Evaluation is not aggregation: assessing material deprivation through partial order theory

Marco Fattore, Filomena Maggino

Florence, September 9-10th, 2011

marco.fattore@unimib.it, **filomena.maggino@unifi.it**

Something taken for granted…

More or less, all the approaches to evaluation of multidimensional deprivation, poverty, quality of life… lean on the following assumptions:

1) To obtain a synthetic evaluation, dimensions must be aggregated (e.g. by averaging, counting…)

2) To account for the different relevance of evaluation dimensions, weights are to be introduced.

Our program…

To show that, when **binary data** (but we could generalize to ordinal variables) are dealt with, a satisfactory methodology for material deprivation assessment can be developed such that:

1. The ordinal nature of binary data (1/0, yes/no) is fully respected
2. No scaling of binary data into numerical variables is needed
3. No aggregation is performed
4. No weights are introduced in accounting for the different relevance of deprivation dimensions
5. Exogenous information (e.g. the choice of the threshold) is introduced in the analysis in a clear and unambiguous way

A very simple example

- 3 binary variables, v1, v2 and v3, e.g.:
  - possessing a house
  - possessing a car
  - possessing a telephone

- 0 means NO DEPRIVATION, 1 means DEPRIVATION

- 8 deprivation configurations or **profiles**:

000, 100, 010, 001, 110, 101, 011, 111

Comparabilities and incomparabilities

v1	v2	v3
1	1	0
↓	↓	↓
1	0	0

2 comparable profiles

v1	v2	v3
1	1	0
↓	↓	↑
0	0	1

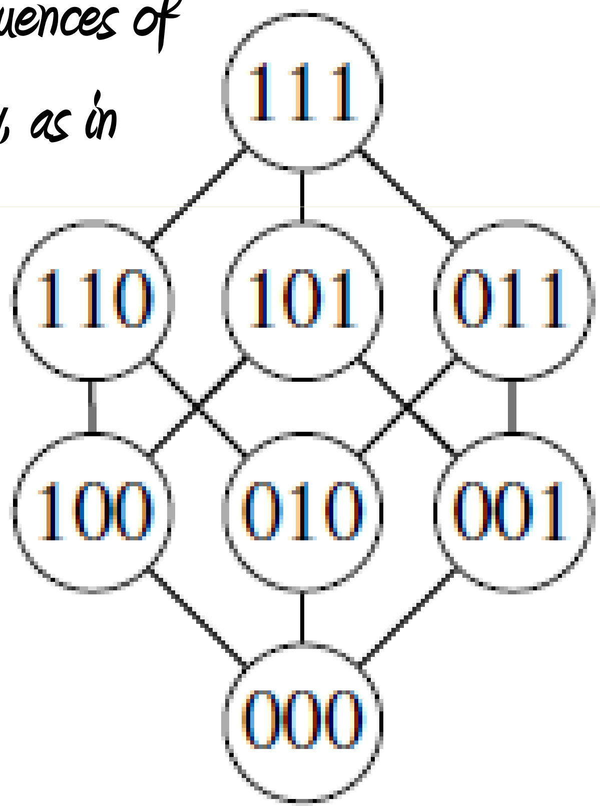
2 incomparable profiles

A very simple example

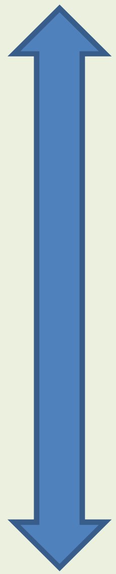
- Some deprivation profiles are **comparable** e.g 110 is more deprived than 100, since 110 shares a deprivation with 100 and has one more.
- Some deprivation profiles are **incomparable**, e.g. 110 and 001, since they do not share any deprivation (even if 110 has two deprivations, it is not correct to directly assert that 110 is more deprived than 001, unless we follow a counting approach).
- So profiles give rise to a **partially ordered set** or **POSET**, for short

Hasse diagrams: graphical representation of a poset

Edges or descending sequences of edges mean comparability, as in 111-110-100-000



Deprivation



Non- deprivation

No edges mean incomparability, as in 110||101; 101||011; 110||011

Where information lies…

The partial order relation is the fundamental algebraic structure for deprivation assessment.

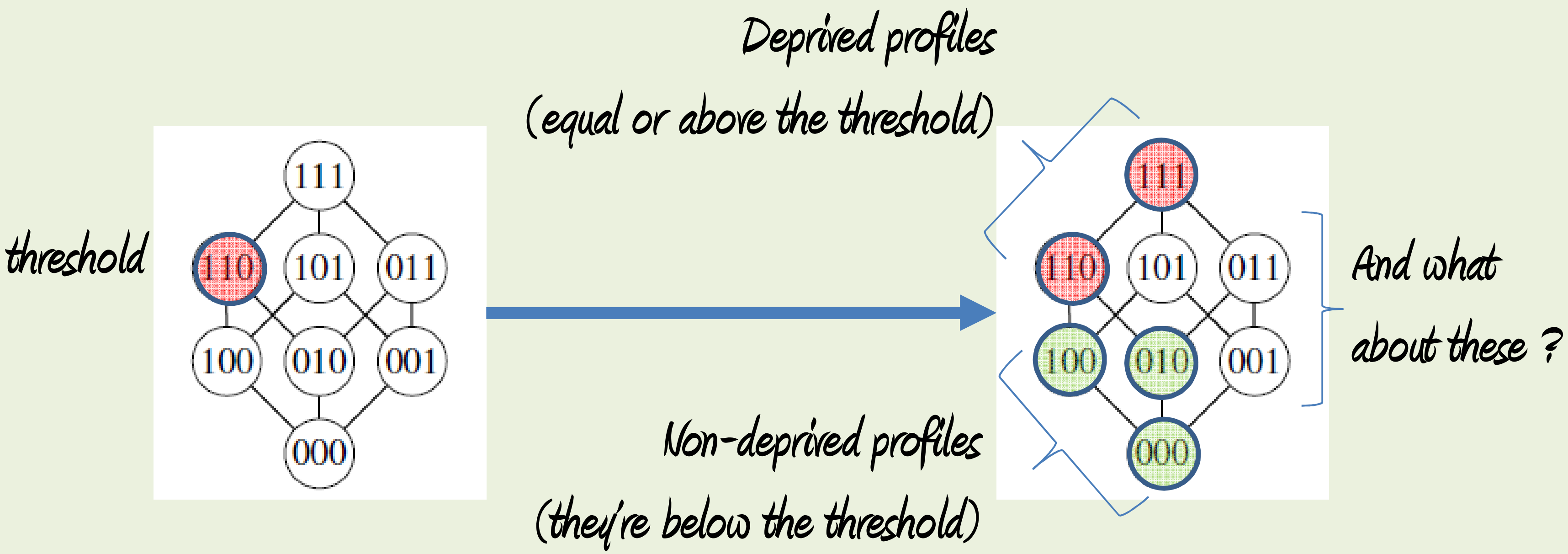
It shares the same role as linear structures in classical multivariate analysis, and the composite indicator approach.

That is, it comprises the information needed for the assessment.

The next goal is to extract such an information, to assign a deprivation degree to each profile (and then to each statistical unit sharing it).

Choosing the threshold

- Exogenous selection of the threshold: 110 (in this simple case, the threshold is composed of only one profile)



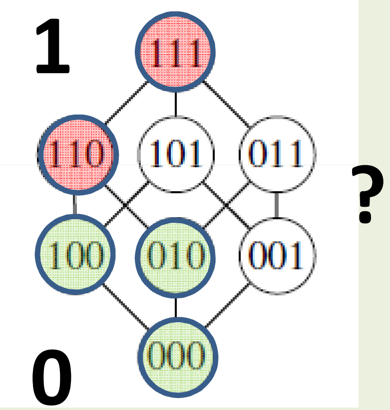
Ambiguities

- Profiles 110 and 011 are neatly classified as deprived (they get deprivation degree 1, say)

- Profiles 100, 010 and 000 are neatly classified as non-deprived (they get deprivation degree 0, say)

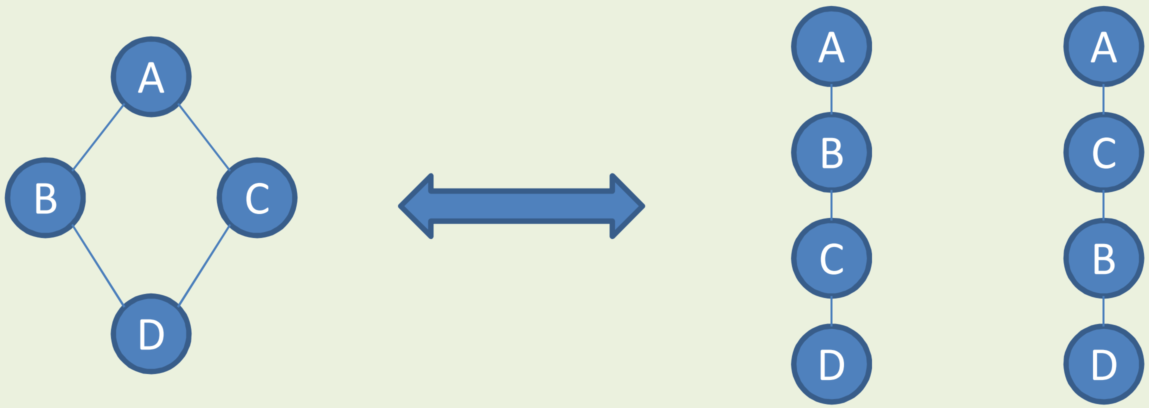
- Profiles 101, 011 and 001 are **not comparable** with the threshold 110. They cannot be neatly classified as either deprived or not: they are "ambiguously" deprived (they get deprivation degrees in]0,1[ ).

- How can we compute the deprivation degrees of such 3 profiles? Analyzing the structure of the poset and assessing the "relational" position of the corresponding nodes within the Hasse diagram, that is…



A simple but fundamental theorem

A finite poset is uniquely specified by the set of its linear extensions, that is by the set of all the complete orders consistent with the poset.



A (different) counting approach…

1. List all the linear extensions of the poset (or sample them, if they are too many…)

2. Count how many times a profile is above the threshold

3. Divide that number by the number of linear extensions of the poset, to get the final degree.

As a result…

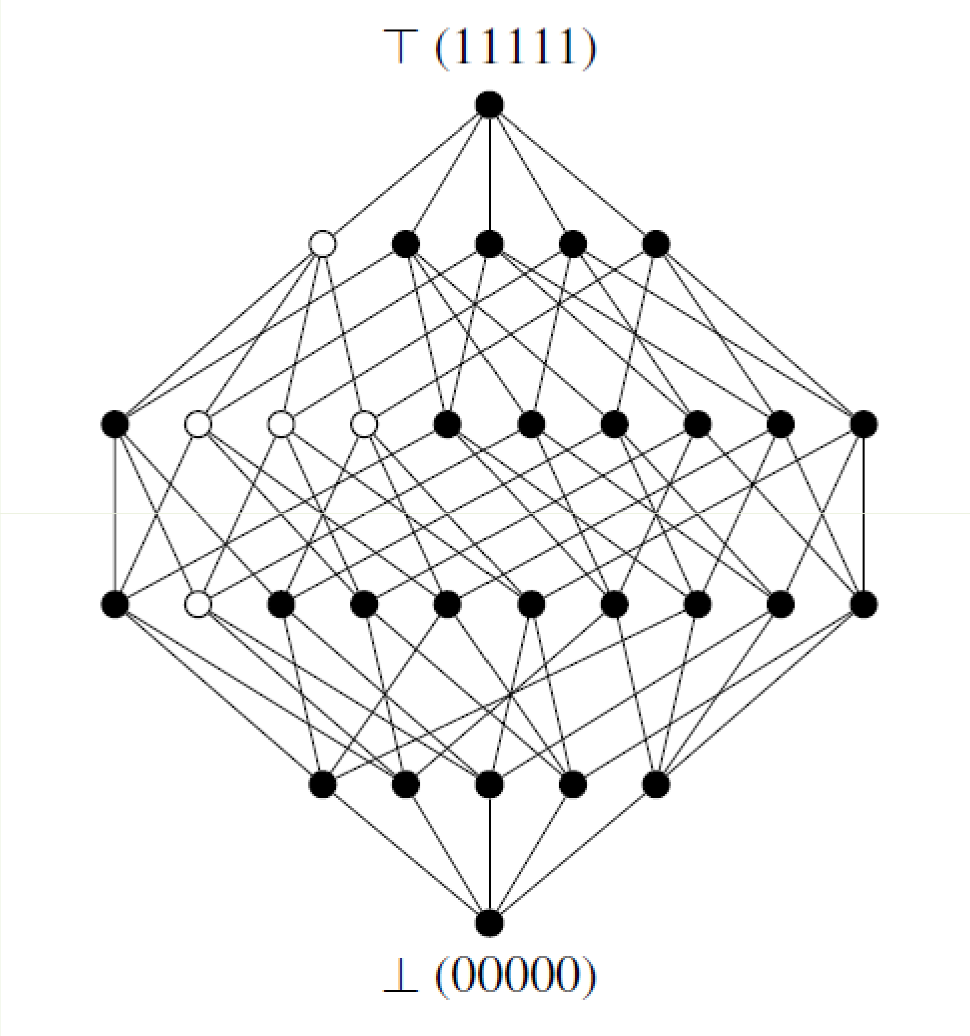
Profile	Depriv. degree
111	1
110	1
101	1/2
011	1/2
100	0
010	0
001	1/12
000	0

Deprivation nuances are accounted better than just counting the number of deprivations

A real example (Eu-Silc 2004, in Italy)

1. HS040 - *Capacity to afford paying for one week annual holiday away from home*;
2. HS050 - *Capacity to afford a meal with meat, chicken, fish (or vegetarian equivalent) every second day*;
3. HS070 - *Owning a phone (or mobile phone)*;
4. HS080 - *Owning a colour TV*;
5. HS100 - *Owning a washing machine*.

(The variables have been turned into binary form, prior to the analysis)



Black nodes represent profiles observed in the data

Selected threshold:

(10000; 01111)

Results

Region	Dep. Degree (min = 0; max = 100)	Intensity*
North – West	4%	76%
North – East	5%	76%
Centre	6%	88%
South	15%	96%
Islands	15%	94%

*Intensity is measured as the average level of deprivation of statistical units with a non-zero degree of deprivation (min = 0; max = 100).

A step forward…

How can the **different relevance** of deprivation dimensions can be accounted for in a purely **ordinal** context?

A step forward…

How can the different relevance of deprivation dimensions can be account foe in an purely ordinal context?

Clearly, there is no room for numerical weights

A step forward…

How can the **different relevance** of deprivation dimensions can be account foe in an purely **ordinal** context?

Clearly, there is **no room** for ~~numerical~~ weights

The basic idea…

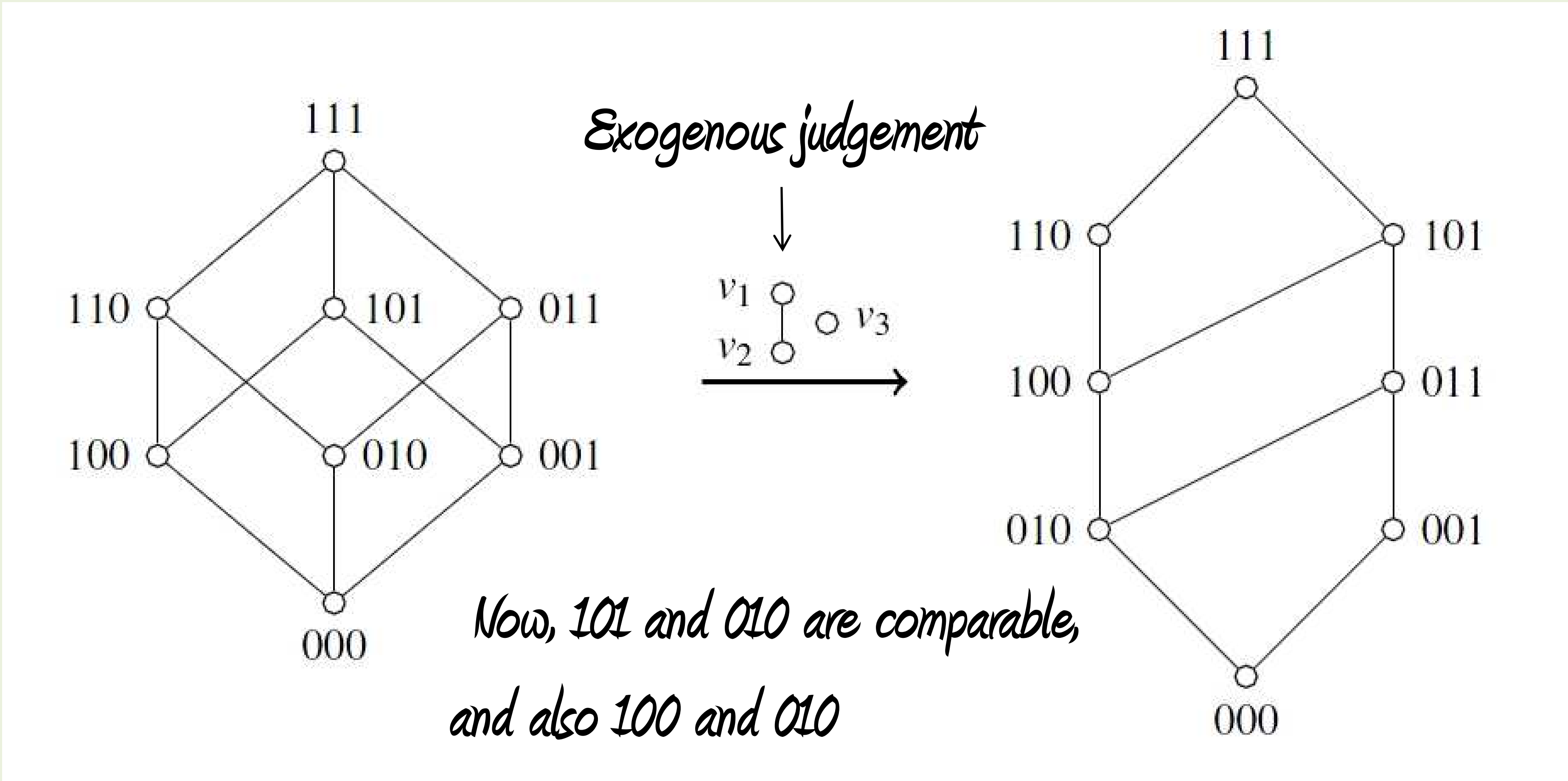
If we state that v1 is more relevant than (say) v2, than some incomparabilities in the poset can be resolved.

For example, we can affirm that 100 is more deprived than 010.

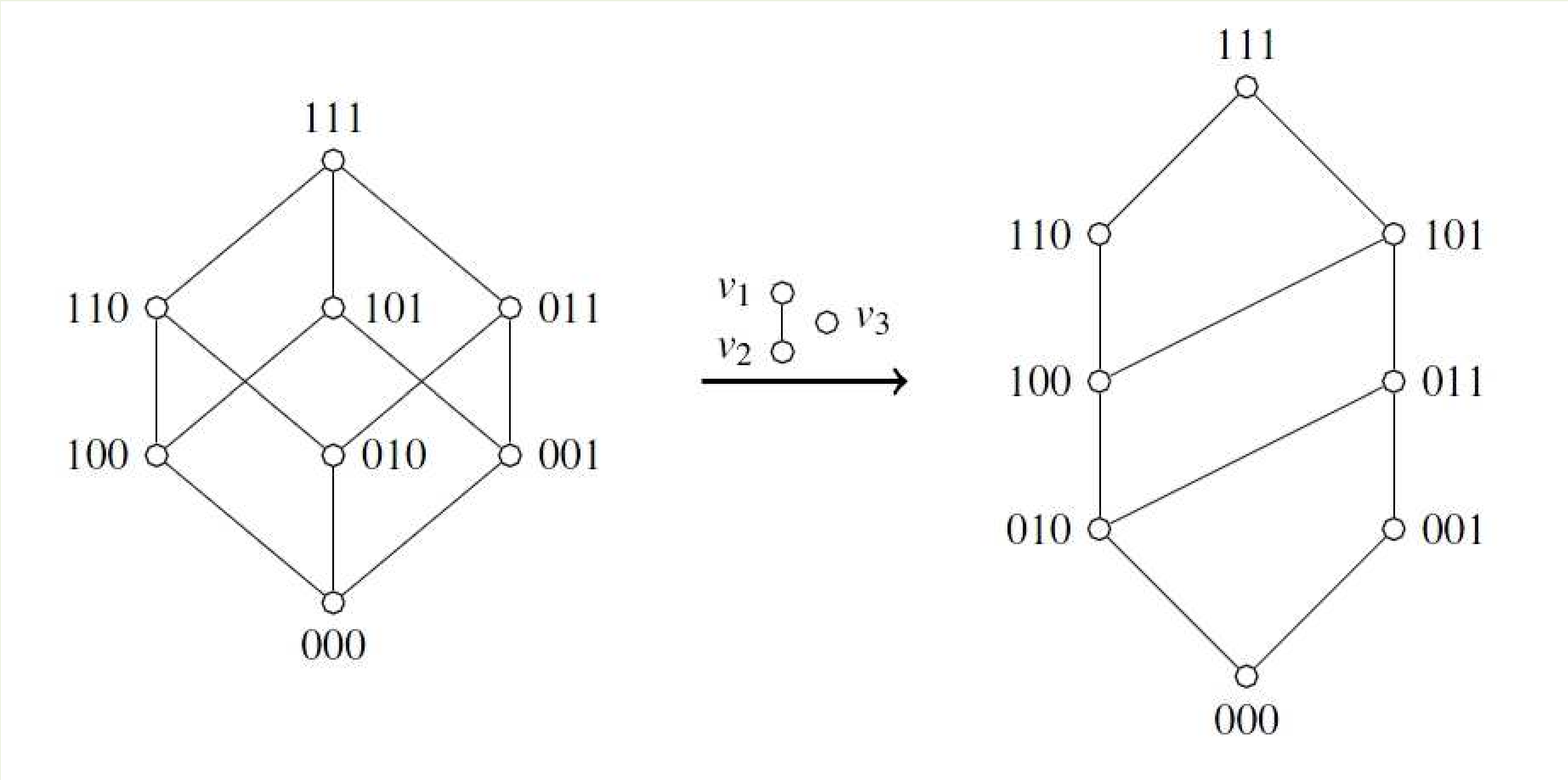
Adding new comparabilities (i.e. extending the poset) is the way exogenous information about the relevance of deprivation dimensions is embedded into the analysis.

For short: in an ordinal framework "weighting" turns into "extending"

An example

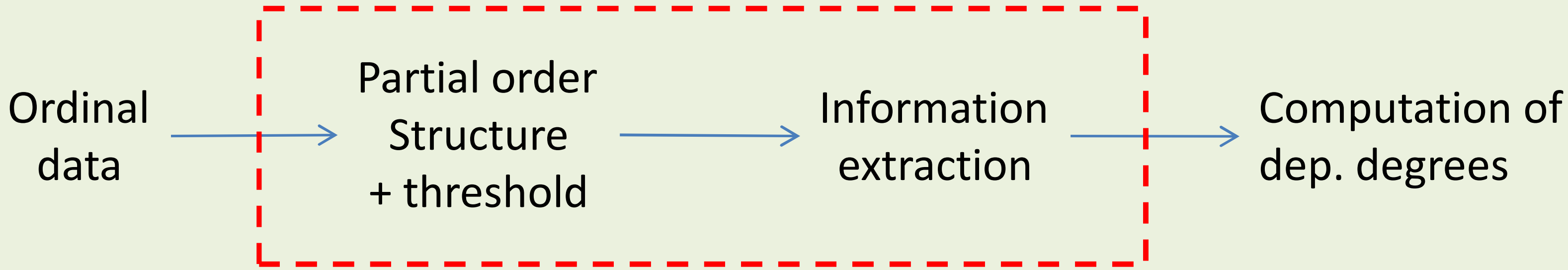

111
110
101
011
100
010
001
000
Exogenous judgement
v_1
v_2
v_3
111
110
101
100
011
010
001
000
Now, 101 and 010 are comparable,
and also 100 and 010

An example

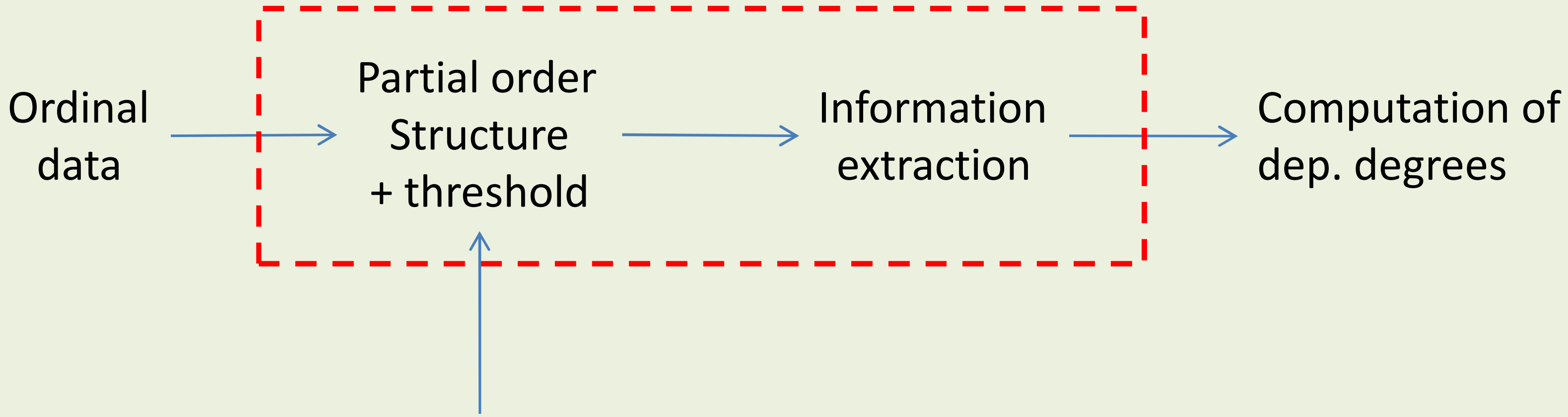


The information conveyed by the partial order among deprivation dimensions has been embedded in the poset and is successively extracted by the evaluation algorithm

The evaluation process

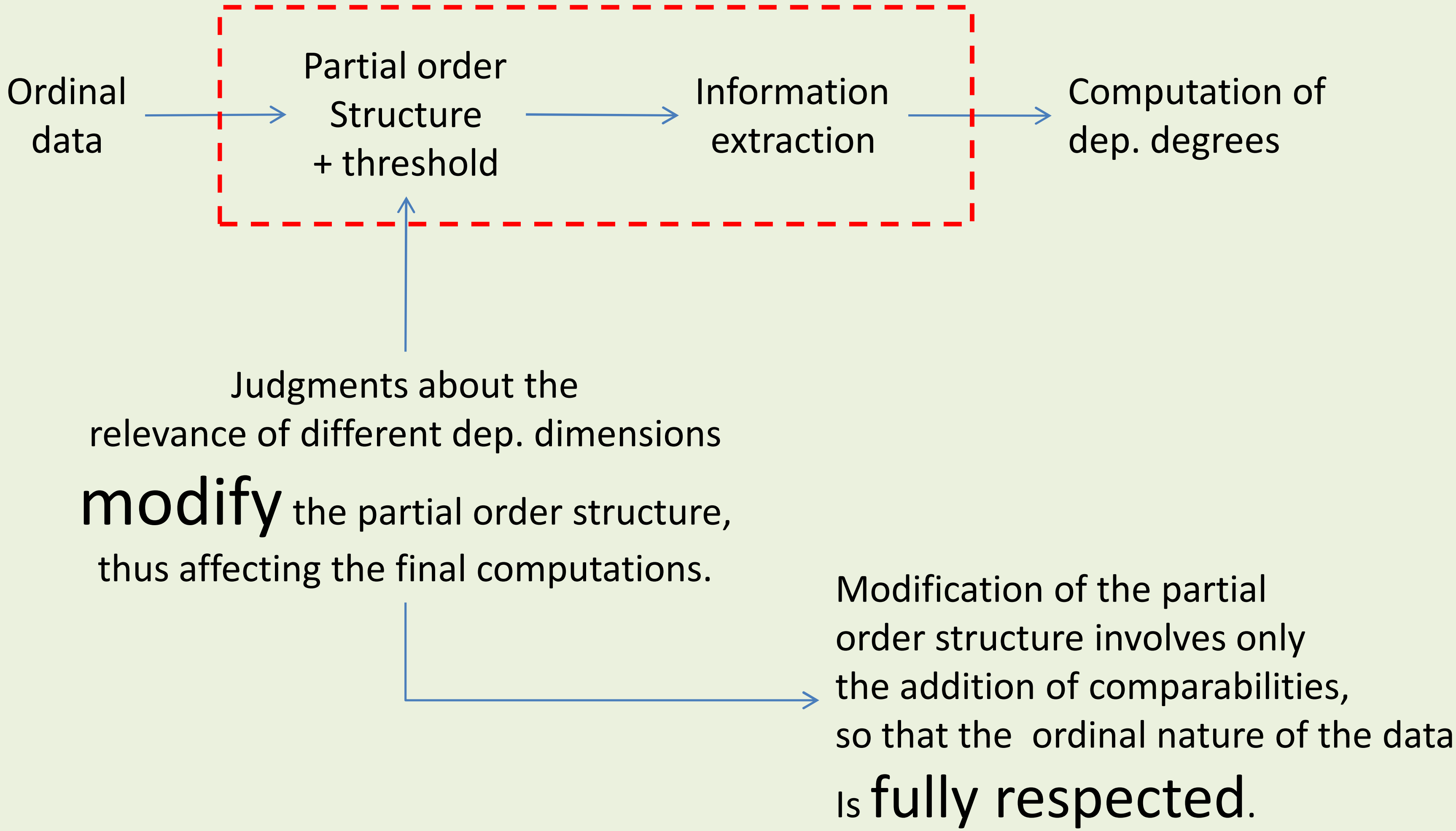


The process



Judgments about the relevance of different dep. dimensions **modify** the partial order structure, thus affecting the final computations.

The process


Ordinal data
Partial order Structure + threshold
Information extraction
Computation of dep. degrees
Judgments about the relevance of different dep. dimensions **modify** the partial order structure, thus affecting the final computations.
Modification of the partial order structure involves only the addition of comparabilities, so that the ordinal nature of the data Is **fully respected**.

Future research

1. Giving an axiomatic basis to the approach (we are close to that).

2. Developing software routines (possibly in R).

3. Applying the approach to real data and comparing it with other approaches.

4. Integrating poset methodologies in larger contexts, also with numerical variables.

Some references

- Fattore M., Maggino F., Colombo E., ‘’From composite indicators to partial orders: evaluating socio-economic phenomena through ordinal data’’, accepted for publication in Maggino F., Nuvolati G., (Eds.) *Quality of life. Reflections, studies and researches in Italy, Social Indicators Research Series, Springer (in press).*

- Fattore M., Maggino F., Greselin F., ‘’Socio-economic evaluation with ordinal variables: integrating counting and poset approaches ’’, accepted for publication in *Statistica & Applicazioni.*

- Fattore M., Brueggemann R., Owsinski J., ‘’Using poset theory to compare fuzzy multidimensional material deprivation across regions”, in Ingrassia S., Rocci R., Vichi M. (eds.) *New Perspectives in Statistical Modeling and Data Analysis, Springer-Verlag* - 2011, ISBN 978-3-642-11362-8.