12th International Conference on Partial Orders in Applied Sciences

Towards an Understanding of Complex Phenomenon: Applying Partial Order Theory to Multi-Indicator Systems

Measuring gender equality: a comparison of different techniques to build synthetic indicators

Enrico di Bella - University of Genoa, Italy

Filomena Maggino – University of Rome «La Sapienza», Italy

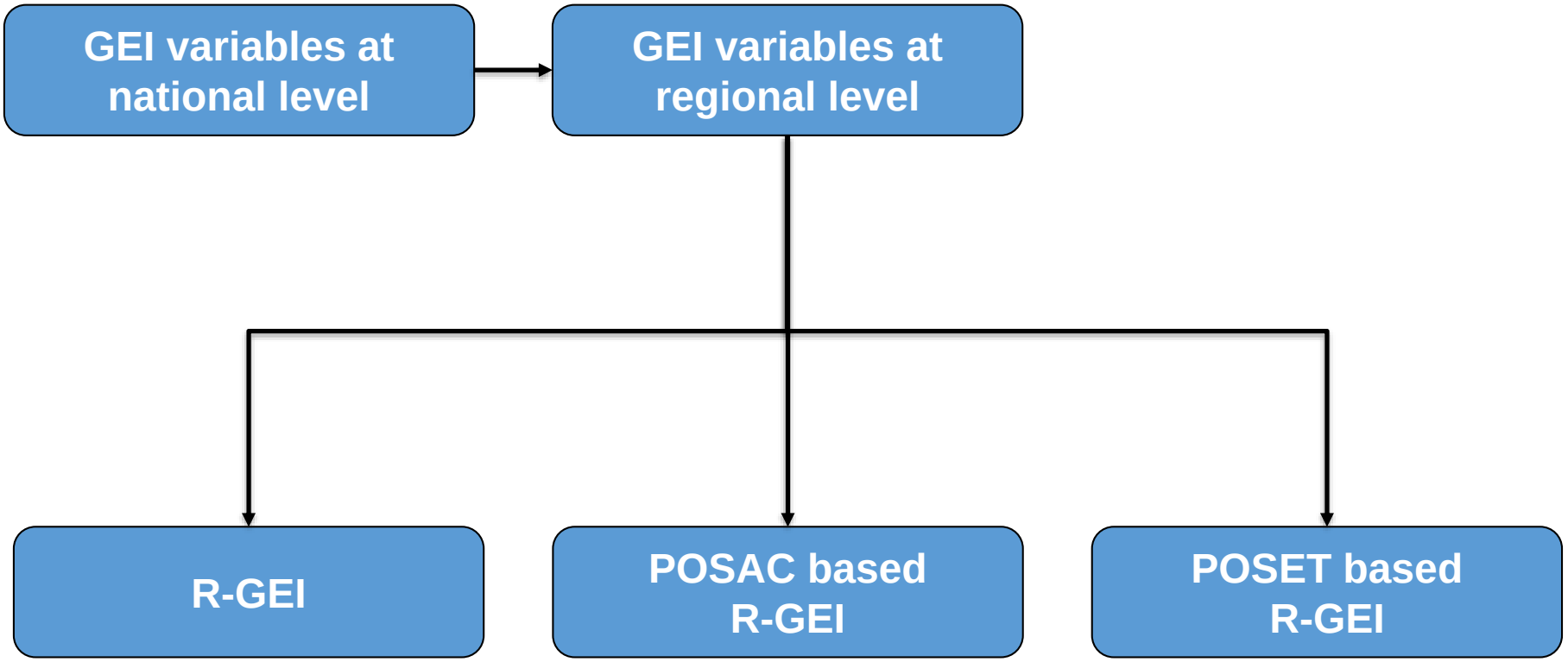
Lucia Leporatti – University of Genoa, Italy

UNIVERSITÀ DEGLI STUDI
DI GENOVA

The starting point

UNIVERSITÀ DEGLI STUDI
DI GENOVA

The different approaches



UNIVERSITÀ DEGLI STUDI
DI GENOVA

The EIGE Gender Equality Index (GEI)

6 domains

14 sub-domains

31 variables

EIGE
European Institute for Gender Equality

UNIVERSITÀ DEGLI STUDI
DI GENOVA

Gender Equality Index

Box 1: Calculating the Gender Equality Index in brief

1. **Selection and processing of indicators.** The Gender Equality Index is composed of 31 indicators, divided between 14 sub-domains, which make up the six domains (work, money, knowledge, time, power and health).

2. **Calcutating gender gaps.** A single measure of gender equality for the indicators is developed. Gender gaps are calculated and transformed so that the value of 1 can be interpreted as full achievement of gender equality, while any value below 1 indicates some degree of gender inequality in a given indicator. The value of 0 theoretically refers to full inequality.

3. **Calculating the correcting coefficient.** Correcting coefficients are calculated and applied to each gender gap. This means that Member States with similar gender gaps are treated differently if their levels of achievement differ. The higher the level of achievement, the lower the correction of the gender gap.

4. **Calculating the gender gap metric.** The final metric for each indicator is a combination of the gender gap and the correcting coefficient. It is dimensionless (allowing comparability since measurement units of variables have been eliminated), and bound between [1; 100].

5. **Calculating the Index (aggregating, weighting, and normalisation)**

   1. Aggregation of variables of each sub-domain, creating indices at the subdomain level (value bound [1; 100]), and using arithmetic mean of the metrics of the indicators.

   2. Aggregation of the sub-domains into domains, using geometric means of the scores of sub-domains (value bound [1; 100]).

   3. Aggregating the scores of the domains into overall Gender Equality Index, using geometric means of the six scores of the domain, by applying experts' weights to the domains, obtained through the analytic hierarchy process (AHP). The Gender Equality Index takes a value on a scale of 1 to 100, where value of 100 stands for complete gender equality, and 1 for full gender inequality.

$$1 - \Upsilon_{(x_{it})} = 1 - \left| \frac{\bar{X}_{it}^{W}}{\tilde{X}_{it}^{a}} - 1 \right|$$

$$\alpha_{(x_{it})} = \left(\frac{\tilde{X}_{it}^{T}}{\max(\tilde{X}_{it}^{T})} \right)^{1/2}$$

$$\Gamma_{(x_{it})} = 1 + \left[\alpha_{(x_{it})} * \left(1 - \Upsilon_{(x_{it})} \right) \right] * 99$$

$$I_i^t = \prod_{d=1}^{6} \left\{ \prod_{s=1}^{nsd} \left[\sum_{v=1}^{n_s} \frac{\Gamma_{(x_{it})}}{n_s} \right]^{\frac{1}{nsd}} \right\}^{w_{AHP_d}}$$

Source: EIGE (2017), Methodological Report

UNIVERSITÀ DEGLI STUDI
DI GENOVA

From GEI to R-GEI

1. Assessment of the original GEI variables in terms of data sources and relevance at a regional level;
2. Computation of single indicators at a NUTS2 level using GEI survey microdata (when representative), alternative surveys (when original surveys are not representative) or other official database;
3. Substitution of meaningless variables with others more consistent with the regional perspective;
4. Use of the GEI methodology (third edition) to build the R-GEI.

	Description	GEI	R-GEI
Survey	Labour Force Survey - LFS (Eurostat)	✓	✓
	European Working Conditions Surveys – EWCS (Eurofound)	✓	✗
	European Union Statistics on Income and Living Conditions - EU-SILC (Eurostat)	✓	✓
	Aspects of Daily Life – ADL (Istat)	✗	✓
	European Health Interview Survey – EHIS (Eurostat)	✓	✗
Official Database	Eurostat Education Statistics	✓	✗
	EIGE Gender Statistics database	✓	✗
	Eurostat Mortality data	✓	✗
	Ministero dell'interno	✗	✓
	INPS	✗	✓

UNIVERSITÀ DEGLI STUDI
DI GENOVA

From GEI to R-GEI

DOMAIN	NUMBER OF VARIABLES	
	GEI	R-GEI
WORK	5	3
MONEY	4	4
KNOWLEDGE	3	3
TIME	4	3
POWER	8	7
HEALTH	7	5
TOTAL	**31**	**25**

- 10 out of the 31 original variables are exactly based on GEI definition and data,
- 15 are based on a definition «as close as possible» to those chosen by EIGE but using data representative at a regional level,
- 6 variables could not be properly substituted.

UNIVERSITÀ DEGLI STUDI
DI GENOVA

From GEI to R-GEI

Domain	Sub-domain	Description of GEI variable	GEI Source	Description of R-GEI variable	R-GEI Source
WORK	participation	Full-time equivalent employment rate (% 15 + population)	LFS	Same as GEI	Same as GEI
		Duration of working life (years, 15+ population)	LFS	Not available at regional level	
	segregation and quality of work	Employed people in education, human health, and social work activities (%, 15 + employed)	LFS	Same as GEI	Same as GEI
		Ability to take an hour or two off during working hours to take care of personal or family matter (%, 15+workers)	EWCS	Not available at regional level	
		Career Prospects Index (points, 0-100)	EWCS	Subjective measure between 0-10 based on question "In the current work, how satisfied are you with past and future career opportunities?"	LFS
MONEY	financial resources	Mean monthly earnings (PPS, working population)	Eurostat SES	Same as GEI	EU-SILC
		Mean equivalised net income (PPS, 16+ population)	EU SILC	Same as GEI	Same as GEI
	economic situation	Not-at-risk-of-poverty, >= 60% of median income (%, 16 + population)	EU SILC	Same as GEI	Same as GEI
		Income distribution S20/S80 (16+ population, %)	EU SILC	Same as GEI	Same as GEI
KNOWLEDGE	attainment and participation	Graduates of tertiary education (%, 15+ population)	LFS	Same as GEI	Same as GEI
		People participating in formal or non-formal education and training (%, 15+ population)	LFS	Same as GEI	Same as GEI
	segregation	Tertiary students in the fields of education, health and welfare, humanities and art (tertiary students) (% 15+ population)	Education statistics	Same as GEI	Istat database
TIME	care activities	People caring for and educating their children or grandchildren, elderly or people with disabilities, every day (%, 18+ population)	EWCS	Mean of the number of minutes devoted to house work and family keeping (18 + population)	ADL
		People doing cooking and/or housework, every day (%, 18+ population)	EWCS	Not available at regional level	
	social activities	Workers doing sporting, cultural or leisure activities outside of their home, at least daily or several times a week (%, 15 + workers)	EWCS	People who don't smoke and are not involved in harmful drinking (i.e. not drinking often between meals) (%, 16+ population)	EU-SILC
		Workers involved in voluntary or charitable activities, at least once a month (%, 15+ workers)	EWCS	People doing physical activities and/or consuming fruits and vegetables more than once a day (%, 16+ population)	ADL

Domain	Sub-domain	Description of GEI	GEI Source	Description of R-GEI	R-GEI Source
POWER	political	Share of ministers (% W, M)	EIGE database	Share of city major and municipality assessors (<15.000 population) (% W, M)	Italian Ministry of Interior
				Share of city major and municipality assessors (>15.000 population) (% W, M)	Italian Ministry of Interior
		Share of members of parliament (% W, M)	EIGE database	Member of the regional assemblies / municipal assembly (< 15.000 population) (% W, M)	Italian Ministry of Interior
				Member of the regional assemblies / municipal assembly (> 15.000 population) (% W, M)	Italian Ministry of Interior
		Share of regional assemblies (% W, M)	EIGE database	Share of presidents of regional board and regional assessors (% W, M)	Italian Ministry of Interior
	economic	Share of boards in largest quoted companies, supervisory board or board of directors (% W, M)	EIGE database	Share of women working in managerial positions	INPS
		Share of board members of Central Banks (% W, M)	EIGE database	Share of managerial positions covered by men and women (% W, M)	INPS
	social	Share of board members of research funding organizations (% W, M)	EIGE database	Not available at regional level	
		Share of board members in publically owned broadcasting organizations (% W, M)	EIGE database	Not available at regional level	
		Share of board members of highest decision-making body of the national Olympic sport organizations (% W, M)	EIGE database	Not available at regional level	
HEALTH	status	Self-perceived health, good or very good (%, 16+ population)	EU SILC	Same as GEI	Same as GEI
		Life expectancy in absolute value at birth (years)	EUROSTAT Mortality Tables	Same as GEI	ISTAT database
		Healthy life years in absolute value at birth (years)	EU SILC	Not available at regional level	
	behaviour	People who don't smoke and are not involved in harmful drinking (%, 16+ population)	EHIS	People who don't smoke and are not involved in harmful drinking (%, 16+ population)	ADL
		People doing physical activities and/or consuming fruits and vegetables (%, 16+ population)	EHIS	People doing physical activities and/or consuming fruits and vegetables (%, 16+ population)	ADL
	access	Population without unmet needs for medical examination (%,16+ population)	EU SILC	Same as GEI	Same as GEI
		People without unmet needs for dental examination (%, 16+ population)	EU SILC	Same as GEI	Same as GEI

UNIVERSITÀ DEGLI STUDI
DI GENOVA

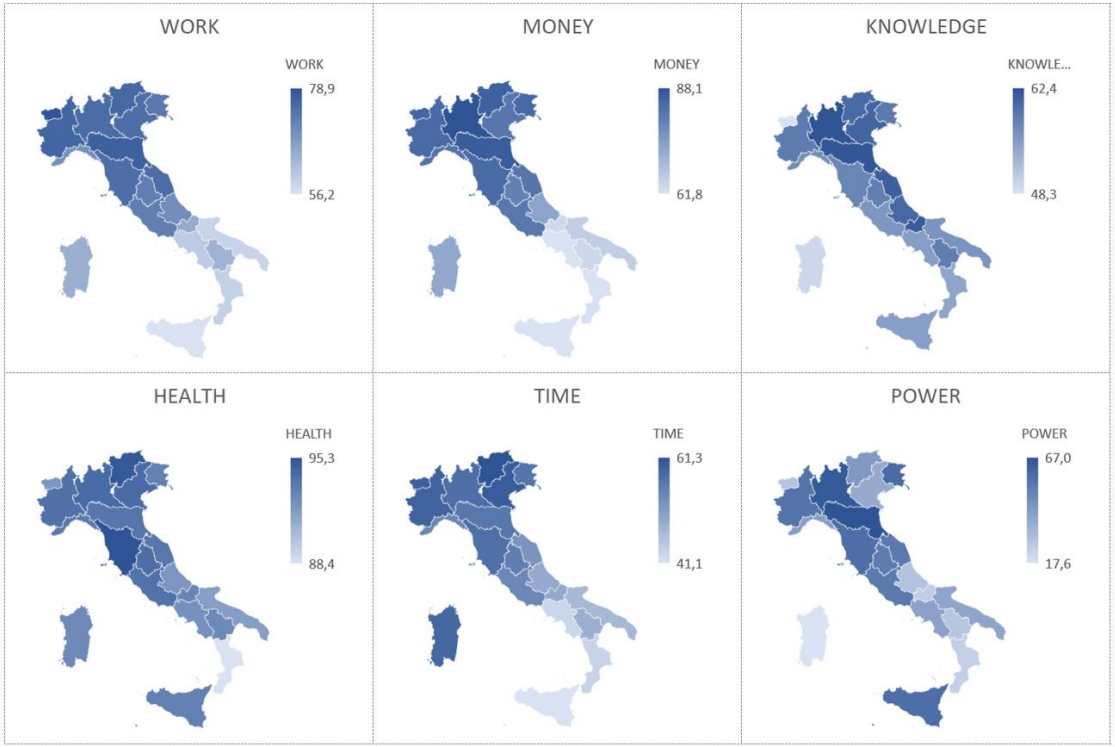
R-GEI - RESULTS

REGION	WORK	MONEY	KNOWLEDGE	HEALTH	TIME	POWER
LOM	52.24	89.07	80.84	77.87	53.53	60.37
EMR	54.08	87.36	75.46	77.45	53.73	60.63
TOS	53.03	86.71	75.13	78.64	55.76	57.75
PIE	51.10	85.14	75.98	77.70	57.61	55.80
FVG	49.22	86.55	76.64	78.06	58.25	49.26
TAA	50.64	81.76	79.45	79.18	58.73	43.21
LAZ	51.57	81.83	72.37	76.21	49.14	55.16
VEN	50.30	83.41	66.28	77.01	56.33	51.52
LIG	45.95	84.06	72.33	76.07	53.70	52.01
MAR	49.10	82.78	74.70	77.32	46.98	50.35
UMB	47.96	78.57	74.42	76.69	49.69	50.66
SAR	39.07	72.95	68.67	76.73	58.53	50.19
VAO	51.88	87.64	47.01	78.04	59.85	49.34
ABR	47.40	75.27	69.51	77.09	45.25	38.68
MOL	42.61	69.97	78.96	75.19	41.08	42.03
PUG	38.63	65.50	67.92	74.50	43.65	40.29
BAS	45.03	61.34	67.62	74.69	37.79	36.82
CAL	37.20	57.63	69.24	72.26	39.54	40.67
CAM	37.50	59.78	67.09	72.68	31.18	41.07
SIC	33.99	58.45	62.52	75.99	32.70	46.61
ITALY	**47.18**	**79.99**	**73.43**	**76.76**	**50.69**	**53.59**

REGION	R-GEI
LOM	66.88
EMR	66.18
TOS	65.65
PIE	64.99
FVG	63.43
TAA	62.33
LAZ	62.26
VEN	61.48
LIG	61.27
MAR	60.83
UMB	60.56
SAR	57.90
VAO	57.89
ABR	55.44
MOL	55.20
PUG	51.90
BAS	50.87
CAL	49.96
CAM	48.33
SIC	48.23
ITALY	**61.20**

UNIVERSITÀ DEGLI STUDI
DI GENOVA

R-GEI - RESULTS



The compensation problem

The construction of a composite indicator is generally developed through a series of subsequent steps (OECD 2008):

1) definition of the phenomenon to be measured;
2) selection of indicators;
3) normalization of individual indicators;
4) **weighting and aggregation of single indicators.**

Although all the above-mentioned phases require a series of subjective decisions by the researcher the aggregation step is the one responsible for the compensation issue.

The compensation problem

Among the various issues that may arise when dealing with multi-indicator systems, one particularly relevant is the compensation problem.

Compensation: When a synthetic indicator is, for instance, a weighted sum of the elementary indicators, compensation means that a good value of such an indicator may be the results of a very good value for some indicators which masks potentially critical values for other indicators.

The compensation problem

Different approaches have been proposed to mitigate the compensability drawback:

1. Geometric aggregation (e.g. HDI, UNDP 2010)
2. Mazziotta-Pareto Index (MPI; Mazziotta and Pareto 2011)
3. Mean-Min function (Mazziotta and Pareto 2015)
4. Multi-criteria Decision Analysis (MCDA; Munda and Nardo 2005)
5. Poset theory (Brüggemann and Patil 2011)
   - POSAC: Partial Order Scalogram Analysis with base Coordinates (Shye, 1985)

The POSAC procedure

Given a scalogram $\mathbf{P}_{n\times m}$, the POSAC procedure (Shye, 1985) produces a representation of partial ordering of profiles by two coordinates “as good as possible”.

The POSAC technique uses an iterative procedure to assign two scores to each profile p_i (i = 1, …, n), X and Y (called base coordinates), so that the location of the points in the space reflects their partial ordering respect to the indicators mapping $p_i \rightarrow (x_i\ ,\ y_i)$ such that:

$$p_v > p_w \leftrightarrow x_v \geq x_w \text{ and } y_v \geq y_w \tag{1}$$

$$p_v || p_w \leftrightarrow \begin{cases} x_v \geq x_w \text{ and } y_v \leq y_w \\ or \\ x_v \leq x_w \text{ and } y_v \geq y_w \end{cases} \tag{2}$$

The POSAC procedure

The POSAC algorithm starts by computing the matrix of weak monotonicity coefficients among all the indicators and it identifies the two indicators that are the least positively correlated. Then, the initial (x_i , y_i) coordinates of *i*-th profile p_i result from the following conditions:

$$x_i + y_i = \sum_{j=1}^{m} \boldsymbol{p}_{ij} \tag{3}$$

$$x_i - y_i = \boldsymbol{p}_{ia} - \boldsymbol{p}_{ib} \tag{4}$$

being $p_{ia}$and p_{ib} the scores of the two aforesaid least positively correlated indicators for the *i*-th unit.

UNIVERSITÀ DEGLI STUDI
DI GENOVA

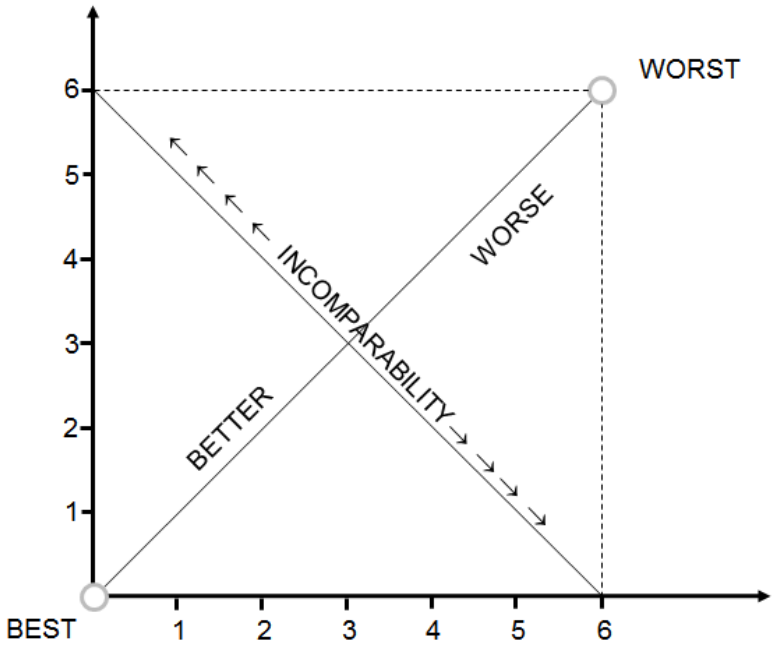
The POSAC procedure

All the values computed using conditions (3) and (4) are transformed to place all profiles within the unit square.

This initial approximation is improved minimizing a loss function defined on conditions (1) and (2).

A steepest-descent process is carried out in the *XY* coordinates until it is not possible to improve the solution.

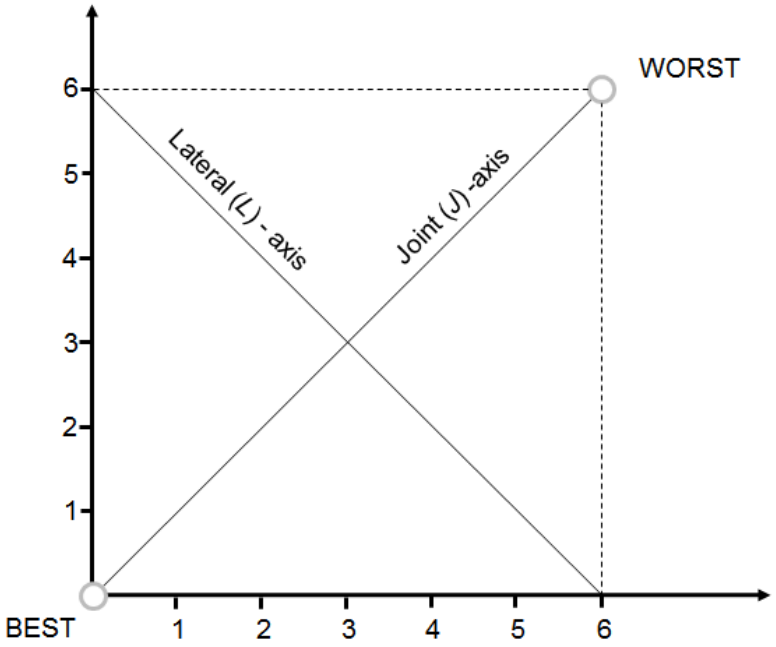
The final result is a Cartesian space the top right corner of which represents the best theoretical profile and the bottom left corner represents the worst theoretical profile



The POSAC procedure

The line joining the best and the worst theoretical profiles (called Joint axis, or *J*-axis: $J = X + Y$) is the main dimension of the resulting two-dimensional space and its interpretation is straightforward: as we move along it, growing values of the coordinates indicate strict improvement in all rankings at the same time.

The line joining the two remaining corners is called Lateral axis or *L*-axis ($L = X - Y$) and it represents the incomparability element of the profiles (similar to “horizontal variability” of the MPI).

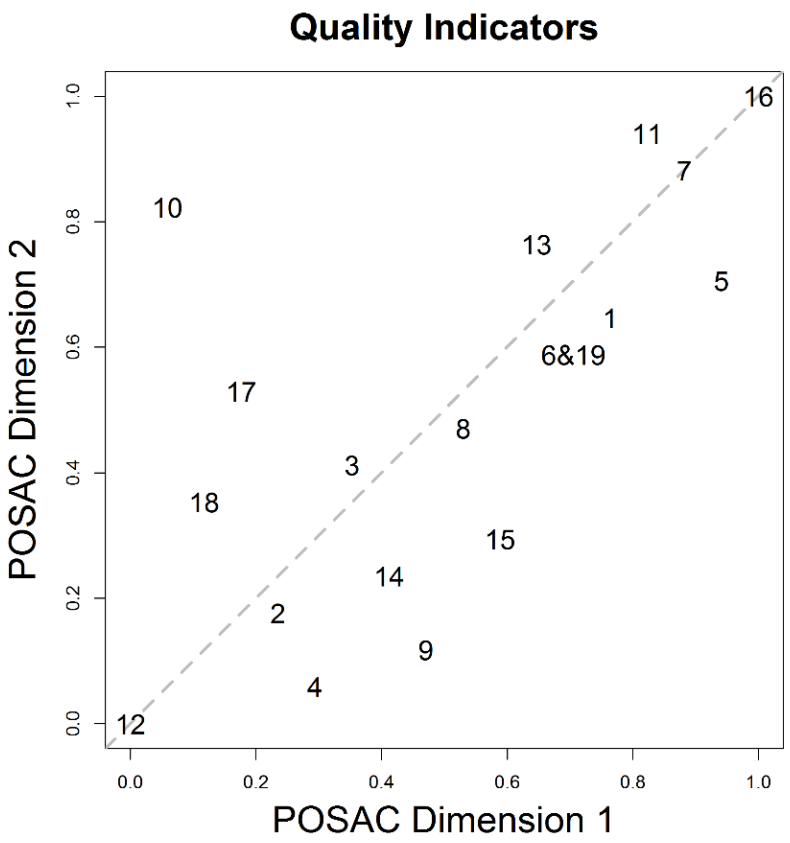


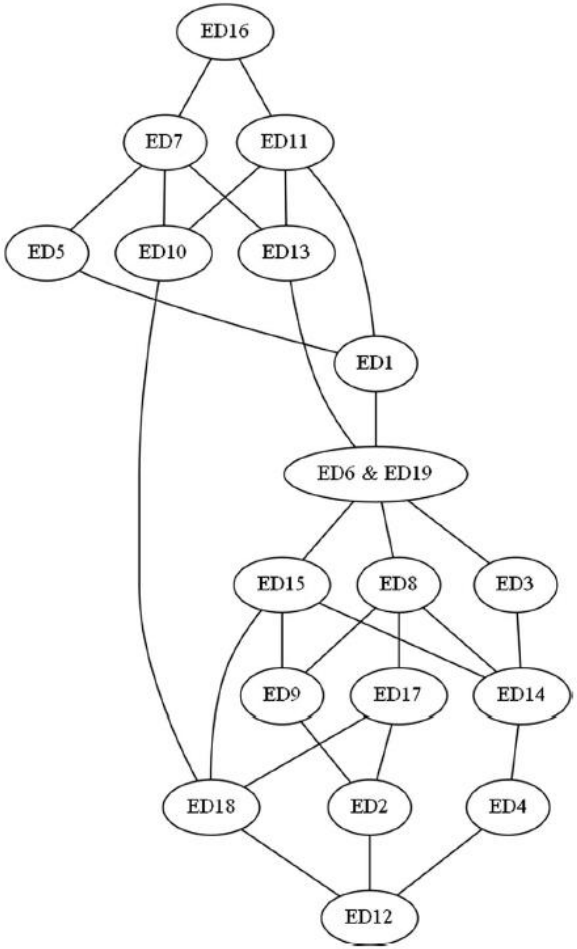
The POSAC procedure

The main «objects» that should be taken into account in a POSAC analysis are:

- The POSAC plot
- The POSAC gof measure (measured as the proportion of order relations out of all profile pairs, correctly represented by their twofold coordinates; see “stress measure” in MDS)
- The correlation matrix of indicators
- The correlation table of indicators with *J*-axis
- The correlation table of indicators with *L*-axis
- …

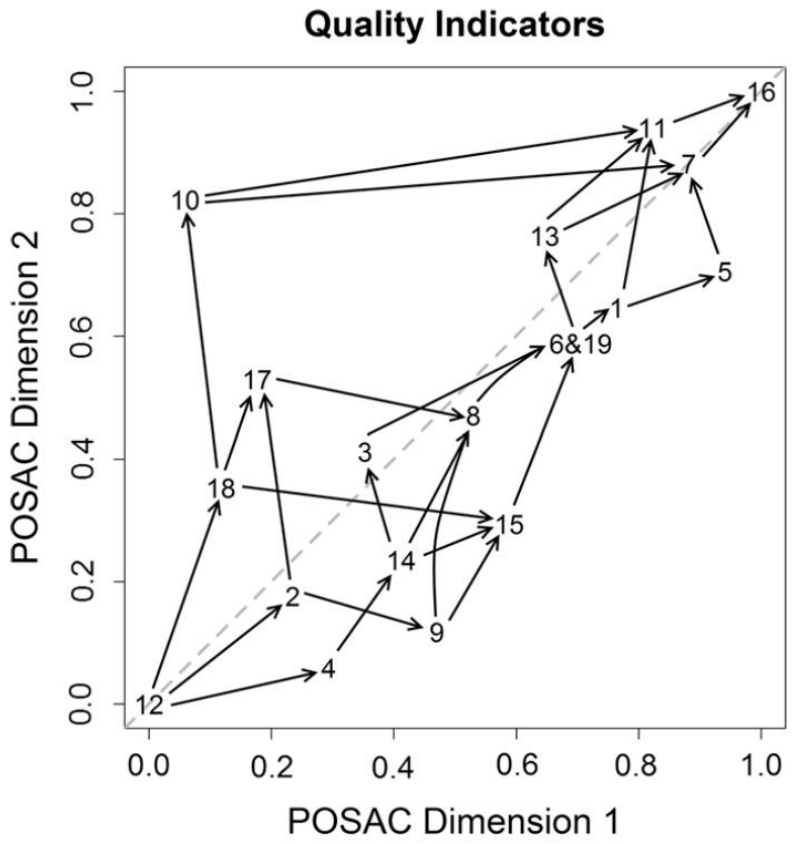
POSAC and Hasse diagrams

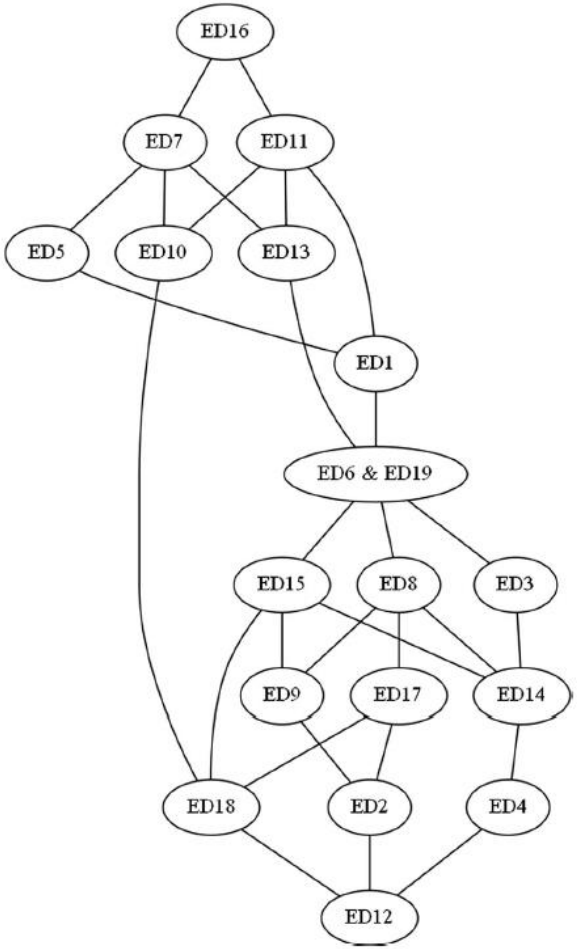




UNIVERSITÀ DEGLI STUDI
DI GENOVA

POSAC and Hasse diagrams





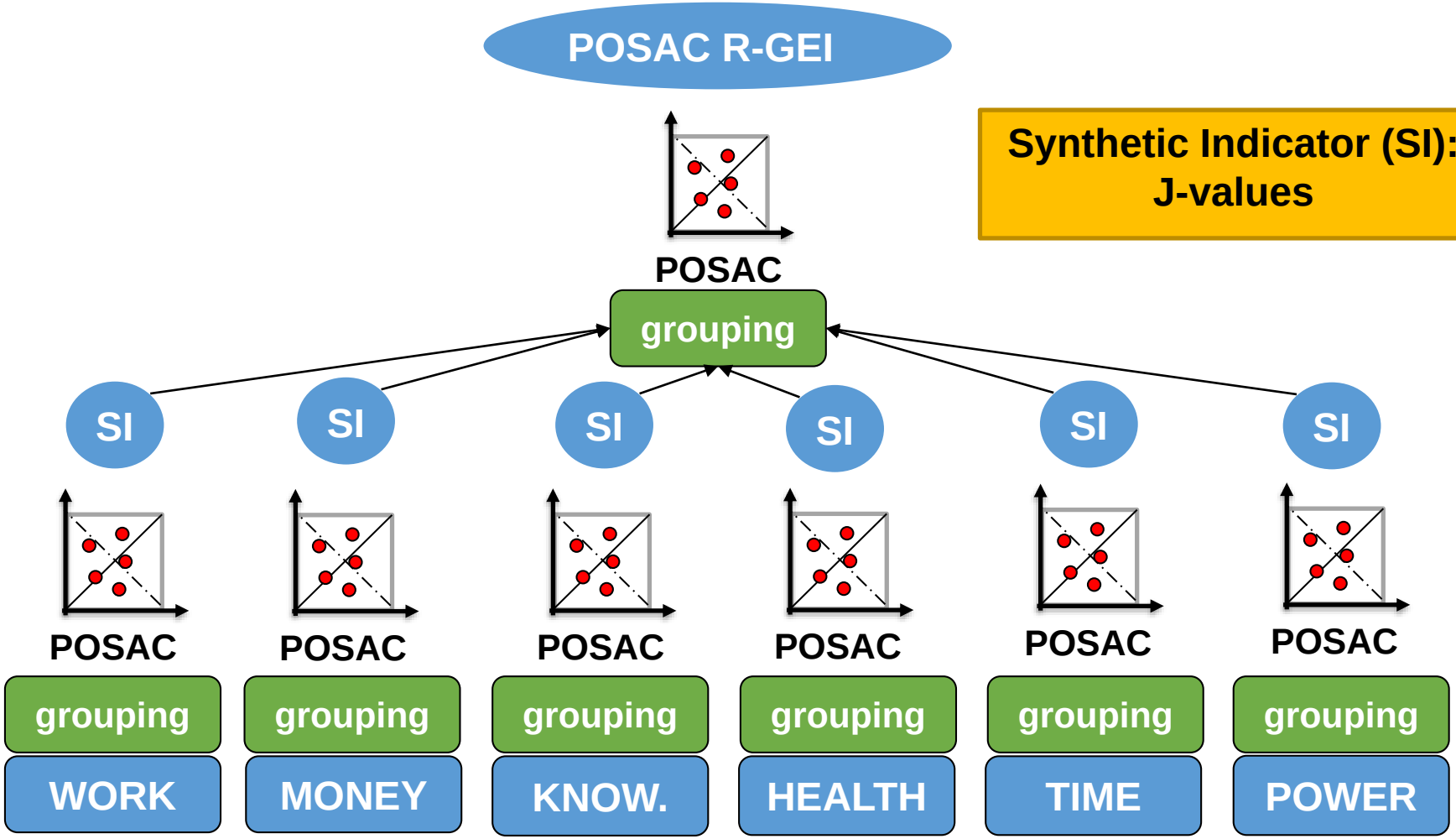
Incomparable: what now?

Non random incomparability is a really interesting condition to compare (!) observations:

1. incomparable situations (opposite to the *J*-axis) may suggest different policy actions;
2. single deviations from the J-axis may identify compensation situations;
3. correlations of indicators to the *L*-axis may help to identify the causes of different situations.

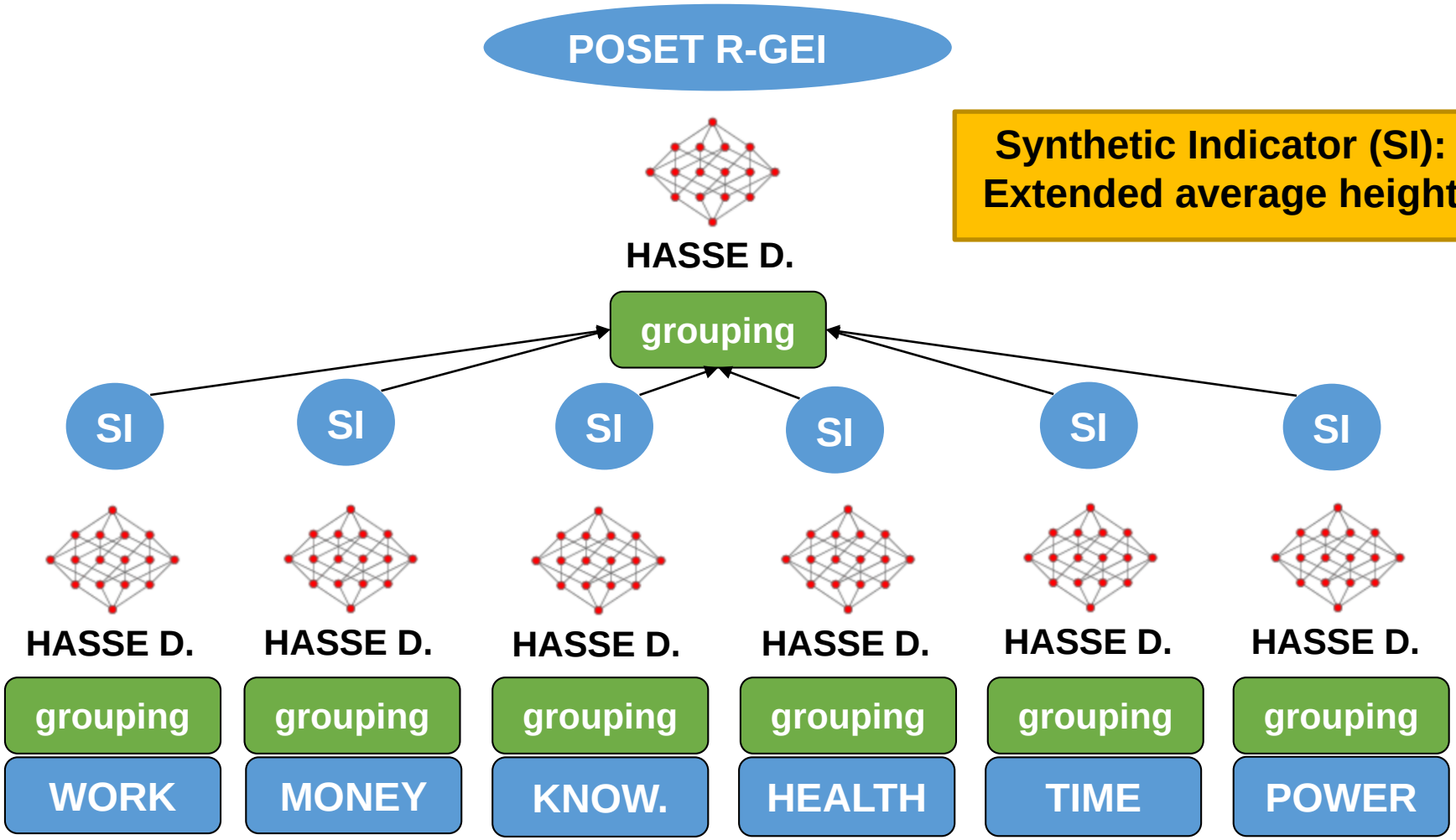
UNIVERSITÀ DEGLI STUDI
DI GENOVA

Posetic R-GEI 1/2 (POSAC R-GEI)



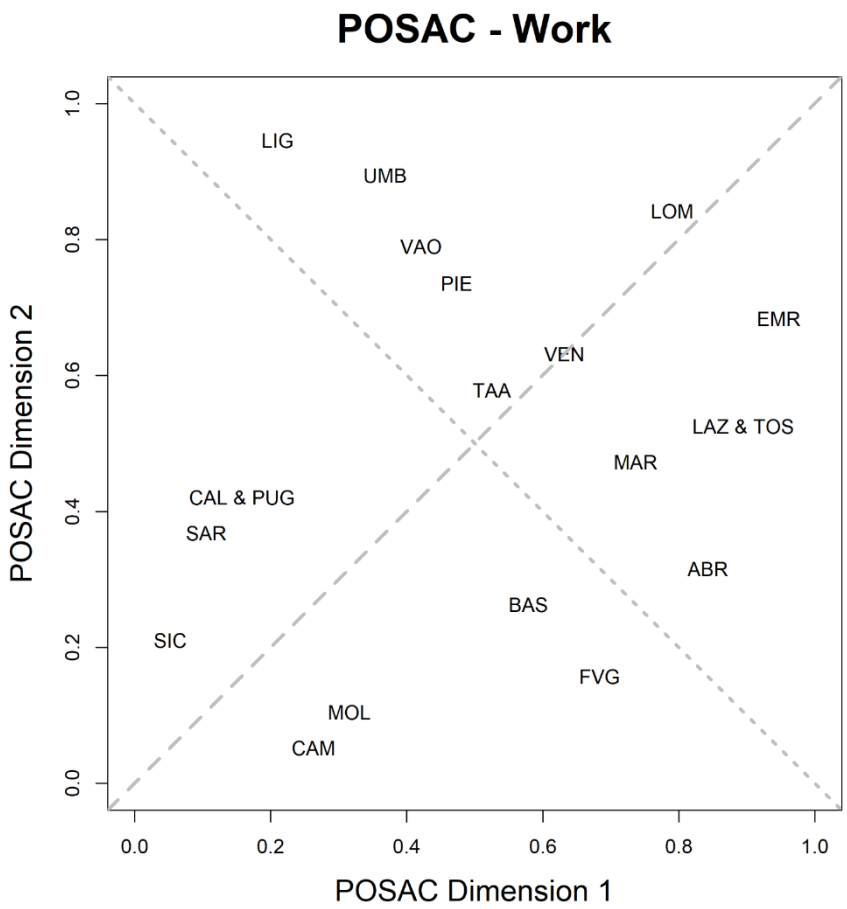
UNIVERSITÀ DEGLI STUDI
DI GENOVA

Posetic R-GEI 2/2 (POSET R-GEI)

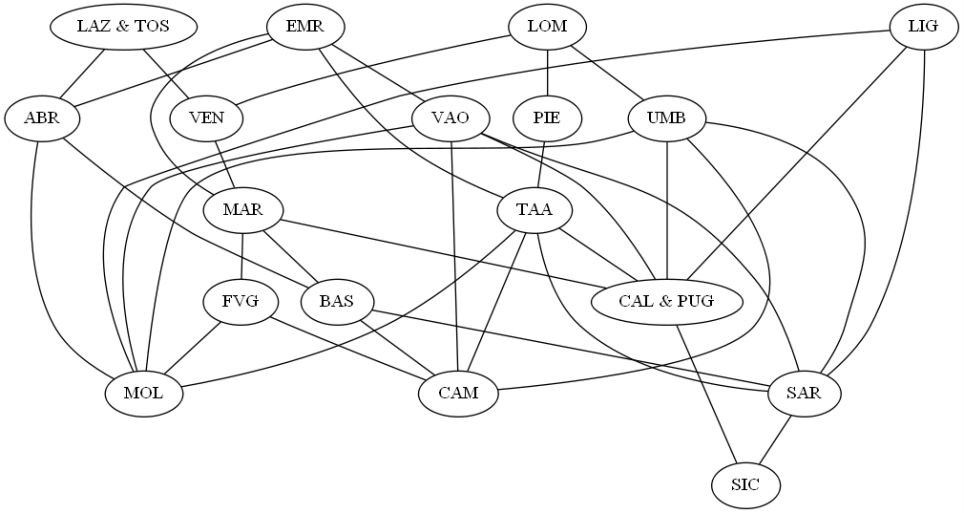


UNIVERSITÀ DEGLI STUDI
DI GENOVA

POSET R-GEI : Work

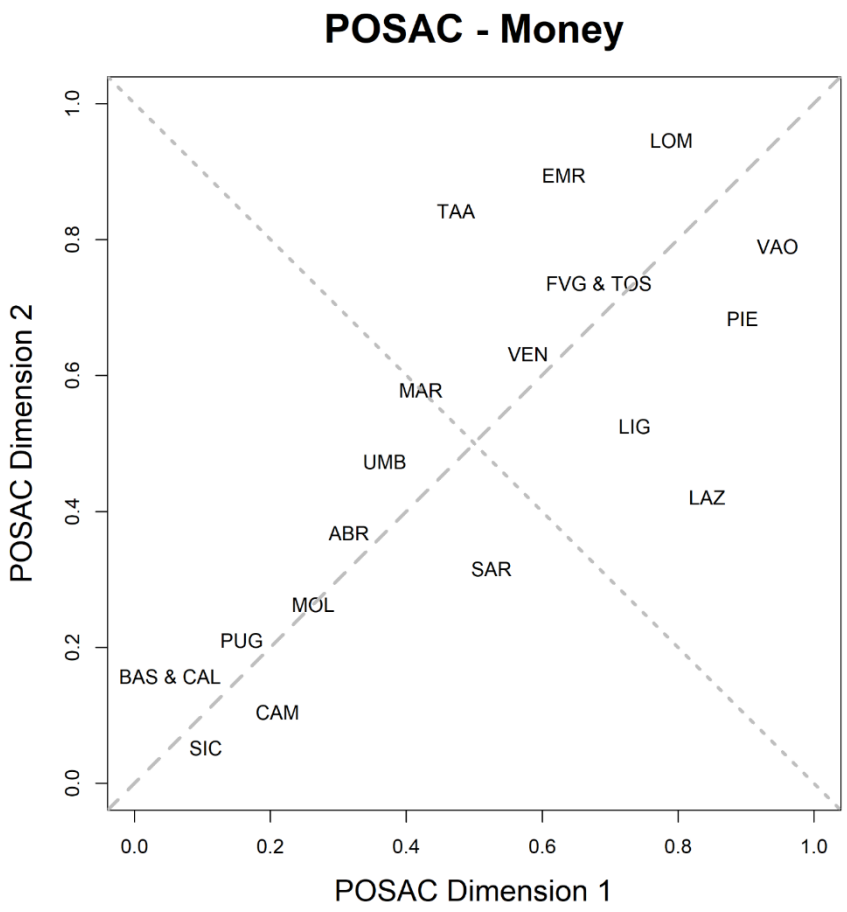


# Indicators	3
Cronbach's Alpha	0,6529
Proportion of Profile Pairs Correctly Represented	0,869
VAR(L-axis values)	0,016

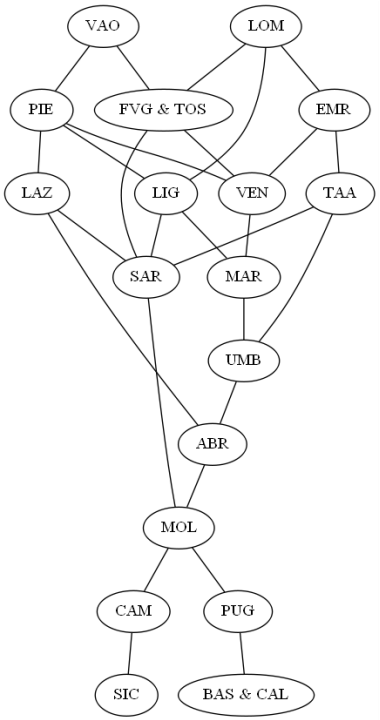


UNIVERSITÀ DEGLI STUDI
DI GENOVA

POSET R-GEI : Money

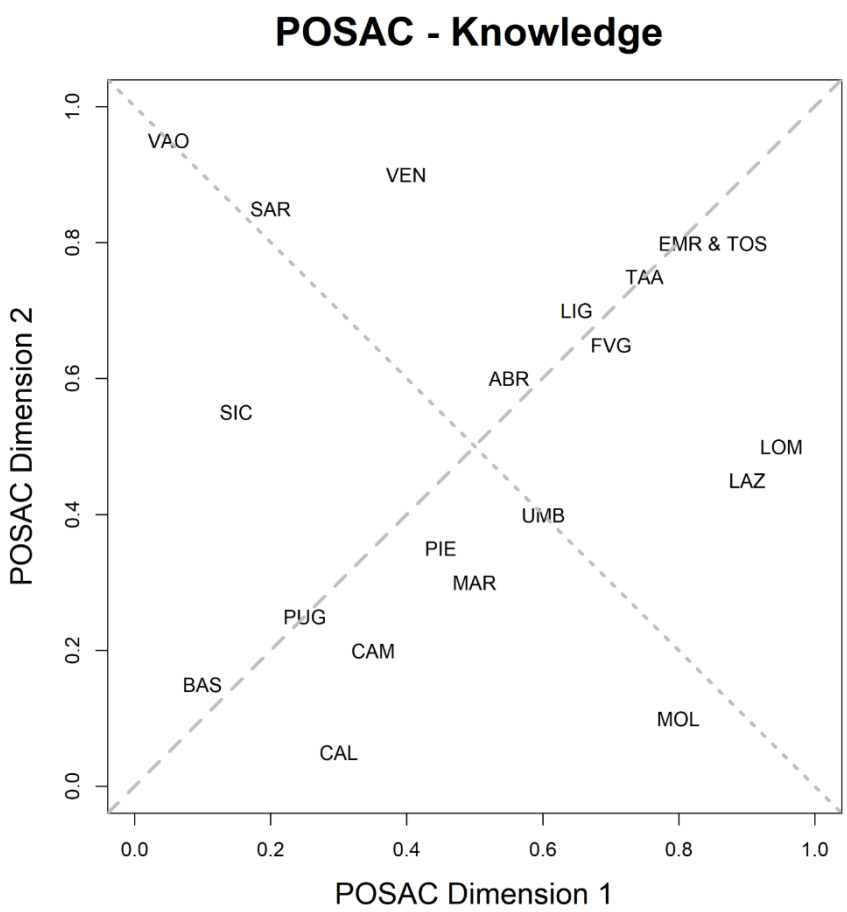


# Indicators	4
Cronbach's Alpha	0,9428
Proportion of Profile Pairs Correctly Represented	0,965
VAR(L-axis values)	0,004

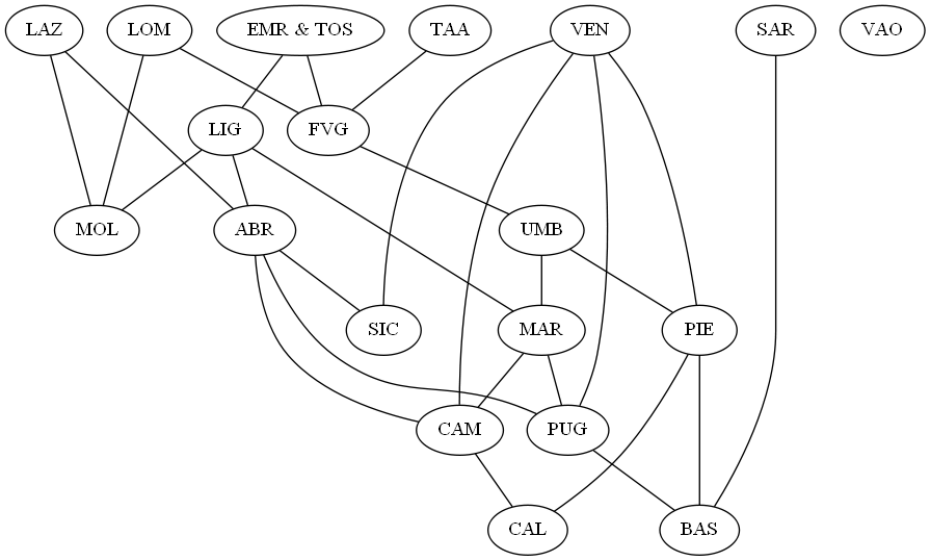


UNIVERSITÀ DEGLI STUDI
DI GENOVA

POSET R-GEI : Knowledge

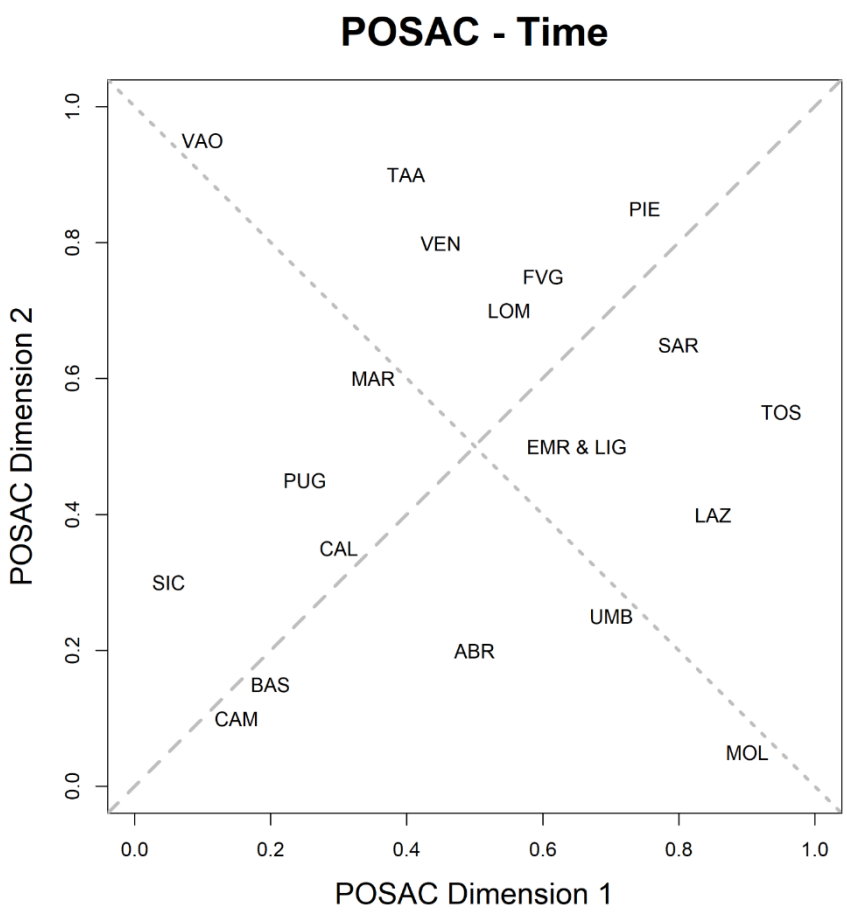


# Indicators	3
Cronbach's Alpha	**0,3194**
Proportion of Profile Pairs Correctly Represented	0,787
VAR(L-axis values)	0,02

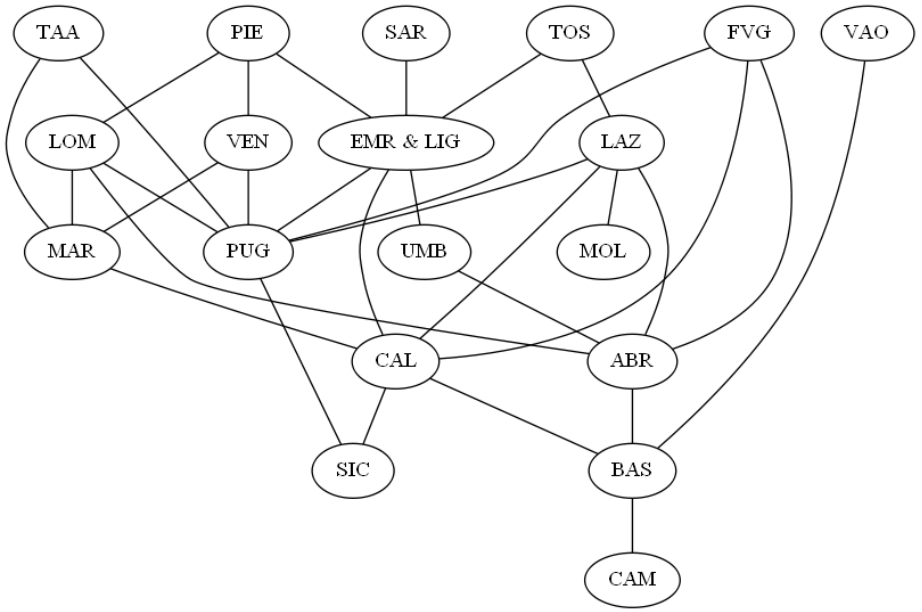


UNIVERSITÀ DEGLI STUDI
DI GENOVA

POSET R-GEI : Time

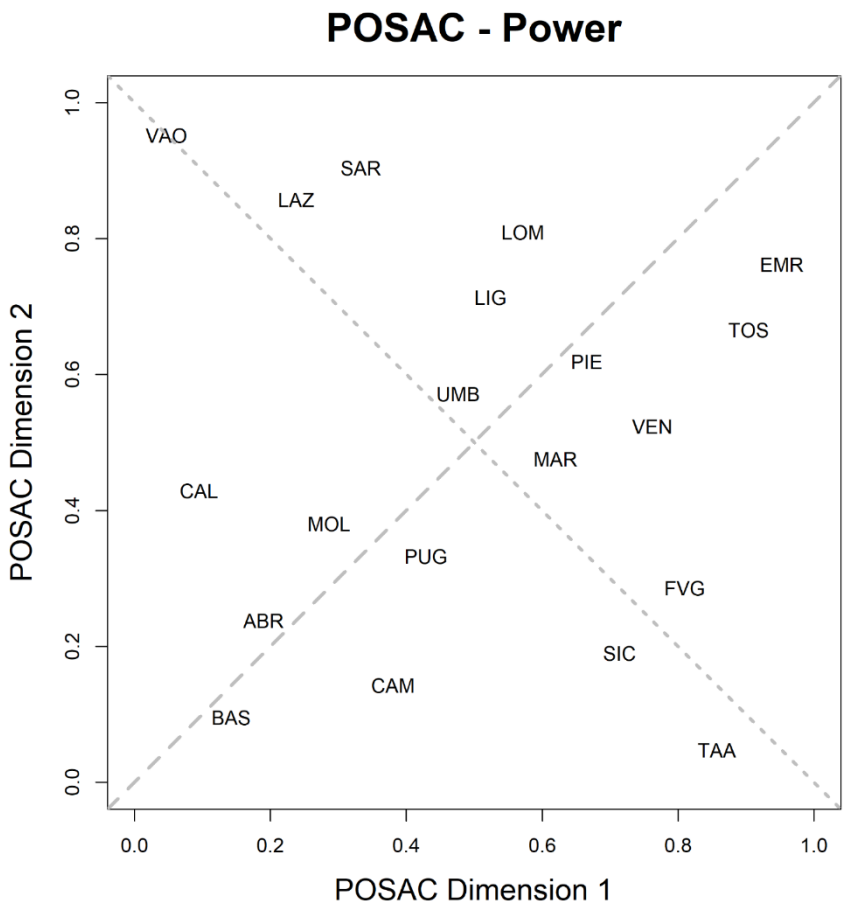


# Indicators	4
Cronbach's Alpha	0,6984
Proportion of Profile Pairs Correctly Represented	0,896
VAR(L-axis values)	0,02

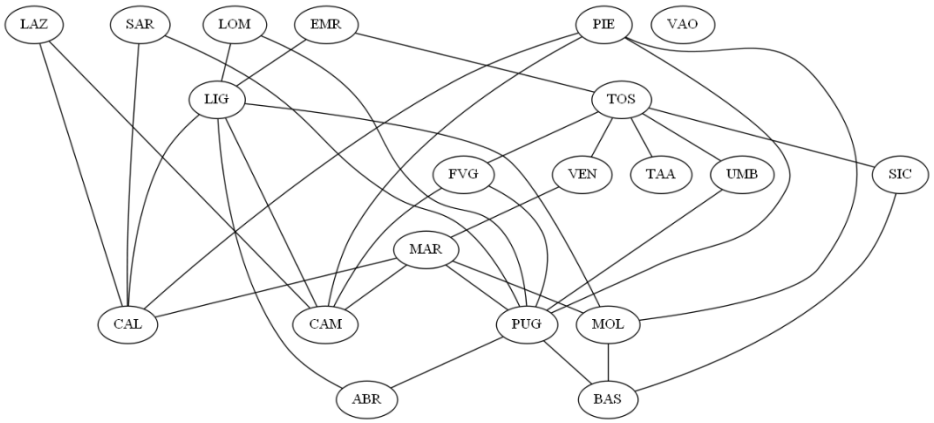


UNIVERSITÀ DEGLI STUDI
DI GENOVA

POSET R-GEI : Power

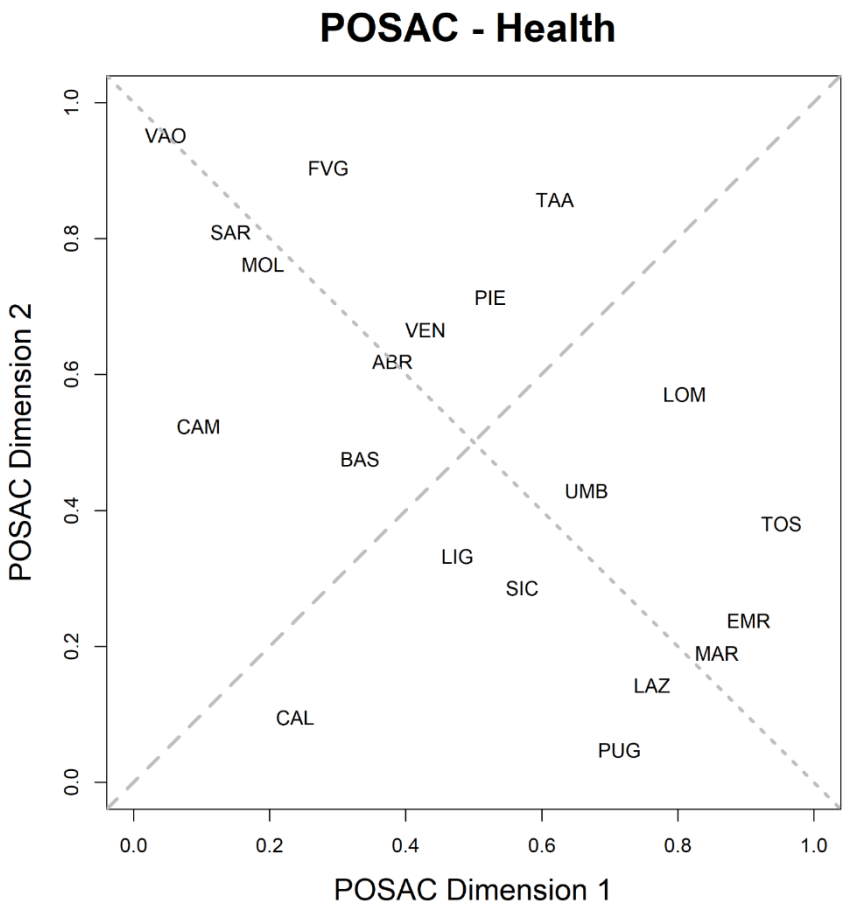


# Indicators	7
Cronbach's Alpha	0,8506
Proportion of Profile Pairs Correctly Represented	0,994
VAR(L-axis values)	0,024

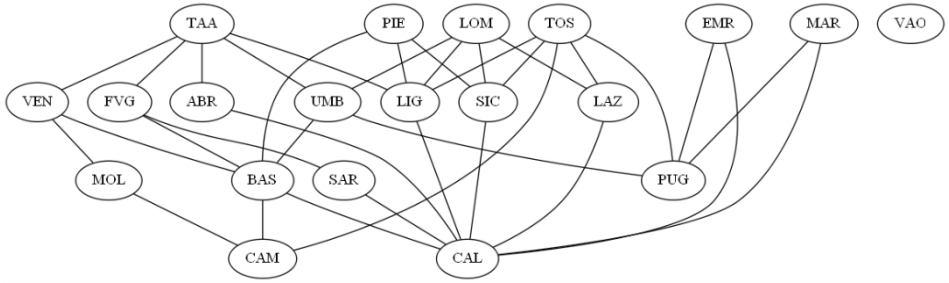


UNIVERSITÀ DEGLI STUDI
DI GENOVA

POSET R-GEI : Health

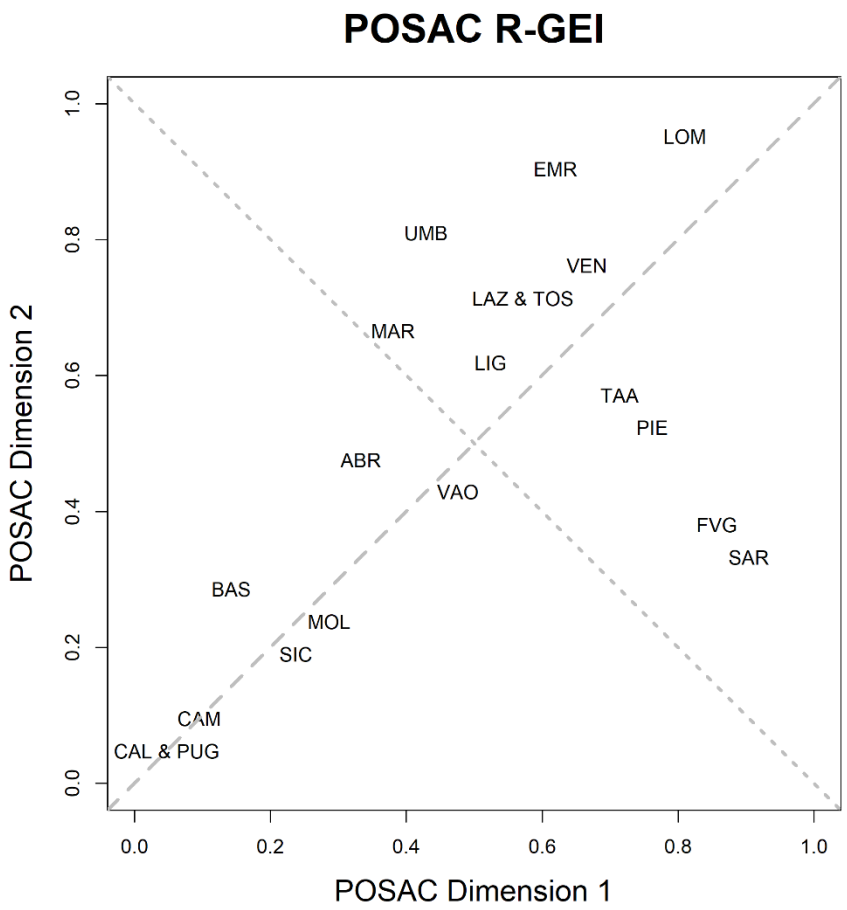


# Indicators	5
Cronbach's Alpha	0,6877
Proportion of Profile Pairs Correctly Represented	0,801
VAR(L-axis values)	0,034

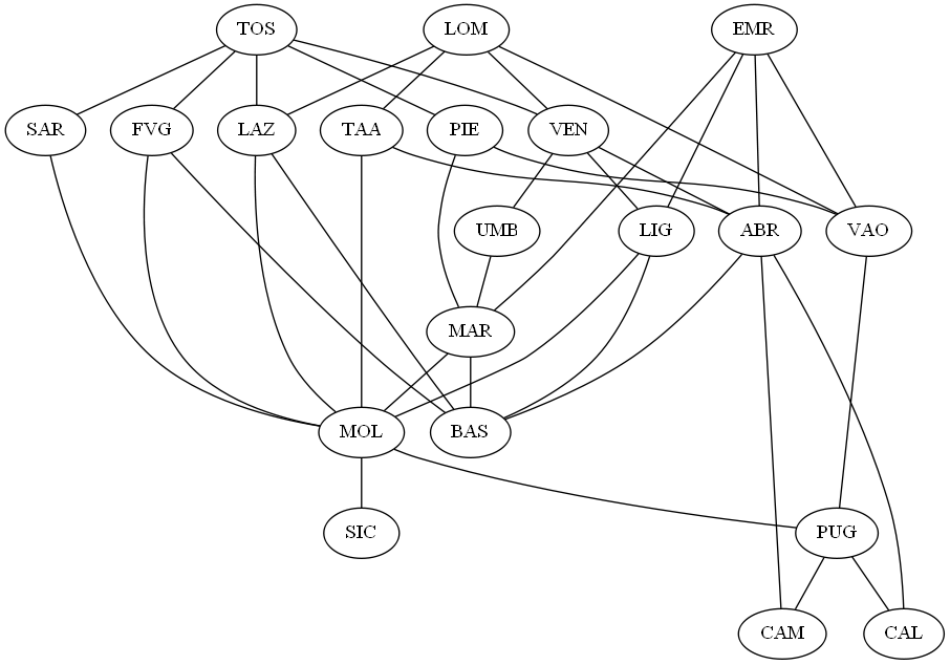


UNIVERSITÀ DEGLI STUDI
DI GENOVA

POSET R-GEI : GLOBAL INDEX

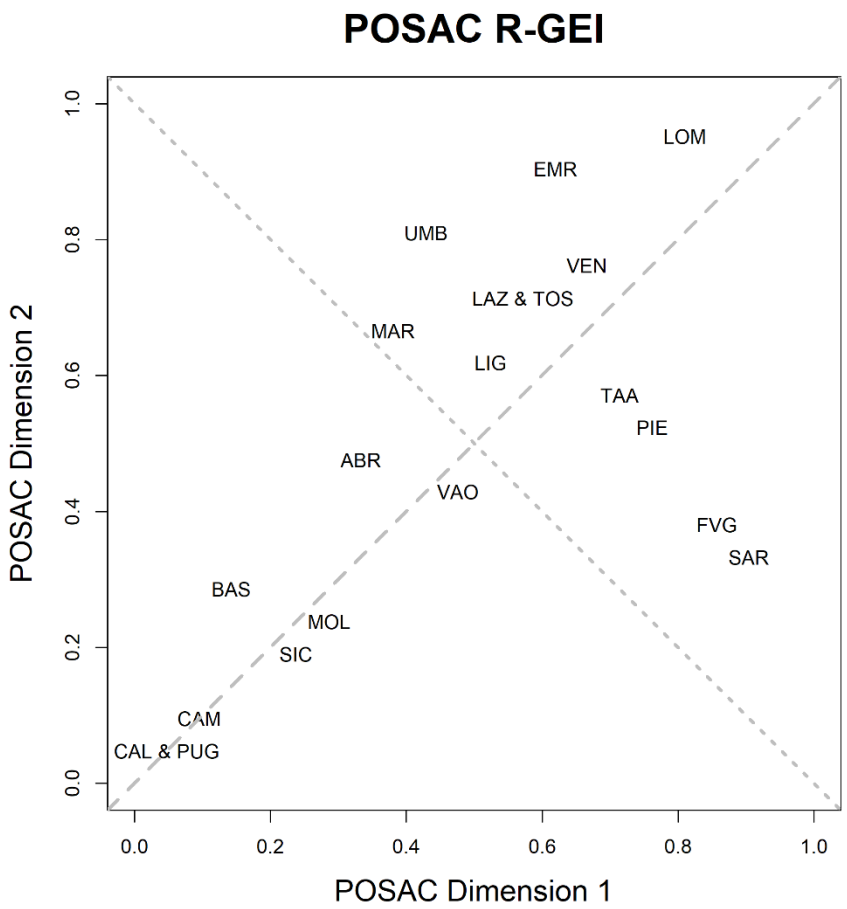


# Indicators	6
Cronbach's Alpha	0,9155
Proportion of Profile Pairs Correctly Represented	0,732
VAR(L-axis values)	0,006



UNIVERSITÀ DEGLI STUDI
DI GENOVA

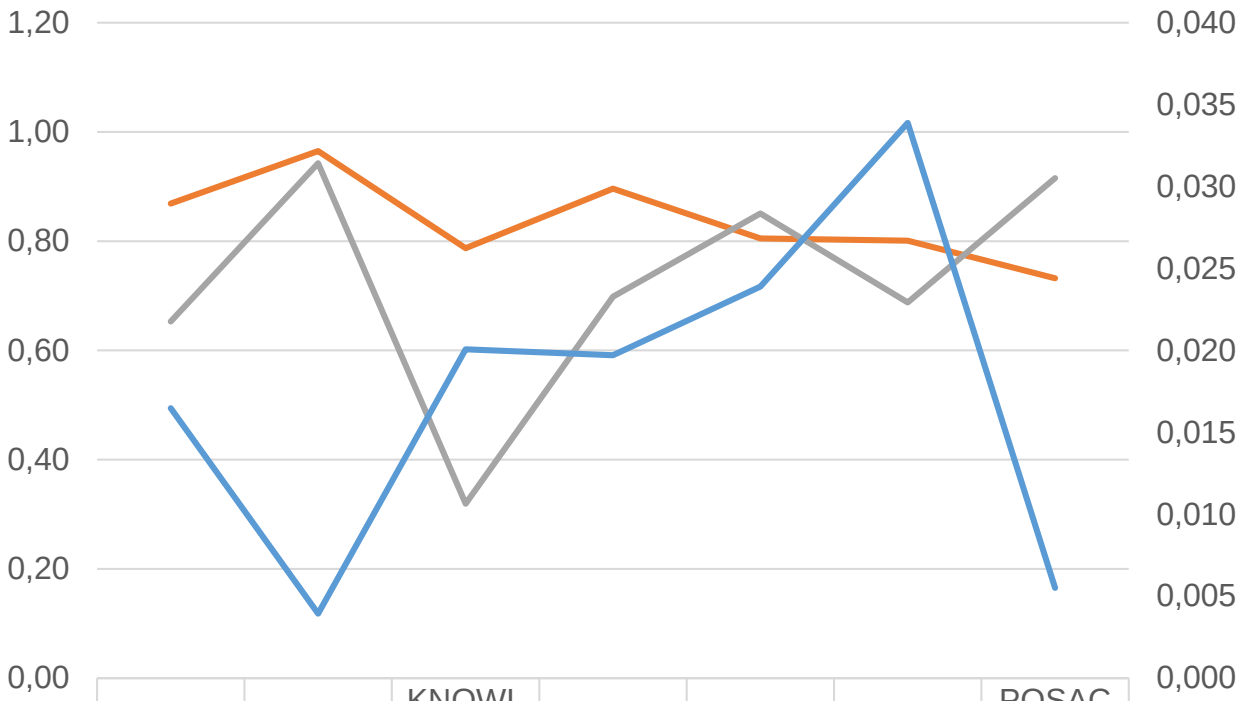
POSET R-GEI : GLOBAL INDEX



# Indicators	6
Cronbach's Alpha	0,9155
Proportion of Profile Pairs Correctly Represented	0,732
VAR(L-axis values)	0,06

Correlations	POSAC R-GEI	
	J axis	L axis
Health	0,84	0,01
Knowledge	0,87	0,03
Money	0,89	0,08
Power	0,78	0,08
Time	0,84	0,31
Work	0,78	-0,44

POSET R-GEI : GLOBAL INDEX



	WORK	MONEY	KNOWLEDGE	TIME	POWER	HEALTH	POSAC R-GEI
proportion of correct pairs	0,87	0,97	0,79	0,90	0,81	0,80	0,73
Cronbach's Alpha	0,65	0,94	0,32	0,70	0,85	0,69	0,92
VAR(L) - right scale	0,016	0,004	0,020	0,020	0,024	0,034	0,006

RANKINGS COMPARISONS

REGION	R-GEI	POSAC R-GEI	POSET R-GEI
LOM	1	2	2
EMR	2	3	3
TOS	3	1	1
PIE	4	7	5
FVG	5	9	7
TAA	6	5	6
LAZ	7	5	10
VEN	8	4	4
LIG	9	11	11
MAR	10	12	12
UMB	11	8	9
SAR	12	10	8
VAO	13	13	13
ABR	14	14	14
MOL	15	15	15
PUG	16	18	16
BAS	17	17	17
CAL	18	20	19
CAM	19	19	19
SIC	20	16	18

Spearman rank correlation matrix

	R-GEI	POSAC R-GEI
POSAC R-GEI	0,929	
POSET R-GEI	0,947	0,968

CONCLUSIONS

1. Posetic based synthetic indicators keep all inherent information of the indicators separate avoiding compensation.

2. In the posetic approach (POSET/POSAC R-GEI) there is no need to specify any weighting of variables/indicators to construct synthetic indicators.

3. Sources of incomparability are detectable and became sources of information.

4. POSAC is helpful to map regions and to define policy actions.

Key references

OECD (2008). Handbook on constructing composite indicators. Methodology and user guide. Paris: OECD Publications.

Maggino F. Zumbo B.D. (2016) Measuring the Quality of Life and the Construction of Social Indicators, in: Land K.C., Sirgy M.J. and Michalos A.C. (Edts), Handbook of Social Indicators and Quality of Life Research, Springer.

Brüggemann, R., Patil, G. P. (2011). Ranking and prioritization for multi-indicator systems. New York: Springer.

Shye, S. (1985). Multiple scaling. The theory and application of partial order scalogram analysis. Amsterdam: North-Holland.