New tools for the construction of ranking and evaluation indicators in multidimensional systems of ordinal variables

Filomena Maggino
Università degli Studi
di Firenze

Marco Fattore
Università degli Studi
di Milano-Bicocca

OUTLINE

1. → **premise**

2. → **partial order theory**

i. **basic elements**
ii. **an application**

3. → **final remarks**

OUTLINE

1. → **premise**

2. → **partial order theory**

i. **basic elements**
ii. **an application**

3. → **final remarks**

Methodology aimed at constructing indicators

↓

consolidated tradition

however

critical issues remained unsolved and unsettled

Critics point out conceptual, methodological and technical issues, especially with reference to difficulty in dealing with data which

- refer to a complex reality
- are ambiguous and softened
- are multidimensional
- are dynamic and evolutionary
- are qualitative also when quantitatively measured
- contain errors and approximations
- are sensitive

A REAL DIALOGUE …

Hi XYZ,. . .
In our [. . .] *Wellbeing Index* project we standardize our data (collected through a1-4 point scale), onto a 0 –100 scale. [. . .]. When this formula is applied […] the result is 70.43 points.

Our […] results are, in fact, very different from yours. […]

To be more precise, when we use the survey mean scores as data (N = 21) the mean is 77.57 points, the standard deviation is 0.83 points, so the normal range (2*SDs around the mean) is 75:91 to 79:23 points. So your figure of 70 points falls about six standard deviations below the bottom of our normal range.

Dealing with ordinal variables in social measurement raises many epistemological, methodological and statistical problems which are still open and unsolved.

Some reflections ...

1. METHODOLOGICAL APPROACHES: between objectivity, subjectivity and arbitrariness

it is important to distinguish between

- a necessary "objectivity" for research methodology (e.g., observation and data collection procedures)
- an unavoidable "subjectivity" related to
    - o definition of the conceptual framework and
    - o choice of the analytical approaches.

1. METHODOLOGICAL APPROACHES: between objectivity, subjectivity and arbitrariness

Real methodological issue

⇩

not

removing subjectivity

rather

⇩

building a sound statistical process,

where

- subjective choices are stated and
- their consequences are worked out in a formal and unambiguous way.

2. ORDINAL DATA: between accuracy and ambiguity

great part of the methodological and statistical efforts is devoted to making measures quantitatively *more precise*

Those procedures
are interesting and may lead to useful results,
but
are often quite questionable

2. ORDINAL DATA: between accuracy and ambiguity

efforts for getting more precise measures

⇩

forcing the true nature of socio-economic phenomena

🡻

socio-economic phenomena are characterized by nuances and "ambiguities",
which are not obstacles to be removed,
but represent what really matters

3. ORDINAL DATA: technical issues

ordinal data are generally

- designed for **quantitative** data analysis and
- based on the analysis of **linear** structures

3. ORDINAL DATA: technical issues

data are forced into a conceptual and technical framework which is not consistent

↓

arbitrary and questionable results

3. ORDINAL DATA: technical issues

the issue of ranking and evaluation in an ordinal setting is still an **open problem**, even from a pure data treatment point of view.

new challenges and perspectives
to improve technical tools strategies

by taking into account

- nature of data → generally ordinal
- process and trends of phenomena → monotonic

↓

Partially Ordered SEt Theory (POSET)
part of Discrete Mathematics

↓

Tools to explore and analyze of discrete datasets structure

OUTLINE

1. → **premise**

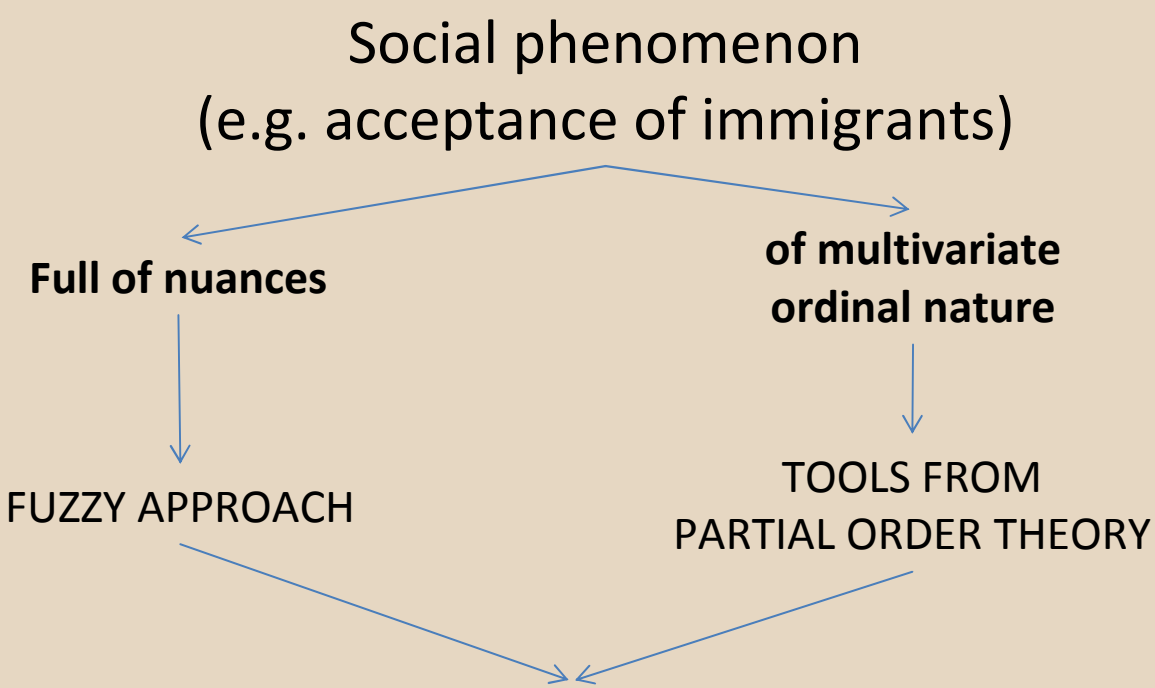
2. → **partial order theory**

i. **basic elements**

ii. **an application**

3. → **final remarks**

Searching for new formal languages…



OUTLINE

1. → **premise**

2. → **partial order theory**

i. **basic elements**
ii. **an application**

3. → **final remarks**

The application

shows how POSET theory can be used to compute indicators out of ordinal data, without turning them into numerical scores.

It aims at

- illustrating the procedure
- comparing the traditional and alternative approach

- in reducing the data structure,
- by using subjective data (from European Social Survey project)

We selected

- the first eight countries (alphabetical order) in the dataset (AT, BE, CH, CZ, DE, DK, ES, FI)
- the following variable (and corresponding items)

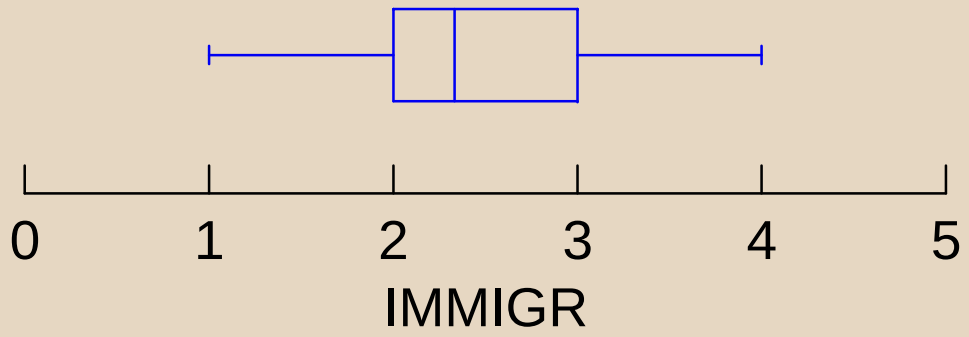
European Social Survey					
Area	Variable	Items	Item number R1 (2002)	Scaling technique	Model of measurement
Immigration and asylum issues	Acceptance of immigration: allow	many/few immigrants of same race/ethnic group as majority	D4 IMSMETN	1. allow many 2. allow some 3. allow a few 4. allow none to come and live here	reflective
		many/few immigrants of different race/ethnic group from majority	D5 IMDFETN		
		many/few immigrants from richer countries in Europe	D6 EIMRCNT		
		many/few immigrants from poorer countries in Europe	D7 EIMPCNT		
		many/few immigrants from richer countries outside Europe	D8 IMRCNTR		
		many/few immigrants from poorer countries outside Europe	D9 IMPCNTR		

First stage: synthesizing indicators at individual level

Goal: synthesizing indicators related to each variable consistently with the adopted model of measurement (reflective or formative). The basic indicators have been aggregated (Cronbach's alpha = .94)

Non-acceptance of immigration

Synthetic score (IMMIGR)	
Minimum	1.00
Maximum	4.00
Median	2.33
Mean	2.40
Standard Dev	0.69
Skewness	-0.03
Kurtosis	-0.16



Second stage: defining macro-units

Goal: synthesizing indicators observed at individual level in order to ascribe a synthetic value to groups.

Country level of non-acceptance

Country	Acceptance mean score
AT	2.61 (rank → 8)
BE	2.41 (rank → 5)
CH	2.18 (rank → 1)
CZ	2.46 (rank → 6)
DE	2.32 (rank → 3)
DK	2.31 (rank → 2)
ES	2.38 (rank → 4)
FI	2.53 (rank → 7)
Overall	2.40

Many ordinal basic indicators

⇩

individuals cannot be directly ordered,

since each indicator can show a different ranking

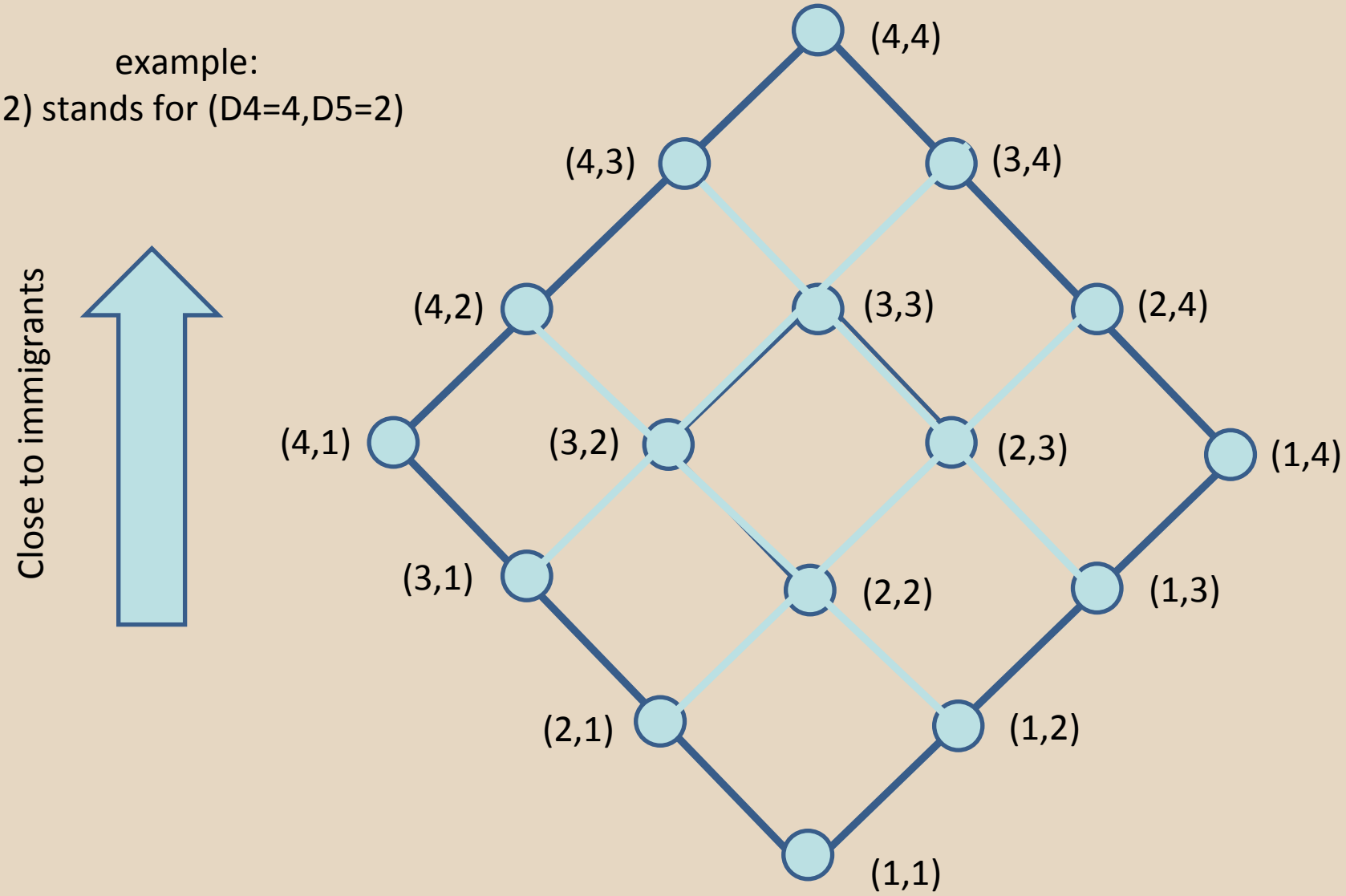
The most natural way to represent such data is through a partial order.

We consider just two among the six basic indicators:

- D4: acceptance of many/few immigrants of same race/ethnic group as majority;
- D5: acceptance of many/few immigrants of different race/ethnic group from majority.

Acceptance configurations on D4 and D5

example:
(4,2) stands for (D4=4,D5=2)



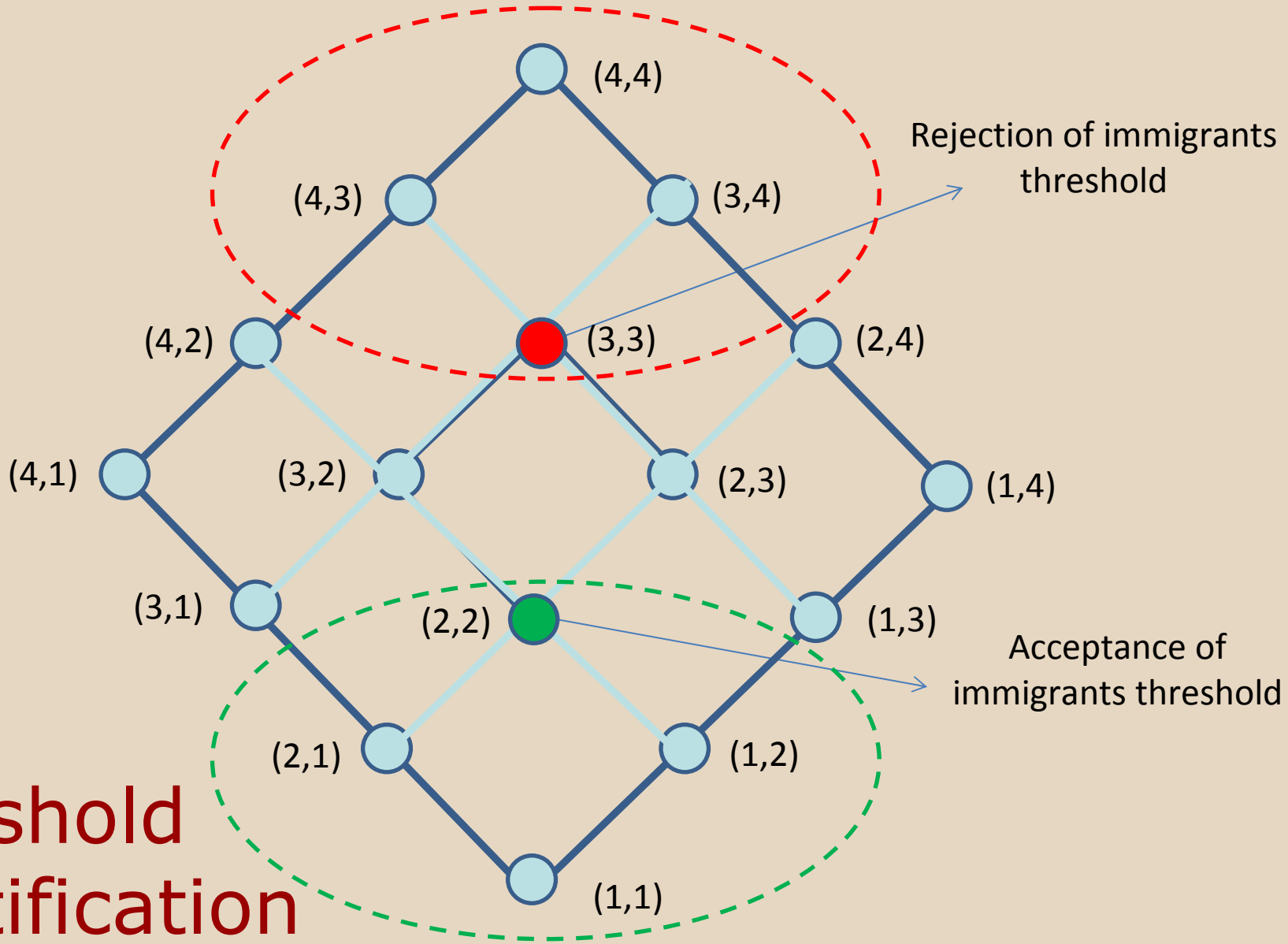
Assessing the degree of acceptance

- Is it possible to assess The degree of rejection of immigration of each combination?
- To what extent?
- Is it possible to assign to each configuration the corresponding degree?

Yes, if we can identify critical thresholds.

Here subjectivity enters but all the implications of the choice of such thresholds are then derived based only on the data structure.

Threshold identification



Let us agree that

- nodes in the red ellipse has degree of rejection of immigration equal to 1 (the maximum)
- nodes in the green ellipse has degree of rejection of immigration equal to 0 (the minimum).

All other nodes should receive a degree of rejection between 0 and 1, reproducing the ambiguities in the phenomenon.

Computation of degrees

based only on

analysis of the partial order structure of the poset

(analysis of the different relational position of each node, with respect to the thresholds selected)

degree of acceptance/rejection of immigration

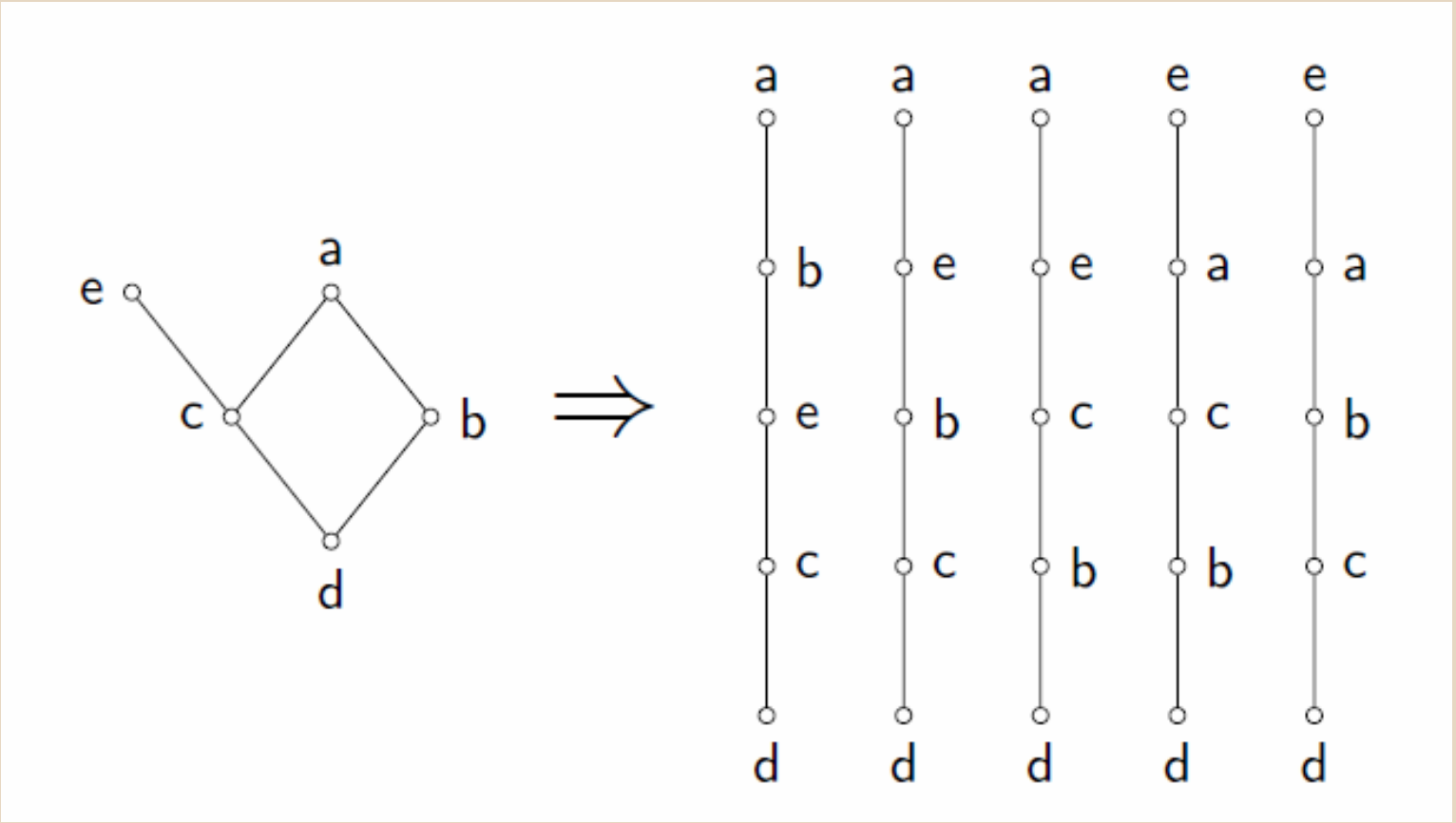
⬇

information extracted

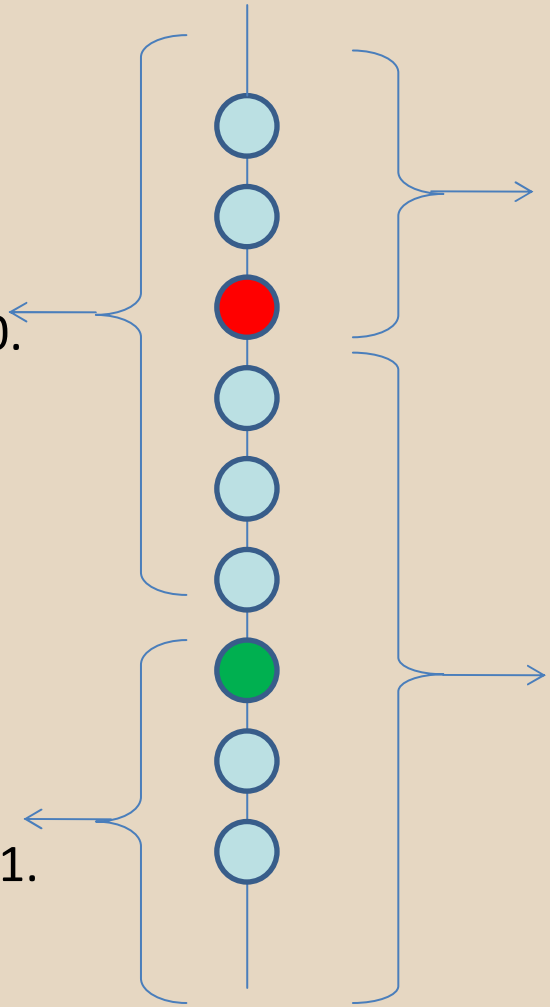
→ from poset's structure and

→ not from the aggregation of variable scores (which are treated as they are, ordinal variables).

Linear extensions of a poset



The basic idea: pick up a linear extension …



These states receive degree of immigrant acceptance equal to 0.

These states receive degree of closeness to immigrants equal to 1.

These states receive degree of immigrant acceptance equal to 1.

These states receive degree of closeness to immigrants equal to 0.

For each state we get two degrees:

deg1 → measures to what extent it can be classified as belonging to the group of states representing people who do not accept immigrants

deg2 → measures to what extent it can be classified as belonging to the group of states representing people who do accept immigrants

Turning deg1 into 1-deg1 we get an alternative measure of acceptance of immigrants (in terms of non-rejection of them).

In other words through

- rejection threshold
- acceptance threshold

two different assessments of the degree of acceptance of immigrants, corresponding to each node

Final degree of *immigration acceptance*

for each node we compute the average of 1-deg1 and deg2

State	Acceptance degree: Acceptance threshold (2,2)	Rejection degree: Rejection threshold (3,3)	Final acceptance degree
(1,1)	1,00	0,00	1,00
(1,2)	1,00	0,00	1,00
(1,3)	0,71	0,00	0,86
(1,4)	0,42	0,09	0,66
(2,1)	1,00	0,00	1,00
(2,2)	1,00	0,00	1,00
(2,3)	0,00	0,00	0,50
(2,4)	0,00	0,39	0,31
(3,1)	0,71	0,00	0,86
(3,2)	0,00	0,00	0,50
(3,3)	0,00	1,00	0,00
(3,4)	0,00	1,00	0,00
(4,1)	0,42	0,09	0,66
(4,2)	0,00	0,39	0,31
(4,3)	0,00	1,00	0,00
(4,4)	0,00	1,00	0,00

Country level of acceptance

Country	Acceptance degree (D4 and D5)
AT	0.49 (rank → 7.5)
BE	0.62 (rank → 4)
CH	0.74 (rank → 1)
CZ	0.52 (rank → 6)
DE	0.65 (rank → 2)
DK	0.63 (rank → 3)
ES	0.54 (rank → 5)
FI	0.49 (rank → 7.5)
Overall	0.62

Comparing the two approaches

With reference to the effectiveness of the two methodologies in extracting information out of data

⇩

CVs of the distributions pertaining the acceptance degrees.

	CV
Traditional approach	0.05
Alternative approach (D4, D5)	0.14
Alternative approach (D6, D7, D8, D9)	0.19

CV shown by D4 and D5 (16 nodes) is smaller than D6-D7-D8-D9 (256 nodes).

The computed numbers depend upon the choice of the thresholds and some sensitivity analysis should be added.

Results

Capacity of discriminating among countries

poset approach ←→ traditional approach

↳ far better

(Traditional approach's nature is mainly compensative / aggregative)

identifying the thresholds and extracting information out of the relational structure of data result in a great increase of the informative content of the computations

OUTLINE

1. → **premise**

2. → **partial order theory**

i. **basic elements**
ii. **an application**

3. → **final remarks**

State of the art

- Approach applied also to other social phenomena (deprivation).
- Computations performed without relying on heavy and complex numerical algorithms.
- Possibility to define thresholds composed of more than a single node (→ more flexibility to actual situations).

Future perspectives → possibility to

- integrate POSET and Structural Equation Modeling.
- define algorithms to help identifying thresholds.
- define "weighting" schemes without introducing numerical weights
- define clustering algorithms, for reducing the dimension of posets, when the number of variables and/or the number of possible scores for each variable is high.

Filomena Maggino
filomena.maggino@unifi.it

Marco Fattore
marco.fattore@unimib.it

Presentation designer: trapani.marco@gmail.com