Visualizing Partially Ordered Sets for Socioeconomic Analysis

Visualización de conjuntos ordenados parciales para análisis socioeconómicos

Marco Fattore[1,a], Alberto Arcagni[1,b], Stefano Barberis[2,c]

[1]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy

[2]Company SmartStat, Italy

Abstract

In this paper, we develop a visualization process for partial orders derived from considering many numerical indicators on a statistical population. The issue is relevant, particularly in the field of socio-economic evaluation, where explicitly taking into account incomparabilities among individuals proves much more informative than adhering to classical aggregative and compensative approaches, which collapse complexity into unidimensional rankings. We propose a process of visual analysis based on a combination of tools and concepts from partial order theory, multivariate statistics and visual design. We develop the process through a real example, based on data pertaining to regional competitiveness in Europe.

Key words: Partial Order, Hasse Diagrams, Self-Organizing Map, Visualization.

Resumen

En este artículo, se desarrolla un proceso de visualización para órdenes parciales derivados al considerar varios indicadores numéricos de una población estadística. Esto es relevante parcialmente en el área de la evaluación socioedonómica donde se toma en cuenta explicitamente incomparabilidades entre los individuos y resulta ser más informativo que la agregación clásica y compensativa que colapsa la complejidad en rankings unidimensionales. Se propone un proceso de análisis visual basado en la combinación herramientas y conceptos de la teoría de orden parcial, estadítica multivariada y diseño

[a]Professor. E-mail: marco.fattore@unimib.it

[b]Professor. E-mail: alberto.arcagni@unimib.it

[c]Researcher. E-mail: stefanobarberis@hotmail.it

usual. Se desarrollo el proceso a través de un ejemplo real, basado en datos de competitividad regional en Europa

Palabras clave: diagramas Hasse, mapa autoorganizado, orden parcial, visualización.

1. Introduction

A detailed study on the economic competitiveness of European regions has been recently published by the Joint Research Centre (Annoni & Dijkstra 2013), to provide insights into the differences and the similarities of regional economic performances. A composite indicator, named RCI (Regional Competitiveness Index), has been computed based on a set of 73 elementary indicators, selected as relevant from a socio-economic point of view[1]. The RCI computation proceeds through a hierarchical process, where elementary indicators are weighted and progressively aggregated in so-called "pillars", until a single composite indicator is obtained. Once RCI is computed, inter-regional comparisons can be made and a final competitiveness ranking is achieved. This kind of aggregative process is prototypical of the way economists and social scientists usually address the assessment of multidimensional socio-economic issues, like competitiveness, well-being, quality-of-life and the like. In fact, multidimensional assessments are very often designed with the aim to return clear and unambiguous rankings of statistical units. Common practice shows that this can be achieved only at the cost of losing a great deal of information. Competitiveness, well-being, quality-of-life and many other similar socio-economic topics are complex, multidimensional, full of ambiguities, nuances and uncertainties. Turning them into unidimensional rankings is burdensome and not necessarily leads to clearly interpretable results. In essence, the problem resides in the fact that these issues are truly multidimensional. This is often confirmed by the absence of strong interrelations among elementary indicators, so that multidimensionality-reducing tools based on correlations (e.g. structural equation models) prove mostly ineffective in achieving any meaningful synthesis. Likewise, it must be noted that RCI primarily aims at measuring the level of competitiveness, despite no natural scale against which can be compared. More properly, competitiveness of a region can be compared to that of other regions, rather than assessed on an absolute scale. Due to multidimensionality, however, such comparisons generally do not lead to complete rankings but to partial orderings, since conflicting indicators in regional competitiveness profiles lead to incomparabilities. The impossibility of obtaining meaningful and unambiguous rankings is typical of multi-criteria decision problems and the relevance of taking this feature into account has been also noted by Nobel prize Sen, in his book on inequality (Sen 1992). It is thus very important for social scientists to get acquainted with this kind of data structures, that is, in technical terms, with partially ordered sets (Barthélemy, Flament & Monjardet 1982). In fact, one can easily figure out the consequences in policy decisions, when a policy-

---

[1]Elementary indicators come from many and heterogeneous different sources (Dijkstra, Annoni & Kozovska 2011). However, all of them are measured on numerical scales.

maker looks at regional competitiveness (or well-being, or quality of life...) in terms of unidimensional rankings, without realizing that different and incomparable competitiveness patterns do exist. Partially ordered sets have their drawbacks too, in that metric information gets lost. But this issue can be, at least partially, solved by exploiting suitable visualization tools, as shown below (see also Al-Sharrah (2014), for analogous attempts to introduce metric information in a partial order context). Generally speaking, the mathematical theory of partially ordered sets is well-established, but its application to socio-economic problems is at a beginning stage (Fattore, Bruggemann & Owsiński 2011, Fattore, Maggino & Greselin 2011, Fattore, Maggino & Colombo 2012). This motivates the present attempt to develop graphical and software tools devoted to the visualization of partial orders, to incline social scientists towards this way of looking at socio-economic data. The paper is organized as follows. In Section 2, we describe the structure of RCI more deeply and introduce the example used to illustrate the visualization tool. In Section 3, we present some elements of partial order theory and introduce Hasse diagrams, the basic visualization tool for partial orders. In Section 4, we provide some details on Self-Organizing Maps, the tool used to cluster statistical units prior to visualization. Section 5 develops the visualization tool. Section 6 provides a conclusion.

2. Regional Competitiveness Data

The Regional Competitiveness Index (RCI) proposed by the Joint Research Centre in its 2013 Report aims at providing a synthetic measure of the socioeconomic attractiveness of 262 European regions, mainly at NUTS 2 level. To build RCI, 73 elementary indicators[2] are first aggregated into 11 so-called "sub-pillars"; in turn, these are aggregated into 3 "pillars", whose final aggregation produces the RCI index. Each aggregation step is performed through simple weighted means (see Annoni & Dijkstra (2013) for details). A scheme of the index architecture is represented in Figure 1.

The structure of pillars in terms of subpillars is as follows:

1. Basic pillar. Subpillars: (i) Institutions, (ii) Infrastructure, (iii) Health, (iv) Macroeconomic stability, (v) Basic education.

2. Efficiency pillar. Subpillars: (i) Market Size, (ii) Labour Market Efficiency, (iii) Higher Education / Training and Lifelong Training.

3. Innovation pillar. Subpillars: (i) Innovation, (ii) Business Sophistication, (iii) Technological Readiness.

This paper is not devoted to the analysis of the RCI in itself, so we focus just on data pertaining to one pillar (the Basic pillar), in order to show the visualization

---

[2]Some elementary indicators are preliminarily transformed, but the published report does not contain the details of the procedure. So it is not possible to work out all the computational details.
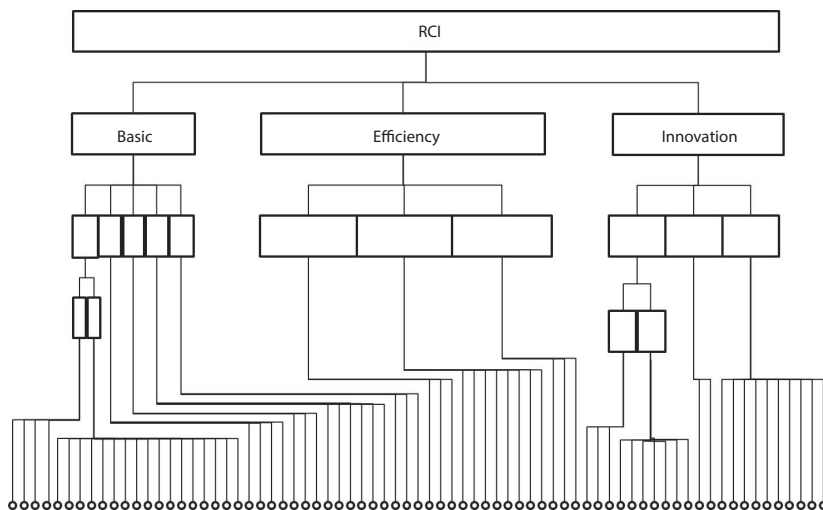
FIGURE 1: Global architecture of the RCI. Circles in the bottom represent elementary indicators. Rectangles represent aggregation of indicators.

tool in action, on a simpler example. Regional Basic pillar scores are built as averages of the corresponding 5 subpillars scores. Such 5 scores, seen as an ordered set of indicator values, constitute what we call the "profile" of the region. If one attempts to compare regions based on their profiles, a lot of "undecidable" comparisons occur, whenever a profile is higher than another on a subpillar and it is lower on another. As explained in the next Section, the set of profiles is technically a "partially ordered set". All of these undecidable comparisons disappear when the aggregated Basic pillar score is computed, but at the cost of losing much information on differences in competitiveness profiles.

3. Partial Orders and Hasse Diagrams

Partial order is a very natural concept and examples of partially ordered sets can easily be found in daily life, for example whenever one has to assign preferences to a set of objects. In the following, we provide the formal definitions of partial order and partially ordered set and introduce Hasse diagrams, the fundamental tool for visualizing them (Rival, 1984, 1985, 1989; Freese, 2004). A *partially ordered set* (or a *poset*) $P = (X, \leq)$ is a set X equipped with a *partial order* $\leq$, that is with a binary relation satisfying the properties of *reflexivity*, *antisymmetry* and *transitivity* (Davey & Priestley 2002):

1. $x \leq x$ for all $x \in X$ (reflexivity);

2. if $x \leq y$ and $y \leq x$ then $x = y$, $x, y \in X$ (antisymmetry);

3. if $x \leq y$ and $y \leq z$, then $x \leq z$, $x, y, z \in X$ (transitivity).

If $x \leq y$ or $y \leq x$, then x and y are called *comparable*, otherwise they are said *incomparable*, written $x \parallel y$. A partial order P where any two elements are comparable is called a *chain* or a *linear order*. On the contrary, if any two elements of P are incomparable, then P is called an *antichain*. Given $x, y \in P$, y is said to *cover* x (written $x \prec y$) if $x \leq y$ and there is no other element $z \in P$ such that $x \leq z \leq y$. A finite poset P (i.e. a poset defined on a finite set of elements) can be easily depicted by means of a *Hasse diagram*. Hasse diagrams are graphs drawn according to the following two rules: (i) if $x \leq y$, then node y is placed above node x; (ii) if $x \prec y$, then an edge is inserted linking node y to node x. By transitivity, $x \leq y$ in P, if and only if in the Hasse diagram there is a descending path linking the corresponding nodes. An example of a Hasse diagram is included in Figure 2. In the case of partial orders built upon a large set of numerical profiles, classical Hasse diagrams have two main drawbacks. First, in general the resulting graph is very cumbersome and hardly readable, due to the "density" of nodes and edges; secondly, any metric information is absent (as in the concept of partial order itself), since Hasse diagrams just represent comparabilities and incomparabilities among statistical units (later, we will take advantage of the flexibility in which Hasse diagrams can be drawn, to graphically introduce some metric information). Consider, for example, the Hasse diagram of 262 European regions assessed on the competitiveness covariates previously introduced (Figure 3). As can be seen, the diagram is very complicated. Moreover, Euclidean (i.e. visual) distances between nodes do not imply any similarity between regional profiles. Only comparabilities and incomparabilities are meaningful, but from the diagram one cannot assess whether these are due to large or small (and possibly statistically non-significant) differences between corresponding components of statistical unit profiles.
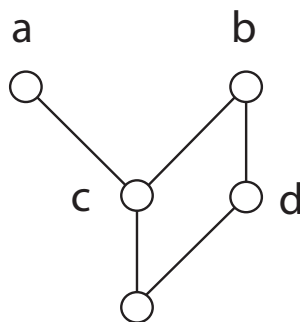

FIGURE 2: Example of a Hasse diagram.

Although the diagram of Figure 3 is very cumbersome, at the same time it reveals a large number of incomparabilities among regions (red dots). As above mentioned, these incomparabilities disappear in the aggregation leading to the Basic pillar index. It should be quite clear that a great deal of information gets lost in this unidimensional reduction (and similarly in the whole aggregative computation of the RCI). It is our opinion that information pertaining to incomparabilities, i.e. to competitiveness patterns, should be preserved and conveyed to those who

address the topic. Some complexity reduction is indeed necessary, to make the diagram of Figure 3 more readable. This is the reason why we implement a clustering analysis process, namely a Self-Organizing Map, prior to Hasse diagram visualization (Tsakovski & Simeonov, 2008, 2011).
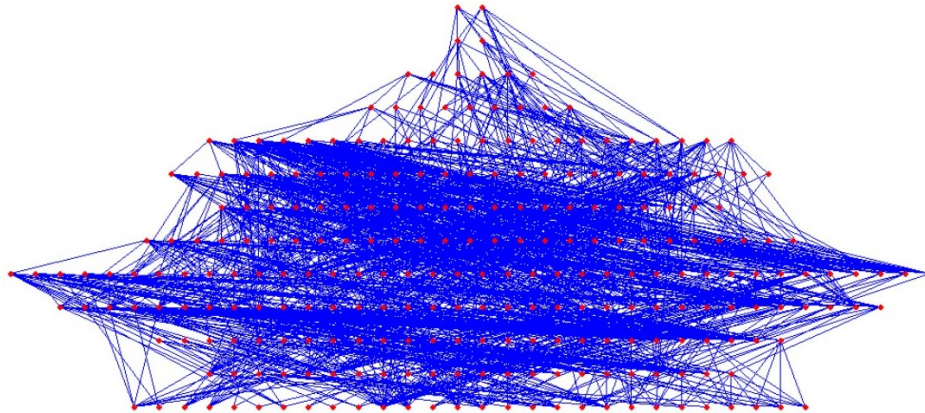

FIGURE 3: Hasse diagram for the Basic pillar data.

4. Self-Organizing Maps

As already suggested by other authors (Bruggemann & Carlsen 2014), to simplify large Hasse diagrams, the original dataset has to be preliminary processed through some cluster analysis. After clusters have been generated, a representative for each is selected and a (smaller) Hasse diagram is built on these elements. In Bruggemann & Carlsen (2014), a hierarchical cluster analysis is implemented. Here we prefer to rely on a more sophisticated tool, namely Self-organizing map (Kohonen 2001). The Self-Organizing Map (SOM) can be viewed as a non-linear projection of a multidimensional density distribution on a bidimensional grid, such that the topology of the input space is preserved as much as possible. The main advantage of SOMs over classical clustering algorithms (e.g. hierarchical cluster analysis or k-means) is that it can fit complex frequency distributions in an adaptive way. The resulting clusters are arranged on a regular euclidean grid in such a way that regions next to each other in the input space are mapped to clusters next to each other in the grid. The grid is thus a planar "image" of the input space. Notice that in the application proposed in the paper, a limited number of clusters will be produced, since the aim is to obtain an easy-to-read visualization. In this respect, SOMs are not directly used as a visualization tool, but for their effectiveness in extracting clusters, "exploring" the input density in an effective way. SOMs are implemented in many programming languages. Here we rely on R[3] package "kohonen" that provides an effective and easy-to-use tool for practical computations. As explained in the package documentation (Wehrens & Buydens 2007),

---

[3]R Core Team (2013).

and leaving aside more technical details, to apply the SOM algorithm, one must previously determine the number of clusters and define how they will be arranged in the bidimensional rectangular grid. Usually one performs several attempts, balancing between two conflicting needs, namely having a number of clusters (i) large enough to assure for their internal homogeneity, but (ii) not too large, to avoid losing interesting density patterns. All in all, setting the right grid and the right number of clusters is an empirical task. After the grid is defined, to each cluster an initial reference profile (called "codebook" in the SOM literature) is associated, randomly extracted from the dataset. Then the SOM algorithm is launched. As the algorithm proceeds, codebooks are updated until a smooth map is obtained, where final codebooks are arranged in an ordered fashion. We refer to specialized literature (Kohonen 2001) for details on the SOM algorithm and limit ourselves to some examples, so as to show what kind of outputs are provided. Consider the data pertaining to the Basic pillar of the Regional Competitiveness Index. As a first example, we cluster European regions into 9 clusters arranged in a 3×3 square grid. Clusters are depicted as circles and the corresponding codebooks are represented by the colored slices within each (this is the standard output of R package "kohonen"). The larger the radius of a slice, the higher the corresponding profile component. Statistical units are then assigned to the cluster whose codebook is most similar to their profile. Figure 4 reports the result of the computations. The left map reproduces the clusters and their codebooks; the right map associates each statistical unit (represented as a dot) to its cluster (notice that some jittering has been added, so as to avoid dot overlapping and give a visual impression of the number of units in the clusters). Similar computations have been performed increasing the number of clusters of the square grid to $5 \times 5 = 25$ and $6 \times 6 = 36$. Results are depicted in Figures 5 and 6. Some remarks are in order. In each example, clusters are arranged on the square grid in such a way that similar codebooks are placed next to each other. This is the main effect of the self-organization process implemented by the SOM. As the number of clusters increases, the SOM reproduces more nuances "selecting", in an adaptive way, which part of the input density to reproduce with more details. Notice that the map orientation has no absolute meaning and that some clusters may be empty. This is not a fault of the algorithm, but a consequence of SOM topology preserving nature. Codebooks of empty clusters may be seen as "bridges" between densely populated regions, needed to preserve the smoothness of the map.

5. The Visualization Tool

The principal aim of a visualization tool for multidimensional and partially ordered datasets is to provide a direct representation of the data structure, reducing its complexity, but retaining the essential patterns in it. Hasse diagrams, the "official" partial order graphical representation, convey a great deal of information on the partial order structure of the data, but they are not easy to read as the number of elements increases and, as noticed before, do not provide any metric information, when this is available in the original data. Cluster analysis, on the other side,
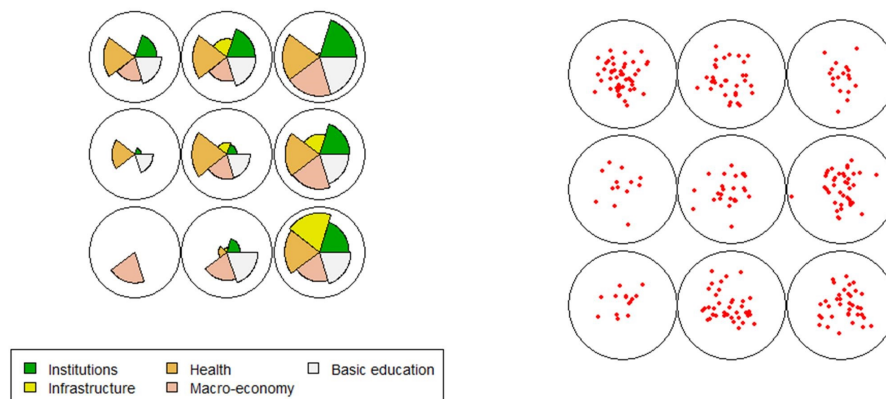
FIGURE 4: 3×3 SOM for the Basic pillar data.
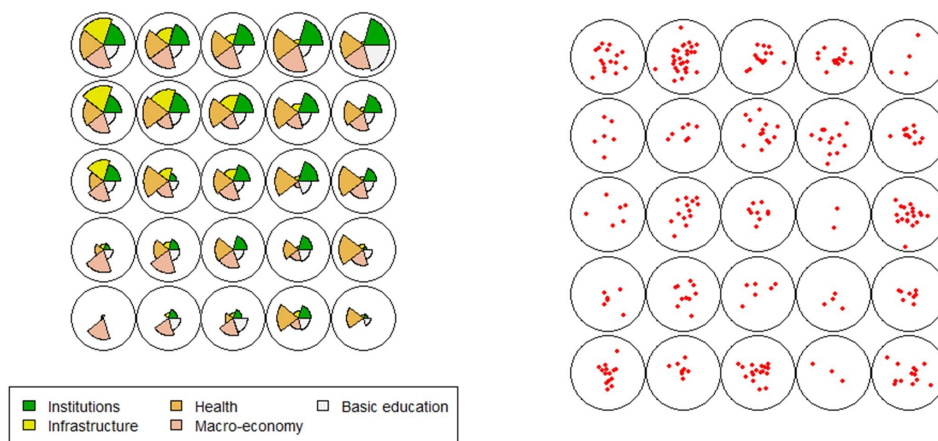
FIGURE 5: 5×5 SOM for the Basic pillar data.

reduces the complexity of the data, but it is not designed to preserve information on comparabilities and incomparabilities. Following a suggestion Bruggemann & Carlsen (2014), we combine Hasse diagrams and Cluster analysis in a complexity-reduction process, producing a visual output allowing final users to jointly grasp the partial order and the metric structure of the data. The visualization process is composed of three main steps:

1. Reducing dataset complexity through a clustering process based on a SOM.

2. Building a classical Hasse diagram on the population of clusters, that is on SOM codebooks.

3. Visually adding information on statistical units and clusters (particularly, information pertaining to the value of the covariates).
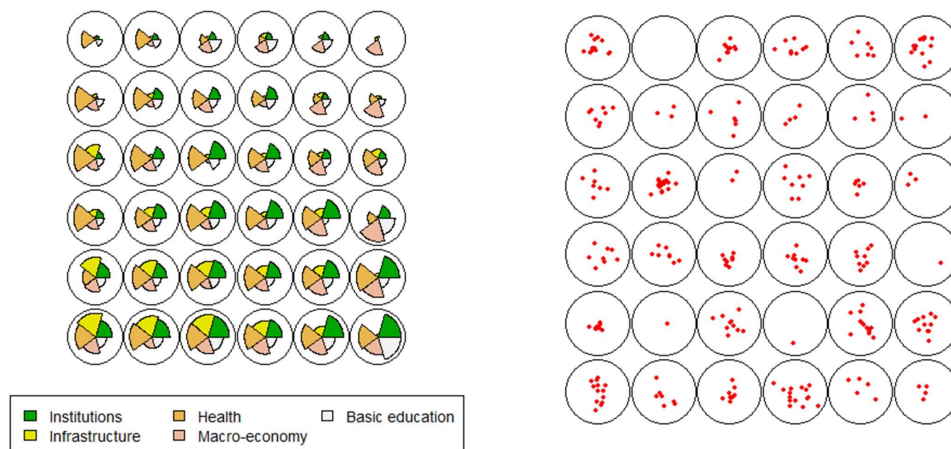
FIGURE 6: 6×6 SOM for the Basic pillar data.

We now build the visualization step-by-step, on the RCI data pertaining to the Basic pillar. To make things easier, we reduce the population of 262 European regions to 9 clusters. Running the SOM algorithm identifies 9 codebooks, represented by the colored slices in the circles of Figure 4. As a second step, we draw the corresponding Hasse diagram on the codebooks, keeping the same color codes (see Figure 7). Considering this image and Figure 3, we can see that clusters are partially ordered and that there are many incomparabilities, i.e. essentially different competitiveness patterns.
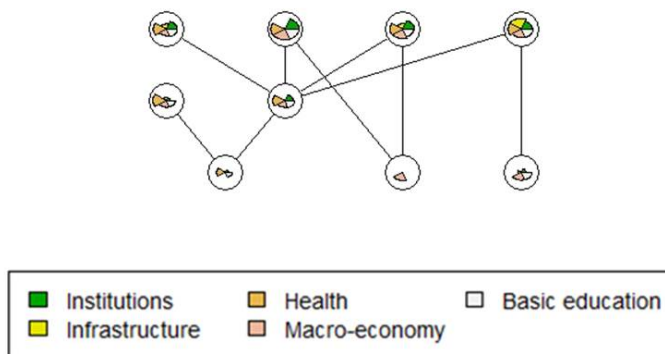

FIGURE 7: Hasse diagram for SOM codebooks (3×3, Basic pillar data).

The diagram of Figure 7 does not provide any metric information yet and we cannot easily guess the values of different competitiveness scores, for each cluster. As a third step, we thus move clusters vertically, based on the values of different profile components. Figure 8, depicts the Hasse diagram of Figure 7 where y coordinates correspond to the Institutions score for each codebook. Here we see that clusters have different levels along the Institutions subpillar. Now the diagram
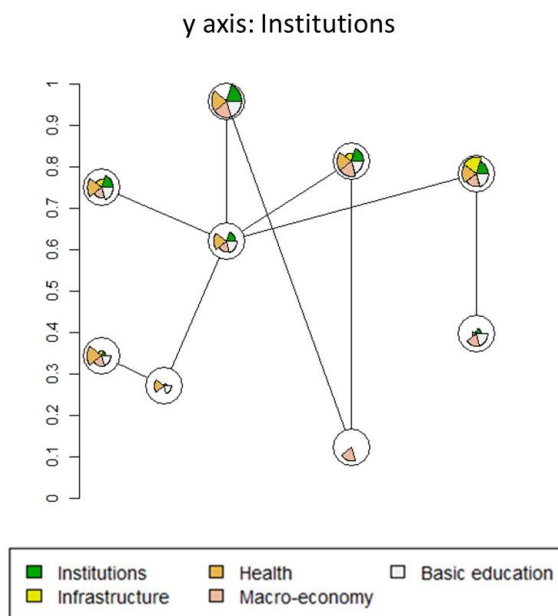
y axis: Institutions

FIGURE 8: Hasse diagram for SOM codebooks (3×3, Basic pillar data). y axis: Institutions.

has a metric meaning, at least along the vertical axis. Notice that moving clusters vertically as done in Figure 8 does not affect the global partial order relation: if a cluster is "greater than" another in the original Hasse diagram, then it is also "greater than" another in the modified one. The same kind of diagram can be obtained for other components. Sometimes, clusters overlap, as in the case of the Health subpillar (Figure 9). Although not pretty from a visual standpoint, this indeed conveys some information, i.e. that some clusters may be very similar with respect to a component of the competitiveness profile. In our computations, all of the five components of the Basic pillar are scaled to $0 - 1$ (simply subtracting the minimim and dividing by the range), so that y distances in the Hasse diagrams are comparable and one can get an impression of the differences among score distributions of the five subpillars, at cluster level. This is made easier if one arranges all of the diagrams side by side, as in Figure 10, where the vertical axis of the first Hasse diagram reports the profile mean of each cluster. Visual comparison reveals many features of the data. For example, one sees that the Hasse diagram reporting the mean value of the cluster profiles is very similar to the diagram reporting the Institutions value. This is somewhat interesting, since Institutions data are collected at national and not at regional level (i.e. each region in the same Country shares the same Institutions score). So it seems that the mean profile value has the same behavior of a national feature (at least at cluster level) and that the metric structure associated to Hasse diagrams is the same for both the mean and the Institutions scores. It may also be observed that the Hasse diagram relative to Infrastructure has a "metric shape" very different from the others, with

great variations in score levels. It is also interesting to look at the different vertical positions of clusters in different diagrams. These reveal the existence of conflicting indicators, explain the existence of incomparabilities and help in assessing whether due to big or small differences in score values. These are some suggestion arising from this kind of visualization and that may deserve further scrutiny, through more technical, and less intuitive, statistical procedures and data analysis. As usual, visualizations give the hint and suggest interesting directions to investigate.
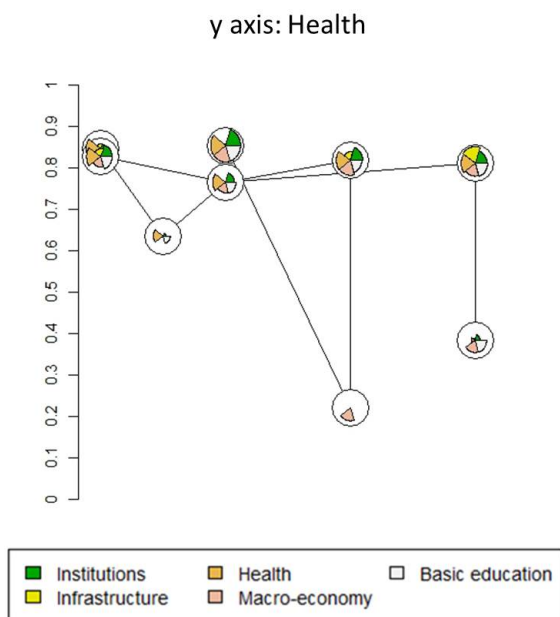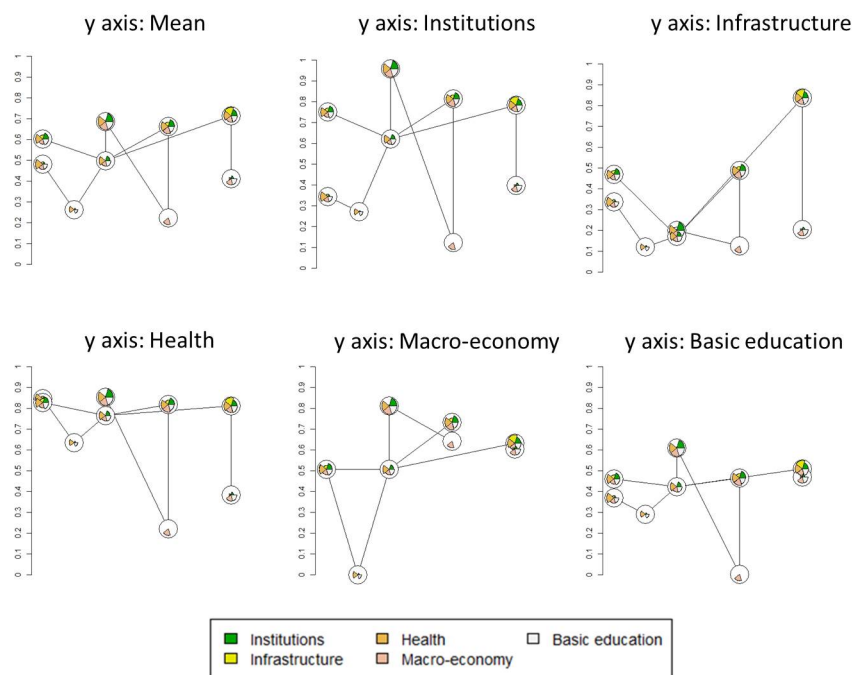

FIGURE 9: Hasse diagram for SOM codebooks (3×3, Basic pillar data). y axis: Health.

***Note* 1.** Other graphical features could be used to add additional information to the modified Hasse diagrams. One could link the dimension of circles to other covariates or use the background color to plot the value of other profile components or even plot dot clouds in the circle areas (as in Figure 4) to give an idea of population distribution. Benchmarks (e.g. the population mean value of the subpillar measured on the y axis) may be graphically added to the diagrams, to see whether a cluster is below or above it. Alternatively, the values of two covariates could be jointly considered, moving nodes both vertically and horizontally, to produce a bivariate "metric" Hasse diagram (i.e. combining a Hasse diagram and a scatterplot). All of these options can be easily implemented with many software languages, also adding interactions to ease user experience. Here we limit ourselves to identifying basic visualization structures that may be improved using classical infovis tools.

FIGURE 10: Hasse diagrams for SOM codebooks (3×3, Basic pillar data).

6. Conclusion

In this paper, we have proposed a simple way to visualize partially ordered datasets. Partial orders arise typically when multicriteria evaluation problems are addressed. They constitute an alternative to classical aggregative compensative procedures, that solve multidimensional evaluation problems computing unidimensional rankings, usually through some composite indicator. Admittedly, the final output is more complex than a simple ranking, but at the same time it is much more informative, reflecting the true nature of the data and helping final users to realize the existence of complex patterns in the data. The procedure integrates the Self-organizing map with Hasse diagrams and simple graphic design. It is planned to develop an R package to make the visualization tool freely available, adding also some interactive functions. The proposed way to combine partial orders and metric information is indeed quite simple. More sophisticated approaches could be explored. In particular, it would be very interesting to try to integrate partial order structures within the SOM algorithm, so as to get the final Hasse diagram through an adaptive process. The application of partial order theory to socio-economic evaluation problems is still at an early stage, although some methodologies have been already proposed, mainly in connection with multidimensional poverty evaluation (Fattore, Bruggemann & Owsiński 2011, Fattore, Maggino & Greselin 2011, Fattore et al. 2012). An R package, named PARSEC (PARtial order in Socio-EConomics Fattore & Arcagni, 2014), is also being re-

leased to the scientific community. The proposed visualization enriches the set of tools available to researchers, and we hope this will promote the use of partial orders in socio-economic studies.

References

Al-Sharrah, G. (2014), Ranking Hazarduous chemicals with a heuristic approach to reduce isolated objectds in Hasse Diagrams, *in* R. Bruggemann, L. R. Carlsen & J. Wittmann, eds, 'Multi-indicator Systems and Modelling in Partial Order', Springer-Verlag, New York, pp. 217–235.

Annoni, P. & Dijkstra, L. (2013), *EU Regional Competitiveness Index - RCI 2013*, Joint Research Centre of the European Commission, Luxembourg.

Barthélemy, J., Flament, C. & Monjardet, B. (1982), 'Ordered sets and social sciences', *Ordered Sets* **83,**, 721–758.

Bruggemann, R. & Carlsen, L. (2014), 'Incomparable - what now? Absorption of incomparabilities by cluster and fuzzy-methods', *Quality and Quantity* . forthcoming.

Davey, B. A. & Priestley, H. A. (2002), *Introduction to Lattices and Order*, Cambridge University Press.

Dijkstra, L., Annoni, P. & Kozovska, K. (2011), A new European Regional Competitiveness Index: Theory, Methods and Findings, Technical Report 02/2011, European Union Regional Policy Working Papers. (RCI 2010).

Fattore, M. & Arcagni, A. (2014), PARSEC: an R package for poset-based evaluation of multidimensional poverty, *in* Bruggemann, R. and Carlsen, L. R. and Wittmann, J., ed., 'Multi-indicator Systems and Modelling in Partial Order', Springer-Verlag, New York, pp. 317–330.

Fattore, M., Bruggemann, R. & Owsiński, J. (2011), Using poset theory to compare fuzzy multidimensional material deprivation across regions, *in* Ingrassia, S. and Rocci, R. and Vichi, M., ed., 'New Perspectives in Statistical Modeling and Data Analysis', Springer-Verlag.

Fattore, M., Maggino, F. & Colombo, E. (2012), From composite indicators to partial order: Evaluating socio-economic phenomena through ordinal data, *in* Maggino, F. and Nuvolati, G. , ed., 'Quality of Life in Italy: Research and Reflections, Social Indicators Research Series 48', Springer.

Fattore, M., Maggino, F. & Greselin, F. (2011), 'Socio-economic evaluation with ordinal variables: Integrating counting and poset approaches', *Statistica & Applicazioni* pp. 31–42. Special Issue.

Freese, R. (2004), Automated lattice drawing, *in* P. Eklund, ed., 'Concept Lattices', Springer-Verlag, Berlin, pp. 112–127.

Kohonen, T. (2001), *Self-Organizing Maps*, Springer.

R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
*http://www.R-project.org/

Rival, I. (1984), *Graphs and Order. The Role of Graphs in the Theory of Ordered Sets and its Applications*, Vol. 147 of *Nato Science Series C*, Springer, Dordrecht.

Rival, I. (1985), The diagram, *in* 'Graphs and Order', D.Reidel Publishing Company, pp. 103–133.

Rival, I. (1989), 'Graphical data structures for ordered sets', *Algorithms and Order* **255**, 3–31.

Sen, A. (1992), *Inequality Reexamined*, Oxford University Press.

Tsakovski, S. & Simeonov, V. (2008), Hasse Diagrams as Explanatory Tool in Environmental Data Mining: A Case Study, *in* J. Owsinski & R. Bruggemann, eds, 'Multicriteria Ordering and Ranking: Partial Orders, Ambiguities and Applied Issues', Systems Research Institute Polish Academy of Sciences, Warsaw, pp. 50–68.

Tsakovski, S. & Simeonov, V. (2011), 'Hasse diagram technique as exploratory tool in sediment pollution assessment', *Journal of Chemometrics* **25**(5), 254–261. doi: 10.1002/cem.1381:1-8.

Wehrens, R. & Buydens, L. M. C. (2007), 'Self- and super-organizing maps in R: The Kohonen Package', *Journal of Statistical Software* **21**(5).