February 2021

FEEM
FONDAZIONE ENI ENRICO MATTEI

Working Paper
001.2021

Partial order algorithms for the assessment of Italian cities sustainability

Alberto Arcagni, Laura Cavalli, Marco Fattore

Partial order algorithms for the assessment of Italian cities sustainability

By Alberto Arcagni, University of Rome La Sapienza
Laura Cavalli, ENI Enrico Mattei Foundation
Marco Fattore, University of Milano-Bicocca

Summary

In this paper, we introduce some recent non-aggregative algorithms, for the construction of synthetic indicators on multi-indicators systems, and exemplify them on data pertaining to the sustainability of main Italian cities. The procedure employs tools from Partial Order Theory and computes the final synthetic scores, without the compensation effects of classical composite indicators, so better preserving the nuances of city sustainability. Since turning multi-indicator systems into single synthetic indicators is unavoidably forcing, the procedure provides also ways to assess the degrees of "rankability" of statistical units, so to help researchers evaluating the soundness of the synthesis process. The algorithms introduced in the paper can be easily implemented, using the package Parsec, freely available in the R ecosystem.

Keywords: Partially Ordered Set, Syntethic Indcators, Multi-indicator Systems

JEL Classification: C14, C43, D81

Address for correspondence:
Marco Fattore
University of Milano-Bicocca
Department of Statistics and Quantitative Methods
Via Bicocca degli Arcimboldi, 8
20100 Milano
Italy
E-mail: marco.fattore@unimib.it

Partial order algorithms for the assessment of Italian cities sustainability

Alberto Arcagni - University of Rome La Sapienza,

Laura Cavalli - ENI Enrico Mattei Foundation,

Marco Fattore - corresponding author:
University of Milano-Bicocca
Department of Statistics and Quantitative Methods
Via Bicocca degli Arcimboldi, 8
20100 Milano MI
Italy
E-mail: marco.fattore@unimib.it

Abstract

In this paper, we introduce some recent non-aggregative algorithms, for the construction of synthetic indicators on multi-indicators systems, and exemplify them on data pertaining to the sustainability of main Italian cities. The procedure employs tools from Partial Order Theory and computes the final synthetic scores, without the compensation effects of classical composite indicators, so better preserving the nuances of city sustainability. Since turning multi-indicator systems into single synthetic indicators is unavoidably forcing, the procedure provides also ways to assess the degrees of "rankability" of statistical units, so to help researchers evaluating the soundness of the synthesis process. The algorithms introduced in the paper can be easily implemented, using the package Parsec, freely available in the R ecosystem.

Keywords. Partially ordered set; Syntethic indcators; Multi-indicator systems.
JEL Codes. C14, C43, D81.

1 Introduction

The aim of this paper is to illustrate a recently developed statistical approach for the construction of synthetic indicators on multi-indicator systems and to show it in action, in the assessment of the sustainability of main Italian cities. The distinctive feature of the procedure is that no aggregation among elementary indicators is performed, so avoiding the mix of "apples and oranges" typical of more classical composite indicators, which are compensative in nature and quite problematic in interpretation, when dealing with nuanced and multifaceted socio-economic or environmental traits. Instead of collapsing the elementary indicators into final figures, trying to measure sustainability against an absolute scale that does not actually exist, the synthetic scores are here computed by building a global system of comparisons among the multidimensional sustainability profiles of the cities and by quantifying their relative "dominance degrees", given the evaluation context provided by the input indicator system. The quantification of multidimensional comparison systems is possible thanks to the concepts and the tools provided by Partially Ordered Set theory, a branch of discrete mathematics designed to deal with order relations and particularly suitable to address multi-criteria decision problems. Indeed, the theory of partial orders has been recently and successfully applied to the construction of synthetic indicators in various socio-economic contexts ([1]; [2]; [3]; [5]; [6] [7]; [9]; [10]; [11]; [16]; [17]; [18]; [19]; [22]; [26]; [27]).

In the following, we first introduce the data used along the paper; we then give some basic concepts of partial order theory and describe the algorithms for the computation of synthetic scores, applying them to the evaluation of city sustainability, in Italy. To avoid too long abstract sections, we intertwine technical results and practical applications, so as to ease readers in catching the logic thread and appreciating the effectiveness of the tool. Finally, we provide a general comment on the use of partial order theory for the investigation of multi-indicator systems, highlighting open issues and research opportunities.

2 The data: sustainability of Italian cities

The data used in this paper refer to the sustainability of 103 Italian cities (capitals of the respective provinces). They come from various sources, mainly from the Italian National Statistical Bureau, and are collected by the ENI Enrico Mattei Foundation (FEEM), which publishes reports on the sustainability of main Italian cities (the data used in the paper have been taken from the 2020 update report, see [12] and [13], where full details on them can be found). The FEEM database comprises 46 elementary indicators, subdivided in the 17 Sustainable Development Goals (SDGs) of the UN Agenda 2030 ([29]). Since the aim of this paper is mainly to illustrate a new way of constructing sustainability synthetic indicators, here we focus on just 15 indicators, organized into two subgroups, namely *socio-economic* and *environment* (see Table 1), to be treated separately. The indicators are continuous in nature, but they will be later turned into discrete scales, prior to the application of partial order algorithms.

Table 1: Sustainability indicators selected for the analysis. The table is meant to describe the kind of information used in the analysis; the formal definition of the indicators and other details can be found in cited references.

Indicator	Reference SDG	Subgroup
Economic suffering index	1 - No poverty	Socio-economic
Taking charge index for childhood kindergarten	4 - Quality education	Socio-economic
Linguistic literacy (II year, high-school)	4 - Quality education	Socio-economic
Math literacy (II year, high-school)	4 - Quality education	Socio-economic
Individuals with only middle school degree	4: Quality education	Socio-economic
NEET incidence (15-29 years old)	8 - Decent work and economic growth	Socio-economic
School dropout	8 - Decent work and economic growth	Socio-economic
Number of social cooperatives per 10.000 inhabitants	17 - Partnership for the goals	Socio-economic
Water supply loss	6 - Clean water and sanitation	Environment
Residents connected to wastewater purification plants	6 - Clean water and sanitation	Environment
Solar panels per inhabitant	7 - Affordable and clean energy	Environment
Meters of bicycle lanes per 100 inhabitants	11 - Sustainable cities and communities	Environment
Average PM2.5 concentration	11 - Sustainable cities and communities	Environment
Separate waste collection on waste production	12 - Responsible consumption and production	Environment
Urban waste production	12 - Responsible consumption and production	Environment

3 Finite partially ordered sets and their matrix representations

In this section, we collect some basic notions of partial order theory, to make the paper self-consistent, in view of the introduction and the application of the scoring procedure. Informally speaking, partially ordered sets are just ordinary sets with the property that some, but in general not all, of their elements can be pairwise compared and ordered, other pairs of elements being instead *incomparable*. As such, they provide the most natural mathematical structure, to represent data which cannot be necessarily ordered in a complete way on a "low-high" axis, as it is the case of statistical units scored against multidimensional indicator systems. Notwithstanding the partiality of the order relation, partially ordered sets often comprise a big amount of information on the degree of dominance among their elements and proper tools can be employed to extract and turn it into synthetic scores and rankings. Notice that the existence of elements that cannot be compared is not necessarily the consequence of insufficient information on the trait of interest; more deeply, incomparability captures the existence of intrinsically different "shapes" of it, so providing a more realistic picture of the phenomenon under investigation. In this respect, the impossibility to compare the profiles of different cities, due to the existence of dimensions where they perform in conflicting ways, reveals the irreducible complexity of sustainability. Indeed, as it will be clarified along the paper, it is the capability of accounting for both comparabilities and incomparabilities that makes partially ordered sets so useful in the treatment of multi-indicator systems.

Main definitions. A *partial order relation* $\trianglelefteq$ over a finite set X is a *reflexive*, *antisimmetric* and *transitive* binary relation on X ([24]; [28]). The pair $\pi = (X, \trianglelefteq)$ is called a (finite) *partially ordered set* (or a *poset*, for short). If $\boldsymbol{x}_i \trianglelefteq \boldsymbol{x}_j$, or $\boldsymbol{x}_j \trianglelefteq \boldsymbol{x}_i$, then $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are said to be *comparable*, otherwise they are said to be *incomparable* (written $\boldsymbol{x}_i || \boldsymbol{x}_j$). If $\boldsymbol{x}_i \triangleleft \boldsymbol{x}_j$ (i.e. $\boldsymbol{x}_i \trianglelefteq \boldsymbol{x}_j$ and $\boldsymbol{x}_i \neq \boldsymbol{x}_j$), and there is no other element $\boldsymbol{x}_k$ such that $\boldsymbol{x}_i \triangleleft \boldsymbol{x}_k \triangleleft \boldsymbol{x}_j$, then $\boldsymbol{x}_j$ is said to *cover* $\boldsymbol{x}_i$

(written $\boldsymbol{x}_i \prec \boldsymbol{x}_j$). For finite posets, the cover relation uniquely determines the partial order relation, by transitivity, and this allows posets to be depicted as *Hasse diagrams*, i.e. as directed acyclic graphs built according to the following two rules: (i) if $\boldsymbol{x}_i \lhd \boldsymbol{x}_j$, then node $\boldsymbol{x}_i$ is placed above node $\boldsymbol{x}_j$ and (ii) if $\boldsymbol{x}_i \prec \boldsymbol{x}_j$ an edge is inserted between the two nodes (see Figure 1). A poset where all of the elements are comparable is called a *linear* (or *complete*, or *total*) order or a *chain*, while a poset whose elements are all incomparable is called an *antichain*.

Extensions of a finite poset. Let $\pi_1 = (X, \unlhd_1)$ and $\pi_2 = (X, \unlhd_2)$ be two posets on the same set X. π_2 is an *extension* of π_1, if $\boldsymbol{x}_i \unlhd_1 \boldsymbol{x}_j$ implies $\boldsymbol{x}_i \unlhd_2 \boldsymbol{x}_j$, i.e. if π_2 is obtained by turning some incomparabilites of π_1 into comparabilities, so extending the set of pairs of elements of X belonging to the order relation. An extension λ of a poset π, which is also a linear order, is called a *linear extension* of π; the set of linear extensions of π is denoted by $\Omega(\pi)$. A fundamental theorem in poset theory states that $\Omega(\pi)$ uniquely determines π by intersection [28] or, equivalently, that $\boldsymbol{x}_i \unlhd \boldsymbol{x}_j$ in π if and only if $\boldsymbol{x}_i \unlhd_\lambda \boldsymbol{x}_j$ in each linear extension λ of π. This result is a cornerstone of the scoring algorithms presented later in the paper.

Matrix representations of a finite poset. Finite posets can be conveniently represented by means of square matrices, namely the so-called *incidence*, *cover*, and *mutual ranking probability* matrices, which convey, in different ways, essential information on the structure of the order relation.

1. **Incidence matrix ($\boldsymbol{Z}$)**. The most natural algebraic representation of a finite poset π is by means of the so-called Z matrix, whose entries are defined as $Z_{ij} = 1$ if $\boldsymbol{x}_i \unlhd \boldsymbol{x}_j$ and $Z_{ij} = 0$, otherwise. Matrix Z is essentially the characteristic function of the partial order relation and simply lists the pairs of elements belonging to it.
2. **Cover matrix ($\boldsymbol{C}$)**. Since by transitivity $\unlhd$ is determined by the associated cover relation $\prec$, the binary matrix C defined as $C_{ij} = 1$ if $\boldsymbol{x}_i \prec \boldsymbol{x}_j$ and $C_{ij} = 0$ otherwise, implicitly provides an equivalent representation of π.
3. **Matrix of mutual ranking probabilities ($\boldsymbol{M}$)**. By "mutual ranking probability" (MRP) of $\boldsymbol{x}_j$ with respect to $\boldsymbol{x}_i$ it is meant the fraction of linear extensions of the input poset π, where $\boldsymbol{x}_j$ is ranked higher than $\boldsymbol{x}_i$, i.e. the probability of picking uniformly at random a linear extension λ such that $\boldsymbol{x}_i \unlhd_\lambda \boldsymbol{x}_j$, out of $\Omega(\pi)$. By collecting all of these probabilities into a matrix, we get the MRP matrix M ([15]), whose entries are formally defined by:

$$M_{ij} = \frac{|\{\lambda \in \Omega(\pi) : \boldsymbol{x}_i \unlhd_\lambda \boldsymbol{x}_j\}|}{|\Omega(\pi)|} \quad (i, j = 1, \ldots, |X|). \tag{1}$$

   By construction, the diagonal of M is composed of 1s and $M_{ij} + M_{ji} = 1$ $(i \neq j)$.

The three matrices Z, C and M can be, at least in principle, obtained one from the other, in fact:

1. $\boldsymbol{C} \Leftrightarrow \boldsymbol{Z}$. The incidence matrix can be obtained from the cover matrix by the formula $Z = Bin(C^{|X|-1})$, where $Bin(\cdot)$ puts to 1 all of the non-null entries of its argument. Viceversa, one gets C from Z as $C = H - Bin(H^2)$, where H is obtained from Z by putting all of its diagonal elements to 0 (i.e. $H = Z - I$). Both formulas are easily proved in [25].

2. $\boldsymbol{Z} \Leftrightarrow \boldsymbol{M}$. There is no closed formula leading from Z to M, but starting from the incidence matrix one can compute all of the linear extensions of the input poset and then get the matrix of mutual ranking probabilities, by direct enumeration. On the contrary, given M one gets immediately Z, by noticing that $Z_{ij} = 1 \Leftrightarrow \boldsymbol{x}_i \trianglelefteq \boldsymbol{x}_j \Leftrightarrow M_{ij} = 1$. In other words, Z is obtained by putting to 0 all of the elements of M less than 1.

By composing the above relations, we can also link matrices C and M, passing through Z.

Toy example. Let v_1 and v_2 be two statistical variables, with possible values 1, 2, 3. By building all possible combinations of these scores, we get 9 profiles which can be quite naturally partially ordered as depicted by the Hasse diagram in the left panel of Figure 1. Two profiles $\boldsymbol{p} = (p_1, p_2)$ and $\boldsymbol{q} = (q_1, q_2)$ linked by a downward sequence of edges are comparable and it holds $\boldsymbol{p} \trianglelefteq \boldsymbol{q} \Leftrightarrow p_1 \leq q_1$ *and* $p_2 \leq q_2$ (called the *product order* of the linear orders associated to v_1 and v_2); the right panel of the figure depicts a linear extension of the poset. The corresponding incidence, cover and MRP matrices are given below:

$Z =$

Prf	(33)	(32)	(23)	(31)	(22)	(13)	(21)	(12)	(11)
(33)	1	0	0	0	0	0	0	0	0
(32)	1	1	0	0	0	0	0	0	0
(23)	1	0	1	0	0	0	0	0	0
(31)	1	1	0	1	0	0	0	0	0
(22)	1	1	1	0	1	0	0	0	0
(13)	1	0	1	0	0	1	0	0	0
(21)	1	1	1	1	1	0	1	0	0
(12)	1	1	1	0	1	1	0	1	0
(11)	1	1	1	1	1	1	1	1	1

$C =$

Prf	(33)	(32)	(23)	(31)	(22)	(13)	(21)	(12)	(11)
(33)	0	0	0	0	0	0	0	0	0
(32)	1	0	0	0	0	0	0	0	0
(23)	1	0	0	0	0	0	0	0	0
(31)	0	1	0	0	0	0	0	0	0
(22)	0	1	1	0	0	0	0	0	0
(13)	0	0	1	0	0	0	0	0	0
(21)	0	0	0	1	1	0	0	0	0
(12)	0	0	0	0	1	1	0	0	0
(11)	0	0	0	0	0	0	1	1	0

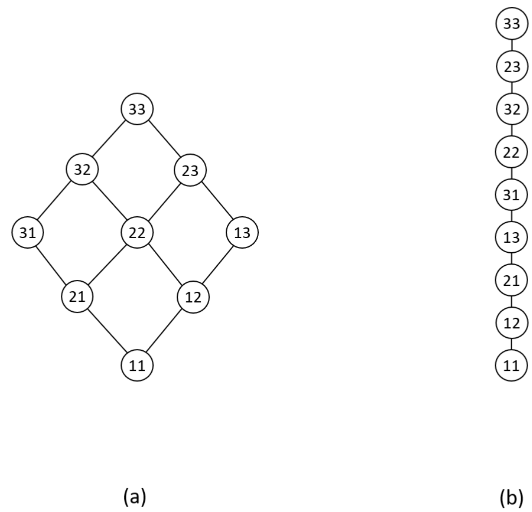

Figure 1: Product order over two variables scored on 1,2, and 3 (left panel) and one of its linear extensions (right panel).

$$
M = \begin{array}{c|ccccccccc}
Prf & (33) & (32) & (23) & (31) & (22) & (13) & (21) & (12) & (11) \\
\hline
(33) & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
(32) & 1 & 1 & 0.5 & 0 & 0 & 0.12 & 0 & 0 & 0 \\
(23) & 1 & 0.5 & 1 & 0.12 & 0 & 0 & 0 & 0 & 0 \\
(31) & 1 & 1 & 0.88 & 1 & 0.5 & 0.5 & 0 & 0.12 & 0 \\
(22) & 1 & 1 & 1 & 0.5 & 1 & 0.5 & 0 & 0 & 0 \\
(13) & 1 & 0.88 & 1 & 0.5 & 0.5 & 1 & 0.12 & 0 & 0 \\
(21) & 1 & 1 & 1 & 1 & 1 & 0.88 & 1 & 0.5 & 0 \\
(12) & 1 & 1 & 1 & 0.88 & 1 & 1 & 0.5 & 1 & 0 \\
(11) & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{array}
$$

Sustainability posets. As mentioned before, the sustainability data described in Section 2 have been subdivided in two datasets, the first comprising socio-economic indicators and the second comprising environmental indicators (see Table 1). Each city gets assigned both a socio-economic and an environmental sustainability profile, producing two posets, whose structure is depicted in Figure 2 (the indicators have been oriented so that higher values correspond to higher sustainability). The Hasse diagrams reveal a quite intricate network of comparabilities and incomparabilities. In both posets there are no “best” and no “worst” cities; there are indeed maximal and minimal units (i.e. cities that are not dominated by others and cities that do not dominate any others) and, interestingly, some cities are incomparable with all of the others. The complexity of the resulting patterns is, quite clearly, at odds with the aggregative-compensative road to synthesis, typical of composite indicators and calls for alternative algorithms.

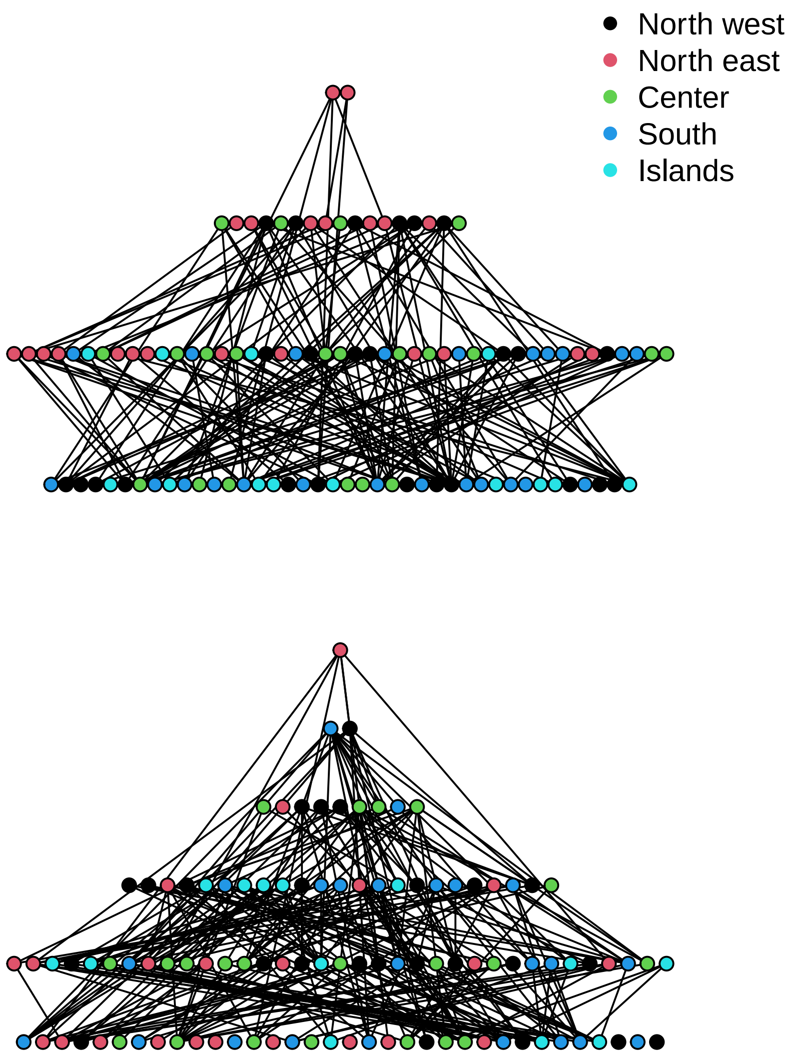

Figure 2: Socio-economic poset (top panel) and environmental poset (bottom panel) for Italian cities. Labels have been omitted and the Hasse diagrams are only meant to provide a general idea of the partial order structure of the data (7 cities with missing data, have been excluded from the diagrams).

4 The scoring procedure

Given a poset π, our first aim is to define a *scoring* function $s(\cdot)$, i.e. a map from π to the non-negative reals, such that whenever $\boldsymbol{x}_i \lhd \boldsymbol{x}_j$, it is $s(\boldsymbol{x}_i) < s(\boldsymbol{x}_j)$ (i.e., the scoring function is required to be *strictly order preserving*, so that when two different profiles can be compared in the input poset, they get scored consistently by $s(\cdot)$). Quite intuitively, the score $s(\boldsymbol{x})$ should reflect the degree of dominance of $\boldsymbol{x}$ over the other profiles in the partially ordered set. For example, in a hypothetical product order built on 5 variables each measured on $[0, 1]$, profiles (1, 1, 1, 1, 0.4) and (0, 0, 0, 0, 0.5) are incomparable, but clearly the first "almost" dominates the second and should be scored higher by any reasonable scoring procedure. In principle, the information on the dominance degrees among poset profiles can be recovered by any matrix representation of the partial order relation. However, the incidence matrix Z, and more even so the cover matrix C, provides a "too implicit" information on the relative position of profiles and would deliver poorly discriminating scores. Much more information on dominances is instead comprised in the MRP matrix, whose generic entry M_{ij} reveals how frequently the j-th profile dominates the i-th profile, in the set of linear extensions of the input poset. We thus compute the scoring function $s(\cdot)$ on the mutual ranking probabilities associated to poset elements; to this goal, the main result is provided by the following proposition (see [21]):

Proposition. Let π be a poset on a set $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, let M be the associated MRP matrix, let $M = UDV^T$ be the singular value decomposition of M and let $\boldsymbol{v} = \{v_1, \ldots, v_n\}^T$ be the first column of V. Then the map

$$s \quad : \quad \pi \rightarrow \mathbb{R}_0^+ \tag{2}$$

$$\quad : \quad \boldsymbol{x}_i \mapsto s(\boldsymbol{x}_i) = v_i \tag{3}$$

is strictly order preserving.

The interpretation of the scoring function $s(\cdot)$ is quite straightforward: essentially, it assigns to each element $\boldsymbol{x}_i$ a weighted mean of the elements of the $i-th$ column of M, i.e. a weighted mean of its dominance degrees over the poset elements (with optimal weights provided by the components of the first row of matrix $D^{-1}U^T$). Notice how the final scores does not depend explicitly upon the figures within the poset profiles, but on the entire network of comparabilities/incomparabilities among them. Indeed, the scoring function does not involve any aggregation of the input variables and extracts the dominance information just out of the structure of the partial order relation.

Sustainability scores of Italian cities. We now apply the above scoring algorithm to the socio-economic and environmental datasets on Italian cities, presented in Section 2. In order to get as much information as possible out of the data, we properly tune the procedure, so as to exploit the feature of the sustainability posets, namely that they are built as product orders, from multi-indicator systems. To ease following the logic thread of the procedure, we organize it in steps.

1. *Indicator discretization and construction of the global product orders*. The input socio-economic and environmental indicators are preliminarily made discrete, by subdividing their observed ranges into 10 equi-spaced intervals,

leading to two product orders, π_{sec} and π_{env}, with 10^8 and 10^7 elements respectively, which comprise all possible profiles built on the discretized indicators. The two global posets are the evaluation spaces within which socio-economic and environmental scores will be computed and their introduction, based on the discretization of the input indicators, can be motivated as follows. First, when partial orders are built on continuous variables, small measurement errors and noise in the data may produce artificial incomparabilities among statistical units, impoverishing the poset structures; to get more robust and meaningful results, it is then advisable to transform the input data into discrete ones, as above. Second, the sustainability profiles of Italian cities are just a subset of all possible (infinite) profiles, that could be built on the (continuous) socio-economic and environmental variables. Computing the mutual ranking probabilities just on such a subset, however, would lose information on the "distances" among the profiles, finally squeezing the distribution of the final scores. To clarify this point, consider a trivial two-element poset comprising two profiles (4,2) and (1,3), built out of two indicators measured on the discrete numerical set $\{1, 2, 3, 4\}$. Such two profiles are incomparable with relative dominance degrees equal to 0.5, the poset having just two linear extensions, namely $(1, 3) \lhd (4, 2)$ and $(4, 2) \lhd (1, 3)$. Suppose, now, to embed the profiles into the "global" product order poset composed of all possible profiles (namely, 16) of the cartesian product $\{1, 2, 3, 4\} \times \{1, 2, 3, 4\}$; in this case, the degrees of dominance of (4,2) over (1,3) and of (1,3) over (4,2) would differentiate, the first becoming higher than the second. This happens, since the global product order has several linear extensions, with additional profiles in-between (4,2) and (1,3), that better resolve the different "relational" positions of the two elements. Not being possible to actually build the infinite posets of all possible socio-economic and environmental profiles and to compute mutual ranking probabilities on them, the input variables must be discretized. In this respect, as mentioned above, the two posets can be considered as two evaluation spaces, in which observed data are embedded and the choice of discretizing the input variables into ten levels can be interpreted as a way to set the "resolution" of such evaluation process (in practice, through the socio-economic and the environmental posets, we can resolve 10^8 and 10^7 different sustainability patterns, respectively).

2. *Computation of the mutual ranking probability matrices*. From the socio-economic and the environmental posets, the respective two mutual ranking probability matrices M_{sec} and M_{env} get computed. Given the high number of elements in both posets, the computations have been performed in an approximate way, by just considering the subset of so-called *lexicographic linear extensions* and using closed analytical formulas (see [23]) which reduce the computational burden of the algorithm. Lexicographic linear extensions are obtained by completely ordering the poset profiles as we usually do with words in a dictionary, but where variable degrees stand for letters, and considering all of the possible variable permutations. So, in a product order built on k variables, there are $k!$ lexicographic linear extensions and these uniquely determine the input poset (for details on the computation of linear extensions and the concept of lexicographic linear extensions, see [8]; [14]; [20]).

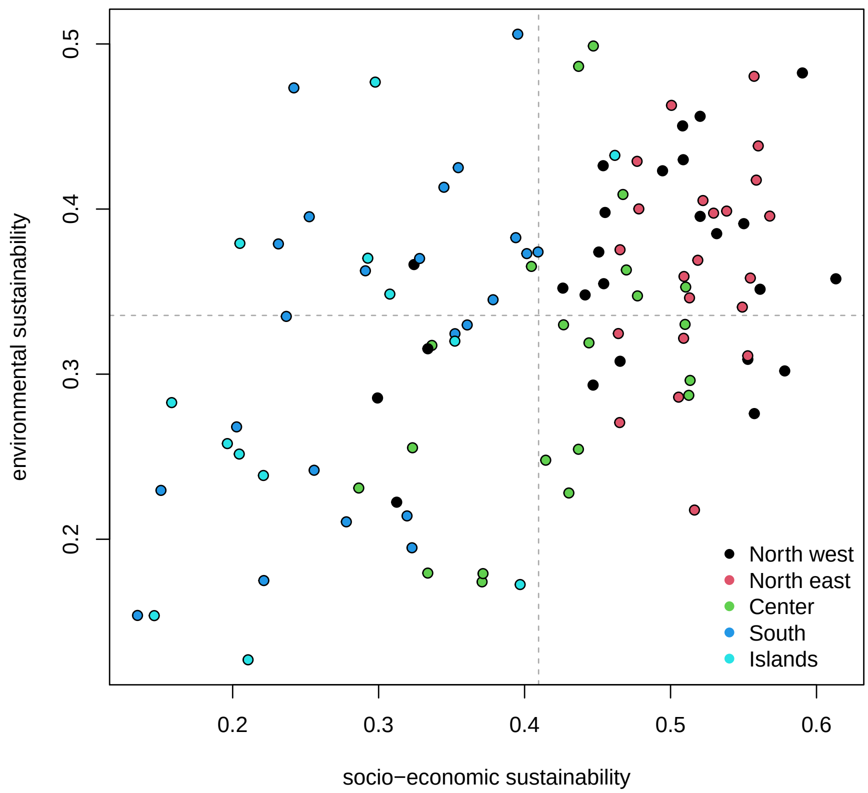

Figure 3: Scatterplot of socio-economic and environment sustainability scores, for Italian cities.

3. *Computation of sustainability scores*. From the matrices M_{sec} and M_{env} obtained above, the rows and columns corresponding to the Italian cities are extracted, getting the matrices M_{sec}^{cities} and M_{env}^{cities}, relative to the observed profiles. The final score vectors $\boldsymbol{v}_{sec}^{cities}$ and $\boldsymbol{v}_{env}^{cities}$ are finally computed, by performing the SVD decompositions of such two matrices. The results plotted in Figure 3 clearly show the typical Italian South-North axis, with cities in the Southern regions in worse situations than those in the Northern part of the Country. The plot also reveals that when socio-economic and environment sustainability scores are unbalanced, Northern cities are better off in the first dimension, while Southern ones prevail in the second. Interestingly, there is some gap, in the middle of the plot, between cities in the bottom-left quadrant and cities in the top-right quadrant, revealing the socio-economic polarization of Italy.

Remark 1. The final scores have been computed on M_{sec}^{cities} and M_{env}^{cities}, and not on M_{sec} and M_{env}, since we are interested in measuring the degree of dominance of one city over the others, excluding non-observed profiles, which have been nevertheless taken into account, in the construction of the MRP matrices, as explained previously.

Remark 2 - Dominance score normalization. The dominance scores have been normalized to [0-1], dividing them by the maximum dominance score in a poset with the same number of elements of the sustainability posets, but where there is one element which dominates all of the others, the these being all incomparable and thus forming an antichain.

Remark 3. Just a few cities have missing values in one or both profiles (namely, Andria, Barletta, Carbonia, Cesena, Fermo, Trani, Urbino) and have been removed from the analysis.

Incomparability scores of Italian cities. Turning multi-indicator systems into synthetic scores is unavoidably forcing, since incomparabilities get transformed into comparabilities and the complexity of the input data get lost into a uni-dimensional space. It is thus important to complement the sustainability scores with some information on the degree of incomparability among profiles, so as to control for the "distortion" implied by the scoring function. This can be achieved quite easily, by building the symmetric *incomparability* matrix INC (see [26]), whose entries measure the degree of incomparability between pairs of poset elements and are formally defined by

$$INC_{ij} = \min(M_{ij}, M_{ji}) \tag{4}$$

(indeed, if the mutual ranking probabilities between two elements are (say) 0.8 and 0.2, their degree of incomparability is 0.2). From the incomparability matrix, a vector of incomparability scores is obtained by performing the same kind of decomposition used for the computation of the dominance scores (however, since INC is symmetric, the singular value decomposition reduces to the spectral decomposition $INC = UDU^T$). Analogously to what done previously, also the incomparability scores have been normalized, dividing them by the incomparability degree of an element of an antichain, with the same cardinality of the sustainability posets. Complementing the sustainability scores with the incomparability ones, for both the socio-economic and the environmental posets, we get the plots depicted in Figures 4 and 5. The plots clearly show that Italian cities are not that comparable, in term of both socio-economic and environmental sustainability (at least, given the selected indicators). Indeed, the normalized dominance scores span a quite limited range in [0-1], with maximum dominance values that in both posets are far from the theoretical maximum. At the same time, a significant number of cities have incomparability scores between 0.4 and 0.5 particularly, as typical in this kind of analysis, for units with sustainability scores in the middle of the distribution. This shows that, while the dominance scores at the top and the bottom of the distribution can be taken reliably, those in the middle are to be considered with care; indeed, the corresponding cities cannot be actually ranked, being largely incomparable with many of the other units. We thus see how the proposed procedure captures the structure of the data and reveals their intrinsic complexity, helping the researchers

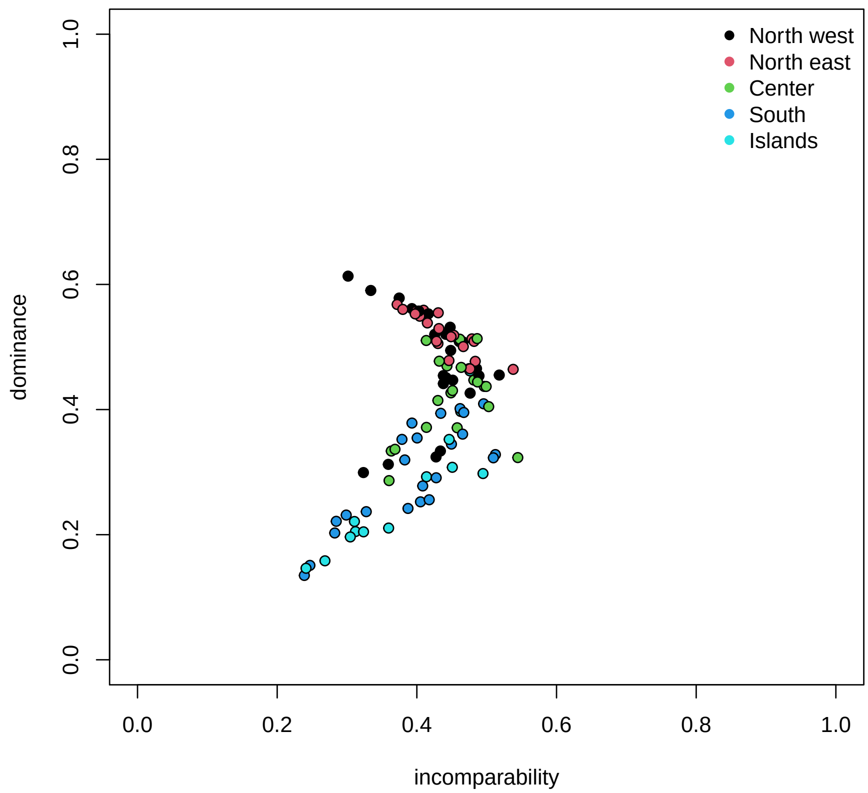

Figure 4: Italian cities plotted in the socio-economic sustainability (dominance) - incomparability plane.

to better realize the reliability of the information conveyed by the data.

5 Conclusion and outlook

In this paper, we have illustrated the application of recent posetic algorithms to the computation of synthetic scores of sustainability, for main Italian cities, along socio-economic and environmental dimensions. The distinctive feature of the procedure is that synthetic scores are obtained without any aggregation of elementary indicators and that sustainability scores get complemented with measures of incomparability of the statistical units, so delivering to decision-makers information on the reliability of the results. The plots presented in the previous paragraphs clearly show the sustainability polarization of Italian cities and, in particular, the existence of a deep

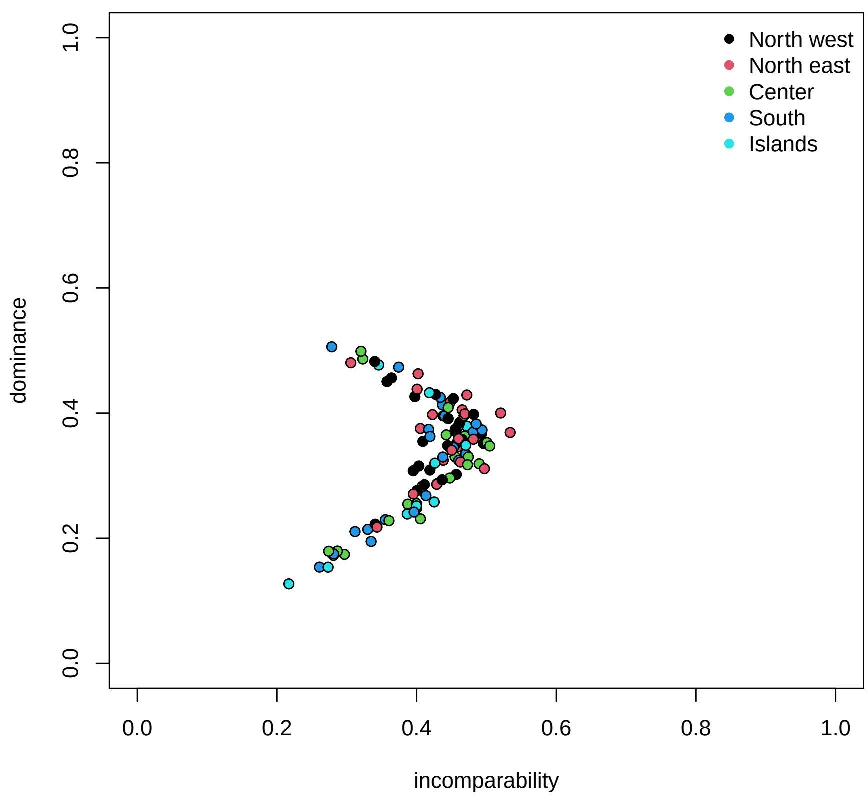

Figure 5: Italian cities plotted in the environment sustainability (dominance) - incomparability plane.

socio-economic gap between the Southern part and the Northern part of the Country. At the same time, they reveal how city sustainability is an inherently intricate trait that can be hardly synthesized into single scores; they do so by quantifying the degree of incomparability among cities and showing that sustainability scores are affected by different levels of uncertainty. The algorithms employed in the paper are based on partial order theory, a branch of discrete mathematics providing concepts and tools, particularly useful to address multi-criteria decision problems. In the last years, posets have been increasingly used in the statistical treatment of multi-indicator systems, on both ordinal and cardinal scales. Interestingly, partial order theory is not only a technical toolbox, but has a deeper conceptual role, in that it provides the proper categories to address and interpret the complexity and nuances of multi-faceted traits, like sustainability. It is in particular the notion of incomparability that is of critical importance, since it allows to catch the inherently diverse shapes of sustainability, instead of looking at them as at a form of noise, to be "hopefully" eliminated. Indeed, the theory of partial order is full of concepts, results and tools that can be conveniently turned into statistical procedures, for the treatment of multidimensional indicator systems. Given the relevance of synthetic indicators, also in terms of decision-making and public communication, the development of such "posetic" tools should attract a great deal of research.

References

[1] Alaimo L., Arcagni A., Fattore M., Maggino F. (2020) "Synthesis of Multi-indicator System Over Time: A Poset-based Approach", Social Indicators Research - doi: 10.1007/s11205-020-02398-5.

[2] Alaimo L., Arcagni A., Fattore M., Maggino F., Quondamstefano V. (2020) "Measuring Equitable and Sustainable Well-Being in Italian Regions: The Non-aggregative Approach", Social Indicators Research - doi: 10.1007/s11205-020-02388-7.

[3] Annoni, P., Fattore, M. & Bruggemann R. (2011). *A Multi-Criteria Fuzzy Approach for Analyzing Poverty structure*. *Statistica & Applicazioni*, Special Issue, 7-30.

[4] Arcagni A., Barbiano di Belgiojoso E., Fattore M., Rimoldi S. (2019) "Multidimensional Analysis of Deprivation and Fragility Patterns of Migrants in Lombardy, Using Partially Ordered Sets and Self-Organizing Maps", *Social Indicators Research* 141(2), 551–579 - DOI: 10.1007/s11205-018-1856-9.

[5] Badinger H., Reuter W. H. (2015). Measurement of Fiscal Rules: Introducing the Application of Partially Ordered Set (POSET) Theory. *Journal of Macroeconomics*, 43, 108-123.

[6] Di Brisco, A. M., Farina P. (2018). Measuring Gender Gap from a Poset Perspective. *Social Indicators Research* 136, 1109 - 1124.

[7] Brueggemann R., Patil G. P. (2011) "Ranking and Prioritization for Multi-indicator Systems ", Springer-Verlag - New York

[8] Bubley R., Dyer M. (1999). Faster random generation of linear extensions. *Discrete Math.*, 201, 81-88.

[9] Carlsen L., Brueggemann R. (2017). Fragile State Index: Trends and Developments. A Partial Order Data Analysis. *Social Indicators Research* 133, 1-14).

[10] Carlsen L. (2017). An Alternative View on Distribution Keys for the Possible Relocation of Refugees in the European Union. *Social Indicators Research.*, 130, 1147-1163

[11] della Queva, S. (2017). Analysis of Social Participation: A Multidimensional Approach Based on the Theory of Partial Ordering. In M. Fattore, R. Bruggemann *Partial Order Concepts in Applied Sciences*, Springer.

[12] Cavalli L., Farnia L.. (2019) "SDSN Italia SDGs City Index", available at $https://www.feem.it/it/ricerca/programmi/firms-and-cities-towards-sustainability/ongoing-projects/indice-di-sviluppo-urbano-sostenibile-per-l-italia-ii-sdg-/$.

[13] Cavalli L., Farnia L., Lizzi G., Romani I., Alibegovic M., Vergalli S. (2020) "L'SDSN Italia SDGs City Index per un'Italia Sostenibile: Report di aggiornamento", available at $https://www.feem.it/m/publications_pages/rpt-2020-city-index.pdf$.

[14] De Loof K., De Baets B., & De Meyer H. (2006). Exploiting the Lattice of Ideals Representation of a Poset. *Fundamenta Informaticae*, 71 (2-3), 309-321.

[15] De Loof K., De Baets B., & De Meyer H. (2008). Properties of mutual rank probabilities in partially ordered sets. In Owsinski J. W., & Bruggemann R. (Eds.). *Multicriteria ordering and ranking: partial orders, ambiguities and applied issues*. Warsaw: Polish Academy of Sciences.

[16] Fattore M., Maggino F., (2015). A new method for measuring and analyzing suffering - Comparing suffering patterns in Italian society. In Anderson R. E. (Eds.). *World Suffering and the Quality of Life*. New York: Springer.

[17] Fattore M., (2016). Partially ordered sets and the measurement of multidimensional ordinal deprivation. Social Indicators Research, 128(2), 835-858.

[18] Fattore M., Maggino F., Arcagni A. (2016) "Non-aggregative assessment of subjective well-being", In G. Alleva, A. Giommi (eds.), Topics in Theoretical and Applied Statistics, Springer International Publishing Switzerland, DOI 10.1007/978-3-319-27274-0_20.

[19] Fattore M. (2018) "Non-aggregated indicators of environmental sustainability", Silesian Statistical Review, 16(22), 7-22.

[20] Fattore M., Arcagni A. (2018) "A reduced posetic approach to the measurement of multidimensional ordinal deprivation", *Social Indicators Research*, 136(3), 1053-1070.

[21] Fattore M., Arcagni A., Maggino F. (2019) "Optimal scoring of partially ordered data, with an application to the ranking of smart cities", Smart Statistics for Smart Applications Book of Short Papers SIS 2019 - ISBN 9788891915108.

[22] Fattore M., Zenga Ma. (2019) "New posetic tools for the evaluation of financial literacy", In Bianco A., Gnaldi M., Conigliaro. P. (eds.), Italian Studies on Quality of Life, Springer, 978-3-030-06021-3.

[23] Fattore M., Arcagni A (2020) "Ranking extraction in ordinal multi-indicator systems ", Book of Short Papers - SIS 2020 - ISBN 9788891910776 - Pearson.

[24] Neggers J., Kim S. H. (1998). *Basic posets*. Singapore: World Scientific.

[25] Patil G.P., Taillie C. (2004). Multiple indicators, partially ordered sets and linear extensions: Multi-criterion ranking and prioritization. *Environmental and ecological statistics*, 11(1).

[26] Rimoldi S.M.L., Arcagni A., Fattore M., Barbiano di Belgiojoso E.. (2020) "Targeting policies for multidimensional poverty and social fragility relief among migrants in Italy, using F-FOD analysis", Social Indicators Research - doi 10.1007/s11205-020-02485-7

[27] Rimoldi S.M.L., Arcagni A., Fattore M., Terzera T. (2020) "Social and material vulnerability of the Italian municipalities: comparing alternative approaches", Social Indicators Research- doi: 10.1007/s11205-020-02330-x.

[28] Schröder. (2002). *Ordered set. An introduction*. Boston: Birkäuser.

[29] https://www.un.org/sustainabledevelopment/development-agenda/

FONDAZIONE ENI ENRICO MATTEI WORKING PAPER SERIES
"NOTE DI LAVORO"
Our Working Papers are available on the Internet at the following address:
http://www.feem.it/getpage.aspx?id=73&sez=Publications&padre=20&tab=1

"NOTE DI LAVORO" PUBLISHED IN 2021

1. 2021, Alberto Arcagni, Laura Cavalli, Marco Fattore, Partial order algorithms for the assessment of italian cities sustainability

Fondazione Eni Enrico Mattei
Corso Magenta 63, Milano – Italia

Tel. +39 02.520.36934
Fax. +39.02.520.36946

E-mail: letter@feem.it
www.feem.it

FEEM
FONDAZIONE ENI
ENRICO MATTEI