Tecnologías del lenguaje para Explainable-AI y su impacto en el soporte a la decisión
Algunas aplicaciones a salud

K. Gibert

Knowledge Engineering&Machine Learning group at

Intelligent Data Science and Artificial Intelligence Res.C. (KEMLG-at-IDEAI).

Universitat Politècnica de Catalunya-BarcelonaTech, Spain

Vice dean of Big Data, Data Science and Artificial Intelligence

Official Chamber of Informatics Engineering of Catalonia

Interest Group IABiomed-Spanish Association of Artificial Intelligence (CAEPIA)

InfoDay sobre tecnologías del Lenguaje en sanidad y Biomedicina
BSC, Barcelona 2, diciembre 2019

Outline

- Introduction: Automatic interpretation of profiles
- Knowledge acquisition tools
  - Prior knowledge bases
  - Ontologies
  - Termometer
  - Super-concept based distance
- Explainability through embedded strategies in Data Science methods
  - Clustering based on rules and ontologies
- Profiles oriented Explainability tools
  - Visual: TLP, a-TLP
  - Conceptual: CCEC, CI-IMS
  - Dinamic: Trajectory map, Adherence map
- Knowldedge production tools
- Other cases: Topic modelling, Explainability in ANN
- Conclusions

Gap Data Mining- Decision making

Lack of data-miners
[Hal Varian 2008]
[AS Pentland 2013]
[Sooraij Shah 2013]

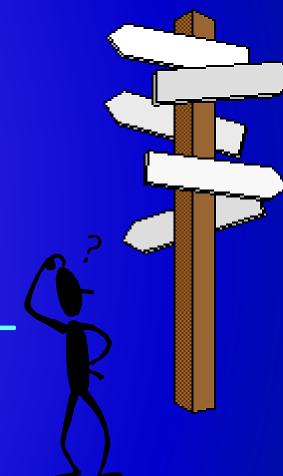
Data

Knowledge
Expertise

Data Mining

Data-driven results

Decision – making



The Fact Gap: The Disconnect Between Data and Decisions ***[Hammond 2004]***

No analysis ***No understandable*** ***No trust***

Explainability

Needs to be general literacy about data interpretation ***[A "Sandy" Pentland]***

keynote Campus Party Europa Sept 4th 2013 Head of MediaLab Enterpreneurship MIT

Data Science concept

- 2018: Gibert, Horsburg, Athanasiadis, Holmes *[ENVSOFT, 2018]*

Data science : emergent multidisciplinary field combining

- *Data analysis*
- *Data processing*
- *Domain expertise*

To transform data into understandable and actionable knowledge

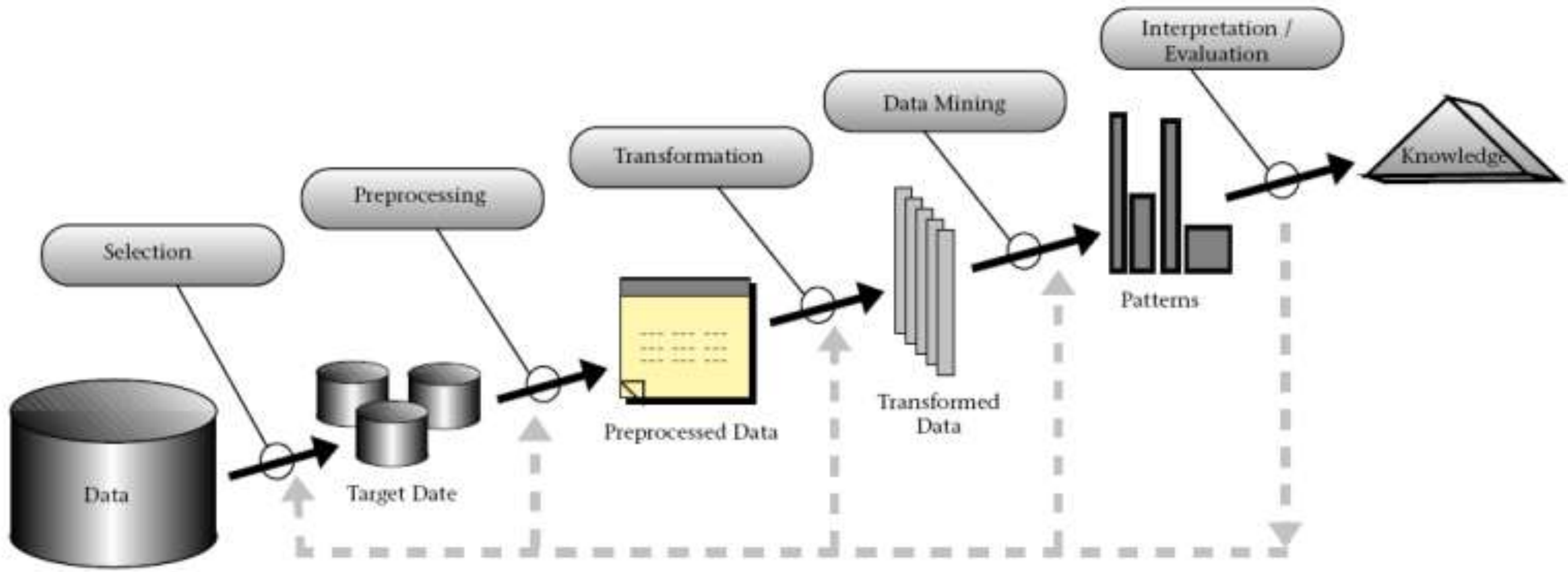
Relevant for informed decision making (reduces the Fact Gap)

- *involves intensive consumption of available and required data*
- *Copes with data heterogeneity*
- *BigData is a tool, not the focus, but domain complexity*

Gibert, K., ,J. Horsburgh, I. Athanasiadis, G. Holmes (2018) "Environmental Data Science." Environmental Modelling & Software 106: 4-12. DOI:10.1016/j.envsoft.2018.04.005

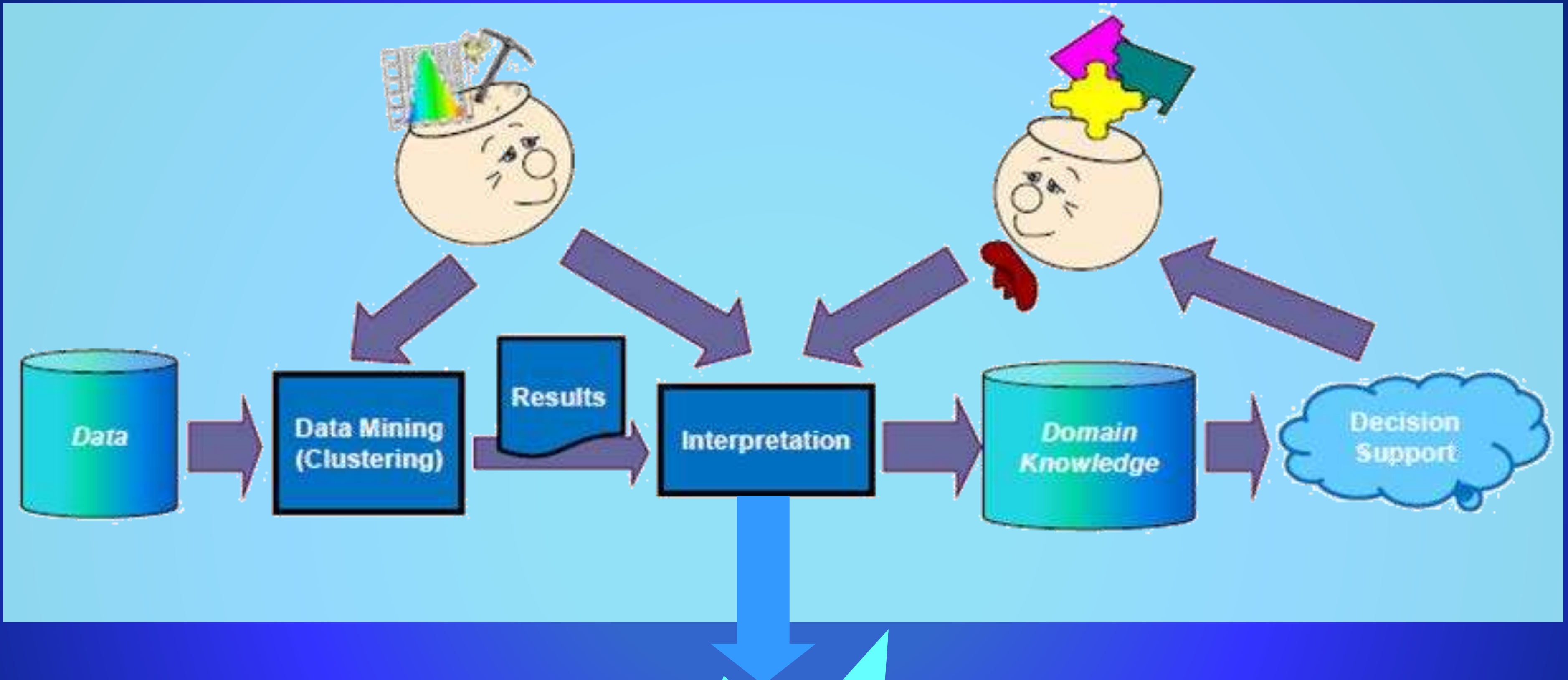
Data Mining and Knowledge Discovery

- Knowledge Discovery System [Fayy96]:



Focus: Clustering/Profiling

Expert-based collaborative Analysis (EbCA)



Include expert interaction in methodology

Gibert, K., García-Alonso, C., & Salvador-Carulla, L. (2010). Integrating clinicians, knowledge and data: expert-based cooperative analysis in healthcare decision support. Health research policy and systems, 8(1), 28 DOI:10.1186/1478-4505-8-28

Profiling mental health systems in LAMIC

Prior Expert Knowledge Acquisition

WHO-AIMS data

Rows: 42 LAMIC
Columns:
19 WHO-AIMS indicators
+ WHO-AIMS region
+ Income group
(World Bank classification)

KLASS

Clustering based on rules [Gib'96]

Ward's criterion
Gibert's mixed Metrics [Gib'97]

Prior Expert Knowledge

r0: Region=euro and income= lower
*-> **PreSoviet***

r1: Region=AMR and income = lower
*-> **poorAmerica***

r2: Region= SEAR and population < 10000000
*-> **smallAsia***

Gibert, K., Izquierdo, J., Sànchez-Marrè, M., Hamilton, S. H., Rodríguez-Roda, I., & Holmes, G. (2018). Which method to use? An assessment of data mining methods in Environmental Data Science. Environmental modelling & software, 110, 3-27.

Profiling mental health systems in LAMIC

Prior Expert Knowledge Acquisition

WHO-AIMS data

Rows: 42 LAMIC
Columns:
19 WHO-AIMS indicators
+ WHO-AIMS region
+ Income group
(World Bank classification)

KLASS
Clustering based on rules *[Gib'96]*

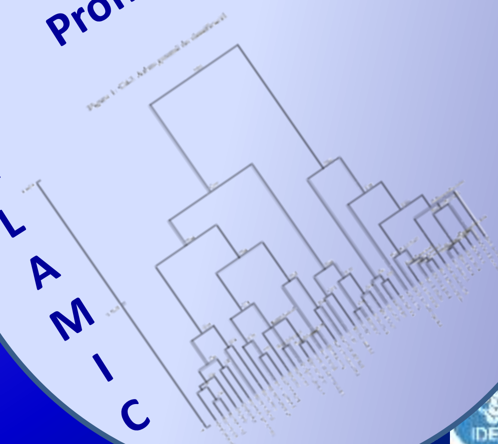
Ward's criterion
Gibert's mixed Metrics [Gib'97]

Prior Expert Knowledge

r0: Region=euro and
income= lower
-> PreSoviet
r1: Region=AMR and
income = lower
-> poorAmerica
R2: Region= SEAR and
population < 10000000
*-> **smallAsia***

MHS in LAMIC

Profiles



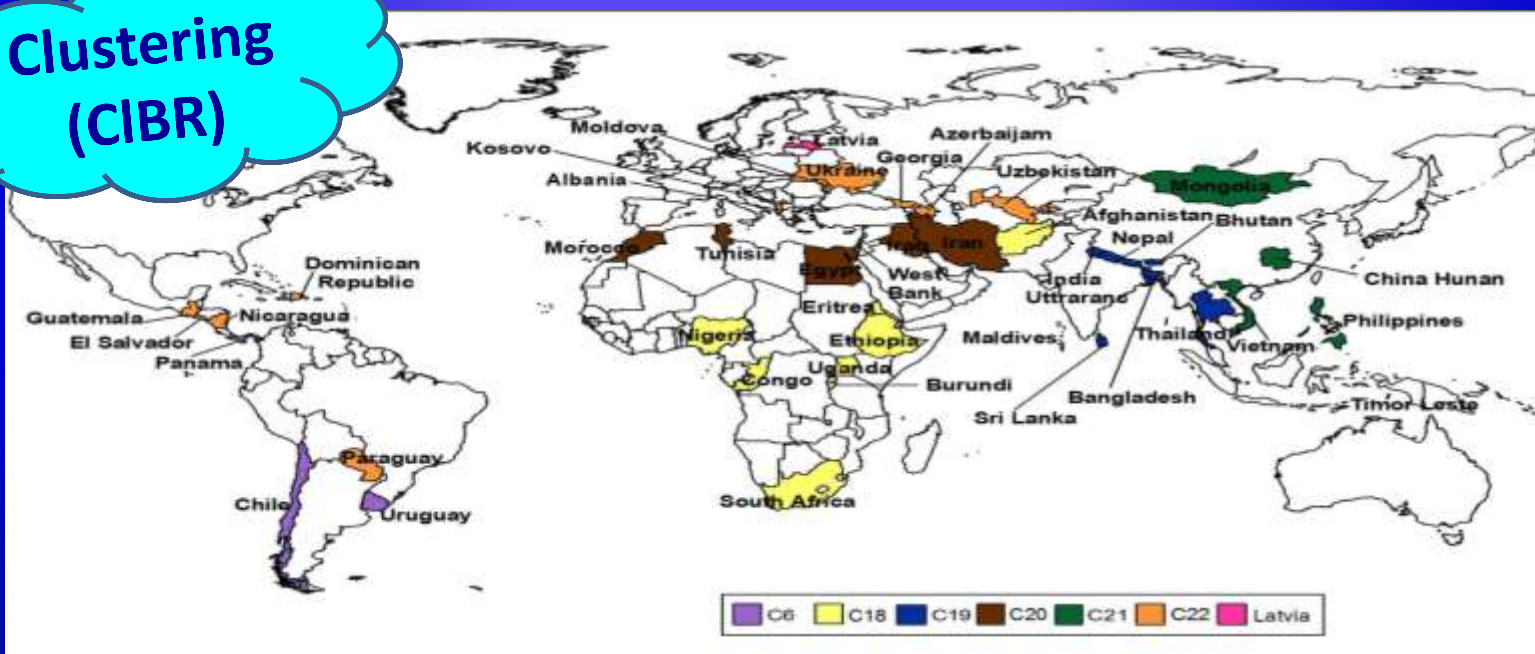
Profiling mental health systems in LAMIC countries for healthcare policy-making at WHO

Use WHO-AIMS DB to learn a **typology** of MHS in LAMIC

- Easy **understanding** of reality
- Assessment to countries
- Intervention design: guidelines, mental health policies....

Postprocessing CPG, TLP

Clustering (ClBR)



Gibert, K, L. Salvador-Carulla, J. Morris, A. Lora, S. Saxena (2017) The data mining approach as the starting point for Mental Health policy-making in low and middle income countries at World Health Organization. In Proc of ISI 2017. Belcasem Abdou et al, Marrakesh, Morocco, Jul 2017

TLP elicits clustering criteria
Conceptualization

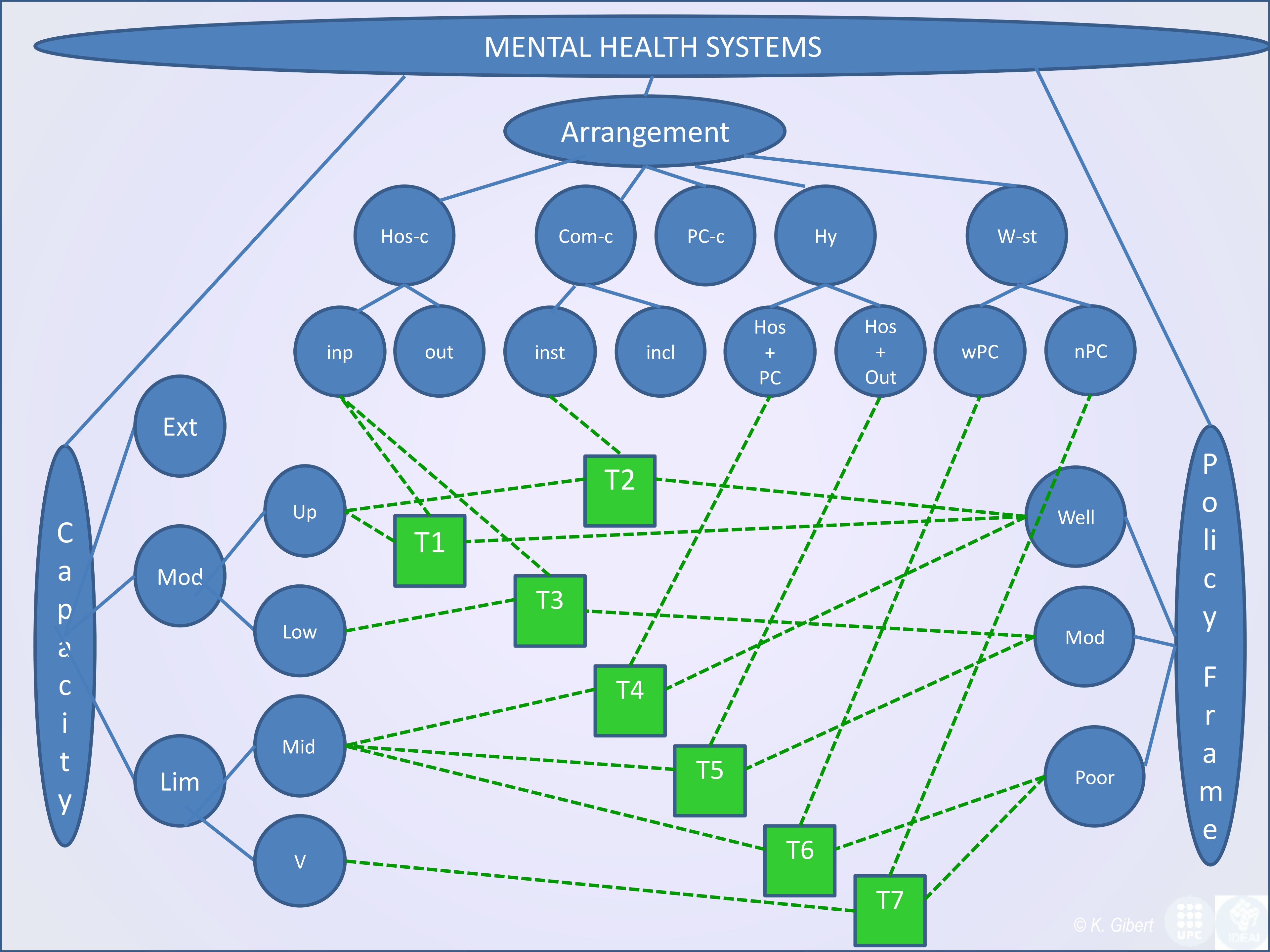
Induces categories of variables and classes

BLOCK		CLASS	CARE CAPACITY						CARE ARRANGEMENT					POLICY FRAMEW		
			Incom	HR	$MHe	Treat pre	Cap-ratio	close beds	%$m-hosp	LTC-pacs	comc arew	Lund	Manua	Legis	Pol-plan	Gov-Rep
I	*Upper-Moderate*	CI	UpMid	Highst	Highst	Highst	Highst	Lowest	High++	High	High	Lowest	No	yes	yes	yes
		C6	UpMid	Mod	Mod	Mod	Low	Low	Mod	Highest	Low	High	Some	yes	yes	Some
	Low-Mod	C22	LMid	Mod	Low	Mod	Mod	Mod	Highest	High	Mod	Mod	No	Some	Some	No
II	*Mid-Limited*	C21	LMid	Low	Low	Low	Mod	Mod	High	High	Mod	Mod	yes	No	yes	Some
		C20	LMid	Low	Low	Low	Low	Mod	High	High	Highest	Highest	yes	Some	yes	Some
		C19	LMid	Low	Low	Lowest	Lowest	High	High	Low	Low	High	Some	No	Most	Most
	Very Lim	C18	Low	Low	Low	Low	Low	Highst	High++	Low	Low	Mod	No	Few	Most	No

Adequate
Moderate
Basic

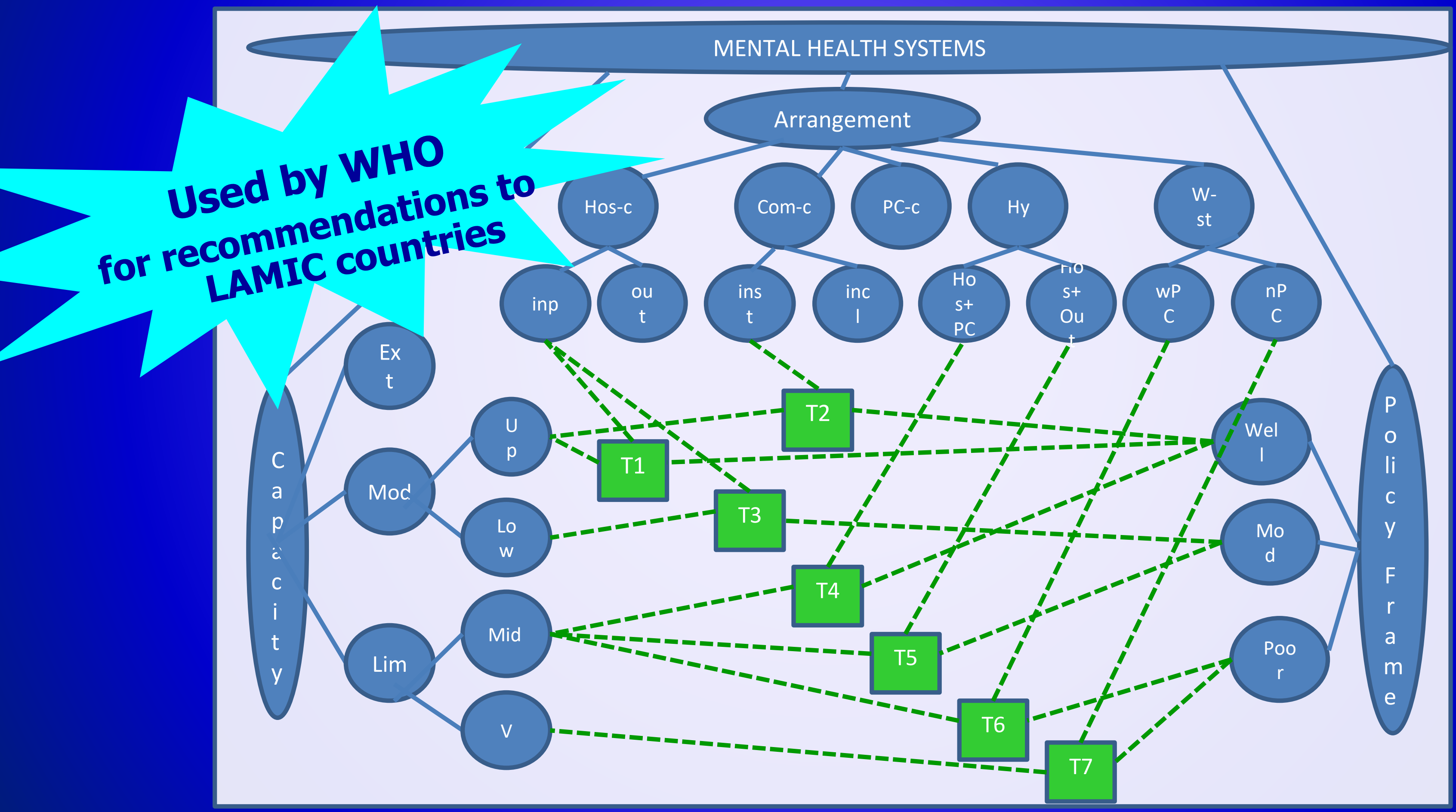
Supports data-driven Ontologies

Gibert K, D. Conti, D. Vrecko (2012) Assisting the end-user in the interpretation of profiles for decision support. An application to wastewater treatment plants. Environmental Engineering and Management Journal 11(5): 931-944


MENTAL HEALTH SYSTEMS
Arrangement
Hos-c
Com-c
PC-c
Hy
W-st
inp
out
inst
incl
Hos + PC
Hos + Out
wPC
nPC
Capacity
Ext
Mod
Lim
Up
Low
Mid
V
T1
T2
T3
T4
T5
T6
T7
Policy Frame
Well
Mod
Poor
© K. Gibert

Knowledge Production

MHS for LAMIC ontology



Intervention plans designed for each type

The KLASS thermometer-tool

hr/100000pop

US$ per capita

treated/100000pop

% (treated pacs/prev)

(%f totbeds) per cap

% from MH budget

%from pacs in MHos

%from treated pacs

Outcontacts/daysHsp

TLP elicits clustering criteria
Conceptualization

Induces categories of variables and classes

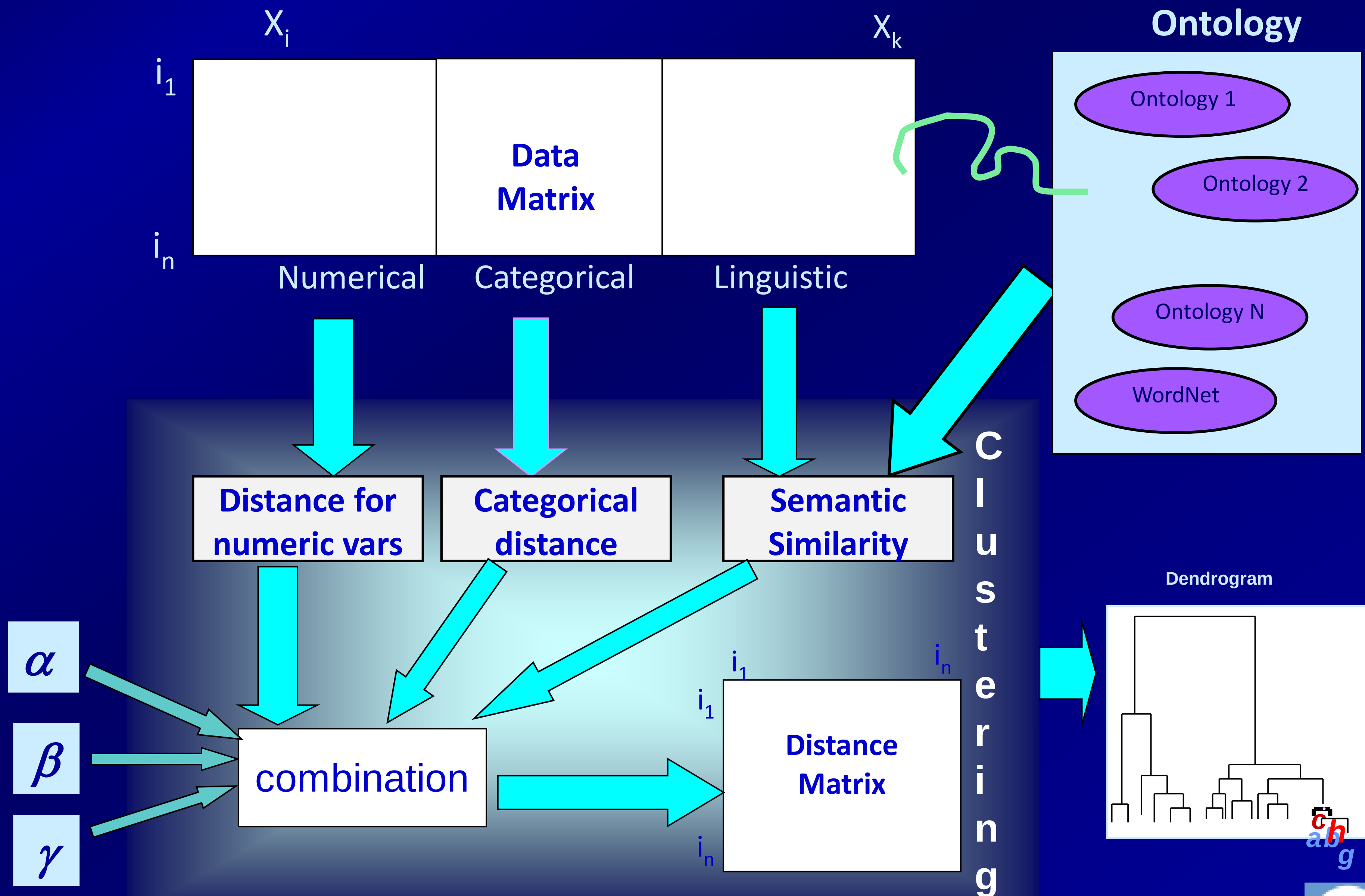
BLOCK		CLASS	CARE CAPACITY						CARE ARRANGEMENT					POLICY FRAMEW		
			Incom	HR	$MHe	Treat pre	Cap-ratio	close beds	%$m-hosp	LTC-pacs	comc arew	Lund	Manua	Legis	Pol-plan	Gov-Rep
I	*Upper-Moderate*	CI	UpMid	Highst	Highst	Highst	Highst	Lowest	High++	High	High	Lowest	No	yes	yes	yes
		C6	UpMid	Mod	Mod	Mod	Low	Low	Mod	Highest	Low	High	Some	yes	yes	Some
	Low-Mod	C22	LMid	Mod	Low	Mod	Mod	Mod	Highest	High	Mod	Mod	No	Some	Some	No
II	*Mid-Limited*	C21	LMid	Low	Low	Low	Mod	Mod	High	High	Mod	Mod	yes	No	yes	Some
		C20	LMid	Low	Low	Low	Low	Mod	High	High	Highest	Highest	yes	Some	yes	Some
		C19	LMid	Low	Low	Lowest	Lowest	High	High	Low	Low	High	Some	No	Most	Most
	Very Lim	C18	Low	Low	Low	Low	Low	Highst	High++	Low	Low	Mod	No	Few	Most	No

Adequate
Moderate
Basic

Supports data-driven Ontologies

Gibert K, D. Conti, D. Vrecko (2012) Assisting the end-user in the interpretation of profiles for decision support. An application to wastewater treatment plants. Environmental Engineering and Management Journal 11(5): 931-944

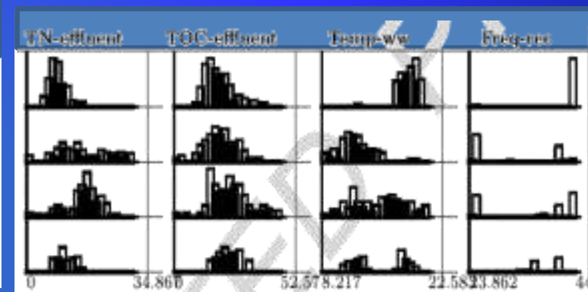
Semantic Distances



Gibert K, Valls A, Batet M. (2014): Introducing semantic variables in mixed distance measures. Impact on hierarchical clustering, Knowledge and Information Systems 40(3):559-593, Springer, DOI: 10.1007/s10115-013-0663-5

a-TLP: going further (WWTP case)

		Influent						2nd An-T	1st A-T			2nd A-T		Effluent			Other
Class	nc	Q	NH4	TN	TOC	Ni Tri tox	FR1-DO TOK	h-ww	Q-air	Val-ve air	O2-1 aero-bic	O2-2 aero-bic	NH4-2 aero-bic	TN	TOC	Temp ww	Frec rec
C360	100																
C358	93																
C353	122																
C357	50																



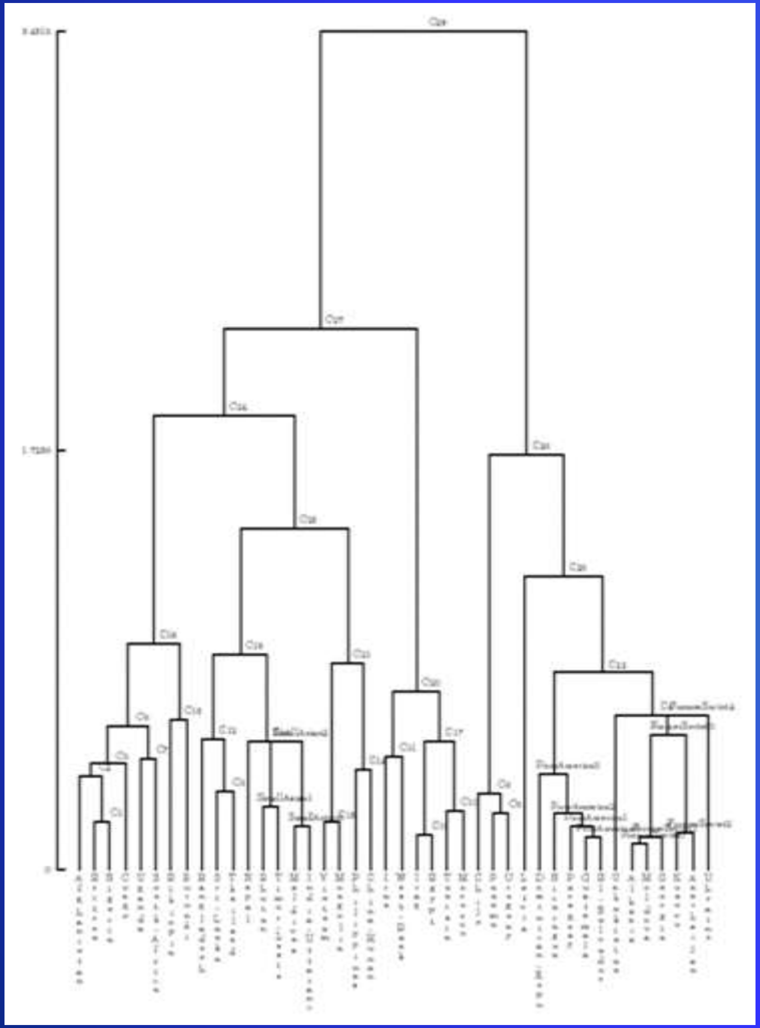
		influent						Aerobic Tanks 1 & 2 - Anoxic Tank 2						Effluent			Other
Class	nc	Q	NH4	TN	TOC	Ni Tri tox	FR1-DO TOK	h-ww	Q-air	Val-ve air	O2-1 aero-bic	O2-2 aero-bic	NH4 aero-bic	TN	TOC	Temp ww	Frec rec
C360	100																
C358	93																
C353	122																
C357	50																

	RGB color	(A,0,0)	(0,A,0)	(A,B,0)	
Coefficient of Variation	A	Colour	Colour	Colour	B
0,00	255				255
0,05	244				255
0,10	233				254
0,15	222				253
0,20	212				252
0,25	202				251
0,30	192				249
0,35	182				247
0,40	173				245
0,45	164				242
0,50	155				239
0,55	146				236
0,60	138				232
0,65	130				227
0,70	122				222
0,75	114				217
0,80	107				211
0,85	100				204
0,90	93				197
0,95	86				189
1,00	80				180

Gibert K, D. Conti (2015) aTLP: A color-based model of uncertainty to evaluate the risk of decisions based on prototypes. Artificial Intelligence Communications 28:113-126, IOSPress

CCEC: Conceptual Caracterization by Embedded Conditioning

Exploits dendrogramm structure to induce classification rules



r1.BC0.−r50r0: ((treatpre∈[18,57,172,77))∧(comcarewor∈[0,0197,0,1098)))∧(Region∈{AFR})−→(NovaClasseBLN7)C18

r2.BC1.−r2−r46−r50r0−r35−r37−r39 : (((treatpre∈[18,57,172,77))∧(comcarewor∈[0,0197,0,1098))) ∧(((Region∈{SEAR})∨(lundpararectrail∈[0,49,0,53)))∨(comcarewor∈[0,0197,0,0255))))∧ ((((Region∈{SEAR})∧(treatpre∈[31,81, 87,59]))∧ (lundpararectrail = 0,49))∧(comcarewor = 0,0197))−→(NovaClasseBLN7)C19

r4.BC3.−r46r53:(treatpre∈[18,57,172,77))∧(comcarewor∈(0,1313,0,624])−→(NovaClasseBLN7)C20

Gibert K 2014 Automatic generation of classes interpretation as a bridge between clustering and decision making International Journal of Multicriteria Decision Making 4(2):154-182 Inderscience

CIMS: Cluster Interpretation based on Integrated Marginal Significance

Same differences with same conceptualizations in all classes

Consistency Inter Classes: Generalized Test –Value

$$\text{Numerical: } \tau_\nu = \frac{\bar{X}^C - \bar{X}}{\sqrt{\left(1-\frac{n_c}{n}\right)\frac{s^2}{\nu}}} \sim t_{\nu-1} \qquad \text{Qualitative: } \pi_\nu = \frac{p_{sc} - p_s}{\sqrt{\left(1-\frac{n_c}{n}\right)\frac{p_s(1-p_s)}{\nu}}} \sim z$$

Sensitivity Analysis

$\downarrow \nu \rightarrow \uparrow$ *p-value*

	ϵ_2 0.2 ←	ϵ_1 0.3 ←	n	ϵ_1 0.3 →	ϵ_2 0.2 →
	0.5n	**0.7n**	**n**	**1.3n**	**1.5n**
Descripion-Power (Π)	ν_1	ν_2	ν_3	ν_4	ν_5
Robust Non-Significant ($\overline{R}$)	✗	✗	✗	✗	✗
Moderate Non-Significant ($\overline{M}$)	✗	✗	✗	✗	✓
Weak Non-Significant ($\overline{W}$)	✗	✗	✗	✓	✓
Weak Significant (W)	✗	✗	✓	✓	✓
Moderate Significant (M)	✗	✓	✓	✓	✓
Robust Significant (R)	✓	✓	✓	✓	✓
Basic Descriptor (B)	B	B	B	B	B

Class Descriptor

< W, C, description-power, sense>

$$W = \begin{cases} X & \text{if } X \text{ numerical} \\ < X, s > & \text{if } X \text{ categorical } \wedge \\ & s \text{ category} \in D_X \end{cases}$$

sense $\epsilon \{\uparrow, \downarrow\}$

Regular Expressions

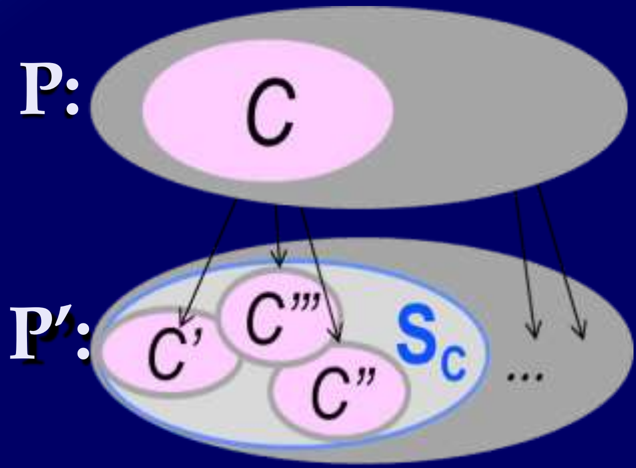
Proportion of Smokers (Tobacco) is higher in C1

Weight is high in class C2

Age is lowest in class C1

K. Gibert

Interpreting X in Nested Partitions



$$C = \bigcup_{C' \in S_C} C' \in P'$$

Relationship between interpretation of X in C and S_C

Sub Classes \ Super Class	Non-Significant	Significant
Non-Significant	Irrelevance $\forall C' \in S_C\colon R(C,C') = Irrelevance$	Inconsistency $\forall C' \in S_C\colon R(C,C') = Inconsistency$
Significant	Specification $\exists\, C' \in S_C\colon R(C,C') = Specification$	Inheritance $\exists\, C' \in S_C : R(C,C') = Inheritance$

Contradiction

NCI-IMS: *Cluster Interpretation based on Integrated Marginal Significance for Nested partitions*

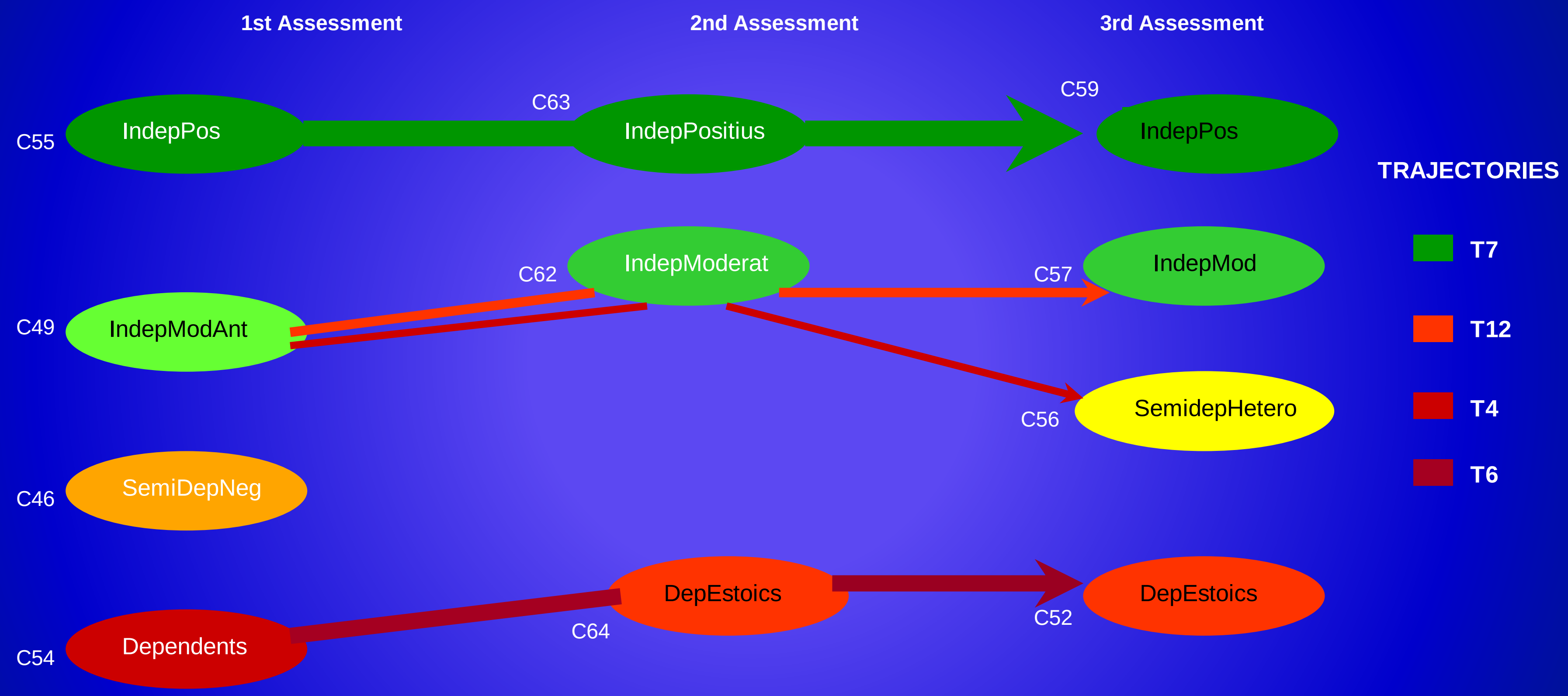
Table $\mathcal{A}$: Actions associated to Table $\mathcal{R}$

<table>
<tr><th>Super / Class / SubClass</th><th>$\overline{R}$</th><th>$\overline{M}$</th><th>$\overline{W}$</th><th>W</th><th>M</th><th>R/B</th></tr>
<tr><td>$\overline{R}$</td><td colspan="4" rowspan="4">W ignored in description of C and C’ ∈ S_C</td><td colspan="2" rowspan="2">W in description of C</td></tr>
<tr><td>$\overline{M}$</td></tr>
<tr><td>$\overline{W}$</td><td colspan="2" rowspan="4">W in description of C and C’ ∈ S_C</td></tr>
<tr><td>W</td></tr>
<tr><td>M</td><td colspan="4" rowspan="2">W in description of C’ ∈ S_C</td></tr>
<tr><td>R/B</td></tr>
</table>

Gibert, K, B. Sevilla-Villanueva, M. Sànchez-Marrè (2016) The role of significance tests in consistent interpretation of nested partitions. Journal of Computational and Applied Mathematics, 292: 623-633, Elsevier, Amsterdam, NL (htpps://doi.org/10.1016/j.cam.2015.01.031)

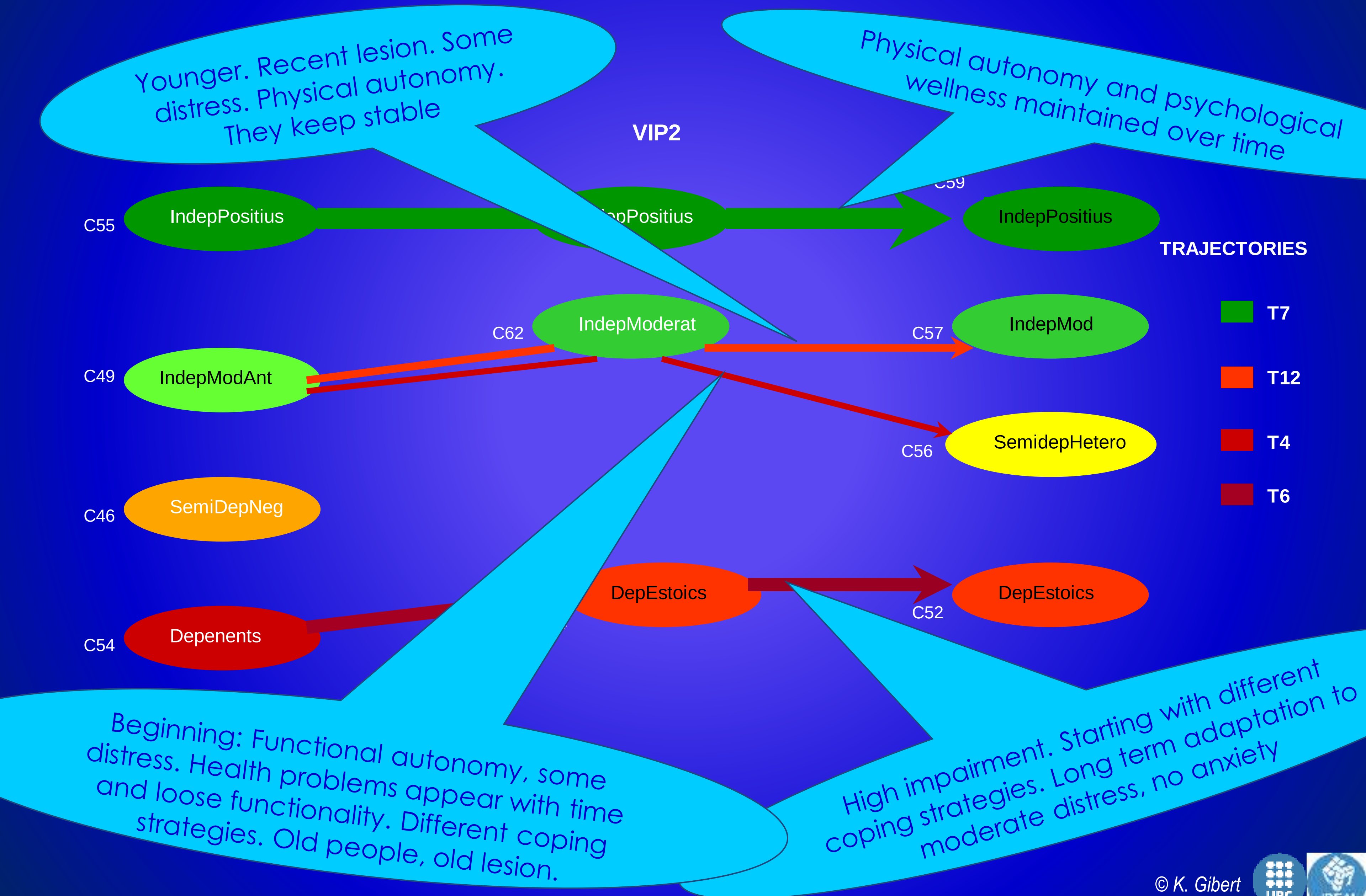
Trajectory maps

More typical patterns ($\gamma \geq 0.05$)

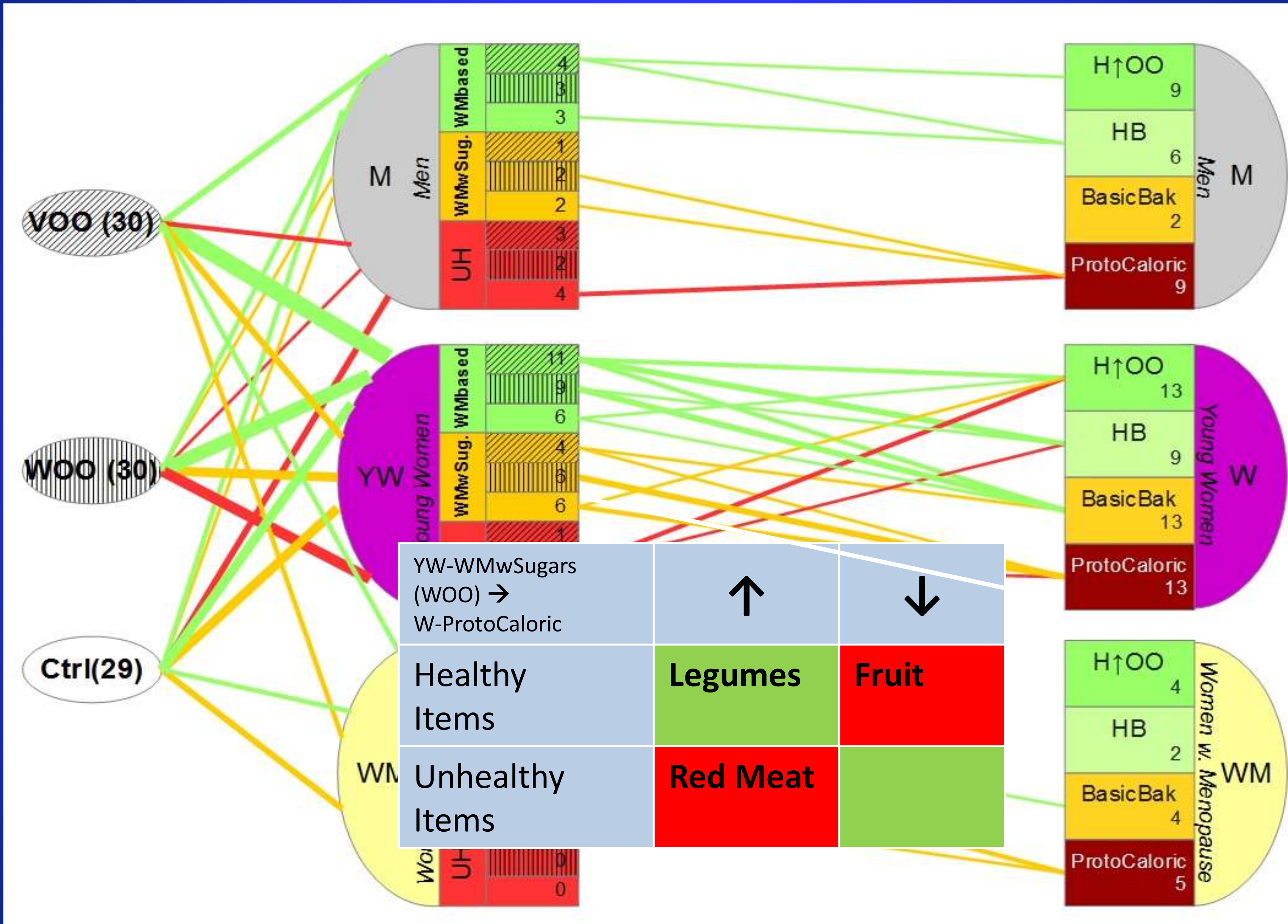


K. Gibert, A. García-Rudolph, L. Curcoll, D. Soler,L. Pla, J. M. TormosKnowledge Discovery about Quality of Life changes of Spinal cord Injury patients: Clustering based on rules by States. In Studies in Health Technology and Informatics, v150: 579—583. IOSSPress

Expert's conceptualization of patterns



Trajectory Characterization. Adherence



YW-WMwSugars (WOO) → W-ProtoCaloric	↑	↓
Healthy Items	**Legumes**	**Fruit**
Unhealthy Items	**Red Meat**	

Sevilla-Villanueva B., K. Gibert, M. Sánchez-Marrè, M. I. Covas, M. Fitó (2016) Evaluation of adherence to nutritional intervention through trajectory analysis. IEEE Journal of biomedical and health informatics, 21(3):628-634, NY (USA) DOI: 10.1109/JBHI.2016.2634698

Assignment of the profile of a new patient

Given a new patient:

Estimate π_{High} by applying equation 1

If p_{High} is $>= \xi$ *then* assign patient to *High* profile

Else, Estimate π_{IntII} by applying equation 2

If p_{IntII} is $>= \xi$ *then* assign patient to *IntermediateII* profile.

Else Estimate π_{IntI} by applying equation 3.

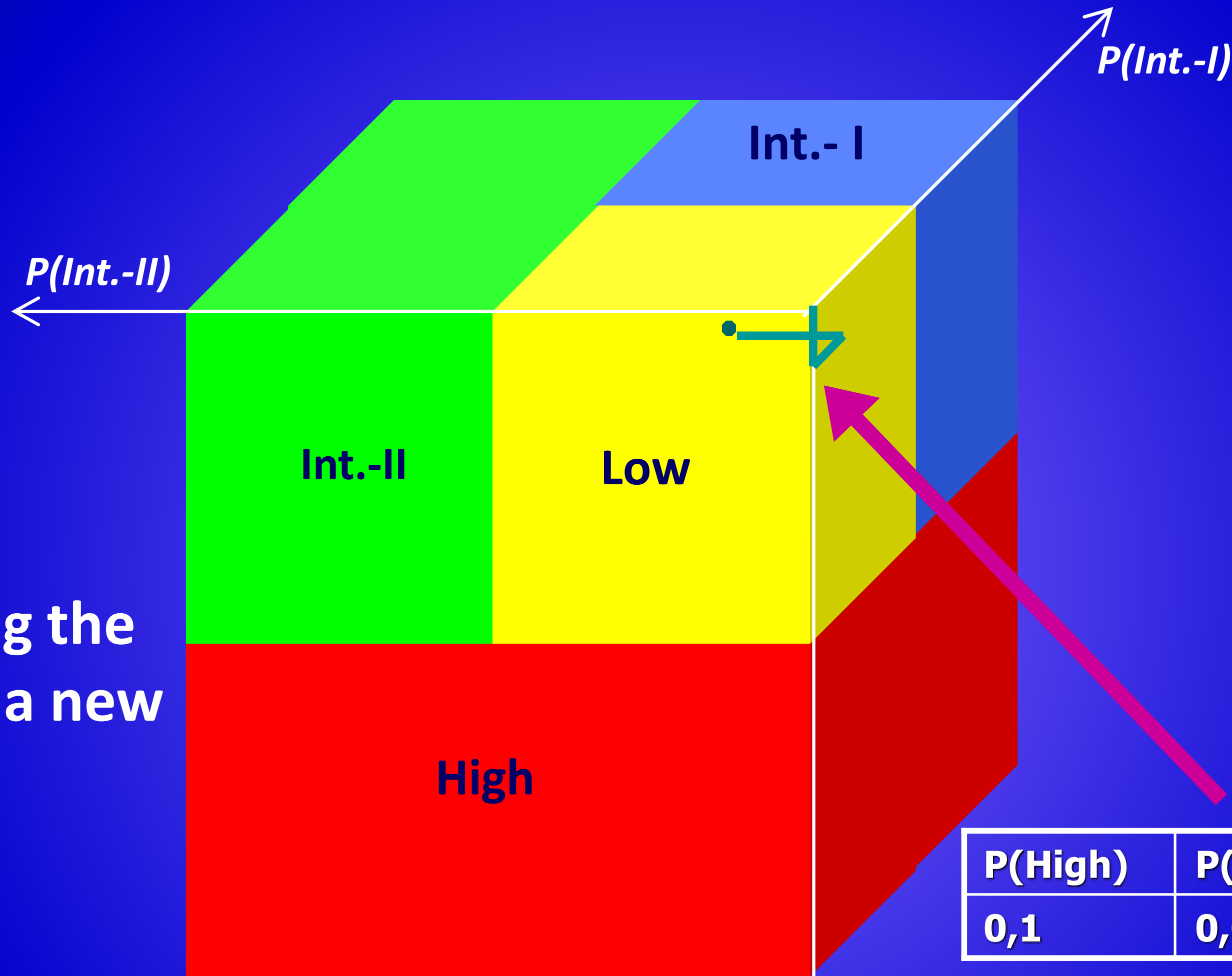
If p_{IntI} is $>= \xi$

then assign patient to *IntermediateI* profile.

Else assign patient to Low profile.

Gibert K, Rodríguez-Silva G, Annicchiarico R (2013) Post-processing: bridging the gap between modelling and effective decision-support. The Profile Assessment Grid in Human Behaviour. Mathematical and Computer Modelling 57(7-8):1633-1639, Elsevier
https://doi.org/10.1016/j.envsoft.2009.11.004

The Profile Assessment Grid



B2	3
B4	2
B9	2
S4	1
S5	2
S9	2

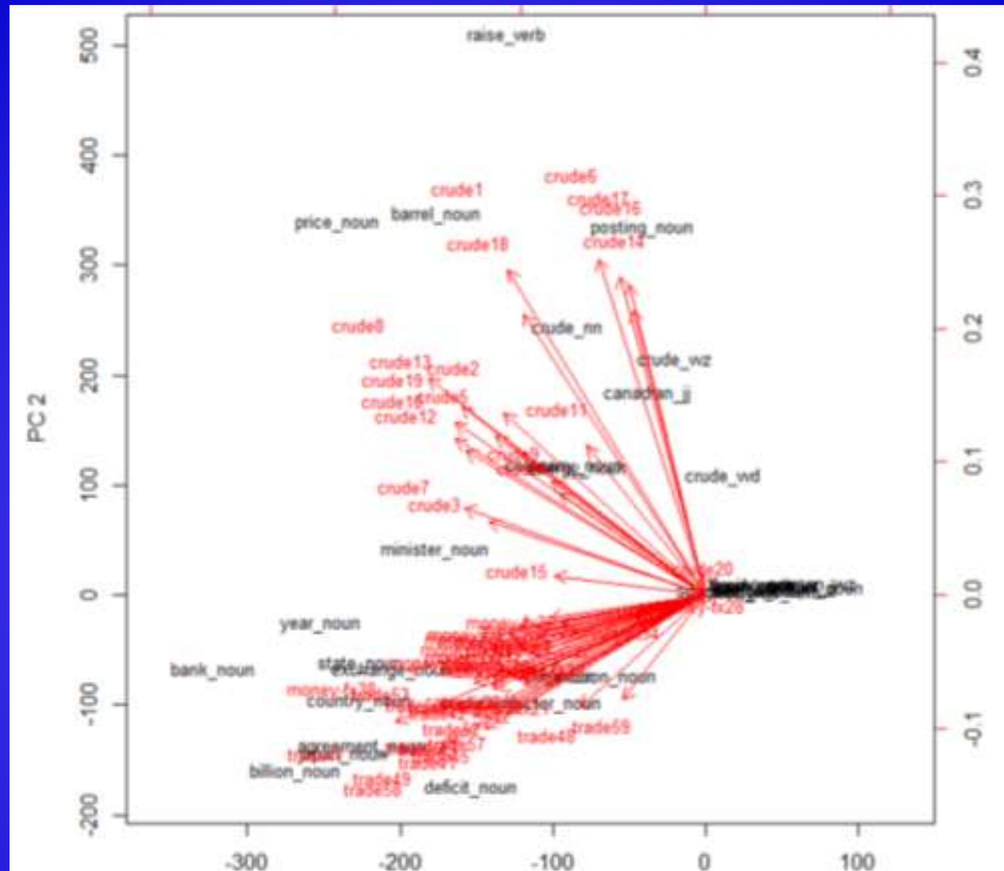
P(High)	P(Int-I)	P(Int-II)
0,1	0,07	0,20

Identifying the profile of a new patient

$$P(High) = \frac{e^{-35.93+1.70*B2+3.35*B4+3.98*B9+2.20*S4}}{(1+ e^{-35.93+1.70*B2+3.35*B4+3.98*B9+2.20*S4})}$$

Misclassification Rate (ε=0.5): 8.3%

PCA for topic modelling

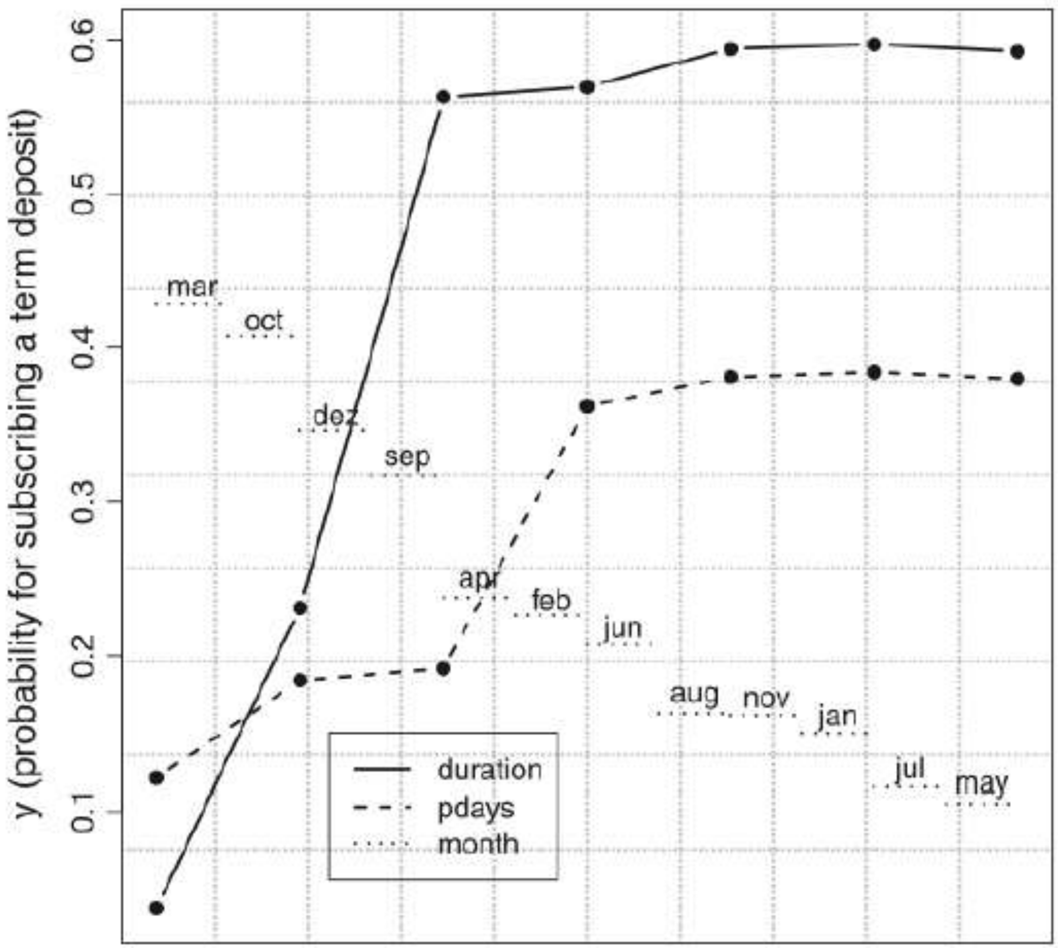


Find terms with significant contributions to axes

Generalize in the reference ontology (Wordnet by default)
Discover the latent variables (automatic interpretation of axes)

Interpreting ANN

Visualization of Input Effect: VEC curve (Bank Marketing)



Cortez, Paulo, and Mark J. Embrechts. "Using sensitivity analysis and visualization techniques to open black box data mining models." Information Sciences 225 (2013): 1-17.

Conclusions

- Explainable models required for trust and decisions
- Post-processing provide explainability
  - Visual tools: TLP/a-TLP (profiling), PAG (predictive models)
  - Conceptual: CCEC, CI-MIS (machine readable)
  - Dynamics: trajectory maps, adherence maps
- Prior knowledge transfer to models increase explainability
  - Termometers (semantics of variables, polarities)
  - Prior Knowledge Bases
  - Ontologies (semantic relations between terms)
- Language technologies play a relevant role in building these tools

Tecnologías del lenguaje para Explainable-AI y su impacto en el soporte a la decisión

Algunas aplicaciones a salud

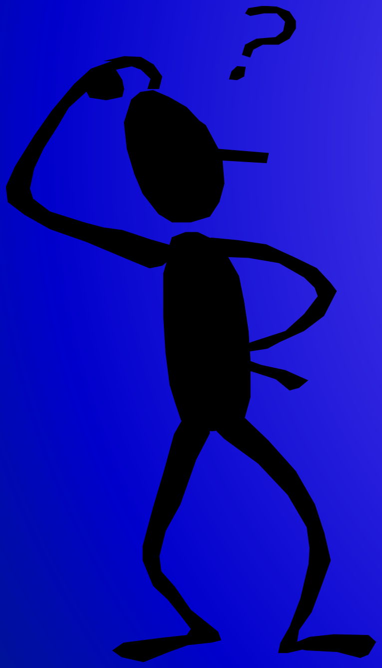
K. Gibert

karina.gibert@upc.edu

https://www.eio.upc.edu/en/homepages/karina

KEMLG-@-IDEAI: Knowledge Engineering and Machine Learning group at Intelligent Data Science and Artificial Intelligence Research Center

Universitat Politècnica de Catalunya, Barcelona



Are there any questions?...