ker(S) \ni r} \langle w, s \rangle.$$
- Then, compute D^+ as
$$D^+ = \max_{r \in V} D_r^+.$$
- All in all, one has to solve $|V|$ linear programs with $|V|$ decision variables and maximally $\mathcal{O}(|V|^2)$ constraints.
- Note: Quantification over $r \in V$ also allows for parallelization.

Computing D_r^+ for some fixed $r \in V$

To compute $D_r^+ = \sup_{S \in \mathcal{S}:\ker(S) \ni r} \langle w, m \rangle$, firstly compute the quasiorder

$$R_r := \{(x, y) \in V^2 \mid (r, y, x) \in Z_{\mathcal{R}}\}.$$

Then, the closure system $\mathcal{S}_r := \{S \in \mathcal{S} : \ker(S) \ni r\}$ can be described by the set $I_r = \{\{x\} \longrightarrow \{y\} \mid (x, y) \in R_r\}$ of implications as

$$\mathcal{S}_r = \{S \subseteq V \mid r \in S \;\&\; S \text{ respects } I_r\},$$

which can be implemented as a linear program as

$$\boxed{\begin{aligned} \langle w, s \rangle \longrightarrow \max \\ w.r.t. \\ s \in [0,1]^k \\ \forall (v_i, v_j) \in R_r : s_i \leq s_j \end{aligned}}$$

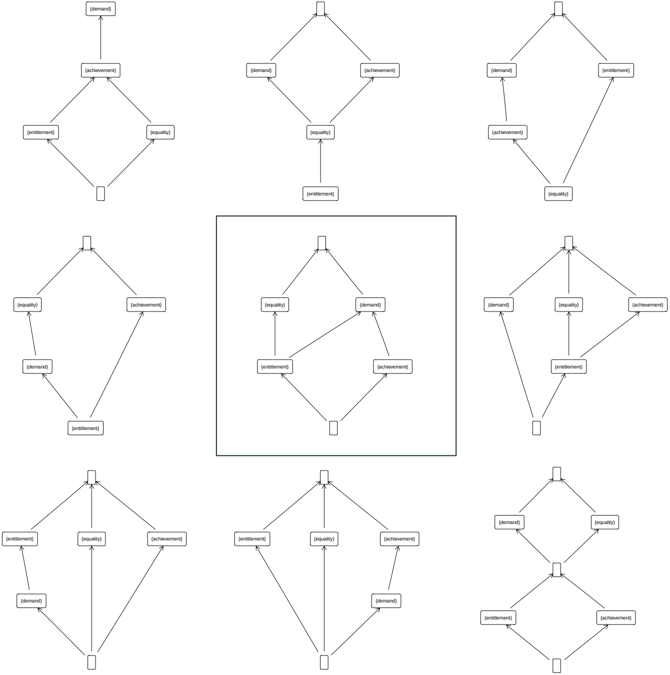
(Instead of R_r, one can also use only a transitive reduction of R_r.)

Application: Star-shaped subgroup discovery i

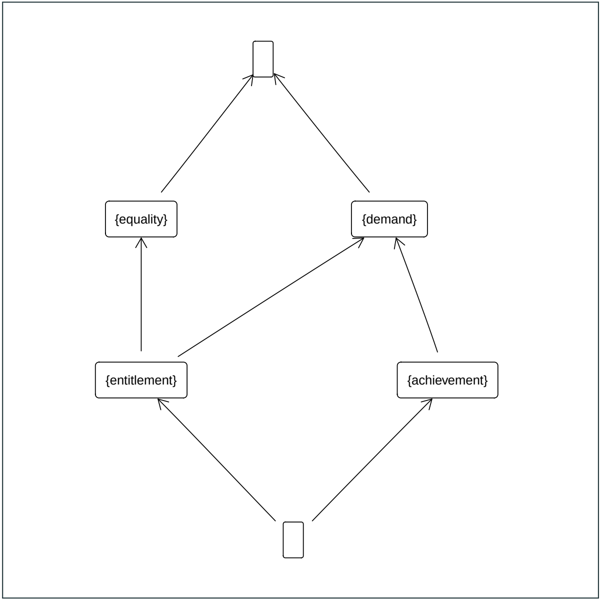
- Allbus 2014: $G = \{g_1, \ldots, g_{874}\}$ respondents from East Germany.
- 4 dimensions of social justice: principle of:
  1. achievement
  2. equality
  3. entitlement
  4. demand
- Measured by 8 constructs (2 for each dimension)
- This gives for every respondent a partial order which represents her/his agreement to the 4 principles of social justice. Concretely: A respondent agrees more to one principle than to another if she/he agrees more to both corresponding constructs, respectively.

Application: Star-shaped subgroup discovery ii

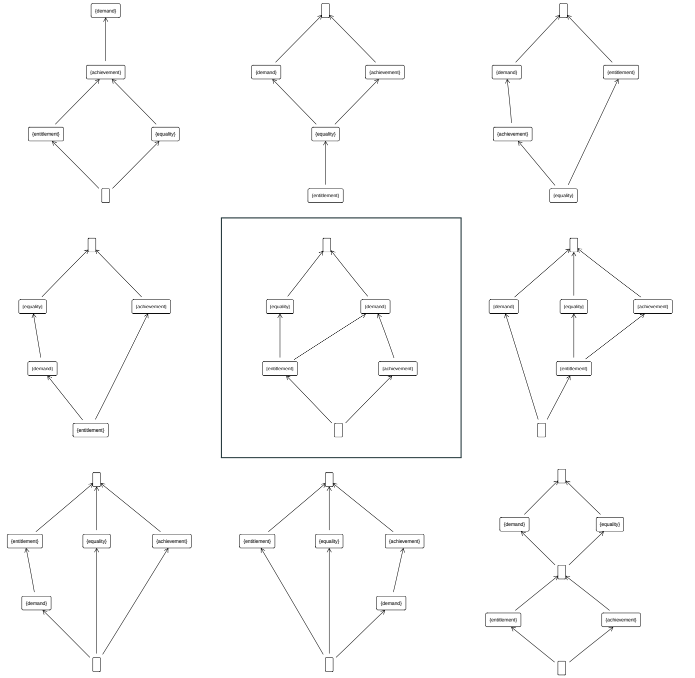
- There are also other ways of analysis thinkable, for example a direct *absolute* anaylsis on the original response scale (e.g., with an interordinal conceptual scaling).
- Target variable: Answer to the question
  "*Do you believe that you will receive your fair share?*"
- Symmetrized canonical betweenness notion (without stylization, cf., later)


{demand}
{achievement}
{entitlement}
{equality}
{demand}
{achievement}
{equality}
{entitlement}
{demand}
{entitlement}
{achievement}
{equality}
{equality}
{achievement}
{demand}
{entitlement}
{equality}
{demand}
{entitlement}
{achievement}
{demand}
{equality}
{achievement}
{entitlement}
{entitlement}
{equality}
{achievement}
{demand}
{entitlement}
{equality}
{achievement}
{demand}
{demand}
{equality}
{entitlement}
{achievement}

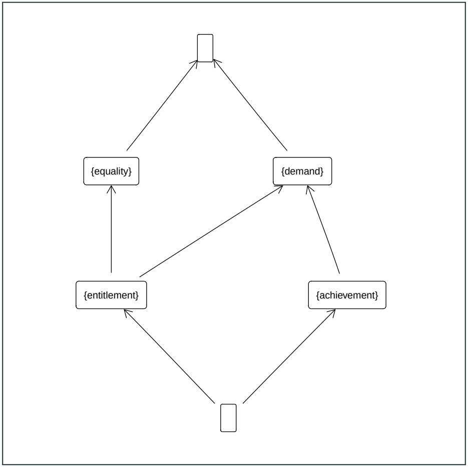
Reference point/Center:

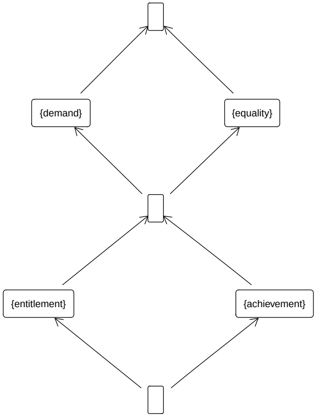


The whole star-shaped set (center plus 9 border points):

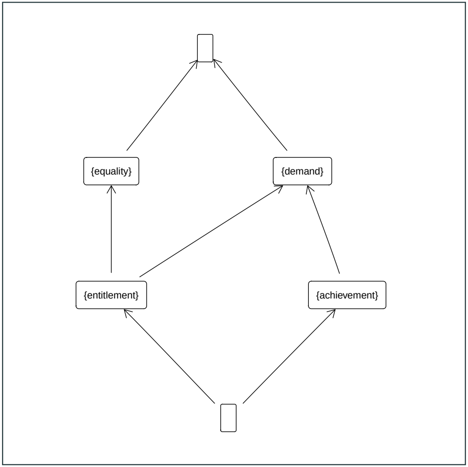


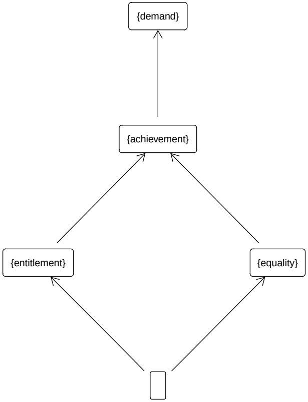
Center plus one selected border point::





Center plus one selected border point::





Stylized betweenness

- Concept of star-shaped sets uses notion of y lying between x and z which can be formalized with a ternary betweenness-relation $Z \subseteq V^3$.
- Sometimes this notion is (statistically) too 'weak' (especially in high dimensions).
- Stylized notion of betweenness: "y lies between x and z". $\rightsquigarrow$ "y lies approximately between x and z ."
- Classical notion of betweenness in $\mathbb{R}^d$: $(x, y, z) \in Z \iff q = \lambda p + (1 - \lambda) r$ for some $\lambda \in [0, 1]$.
- Stylized notion of betweenness (one possibility): $(x, y, z) \in Z \iff$ the angle (x, y, z) is approximately π, say $\in [\pi - \delta, \pi + \delta]$. (There are many other possibilities.)
- Also for the canonical FCA betwenness notions stylization is possible.

Statistical aspects

- Local analysis: Given reference point r, V.C.-dimension of set of all star-shaped sets with reference point r is the width of the binary relation $Z(c,\cdot,\cdot)$ (i.e., the maximal cardinality of an antichain).
- Stylization parameter δ controls the width of $Z(c,\cdot,\cdot)$.
- 'Local' V.C.-dimension of $\mathcal{S}$ is thus controlled with δ.
- Global analysis: Maximally n center points: growth function is controlled.
- 'Uniform' control of the V.C.-dimension possible.
- Variation of 'local' V.C.-dimension is low.
- V.C.-entropy less dependent on P?
- How does this V.C.-analysis driven regularization compare to more classical regularization where some notion of 'smoothness of functions' is used (e.g., total variation for ternary relations)

Open questions

1. (Graphical) Presentation of results, especially for the stylized/regularized case.
2. How is the relation to an analysis where the target variable (personal perception of receiving fair share) is modeled as the covariate and where the poset-valued covaraiate (agreement to dimensions of social justice) is modeled as the contravariate.
3. Analysis under incorporation of possible 'cofounders' (e.g., income)
4. Concrete choice of stylization for canonical FCA betweenness notions
5. Statistical guidence for concrete choice of stylization parameter

References

G. Birkhoff. *Lattice Theory*. American Mathematical Society, 1940.

K. H. Borgwardt. *The Simplex Method: a probabilistic analysis*, volume 1. Springer Science & Business Media, 2012.

B. Ganter. *Diskrete Mathematik: Geordnete Mengen*. springer, 2013.

V. Klee and G. J. Minty. How good is the simplex algorithm? In O. Shisha, editor, *Inequalities*, volume III, pages 159–175. Academic Press, New York, 1972.

S. Liebig and M. May. Dimensionen sozialer Gerechtigkeit. *Aus Politik und Zeitgeschichte*, 59(47), 2009.

D. A. Spielman and S.-H. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *J. ACM*, 51(3):385–463, May 2004. ISSN 0004-5411.

S. Wrobel. Inductive logic programming for knowledge discovery in databases. In *Relational data mining*, pages 74–101. Springer, 2001.