Retrieving Sparser Fuzzy Cognitive Maps Directly from Categorical Ordinal Dataset using the Graphical Lasso Models and the MAX-threshold Algorithm

Zoumpolia Dikopoulou
Research group of Business Informatics, Faculty of Business Economics, Hasselt University
Diepenbeek Campus, 3590 Diepenbeek, Belgium
zoumpolia.dikopoulou@uhasselt.be

Elpiniki I. Papageorgiou
Faculty of Technology, University of Thessaly, Geopolis Campus Ring Road of Larisa-Trikala, 41500, Larisa, Greece
elpinikipapageorgiou@uth.gr

Koen Vanhoof
Research group of Business Informatics, Faculty of Business Economics, Hasselt University
Diepenbeek Campus, 3590 Diepenbeek, Belgium
koen.vanhoof@uhasselt.be

***Abstract*— Learning FCM models from data without any a priori knowledge and expert intervention remains a considerable problem. This research study utilizes a fully data-based learning method (the glassoFCM) for automatic design of Fuzzy Cognitive Maps (FCM) using large ordinal dataset based on the efficient capabilities of graphical lasso (glasso) models. Therefore, glasso represents its structure as a sparser graph, while maintaining a high likelihood, by producing an adjacent weighted matrix, where relationships are expressed by conditional independences. By minimizing the negative log-likelihood indicates that the model fits better to the data under the assumption that the observed data are the most likely data. The principle questioning is which of the observed concepts is the appropriate to trigger the remaining concepts in the map in order to create the glassoFCMs and obtain reasonable results. The answer derives from the FCM structure analysis based on the strength centrality indices. Moreover, the MAX-threshold algorithm based on the FCM scenario analysis is proposed in order to prune edges and retrieve sparser graphs. This algorithm shrinks the meaningless weights of the FCM, without affecting significantly the outcomes in scenario analysis. The whole approach was implemented in a business intelligence problem of evaluating the attractiveness of Belgian companies.**

Keywords—fuzzy cognitive map; graphical lasso model; MAX-threshold algorithm; ordinal data; sparser graph;

I. Introduction

FCM can be determined as a fuzzy digraph that describes the behaviour of an intelligent system in terms of concepts. These concepts are connected by signed and weighted causal relationships. Their structure is essential for modelling complex systems (utilizing existing knowledge and human experience), making decisions/predictions [1]. Through the literature, there are many methods of designing FCMs, based either on expert knowledge, on historical data or on both of them. The expert-based FCM construction relies on a group of experts and their domain knowledge to assign the concepts and the interconnections among them [21]. Generally, the data-based FCM structures require historical data for FCM learning, although new FCM models are creating from such data [2]. The combination of both methods, the Hybrid Learning Methods, produces new FCM structures, inheriting the advantages of domain knowledge and the use of historical data [2, 20].

To our knowledge, few studies have approached the data-driven modelling without expert intervention. In 1998, Schneider, Shnaider, Kandel and Chew [3] described a distance-based method for constructing FCMs based on numerical measurements (data). Each variable is represented by a numerical vector which is transformed into a fuzzy set. Then, the polarity (positive or negative) and the strength of the relations among concepts are defined based on the distances between numerical vectors. The distance-based method has been applied in many studies using a relatively small number (less than 150) of observations [31 - 35]. Nevertheless, Dikopoulou, Papageorgiou and Vanhoof [19] reported the main limitations of this method when it is applied in larger datasets (over 2,900 observations), such as a) the computation of complete weighted matrices (all concepts are connected to each other) which reduce the interpretability of these graphs and b) the calculation of very high weights between concepts (the standard deviation is close to zero) which is impossible to make decisions and policies with the FCM method. Finally, c) experts must decide the direction of causality among concepts to create the final weighted matrix. Therefore, this method is classified as a semi-automated method because it requires limited human intervention.

Another method to model FCMs from numerical data is Structural Equation Modelling (SEM) [10]. In short, SEM is a multivariate statistical technique that estimates simultaneously series of separate multiple regression equations [27]. It models the interrelationships between items with fewer variables (called latent factors) by sharing the amount of variance among a set of items and the latent variable. Moreover, SEM analyses and visualizes the structural relationship between measured variables and latent constructs [29]. First, Kang et al. [36] proposed the combination of SEM and FCM for relationship management in airline service and other studies followed the same procedure [37 - 40]. A key problem with much of the literature on SEM is a) the requirement of sufficient data to model a system since small datasets lead to non-convergence of solutions and the inability to estimate parameters and b) the under-identification of the model, especially with complex systems [30]. Moreover, not all complex problems could be explained as a causal view of latent factors [24, 25]. In many real problems, a common cause could difficult to acquire independent associations with its symptoms that explain their emergence and covariance [26]. Some problems are more appropriate to be represented as a network that models the most important and direct relations among the observed variables.

In our previous study, the *glassoFCM* methodology was proposed as a combination of glasso algorithm with the FCM method to model graphs from large datasets and simulate different decision-making scenarios [19]. The first solution of glasso investigation was adopted from the Markov Random Fields also known as undirected graphical models [9, 11]. This new approach based on generalized covariance matrices to identify the proper neighbourhood for each node [8, 9]. In order to minimize the number of edges, a sparse inverse covariance matrix was estimated using a lasso (ℓ_1) penalty [5], controlled by the Extended Bayesian Information Criterion (EBIC). In the previous paper [19], solving the lasso problem, the estimates for each edge are combined with an OR-rule and finally, the weighted adjacency matrix was defined. The outcome of that study indicated that the graphical lasso (*glasso*) model was capable of producing symmetric ($p \times p$) undirected weighted matrix (W) where p indicates the number of concepts. Due to the lack of experts' knowledge, there was no information about the direction of the edges. For this reason, it was obtained the upper triangular matrix of the observed methods in order to run different scenarios. The graphical model determined significantly sparser graph with valuable connections, higher variability of weights and consequently, reasonable values in the FCM scenarios compared to the distance-based method [19].

Due to the selection of the upper triangular matrix to run FCM scenarios, the first variable in the matrix is actually the transmitter concept. Generally, there are three types of nodes: the transmitter, the receiver and the ordinary [21, 13, 11] which depend on the out-degree and in-degree indices. The out-degree of node i is the total number of outgoing edges and is determined by the sum of the ith column of the adjacency matrix, $k_i^{out} = \sum_{j=1}^{n} a_{ji}$, where a_{ji} represents the edge of the adjacency matrix. On the other hand, the in-degree of node i is the total number of ingoing edges and is defined by the sum of the ith row of the adjacency matrix, $k_i^{in} = \sum_{j=1}^{n} a_{ij}$. Transmitter variables have a positive out-degree, k_i^{out}, and zero in-degree, k_i^{in}. Receiver variables have a positive in-degree, k_i^{in} and zero outdegree, k_i^{out}. Ordinary variables have both a non-zero in-degree and out-degree. Therefore, it is crucial to choose the appropriate node (transmitter) to be placed in the first position of the upper-triangular weighted matrix since it can influence the remaining nodes of the glassoFCM model. Another principal index of a node is the strength-centrality which is defined by the absolute summation of weights of node i and it is formalized as $s_i = \sum_{i,j=1}^{n} |w_{ij}|$, where w_{ij} represents the weighted edge of the weighted adjacency matrix [22].

In this study, we apply the glasso with the EBIC regularization method to estimate the job-satisfaction FCM model directly from a large ordinal dataset (3,262 observations). Ordinal data is a classification of categorical data and it incorporates the rank-ordering operation [28]. In addition, ordinal data are used widely by researchers to study judgments, feelings or emotions of people. A classic example of an ordinal scale on the Likert scale. A Likert scale is usually a 5 to 10-point scale with different opinions that fluctuate from one extreme to the other, such as *Very Bad*, *Bad*, *Mediocre*, *Good*, *Very Good*. Consequently, we model a glassoFCM graph according to employees' perspectives using a 5-point Likert scale on 10 observed variables.

Thus, we investigate which is the best order to place the observed nodes in the upper-triangular weighted matrix. For this reason, we rank the variables according to two instances: i) the strength-centrality of each node of the estimated FCM model (from the glassoFCM method) and ii) the average values of the variables from the initial dataset. First, these rankings are compared if they are significantly different using Kendall's tau coefficient (τ_b) [23]. In order to validate which of these two rankings is more consistent with the initial dataset, we will run different FCM scenarios (IF-THEN cases) using the reordered weighted matrices. The concept values of each scenario after the FCM inference procedure are ranked and then, are compared with the mean values of the initial dataset filtering specific variables with the maximum value according to the FCM scenario. Likewise, for these comparisons, we apply Kendall's tau correlation coefficient. Next, for retaining the most important weights of the model, a new threshold selection algorithm (the *MAX-threshold*) is proposed which shrinks very small weights to zero. Different thresholds are applied in order to obtain the appropriate sparser graph that derives approximately the same scenarios output values. A number of experiments are conducted and the results are validated. Due to space limitations, this paper focuses on our pruning algorithm and we do not compare the acquired results with other methods. In our future works, we plan to optimize the proposed algorithm including statistical measurements in order to fully compare different models.

Accordingly, the contribution of this study is two-fold: (a) comparing the rankings according to strength-centrality of the nodes and mean values of variables and validating after the FCM inference procedure using Kendal's tau correlation coefficients and (b) proposing a new algorithm, the *MAX-threshold* which shrinks very small weights of the weight adjacency matrix to zero, in order to maintain the important weights that affect substantially the inference of the FCM. The outcomes of this work provide a sparser and meaningful FCM model considering the estimated interrelationships among concepts.

The rest of the paper is organized as follows: Section 2 introduces i) the graphical lasso model in order to obtain the

FCM from 3,262 observed cases, ii) the FCM reasoning process, iii) the proposed MAX-threshold algorithm which shrinks very small weights to zero and iv) Kendall's tau correlation coefficient. The next section describes the real-world problem to be addressed. The results of the proposed methods have been displayed in Section 4 and the discussion of results is described in Section 5. Finally, Section 6 outlines the conclusions.

II. METHODOLOGIES

A. The Graphical Models

A Markov Random Field (MRF) or undirected graphical model is a network of undirected links indicating conditional dependence between two nodes; while, two nodes are independent (if there is no edge among them) after conditioning on all other variables [6]. Moreover, MRFs are belonging to families of probability distributions respecting the structure of the symmetric graph $G = (V, E)$. The graph G consists of a group of nodes $V = \{1,2,...p\}$ and edges pairs $E \subseteq V \times V$. The set of nodes $t \in V$ are connected to the nodes s, $N(s) := \{t \in V | (s,t) \in E\}$ is determined as neighborhood and the graph G consists of the collection of all neighborhoods N_s of all $s \in V$. Retrieving the undirected graphical model from ordinal data, it is necessary to estimate the prediction vector X. This indicates the computation of the mean of the conditional distribution $P(X_s \mid X_{\setminus s})$ of node X_s conditioned on all other nodes $X_{\setminus s}$ [8, 9]. In case that X_s is a Gaussian random variable, where □ = *1*, the univariate Gaussian distribution with unit variance (4), where the mean □ is a linear combination of the $N(s)$.

$$P(X_s|X_{\setminus s}) = \frac{1}{\sqrt{2\pi\sigma}} exp\left\{-\frac{(X_s-\mu_s)^2}{2\sigma^2}\right\} \quad (1)$$

In case of estimation of MRF networks from multivariate Gaussian distributions, Haslbeck and Waldorp [8, 9] generalized the covariance matrices.

A1. The algorithm of the graphical lasso model

The steps below, describing the procedure to obtain the weighted adjacency matrix directly from ordinal data of n observations and p variables. The weight of an edge w_{ij} identifying the strength of the association between two nodes after conditioning on all other variables in the network. This algorithm is known as a nodewise estimation algorithm [5, 7, 8]. Parameter λ is a penalty that governs the sparsity of the graph. In practice, λ is a vector of series of values resulting in a series of networks that vary from very dense (λ_{min}) to very sparse (λ_{max}). A λ_{min} can be chosen by multiplying some ratio (0.0001) with the λ_{max} value [4, 8].

The Algorithm I estimates the parameters of a joint distribution from observations by a series of regressions in the Generalized Linear Model (GLM). This signifies that the neighborhoods are combined to identify an estimate of graph G [18]. Therefore, for every s ∈ V, the negative *log-likelihood* $LL(\theta, X)$ and the ℓ_1-norm of the parameter vector $\|\theta\|_1$ are minimized to shrink small parameters exactly to zero:

$$\hat{\theta} = argmin_{\theta}\{LL(\theta, X) + \lambda_{\aleph}\|\theta_1\|\} \quad (2)$$

where $\|\theta_1\| = \sum_{j=1}^{J}|\theta_j|$ is the sum of absolute values of the parameters θ of length vector *J*. Then, a lower bound $\tau_{\aleph}$ [14] is applied to the size of the parameters in the true model to ensure that false and true positive rates for the lasso estimator. For the estimation of the joint distribution, the τ_n is defined as:

$$\tau_{\aleph} = s_0^* \sqrt{\log {}^{p}/_{n}} \leq s_0^* \lambda_{\aleph} \quad (3)$$

where s_0^* represents the true number of neighbours. Nevertheless, the true parameter θ^* and consequently, the number of s_0^* is unknown. Therefore, the estimated number of neighbours $\hat{s}_0$ is replaced with the estimated parameter vector to collect the estimated number of neighbours $\hat{s}_0 = \|\hat{\theta}_0\|$. The weights among two categorical variables are estimated from the pairwise interaction of $k = 2$ order [8]. For instance, the estimated weight ($\hat{w}_{st}$) among nodes s and t are derived from two parameters, the $\hat{\theta}_{s,t}$ and the $\hat{\theta}_{t,s}$ which are combined using the OR-rule (the mean of parameter estimates is calculated). Therefore, the final graph is defined for the specific value of $\lambda_{\aleph}$. Next, the Extended Bayesian Information Criterion (EBIC) is applied to estimate the fit of the model into the data:

$$EBIC_{\gamma}(\hat{\theta}) = -2LL(\hat{\theta}) + \hat{s}_0 \log n + 2\gamma\hat{s}_0 \log p \quad (4)$$

where γ is a tuning parameter ($0 \leq \gamma \leq 1$) which controls the sparsity of the graphs [12]. As the number of γ increases, the sparser the graph will be. However, Foygel and Drton [12] have proven that if γ fluctuating between 0 and 0.25 then the false positives will be decreased, without increasing the false negatives. According to Haslbeck and Waldorp [8], the computational complexity of algorithm I is $\mathcal{O}(p \ \log(2 \cdot p))$.

Algorithm I: Graphical lasso model with EBIC regularization via Neighborhood Regression

Input n × p dataset, vector $\lambda_{\aleph}$ of 100 values,

Output A sparse weighted matrix W (p × p)

Step 1: For each $\lambda_{\aleph}$

Step 2: For each s ∈ V

Step 3: Solve the lasso problem in Equation 2

Step 4: Threshold the estimates at $\tau_{\aleph}$ (Equation 3)

Step 5: Aggregate interactions with several parameters into a single edge-weight

End For

Step 6: Combine the edge-weights with the OR-rule

Step 7: Define the graph G based on the zero/nonzero pattern in the combined parameter vector

Step 8: Calculate the EBIC in Equation 4.

End For

Step 9: Choose the G that minimizes EBIC

Step 10: END

The R package '*mgm*' [9] is used to perform the weighted adjacency matrix (Table I) that described in Section III. In order to visualise the graphs, '*qgraph*" [15] is applied to the weighted adjacency matrix derived from the *mgm* function.

B. Fuzzy Cognitive Maps (FCMs)

Fuzzy Cognitive Maps are fuzzy graph structures which can be presented as an associative single-layer neural network [1]. They describe particular domains using nodes (concepts) and directed edges, which represent the causal relationships between the concepts. Each of FCM's edges is associated with a weight value that reflects the strength impact between the relevant concepts. Each of FCM's edges is associated with a weight value that reflects the strength impact between the relevant concepts. This value is usually normalized in the interval [–1, 1]. Thus, each node quantifies a degree to which the corresponding concept in the system is active at each iteration step.

For the FCM reasoning process, values of the concept C_i at time-step t are represented by the state vector C_i^t and w_{ji} is the causal weight connecting the C_j cause with the C_i effect. The state of the whole fuzzy cognitive map could be described by the state vector C^t=[C_1^t,..., C_p^t], which represents a point within a fuzzy hypercube $I^p = [0,1]^p$ that the system achieves at a certain point. The whole system with an input vector C^0 describes a time trace within a multidimensional space I^p, which can gradually converge to an equilibrium point, or a chaotic point or periodic attractor within a fuzzy hypercube. In general, a scenario assists people to identify different alternatives of the future state. Therefore, the scenario at the initial state vector can be accomplished with either all concepts or a subset of concepts being activated depending on the problem. Consequently, different policy scenarios are determined to estimate the inference of the system [1] answering the "what – if" conditions.

Equation (5) expresses the modified inference rule where concepts consider their past values and the corresponding weights when performing the inference process.

$$C_i^{(t+1)} = f\left(C_i^{(t)} + \sum_{\mathrm{j}\neq\mathrm{i}}^{N} w_{ji} \cdot C_j^{(t)}\right) \tag{5}$$

Another modified updating rule in Equation (6) was the rescale inference to avoid the conflicts emerging in the case of non-active concepts.

$$C_i^{(t+1)} = f\left(\left(2C_i^{(t)} - 1\right) + \sum_{\mathrm{j}\neq\mathrm{i}}^{N} w_{ji} \cdot (2C_j^{(t)} - 1)\right) \tag{6}$$

The iteration stops when a limit vector is reached, i.e., when $C^t = C^{t-1}$ or when $C^t - C^{t-1} \leq e$ where e is a residual (most applications is equal to 0.001).

We have been implemented in R programming language the '*fcm*' package [17] to estimate the inference of the Fuzzy Cognitive Map and accomplish the scenario analysis. This open-source package is available in CRAN and provides the opportunity to everyone to run different scenarios in their weighted matrices using the *fcm.infer* function. Six different inference rules (kosko, modified-kosko, rescale and the clamped versions of these rules) and four threshold functions (bivalent, trivalent, sigmoid and hyperbolic tangent) are provided. (For further information of how the function *fcm.infer* works, including some examples, visit the official CRAN website *https://cran.r-project.org/web/packages/fcm/vignettes/vignettes.html* or the github webpage *https:// github.com/LiaDD/Fuzzy-Cognitive-Maps-FCMs*)

C. The proposed MAX-threshold algorithm

In this study, the MAX-threshold algorithm is proposed in order to obtain a sparser as possible weighted matrix retaining only the important higher weights. This algorithm utilizes a tuning parameter ϑ (*theta*) that is important to be set to control the sparsity of the graph without affecting significantly the values of scenarios in FCM. The tuning parameter ϑ ranged from ϑ_{min} to ϑ_{max} and it is responsible to shrink smaller weights of the initial weighted adjacency matrix exactly to zero. Furthermore, the values of ϑ could be set manually, in respect of the following assumptions, $w_{min} \leq \vartheta_{min} < w_{max}$ and $w_{max} > \vartheta_{max} > w_{min}$, where w_{min} and w_{max} defined as the minimum and the maximum weight in the weighted matrix W, respectively.

Algorithm II: The MAX-threshold

Input W, $\vartheta_{\min}$, $\vartheta_{\max}$, δ, μ

Output A sparser weighted matrix W'

Step 1: Set $\vartheta = \vartheta_{\min}$

Step 2: Apply condition test

while $\vartheta \leq \vartheta_{\max}$ do

Step 3: Apply threshold ϑ to the weighted matrix. The algorithm cuts-off the weights of the matrix W that are lower than ϑ. The updated weighted matrix is stored in W'.

Step 4: Run the FCM using the activation vector C of the selected scenario and the weighted matrix W'. Save the outcome C^{ϑ}.

Step 5: Compare the outputs of the concepts' values C_i of current (ϑ) and previous run ($\vartheta - \mu$).

If $\left|C_i^{\vartheta}\right| - \left|C_i^{\vartheta-\mu}\right| \leq \pm\delta$ then

$\vartheta \leftarrow \vartheta + \mu$

go to Step 2

else

go to END

end if

Step 6: END

For the first run, the algorithm sets the threshold parameter ϑ equal to ϑ_{min} (Step 1). Next, weights w_{ij} that have been estimated lower than ϑ ($w_{ij} < \vartheta$) are removed from the weighted matrix W (Step 3). Afterwards, the FCM scenario is performed using the *fcm.infer* function [17] from the *fcm* R package (Step 4). The output values of the concepts ($C_i^{\vartheta}, i = 1,2,\dots,p$) are compared to the corresponding output values ($C_i^{\vartheta-\mu}$) of the previous weighted matrix using the same activation vector (Step 5). If the difference of every concept in the state vector C_i^{ϑ} and previous state vector ($C_i^{\vartheta-\mu}$) is lower than δ (*delta;* δ is a residual which it is set equal to 0.01) then the threshold is increased by μ (*me*). Moreover, the increment of μ is somewhat arbitrary depending on the values of weights and the researcher's decision. Nevertheless, the default value of μ is 0.005. The algorithm terminates when one of the two assumptions will occur. First, if the threshold ϑ will be higher than the maximum threshold (ϑ_{max}) as Step 2 indicates and second if the differences of the concepts' values between current and previous state (or run) are higher than μ (Step 5).

D. Kendall's tau (τ_b) correlation coefficient

Kendall's tau correlation coefficient [23] is a non-parametric measure of statistical associations ($-1 \le \tau_b \le 1$) between columns of ranked data and it is defined as:

$$\tau_b = \frac{n_c - n_d}{\sqrt{\left(\frac{n(n-1)}{2} - \sum_i \frac{t_i(t_i-1)}{2}\right) \cdot \left(\frac{n(n-1)}{2} - \sum_j \frac{u_j(u_j-1)}{2}\right)}} \quad (7)$$

Where n is the data size, n_c represents the number of concordant pairs which are the number of observed ranks below a particular rank which are larger than that particular rank, n_d denotes the number of concordant pairs which are the number of observed ranks below a particular rank which are below than that particular rank, t_i is the number of tied values in the i^{th} group of ties for the first quality and u_j is the number of tied values in the j^{th} group of ties for the second quality. If there is an absence of association between the observed variables then the correlation coefficient τ_b returns zero value, if the rankings are identical then the τ_b equals to one (positive perfect correlation) and if the τ_b equals with minus one (negative perfect correlation) indicates that the rank of one variable is increased while the rank of the second variable decreased.

III. Data DESCRIPTION

In this paper, a large dataset of 3,262 responses was used to assess the job attractiveness of Belgian companies. Attracting potential candidates is a significant issue in the recruitment process since it involves how companies compete for often scarce skills in the labor market. Therefore, a 10-item questionnaire on scale 1 (strongly disagree) to 5 (strongly agree) was developed in which employees from Informatics & Consulting sector were evaluating their preference associated with job satisfaction. The variables (V_1 to V_{10}) are defined as: "V_1: Competitive salary package", "V_2: Prospects/career opportunities", "V_3: Pleasant working environment ", "V_4: Offers long-term job security ", "V_5: Good balance (private life & work)", "V_6: Financially sound", "V_7: Offers interesting jobs (job description)", "V_8: Offers good quality of training", "V_9: Strong management", "V_{10}: Deliberately handles the environment and society". An example of a question is "*When I am looking for a job position, the job that offers good quality of training is importan*t". From this survey, the preference of participants from different gender, age, education and activity were evaluated, in order to analyze the employees' reasoning, when they are applying for a job position.

IV. RESULTS

In this section, the weighted matrix derived from the lasso graphical models and the corresponding visualization graph is presented. Since the upper triangular matrix of the weighted matrix is used for scenario analysis, the variables are prioritized according to i) the strength-centrality indices [22] obtained from the weighted matrix and ii) the average of variables obtained from the initial dataset. Moreover, the scenarios outputs are validated and the ordinal associations are compared to Kendall's tau coefficient (τ_b) [23]. Finally, the proposed MAX-threshold is applied to the initial weighted matrix, in order to obtain sparser maps that do not affect significantly the results of the FCM's scenarios.

A. Results of graphical lasso model

In this sub-section, the function *mgm* (from the R package mgm) [9] has been applied to a real dataset in order to define relations among the observed variables associated with the job attractiveness in Belgium. The symmetric weighted adjacency matrix is estimated directly from ordinal data using the lasso graphical model with EBIC regularization technique. The edge weights can be interpreted as the strength of conditional dependence between two variables (Table I).

TABLE I. THE SYMMETRIC WEIGHTED ADJACENCY MATRIX USING MGM METHOD (THE UPPER TRIAGULAR MATRIX IS HIGHLIGHTED).

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
V_1	-	0.341	0.052	0	0	0.023	0.133	0.098	0.055	0
V_2	0.341	-	0.089	0.179	0	0.051	0.218	0.106	0.093	0
V_3	0.052	0.089	-	0.067	0.140	0	0.169	0.017	0.113	0.145
V_4	0	0.179	0.067	-	0.226	0.243	0	0.044	0	0.040
V_5	0	0	0.140	0.226	-	0.025	0	0.019	0	0.146
V_6	0.023	0.051	0	0.243	0.025	-	0	0	0.300	0.006
V_7	0.133	0.218	0.169	0	0	0	-	0.356	0.061	0
V_8	0.098	0.106	0.017	0.044	0.019	0	0.356	-	0.084	0.079
V_9	0.055	0.093	0.113	0	0	0.300	0.061	0.084	-	0.149
V_{10}	0	0	0.145	0.040	0.146	0.006	0	0.079	0.149	-

B. Results of Rankings and Scenario analysis

The rankings of variables (Table II) are accomplished according to two cases: i) the strength-centrality that derived from the symmetric weighted matrix (Table I) and ii) the mean values of each variable that are derived from the dataset. It is important to mention that the order of variables differs according to Table II and the visualization of these graphs displayed in Fig.1. Nodes that are depicted closer together are strongly related. Hence, thicker edges determine stronger relations, while positive and negative relations are represented by black and red connections, respectively. Both models were designed by the lasso regularization technique (Section III), using as input the same dataset. A large number of scenarios were conducted for different initial values of concepts for the observed glassoFCM structures.

TABLE II. THE UPPER TRIANGULAR MATRIX OF WEIGHTED ADJACENCY MATRIX USING THE GLASSO ALGORITHM.

Var	Strength	Rank	Mean	Rank
V_1	0.703	7	3.206	4
V_2	1.077	1	3.199	5
V_3	0.792	6	3.044	9
V_4	0.800	5	3.246	2
V_5	0.557	10	3.088	8
V_6	0.648	8	3.384	1
V_7	0.936	2	3.141	6
V_8	0.804	4	3.106	7
V_9	0.854	3	3.225	3
V_{10}	0.565	9	2.888	10
τ_b.531, p = .156				

Correlation is significant at the 0.01** and 0.05* level.

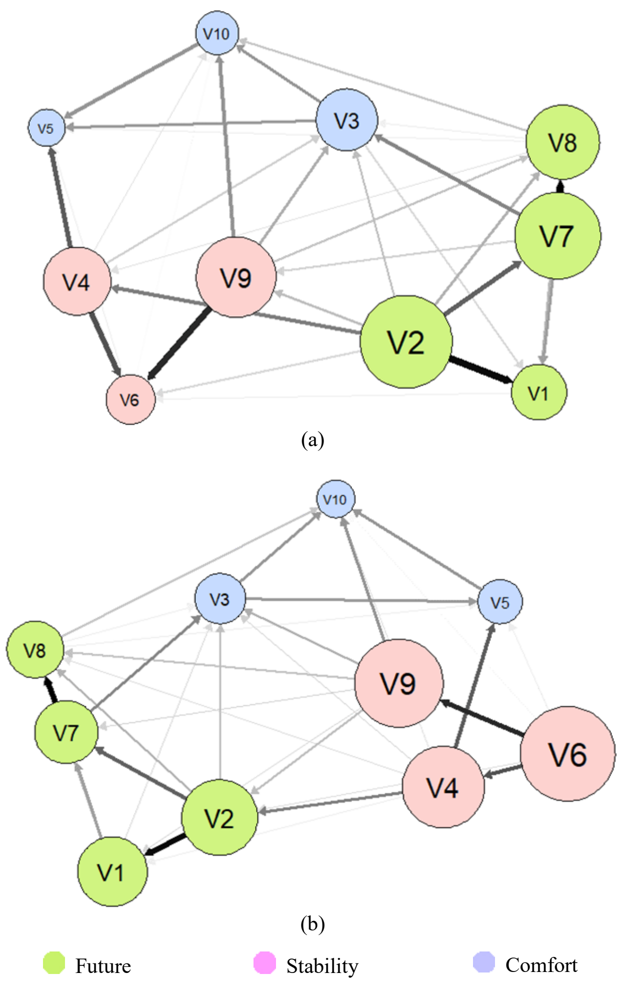

Fig. 1 Graph visualizations of two glassoFCM models estimated from the graphical lasso regularization. The orderings of the nodes are based on the strength centrality indices Fig.1(a) and the average values of the dataset Fig.1(b). The size of each node is different according to the rank place as Table II displays.

For each one of the estimated weighted matrices, we present only some representative results of rankings, considering three different scenarios. In Scenario I the transmitter concept (the concept that stands in the first place) is activated; $V_2 = 1$ (strength rank) and $V_6 = 1$ (Mean rank), respectively. In scenario II and (ii), two concepts are activated $V_2 = V_6 = 1$; while in scenario III and (iii) the three first concepts are activated, $V_2 = V_7 = V_9 = 1$ and $V_6 = V_9 = V_4 = 1$, respectively. The column "Filter.mean" obtains the rankings after filtering the values.

TABLE III. COMPARING THE RANKINGS OF FCM SCENARIOS (STRENGTH – FILTER.MEAN) AND (MEAN- FILTER.MEAN) USING THE KENDALL'S TAU COEFF.

	Scenario I		Scenario II		Scenario III			Scenario i		Scenario ii		Scenario iii	
	Strength	**Filter.Mean**	**Strength**	**Filter.Mean**	**Strength**	**Filter.Mean**		**Mean**	**Filter.Mean**	**Mean**	**Filter.Mean**	**Mean**	**Filter.Mean**
V2	**1**	**1**	**1**	**1**	**1**	**1**	**V6**	**1**	**1**	**1**	**1**	**1**	**1**
V7	3	3	4	4	**1**	**1**	**V9**	3	3	4	6	**1**	**1**
V9	8	6	8	6	**1**	**1**	**V4**	2	2	3	5	**1**	**1**
V8	4	7	5	7	5	5	**V1**	10	4	10	3	10	4
V4	5	5	6	5	9	7	**V2**	4	6	**1**	**1**	4	7
V3	6	8	7	8	7	8	**V7**	9	5	5	4	9	5
V1	2	2	3	3	4	4	**V8**	8	7	6	7	8	6
V6	7	4	**1**	**1**	6	6	**V3**	7	9	9	8	7	9
V10	10	10	10	10	8	10	**V5**	6	8	7	9	6	10
V5	9	9	9	9	10	9	**V10**	5	10	8	10	5	8
τ_b	.667*		.786**		.714*		τ_b	.111		.214		–.524	
p	.012		.006		.024		p	.677		.458		.099	

Correlation is significant at the 0.01** and 0.05* level (2-tailed).

C. Results of the proposed MAX-threshold

Different thresholds values are applied to the glassoFCM model (Fig. 1a) in order to shrink very small weights to zero and obtain an appropriate sparser graph. The validation of these thresholds is accomplished from the consistency of the output values in FCM scenarios. The minimum change among the current (C^{ϑ}) and previous ($C^{\vartheta-\mu}$) vector of concepts' values depend on residual $\delta = 0.01$ (Section III). Thus, the minimum and maximum tuning parameters are set as $\vartheta_{min} = 0.01$ and $\vartheta_{max} = 0.09$, respectively. Additionally, the increment of the threshold ϑ is set equals to 0.01 ($\mu = 0.01$).

TABLE IV. THE INCREMENT OF MAX-THRESHOLD IN FCM SCENARIO

	Scenario Analysis after MAX-threshold						
ξ	*0*	*<0.01*	*<0.02*	*<0.03*	*<0.04*	*<0.05*	*<0.06*
V_2	1	1	1	1	1	1	1
V_7	1	1	1	1	1	1	1
V_9	0.575	0.575	0.575	0.575	0.575	0.575	0.575
V_8	0.710	0.710	0.710	0.710	0.710	0.710	0.710
V_4	0.596	0.596	0.596	0.596	0.596	0.588	0.588
V_3	0.639	0.639	0.636	0.636	0.636	0.636	0.636
V_1	0.731	0.731	0.731	0.731	0.731	0.731	0.726
V_6	0.575	0.575	0.575	0.570	0.570	0.568	**0.543**
V_{10}	0.552	0.551	0.551	0.551	0.551	0.547	0.547
V_5	0.553	0.553	0.549	0.547	0.547	0.545	0.545
Edges	32	31	29	27	27	25	22
Density	71.1%	68.9%	64.4%	60.0%	60.0%	55.6%	51.1%

Table IV demonstrates the increment of the threshold in the weighted matrix where the variables are placed according to the higher strength indices (Section IV.B). Furthermore, Table IV shows which threshold affects the output values of the selected scenario (the activated concepts are $V_2 = V_7 = 1$).

V. Discussion of Results

Table I presented the weighted adjacency matrix estimated from the graphical lasso algorithm. Each node was corresponded to a variable associated with the job satisfaction problem, as described in Section II. All weights in the weighted matrix were signed as positive and varied between $w_{min} = 0.006$ and $w_{max} = 0.356$; while, 36 of 90 edges were estimated as zero, indicating that there was no relation between the relevant nodes. For instance, according to Table I, employees in the Informatics sector consider that there was no association between "Good balance between private life & work (V_5)" and "Offers interesting jobs (V_7)". Thus, 17 edges have been distinguished as very-very low ($w_{ij} < 0.1$), 12 links as low ($0.1 \leq w_{ij} < 0.25$) and 3 edges as a medium.

For accomplishing FCM scenarios, first, it was necessary to transform the symmetric weighted matrix graph (Table I) to asymmetric in order to derive the glassoFCM graphs. Due to the lack of experts' knowledge, there was no information related to the direction of the edges between the relevant nodes. For this reason, it was obtained the upper triangular matrix which in fact, the first variable of this matrix triggered the remaining variables when it was activated. We were searching for the proper variables to activate the remaining variables or in other words for the appropriate ranking of nodes to insert in the upper weighted matrix. As was already reported, the rankings of the observed variables (Table II) were performed according to i) the centrality strength derived from the symmetric weighted matrix of Table I and ii) the mean values of each variable that were derived from the dataset. A Kendall's tau-b correlation was run to determine the relationship between strength-centrality ranking towards mean ranking. The τ_b correlation indicates that there was a moderate, positive relationship that was not statistically significant ($\tau_b = .531, p = .156$).

Afterwards, two new glassoFCM models were created regarding strength-centrality and mean values. It was essential to highlight that despite the reordering of the variables, the weights between the relevant nodes were invariable. The correspondent graphs of the upper triangular matrices were illustrated in Fig.1(a) and Fig.1(b). Furthermore, Fig.1(a) and Fig.1(b) were presented the clusters determined by the K-means method. To estimate the clusters of variables using the k-means method, "cluster" R package [16] had been applied to the dataset. Three clusters were estimated: future (green color): V_1, V_2, V_7, V_8, stability (pink color) V_4, V_6, V_9 and comfort (blue color) V_3, V_5, V_{10}. Next, different FCM scenarios were accomplished using the *fcm.infer* function ('*fcm*' R package [17]) to identify which of the weighted matrices were the most preferable and most consistent to the dataset. The validation derives from the dataset in which the variables (that were activated in FCM scenarios) were 'filtered' with the highest values ('5') and the mean values of the inactive variables were calculated. The results of scenarios and the 'filtered means' were ranked and compared using Kendall's tau (τ_b) coefficient (Table III). The similarities between Strength and Filter.mean were statistically higher and significant ($\tau_b > .677, p \leq .024$) comparing to Mean and Filter.mean ($-.524 \leq \tau_b \leq .214, .099 \leq p \leq .677$). This signifies that the reordering based on Strength centrality indices was more consistent with the original dataset.

As was previously stated, the glassoFCM model was based on lasso regularization which estimates 17 very-very low weights (Table I). Therefore, it was necessary to investigate which of these weights would not cause significant differences in the output of the FCM simulation if they shrunk to zero. In this regard, the MAX-threshold was applied to one scenario to demonstrate how the algorithm works. Therefore, in this scenario, the first two variables were activated ($V_2 = V_7 = 1$). The initial upper triangular weighted matrix was determined to calculate the output of the lasso graphical model in which the variables were ordered by the strength-centrality indices. The density of the graph without trimming the edges was 71.1% (Table IV) and the total number of non-zero edges was 32. As the proposed algorithm increases the threshold ϑ, the density was gradually decreased. In this example, the algorithm stops when ϑ minimizes the weights of the weighted matrix under 0.06 (Table IV) because the difference of the concept V_6 was higher than residual ($\delta = 0.01$), $|V_6^{0.06}| - |V_6^{0.05}| > \pm\delta$. Consequently, the appropriate graph that was not significantly influenced the results of FCM scenario was determined with a density of 55.6% or 25 non-zero edges (a 15.5% decrease of density from the initial FCM). In other words, for the observed FCM, the weights under 0.05 were removed (set equal to zero in the weighted matrix) since they did not significantly change the output values of the system. For reasons of simplicity, the final threshold (the final graph density's decrement) would be determined as the minimum threshold was satisfied for all accomplished scenarios.

VI. Conclusion

To sum up, this paper contributed to the existing literature by presenting an automatic data-driven learning approach, the glassoFCM method to retrieve sparser Fuzzy Cognitive Maps directly from large ordinal dataset using the graphical lasso algorithm. Reordering the variables according to strength centrality indices, the produced glassoFCM structure derived from the upper triangular weighted matrix provided a meaningful FCM model considering the estimated interrelationships among concepts, indicating better planning and decisions. Moreover, the main contribution of this study was the proposed MAX-threshold algorithm which was able to shrink spurious edges to zero without affecting significantly the values of FCM scenarios in the initial weighted matrix. The aim of the proposed algorithm was to obtain the FCM structure in which as few connections as possible were selected to parsimoniously explain different decisions. Moreover, the proposed MAX-threshold algorithm could be useful in complex systems of hundreds of concepts by pruning meaningless weight edges. This indicates that fewer edges among concepts could perform decisions through FCM scenarios. Even the results of this research study are useful, further work is needed to exploit graph-based methods for automatic FCM design from large categorical data sets.

REFERENCES

[1] B. Kosko, “Fuzzy cognitive maps”, International Journal of Man-Machine Studies, 24 (1), pp. 65–75, 1986.

[2] E. Papageorgiou, J. L. Salmeron. “Methods and Algorithms for Fuzzy Cognitive Map-based Modeling”, Intelligent Systems Reference Library 54, pp. 1-28, 2014.

[3] M. Schneider, E. Shnaider, A. Kandel and G. Chew, “Automatic construction of FCMs”, Fuzzy Sets and Systems, 93 (2), pp. 161–172, 1998.

[4] R. Tibshirani, “Regression shrinkage and selection via the lasso”, Journal of the Royal statistical society, B 58, pp. 1436–1462, 2006.

[5] D. Witten, J. Friedman and N. Simon, “New insights and faster computations for the graphical lasso”, Journal of Computational and Graphical Statistics 20(4), pp. 892-900, 2011.

[6] G. R. Grimmett, “A theorem about random fields”, Bull. Lond. Math. Soc. 5, pp. 81–84, 1973.

[7] S.L. Lauritzen, “Graphical Models”, Oxford Univ. Press, New York, 1996.

[8] J.M.B. Haslbeck, L.J. Waldorp, “Structure estimation for mixed graphical models in high-dimensional data”, Anals of Applied Statistics, 2015. (in press) ArXiv: 1510.05677.

[9] J.M.B. Haslbeck, L.J. Waldorp, “How well do Network Models predict Future Observations? On the Importance of Predictability in Network Models”, Behavior Research Methods, 50, pp. 853-861, 2018.

[10] J. Wright, S.: On the nature of size factors. Genetics 3, 367–374, 1918.

[11] E. Harary, F., Norman, R.Z., Cartwright, D., “Structural Models: An Introduction to the Theory of Directed Graphs”, John Wiley & Sons, New York, 1965.

[12] R. Foygel and M. Drton, “Bayesian model choice and information criteria in sparse generalized linear models”, 2011. arXiv preprint arXiv:1112.5635.

[13] M. Bougon, K. Weick and D. Binkhorst, “Cognition in organizations: an analysis of the Utrecht Jazz Orchestra”, Admin. Sci. Quart. 22, pp. 606–639, 1977.

[14] P.L Loh and M.J. Wainwright, “Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses”, The Annals of Statistics 41, pp. 3022-3049, 2013.

[15] S. Epskamp, A.O.J. Cramer, L.J. Waldorp, V.D. Schmittmann, d. Borsboom, “qgraph: Network Visualizations of Relationships in Psychometric Data.” *Journal of Statistical Software*, 48(4), 1–18 , 2012.

[16] M. Maechler, P. Rousseeuw,C A. Struyf, M. Hubert, K. Hornik, “cluster: Cluster Analysis Basics and Extensions”, R package version 2.1.0, 2019.

[17] Z. Dikopoulou and E. Papageorgiou, “Inference of Fuzzy Cognitive Maps (FCMs)”, 2017. https://cran.r-project.org/web/packages/fcm/vignettes/vignettes.html.

[18] N. Meinshausen and P. Buhlmann, “High-dimensional graphs and variable selection with the lasso”, The Annals of Statistics, 34(3), pp. 1436–1462, 2006.

[19] Z. Dikopoulou, E. Papageorgiou, V. Mago, K. Vanhoof, “A new approach using Mixed Graphical Model for automated design of Fuzzy Cognitive Maps from ordinal data”, IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2017

[20] E.I. Papageorgiou, M.F. Hatwágner, A. Buruzs, L.T. Kóczy, “A Concept Reduction Approach for Fuzzy Cognitive Map Models in Decision Making and Management“, Neurocomputing Journal, Special Issue on Fuzzy Cognitive Maps, vol 232, pp 16–33, 2017.

[21] E.U. Özesmi, S.L. Özesmi, “Ecological models based on people’s knowledge: a multi-step fuzzy cognitive mapping approach”, Ecological Modelling, 176(1-2), pp. 43–64, 2004.

[22] M.E.J. Newman, “Analysis of weighted networks”, Physical Review E, 70(5), 2004.

[23] M. Kendall, J. D. Gibbons, “Rank Correlation Methods”, Charles Griffin Book Series (5th ed.). Oxford: Oxford University Press, 1990 [First published 1948].

[24] G. Costantini, S. Epskamp, D. Borsboom, M. Perugini, R. Mõttus, L.J. Waldorp, and A.O.J. Cramer, “State of the aRt personality research: A tutorial on network analysis of personality data in R”, Journal of Research in Personality, 54, pp. 13–29, 2015.

[25] A.O.J. Cramer, S. van der Sluis, A. Noordhof, M. Wichers, N. Geschwind, S.H. Aggen, “Dimensions of normal personality as networks in search of equilibrium: You can’t like parties if you don’t like people European”, Journal of Personality, 26 (4) pp. 414-431, 2012. doi:10.1002/per.1866

[26] R.J. McNally, “Can network analysis transform psychopathology?” Behaviour Research and Therapy, 86, pp. 95-104, 2016.

[27] J.F Hair, W.C. Black, B.J. Babin, and R.E. Anderson, “Multivariate Data Analysis” (new int. ed.). Harlow: Pearson Education, 2014.

[28] S.S. Stevens, “On the Theory of Scales of Measurement”, Science. New Series. 103 (2684) pp. 677–680, 1946.

[29] J. C. Anderson, and D.W. Gerbing, “Structural equation modeling in practice: A review and recommended two-step approach”, Psychological Bulletin, 103(3), pp. 411-423, 1988.

[30] J.P. Craiger, D.F. Goodman, R.J. Weiss, and A. Butler, “Modeling organizational behavior with Fuzzy Cognitive Maps”, Int. J. Comput. Intell. Org., vol. 1, pp. 120–123, 1996.

[31] S. Ahmadi, E. Papageorgiou, C.-H. Yeh, and R. Martin, “Managing readiness-relevant activities for the organizational dimension of ERP implementation”, Computers in Industry, 68, pp. 89–104, 2015.

[32] M. Arvan, A. Omidvar, and R. Ghodsi, “Intellectual capital evaluation using fuzzy cognitive maps: A scenario-based development planning”, Expert Systems with Applications, 55, pp. 21–36, 2016.

[33] M. Zarrin and A. Azadeh, “Mapping the influences of resilience engineering on health, safety, and environment and ergonomics management system by using Z -number cognitive map”, Human Factors and Ergonomics in Manufacturing & Service Industries, 2018.

[34] E. Bakhtavar and Y. Shirvand, “Designing a fuzzy cognitive map to evaluate drilling and blasting problems of the tunneling projects in Iran”, Engineering with Computers, 2018.

[35] C. D. Stylios, E. Bourgani and V. C. Georgopoulos, “Impact and Applications of Fuzzy Cognitive Map”, 2020, pp. 229-246. Methodologies Kosheleva, O., Shary, S. P., Xiang, G., & Zapatrin, R. (Eds.). (2020). Beyond Traditional Probabilistic Data Processing Techniques: Interval, Fuzzy etc. Methods and Their Applications. Studies in Computational Intelligence.

[36] I. Kang, S. Lee, J. Choi, “Using fuzzy cognitive map for the relationship management in airline service”, Expert Systems with Applications, 26 (4), pp. 545-555, 2004.

[37] S. Lee, B.G. Kim and K. Lee, “Fuzzy cognitive map-based approach to evaluate EDI performance: a test of causal model”, Expert Systems with Applications, 27(2), pp. 287–299, 2004.

[38] S. Lee, H. Ahn, “Fuzzy cognitive map based on structural equation modeling for the design of controls in business-to-consumer e-commerce web-based systems”, Expert Systems with Applications, 36 (7) , p. 10447-10460, 2009.

[39] S.-C. Huang, S.-L. Lo and Y.-C. Lin, “Application of a fuzzy cognitive map based on a structural equation model for the identification of limitations to the development of wind power”, Energy Policy, 63, 851–861, 2013.

[40] S. Galehbakhtiari, and T. Hasangholi pouryasouri, “A hermeneutic phenomenological study of online community participation”, Computers in Human Behavior, 48, pp. 637–643, 2015.