From Undirected Structures to Directed Graphical Lasso Fuzzy Cognitive Maps using Ranking-based Approaches

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, GR41500, Larissa, 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*— Fuzzy cognitive maps (FCMs) have gained popularity within the scientific community due to their capabilities in modelling and decision making for complex problems. However, learning FCM models automatically from data without any expert knowledge and/or historical data remains a considerable challenge. For our research, we use the estimated weight matrix from the graphical lasso (glasso) method with the EBIC regulation technique. Particularly, the glasso is a technique originated from machine learning which is used to model a problem by learning the weight matrix directly from a dataset. Moreover, the relationships are expressed by conditional independence among two nodes after conditioning on all the other nodes of the graph. However, the challenging task in this study is the investigation of the suitable transformation of the weight matrix from a symmetric matrix to asymmetric in order to determine the directions of the edges among the concepts and construct the glassoFCM model. For this reason, statistical comparisons are applied to examine if there are significant differences in the value of the output concept when the input concepts are rearranged according to four different cases. The whole approach was implemented in a business intelligence problem of evaluating the willingness of the employees to work in Belgian companies.**

Keywords—fuzzy cognitive map; glassoFCM method; graphical lasso models; data-driven approach; ordinal data;

I. Introduction

As data modelling is referred to the process of creating a model from a dataset that can be useful for recognizing patterns, making decisions and/or predictions and control systems [7]. In other words, data modelling is a visual representation of a problem which often is visualized as a graph. Every graph is composed of two important elements (or vertices or nodes) and edges [8]. There are many methods to estimate a model from a single dataset. The choice of models is depending on the variety of variables, the sample size of data and the available knowledge (if any) from experts [7]. After data modelling (learning the weights between concepts) from experts or from available historical data, it is essential to perform policy scenarios through a simulation method which is known as decision making [27]. One of the well-known methods of modelling and simulating a system is the Fuzzy Cognitive Map (FCM) [1]. FCM is classified as soft computing technique which is trying to mimic humans' reasoning and decision making [1, 2, 3]. It combines the desirable properties of fuzzy logic and neural networks. FCMs have been applied in many scientific areas such as business, social and political sciences, ecological and environmental management, engineering, information technology, robotics, expert systems, medicine, neuroscience, education and so on [3, 5, 10, 11, 12, 26, 48].

Generally, FCM models are categorized into three types: i) *manual*, ii) *semi-automated* and *automated* models. Manual FCMs are produced by experts manually, semi-automated FCMs are constructed by a relatively expert intervention and automated FCMs are estimated from historical data. Expert-based models require relatively fewer data and lead to simpler and sparser interpretable models [25, 26]. On the other hand, data-driven models require a sufficient amount of training historical data; however, they are more expressive since they are able to uncover unexpected patterns of the data comparing to expert-based models [3, 20, 25, 26]. Mainly, the FCM learning methods which construct and train an FCM fall into three categories: i) the *adaptive-based* (usually adapted from the Hebbian law), ii) the *population-based* and iii) the *hybrid* learning algorithms [3, 20, 25, 26]. The adaptive-based algorithm is a semi-automated method since it updates the weights of the connection matrix obtained from experts. Population-based (evolutionary) learning algorithms is an automated method which uses solely historical data to estimate the weights between concepts. The main goal of this algorithm is the estimation of an appropriate model that miming the input data. Finally, the hybrid approach combines the adaptive and population-based learning methods in which the knowledge of an expert and historical data are simultaneously used.

Therefore, most of the studies have been focused on semi-automated and automated methods to learn FCM based on historical data. However, few approaches have been applied to define the relations and the weights among concepts (nodes) from cross-sectional data (not historical data). In general, a cross-sectional study is defined as an observational study in which multiple variables can be studied at a given point in time [28]. Schneider, Shnaider, Kandel and Chew [4] proposed a distance-based method for constructing FCMs based on numerical measurements. The variables are represented as numerical vectors which are transformed into fuzzy vectors. Next, the determination of positive or negative relations among concepts are defined. Afterward, the strength (the weight) of each edge is calculated based on the distances between numerical vectors. At last, experts must decide the direction of causality between concepts indicating the final FCM weighted matrix. Consequently, this method is categorized as a semi-automated method since limited human intervention is required [12]. In the last decade, many studies have been used the distance-based method to model FCMs using small datasets (up to 150 observations) [29 - 33]. However, Dikopoulou, Papageorgiou and Vanhoof [12] noted that this method was incapable to construct efficient FCM models from larger datasets (over 2,900 observations) due to small variability of the weighed edges (all weights of the FCM were calculated very strong) and high density of the weighted matrix (all concepts are connected to each other). As a consequence, the interpretability of that FCM was reduced and the decision-making was impossible to be occurred using the FCM inference procedure.

Nevertheless, an automated FCM model estimation from (cross-sectional) data assessing the direct relationships between concepts was a major challenge. In order to bridge the gap, in 2017, we proposed the automated data-driven method, the *glassoFCM* [12] to model an FCM (concerning the job satisfaction) from 2,903 categorical ordinal observations without expert intervention and/or usage of historical data. The glassoFCM is a combination of two methods, the *graphical least absolute shrinkage and selection operator* (glasso) with *EBIC* (*Extended Bayesian Information Criterion*) regularization for data modelling and the FCM simulation for decision making. Results had shown that glasso method had estimated a sparser graph with a higher standard deviation on edge weights comparing to a distance-based method. Therefore, the glassoFCM model was easier to interpret and the decisions/policies through FCM simulations were meaningful. In 2020, we extended our primary research work proposing the *MAX-threshold* algorithm to cut-off (prune) more small spurious weights of the glassoFCM model without affecting the concepts' values after the FCM inference procedure [13]. Our findings have shown that the density of the automated data-driven FCM was decreased from 71.1% to 55.6% indicating that fewer edges among concepts could perform decisions through FCM scenarios.

Originally, the lasso is a regression analysis method that is used for variable selection and regularization in order to improve not only the prediction accuracy but also the interpretability of the estimated model [6]. Specifically, the ℓ_1 regularization of the lasso algorithm adds a penalty equal to the absolute value of the magnitude of coefficients (ℓ_1-norm). Consequently, some spurious (small) coefficients are shrunk exactly to zero. In fact, the tuning parameter λ controls the sparsity of the ℓ_1 penalty and it is varied between 0 and $+\infty$. The larger the penalty is applied, the further estimates are shrunk towards zero. Accordingly, the value of the appropriate tuning parameter must be chosen (using for instance, the k-fold cross-validation [34]) to select the best fitting model and acquire a more accurate estimate of the model's test error rate [6]. Graphical lasso [9, 37] is an algorithm for learning the undirected structure of a Gaussian Graphical Model (GGM) [35]. In general, graphical models became very popular since they were able to model complex problems and estimate the interconnections among observed variables in different scientific fields such as: statistics, machine learning, neuroscience, psychology, biology and business [12, 13, 36]. The glasso method utilizes the ℓ_1 regularization parameter to control the sparsity of the precision matrix. Consequently, it estimates a group of networks [14] ranging from a fully connected network (λ_{min}) to an empty network (λ_{max}). Therefore, the best network out of this range of networks must be selected by optimizing the fit of the network to the data. According to previous studies [16, 38, 39], minimizing the Extended Bayesian Information Criterion (EBIC) [14, 15], the true network structure was estimated indicating that high specificity has occurred. Moreover, the EBIC added an extra penalty, the hyperparameter γ (gamma) to control (additionally) the sparsity of the model [16, 14].

For the record, the lasso regression algorithm was introduced, as an evolutionary algorithm (named as $\text{LASSO}_{\text{FCM}}$) to learn sparser large-scale FCMs from historical data [40]. Results have shown that $\text{LASSO}_{\text{FCM}}$ was accurately learned sparser FCMs with high accuracy comparing to other evolutionary methods. This primary approach was classified as a semi-automated procedure since a user-defined constant was required to balance the sparsity and measurement error. Later, Wu and Liu [41] proposed an initialization operator based on the $\text{LASSO}_{\text{FCM}}$ for evolutionary algorithms to learn sparse FCMs. In 2019, an automated method was proposed. Specifically, the multitasking multiobjective memetic algorithm for learning FCMs (MMMA-FCMs) [43] was a combination of the $\text{LASSO}_{\text{FCM}}$ using the decomposition strategy into a multiobjective evolutionary algorithm [42] to learn large-scale FCMs. The results have shown high accuracy and efficiency in the learning procedure testing into different numbers of nodes, densities and activation functions.

However, in this paper, our research interest is solely focused on the FCM learning procedure from cross-sectional datasets (not from historical data) including 11 variables (ten input variables and one output). As reported above, glasso method produces symmetric undirected graphs. Generally, in order to transform each model into the glassoFCM model, it is necessary to determine the direction of the edges among concepts. Rearranging the observed concepts in the symmetric weight matrix according to different rankings, we obtain the upper triangular matrix. Then, we examine the values of the observed output concept if they are significantly different (after the usage FCM inference procedure) using a one-way ANOVA method [23]. The variables are ranked according to four cases:

i) the strength-centrality of the concepts ii) the average values of the variables (from the initial dataset), iii) the random order and iv) the inverse strength-centrality ranking. The strength-centrality of node i is defined by the absolute summation of the weights that are linked with the node i [8]. Consequently, four glassoFCM models are constructed, each one for each ranking case. Moreover, due to the selection of the upper triangular matrix, the last variable (the decision output concept) in the matrix is considered as the receiver concept. Mainly, there are three types of a node: the transmitter, the receiver and the ordinary node [44, 45] depending on the in-degree and out-degree indices. 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; while, 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. Transmitter nodes have a positive out-degree and zero in-degree, receiver nodes have positive in-degree and zero out-degree and ordinary nodes have both non-zero in-degree and out-degree indices.

The aim of this study is to examine if the observed rearrangements play a determinant role or not to the values of the output concept under 60 scenarios policies (15 scenarios for each glassoFCM model). It is important to emphasize that in the FCM simulation process, only the results of the decision output concept ("*Willingness to work*") are inspected. This implies that the focal concept is always placed in the last (p^{th}) position of the upper-triangular weight matrix; where p represents the total number of the concepts. The remaining 10 variables represent the input concepts associated with job-satisfaction features.

The rest of the paper is organized as follows: the next Section describes the glassoFCM methodology, Section 3 introduces the real-world job-satisfaction problem considering 11 features. Section 4 presents the results of glasso model, the results of four rankings, the results of FCM scenario analysis applying to the produced glassoFCM models and the statistical results of comparison using one-way ANOVA. The discussion of the results is described in Section 5. Finally, Section 7 outlines the conclusions.

II. METHODOLOGIES

A. The graphical lasso with the Fuzzy Cognitive Map (glassoFCM) method

The *glassoFCM* methodology [12, 13] was proposed as a combination of glasso algorithm with the FCM method to model graphs from large datasets and simulate different decision-making scenarios. Below, the glasso method, the FCM method and the glassoFCM algorithm are described.

A1. The graphical lasso with the EBIC regularization

Principally, an undirected graphical model is a network of undirected links indicating conditional dependence among two nodes. Such models are members of probability distributions respecting the structure of the symmetric graph *G* = (*V*, *E*). Mainly, two nodes are considered as independent (if there is no link among them) after conditioning on all other variables [47]. Therefore, the glasso method is used to estimate the weighted adjacency matrix directly from cross-sectional data (containing ordinal values) of n observations and p variables. This algorithm is known as a nodewise estimation algorithm [35, 36] since it identifies the strength of the association (w_{ij}) between two nodes after conditioning on all other variables in the network. Moreover, vector λ consists of a series of λ values which are able to control the sparsity of the graph. Therefore, a group of networks [14] is estimated ranging from a very dense network (λ_{min}) to a very sparse (λ_{max}). In order to estimate the graph G, the neighborhoods of all p nodes are combined by estimating the parameters of a joint distribution from observations by a series of regressions in the Generalized Linear Model (GLM) [36]. Consequently, 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 produced parameters $\hat{\theta}_{s,t}$ and the $\hat{\theta}_{t,s}$) among nodes *s* and *t* are combined using the OR-rule (the mean value is calculated) to estimate the weight ($\widehat{w}_{st}$).As a result, the final graph is specified for the specific value of $\lambda_{\aleph}$. Afterward, the Extended Bayesian Information Criterion (EBIC) is applied to estimate the fit of the model into the data [15]:

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

where γ is a tuning parameter ($0 \leq \gamma \leq 1$) which controls the sparsity of the graphs [16, 36, 38, 39]. As the number of γ increases, the sparser the graph will be. However, Foygel and Drton [16] 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 [36], the computational complexity of algorithm I is $\mathcal{O}(p\ \log(2 \cdot p))$.

A2. The Fuzzy Cognitive Map (FCM) method

A FCM is a graphical representation of a directed weighted map consisted of concepts and weighted edges [1, 2, 3]. Each concept represents a certain characteristic of the system which are connected to each other using directed weighted edges, expressing the causal-effect relationships among concepts. Each concept carrying a value revealing the degree of activation of the concept in the system at a particular time and it is indicated by $C_i, i = 1,2, \ldots, p$; where p denotes the total number of variables. Each weight w_{ij} of the weighted matrix W indicates the strength of the association between C_i and C_j taking value in the range -1 to 1. Specifically, w_{ij} measures how much C_i affects C_j ($C_i \rightarrow C_j$). There are three possible types of causal relationships among concepts C_i and C_j: positive ($w_{ij} > 0$), negative ($w_{ij} < 0$) and no causality ($w_{ij} = 0$).

The FCM method was primarily proposed by Kosko [1] for understanding, modelling and simulating systems with numerous interconnections between important components. FCMs adapt the desirable properties of fuzzy logic and neural networks like the ability to represent the structured knowledge of the system and the computation of the inference using a numeric matrix operation instead of straightforward IF-THEN rules [2]. The initial vector C^0, at time-step 0 ($t = 0$), includes the values of p concepts: $C^0 = [C_1^0, C_2^0, \dots, C_p^0]$. The modified FCM inference rule (5) was the rescale inference to avoid the conflicts emerging in the case of non-active concepts.

$$C_i^{(t+1)} = f\left(\left(2C_i^{(t)} - 1\right) + \sum_{\substack{j \neq i}}^{N} w_{ji} \cdot \left(2C_j^{(t)} - 1\right) \right) \quad (5)$$

Lastly, $f(\cdot)$ is the threshold or alternatively a transformation function that constrains the results of the initial vector C into the range between $[0, 1]$. In most studies, the sigmoid function (6) is widely applied to obtain the inference of the system [20, 3].

$$f(x) = \frac{1}{1 + e^{-\lambda x}} \quad (6)$$

This process continues until i) the system converges in a fixed equilibrium point in which the difference between two subsequent values of the outputs must be equal, $C_i^t = C_i^{(t+1)}$ or lower to the residual ε (epsilon; in most of the cases is equal to 0.001), $C_i^t - C_i^{(t+1)} \leq \varepsilon$; ii) a limited cycle is reached or iii) a chaotic behavior is revealed.

A3. The glassoFCM algorithm

Below, the modelling and the simulation of the glassoFCM is presented in 16 steps:

Algorithm I: The glassoFCM with EBIC regularization via Neighborhood Regression

Input Dataset ($n \times p$, n: observations, p: variables), $\lambda_{\aleph}$ (vector of 100 values), $t_{\max}$ (maximum number of steps), C^0(initial vector, $1 \times p$), ε (residual error)

Output A converged vector $C^t = [C_1^t, C_2^t, \dots, C_p^t]$)

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

Step 2: **For** each node $s \in V$

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 graph G that minimizes EBIC (the symmetric weight W matrix is estimated)

Step 10: Get the upper-triangular matrix of W

Step 11: **While** $t \leq t_{\max}$ AND ($V_i^t \neq V_i^{(t+1)}$ or $V_i^t - V_i^{(t+1)} > \varepsilon$) **do**:

Step 12: Apply the inference rule (5) to update the initial Vector C^0.

Step 13: Use the threshold function (6) to reduce the unbounded input of V to a strict range that threshold controls.

Step 14: Save the updated $C^{(t+1)}$ (consider that this vector will be the initial vector t for the next iteration).

Step 15: Increase step t ($t \leftarrow t + 1$)

End while

Step 16: Return the converged vector C^t

END

III. Data Description

Job satisfaction is the evaluation of various concepts that can describe how employees think about their job [19]. Improving job satisfaction features, more employees will be satisfied; therefore, their loyalty will be increased and workers will be 'ambassadors' for the company. Furthermore, recruitment costs and employee training will be reduced, while HR (Human Resources) will able to recruit talented workers and save money.

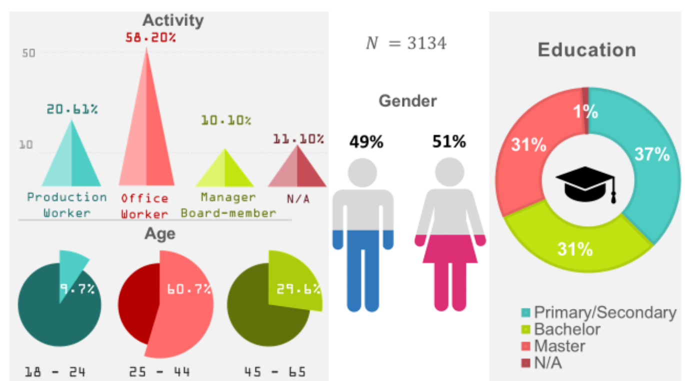

Fig. 1. The demographic information (gender, age, activity and education) of the cross-sectional ordinal dataset including 3134 employees that work in Business sectors in Belgium.

In this study, a dataset of 3,134 records (from the Business sector) was used to evaluate on scale 1 (strongly disagree) to 5 (strongly agree). All variables are associated with a job satisfaction problem in Belgium. The input variables are: "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". The output variable or the decision output concept is the 11th variable " V_{11}: Willingness to work to company z", where z is one of the 349 observed Belgian companies. An example of a question is "When I am looking for a job position in company X, the job that offers competitive salary package (V_1) is important". Moreover, the questionnaire includes demographic questions (Fig. 1) that each participant has to complete such as: *gender*: male (48.56%), female (51.44%), *age*: 18 – 24 (9.7%), 25 – 44 (60.66%), 45 – 65 (29.64%), *education*: primary/secondary (37.43%), bachelor (31.08%), master (30.44%), N/A (1.05%) and *activity*: production worker (20.61%), office worker (58.20%), manager (10.10%), N/A (11.10%).

IV. Results

As mention in the introduction, our research of interests are i) to transform an undirected weight matrix (which is estimated from the glasso algorithm) into a directed weight matrix in order to run the FCM simulation process and ii) to investigate if any rearrangement of the input concepts affects or not the values of the output concept V_{11}. More details on the glasso algorithm i.e. the steps to produce the weighted matrix can be found in our previous work [12, 13].

A. Results of the graphical lasso model

In this sub-section, the *bootnet* R package [46] was used to integrate the *lasso* algorithm with the EBIC model selection to define relations among the observed variables associated with the job attractiveness in Belgium. Specifically, the *estimateNetwork* function utilizes the part of the glasso algorithm [8] from the *bootnet* package [46]. This algorithm returns a symmetric weighted adjacency matrix from ordinal data. The edge weights can be interpreted as the strength of conditional dependence between two variables (Table I).

TABLE I. THE SYMMETRIC WEIGHTED ADJACENCY MATRIX IS ESTIMATED FROM THE GLASSO METHOD

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}
V_1	-	0.39	0	0	0	0	0.13	0.10	0.06	0	0
V_2	0.39	-	0.05	0.25	0	0	0.12	0.16	0.12	0	0.08
V_3	0	0.05	-	0.04	0.19	0.07	0.19	0.07	0.08	0.19	0.12
V_4	0	0.25	0.04	-	0.27	0.29	0	0.07	-0.08	0.10	0
V_5	0	0	0.19	0.27	-	0.09	0.05	0	-0.06	0.13	0
V_6	0	0	0.07	0.28	0.09	-	0.04	0	0.38	0	0
V_7	0.13	0.12	0.19	0	0.05	0.04	-	0.32	0	0	0.25
V_8	0.10	0.16	0.07	0.07	0	0	0.32	-	0.20	0.17	0
V_9	0.06	0.12	0.08	-0.08	-0.06	0.38	0	0.20	-	0.10	0
V_{10}	0	0.00	0.19	0.10	0.13	0	0	0.17	0.10	-	0
V_{11}	0	0.08	0.12	0	0	0	0	0.25	0	0	-

Moreover, the *qgraph* package [18] plots the sparser graphical models of the observed 11 nodes (10 input and 1 output). The input nodes ($V_1 - V_{10}$) in the graph correspond to a job-satisfaction factor; while the output concept (V_{11}) represents the "Willingness to work" feature. Thus, each edge depicts the partial correlation between the features controlling for all other connections in the network. Fig. 2 illustrates the partial correlations of the Business sector of Table I. Nodes that are depicted closer together are strongly related. Hence, wider and more saturated edges determine stronger relations; while, positive and negative relations are depicted by blue and red connections, respectively.

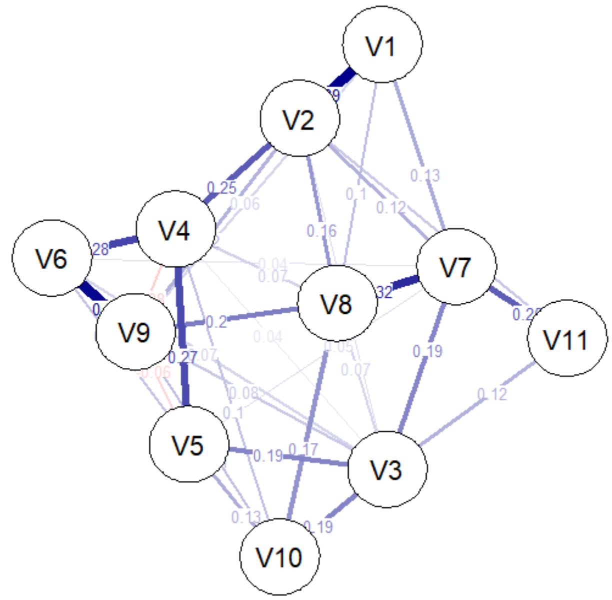

Fig. 2. Graph visualizations of 11 variables associated with job satisfaction problem. The graph displays the symmetric weight matrix which is estimated from the graphical lasso using the EBIC regularization technique.

B. Results of Rankings

After the estimation and the visual inspection of the data-driven network, it is necessary to transform the symmetric weighted matrix graph (Table I) to a directed weighted matrix to accomplish the FCM scenarios.

TABLE II. REORDERING CASES OF THE CROSS-SECTIONAL DATASET IN BUSINESS SECTOR

Type of node	Rank	Strength	Mean	Random	Reverse Strength
Transmitter	1st	V2	V6	V1	V10
Ordinal	2nd	V7	V9	V2	V1
Ordinal	3rd	V4	V4	V3	V5
Ordinal	4th	V8	V2	V4	V6
Ordinal	5th	V9	V1	V5	V3
Ordinal	6th	V3	V7	V6	V9
Ordinal	7th	V6	V8	V7	V8
Ordinal	8th	V5	V3	V8	V4
Ordinal	9th	V1	V5	V9	V7
Ordinal	10th	V10	V10	V10	V2
Receiver	11th	V11	V11	V11	V11

Because expert knowledge is not available to indicate the causal effect connections among variables, the input variables will be reordered and the upper weighted triangular matrix will be obtained. The variables are ranked according to four cases: i) the strength-centrality, ii) the average values of the variables (from the dataset), iii) the random order and iv) the inverse strength-centrality.

Table II depicts the ranked cases of strength, mean, random and inverse strength that will be used to arrange the weights. Afterwards, the weights of the upper triangular matrix are reordering according to each of the observed cases. For instance, Table III illustrates the upper weight matrix rearranging the variables according to the strength-centrality indices. As it is observed, the output concept (V_{11}) is placed always in the last position. Additionally, Fig. 3 visualizes the directed weighted matrix of Table III. The activator concept V_2 (the transmitter) is highlighted with red colour; while, the output concept V_{11} (the receiver) with blue colour.

TABLE III. THE DIRECTED WEIGHTED MATRIX. THE CONCEPTS ARE ORDERED BY THE STRENGTH-CENTRALITY INDICES

	V_2	V_7	V_4	V_8	V_9	V_3	V_6	V_5	V_1	V_{10}	V_{11}
V_2	-	0.12	0.25	0.16	0.12	0.05	0	0	0.39	0	0.08
V_7		-	0	0.32	0	0.19	0.04	0.05	0.13	0	0.25
V_4			-	0.07	-0.08	0.04	0.29	0.27	0	0.10	0
V_8				-	0.20	0.07	0	0	0.06	0.17	0
V_9					-	0.09	0.38	-0.07	0	0.10	0
V_3						-	0.07	0.19	0	0	0.12
V_6							-	0.09	0	0	0
V_5								-	0	0	0
V_1									-	0	0
V_{10}										-	0
V_{11}											-

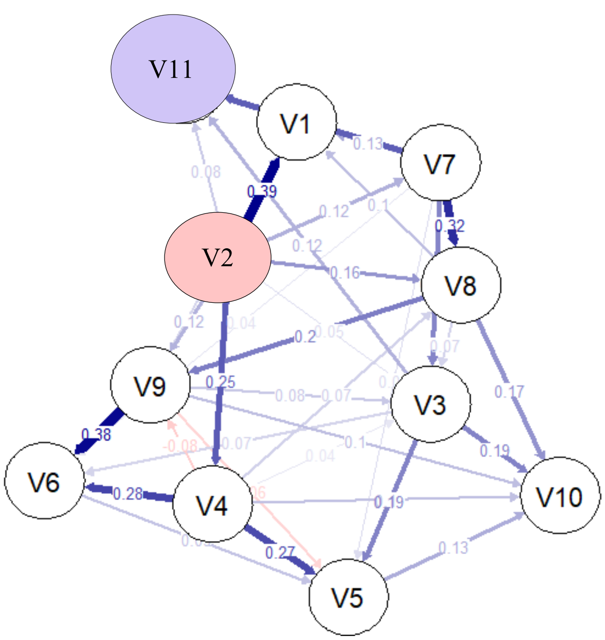

Fig. 3. The upper-triangular weighted matrix of 11 concepts which are ranked according to strength-centrality indices in Table II. The transmitter concept V_2 is highlighted with red colour; while the output concept V_{11} of the glassoFCM model is marked with blue colour.

C. Results of FCM Scenario analysis

In order to estimate the inference of the FCM, we have been implemented in R programming language the '*fcm*' package [24] to estimate the inference of the Fuzzy Cognitive Map and accomplish the scenario analysis. Particularly, the function *fcm.infer()* is applied. The main inputs of this function are: the initial vector (or else the policy scenario) and the weighted matrix W. Additionally, 6 arguments must be set such as the number of iterations, the inference rule, the threshold function, the lambda and the epsilon parameter. The fcm.infer function returns three objects: (a) a data frame which contains the concepts' values of each iteration, (b) an FCM map and (c) the convergence plot of κ steps. (The usage of the function fcm.infer 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). Finally, for the simulations of this study, the rescale-clamped inference function and the sigmoid transformation function are applied.

TABLE IV. THE VALUES OF V_{11} (WILLINGNESS TO WORK) AFTER THE FCM SIMULATION FOR THE 4 OBSERVED RANKINGS IN THE BUSINESS SECTOR, THE NORMALITY TEST AND THE ANOVA ANALYSIS.

Scenarios	**Strength**	**Mean**	**Random**	**Reverse Strength**
V2=1	0.561	0.566	0.559	0.539
V7=1	0.628	0.628	0.618	0.622
V4=1	0.502	0.518	0.502	0.509
V8=1	0.504	0.503	0.499	0.545
V9=1	0.504	0.512	0.499	0.512
V2=V7=1	0.666	0.665	0.656	0.653
V2=V4=1	0.562	0.568	0.562	0.539
V2=V8=1	0.564	0.568	0.559	0.576
V7=V4=1	0.630	0.639	0.618	0.630
V7=V8=1	0.631	0.630	0.618	0.628
V2=V7=V4=1	0.667	0.667	0.656	0.653
V2=V7=V8=1	0.668	0.667	0.656	0.653
V7=V4=V8=1	0.632	0.641	0.618	0.634
V2=V3=V7=1	0.699	0.699	0.699	0.699
ALL Vi	0.698	0.698	0.698	0.698
Shapiro-Wilk Statistics	.895	.908	.914	.911
Sig.	.081	.127	.155	.143
One-way Anova	$F(3,56) = 0.059, p = .981$			

In total, four models will be created. Specifically, the order of variables differs according to Table II. Therefore, the transmitter variable that is located in the first place of the weighted matrix is different. In particular, for the Strength ranking, the transmitter variable is the V_2, for the Mean is the V_6, for Random is V_1 and for Reverse Strength is V_{10}. Thus, 60 experiments (15 for each case) are conducted and presented in Table IV. The results are validated in order to justify if the observed rearrangements play a significant role in the output (V_{11}) values. For this reason, *one-way ANOVA* is applied to examine if there are significant differences among the values of the output. The one-way ANOVA is selected since the Shapiro-

Wilk test indicate that the values of Table IV are normally distributed.

V. DISCUSSION OF RESULTS

For our research, we used the cross-sectional dataset consisted of 11 variables and 3.134 records, 11×3.134. The input nodes corresponded to variables associated with the job satisfaction problem, as described in Section II; while, the output variable was indicated the "Willingness to work (V_{11})". In order to define the job-satisfaction model from the data, glasso method with the EBIC regularization was used (Table 1 and Fig. 2). This network structure obtained 34 connections (out of 45) in which 32 were assigned with positive weights and 2 with negative weights. Specifically, the highest positive weights were observed between Competitive package and Career Opportunities ($V_1 - V_2, w_{12} = .389$) and between Interesting jobs and Quality of training ($V_7 - V_8, w_{7,8} = .389$). On the contrary, the negative weights were emerged among Long term job security and Strong management ($V_4 - V_9, w_{4,9} = -.076$) and among Good balance between private life and work and Strong management ($V_5 - V_9, w_{5,9} = -.065$) indicating that the increase of V_9 causes a very low decrement of V_4 and V_5 and vice versa.

As reported previously, glasso method was returned the asymmetric weighted matrix (Table I). For this reason, the upper triangular part of the weighted matrix was used after rearranging the nodes according to four different cases (Table II). In such situations, the first variable was the transmitter variable (the concept that could influence other concepts since the in-degree is zero). Thus, the last node was determined as the receiver variable (the concept that was affected by other variables and it could not affect others and the out-degree is zero) [21]. Fig. 3 was visualized the upper weight matrix of Table III reordering the input variables according to strength-centrality indices. It was notable to mention that reordering the variables in the dataset, the weights among the nodes derived from glasso method were not changing. Therefore, the glasso with EBIC regularization technique would be applied only once. However, some directions of the edges among the concepts could be changed. For example, the connection between V_4 and V_6 is 0.29. Reordering the variables according to Strength and Random orderings then V_4 affects V_6 ($V_4 \rightarrow V_6$). On the other hand, if we consider the rearrangements according to Mean and Reverse strength rankings then V_6 will affect V_4 ($V_6 \rightarrow V_4$).

In order to investigate if the rearrangements of the variables (Table II) were sensitive to the output value of V_{11}, different FCM scenarios were accomplished using the *fcm.infer* function ('fcm' R package) [24]. Before the comparisons, it was crucial to measure the normality of the data in order to select the appropriate comparison test. For small sample sizes (< 50 samples), the Shapiro-Wilk test [22] is a more relevant test to assess normality. If the p-value is less than or equal to 0.05 then the test rejects the hypothesis of normality and the non-parametric test will be used. Otherwise, if the p-value is greater than 0.05, then the outputs are normally distributed and a parametric test will be applied. Table IV showed the results from the Shapiro-Wilk test. The output values derived from FCM of the four observed cases were determined that the values were normally distributed because of the Sig. value of the Shapiro-Wilk test was greater than 0.05.

Therefore, the appropriate method to determine whether there were any statistically significant differences between the means of four orderings was the one-way analysis of variance (ANOVA) [23]. Table IV was included the results of ANOVA table. Specifically, the significance value in Business ($F(3,56) = 0.059, p = .981$) was above 0.05. Consequently, it was concluded that in hierarchical structures or in decision-making problems with a single output, if the input variables were placed in the weight matrix in any order, the result of the observed output concept in the FCM simulation would be close to the true value.

VI. CONCLUSION

The fundamental challenge of this study was the learning procedure of the weights among the observed concepts straight from the available data without any intervention of experts or the usage of historical data. For this reason, we introduced an automated data-driven learning algorithm, the glassoFCM which modelled (using the *glasso with EBIC regularization* method) and simulated (using the FCM inference method) a job-satisfaction problem. Specifically, the resulting FCM models were used to analyse and simulate the influence of 10 concepts in the output concept in order to increase the willingness of the employees to work. Due to the estimation of the symmetric weight matrix from the glasso method, we obtained the upper part of the symmetric weighted matrix in order to construct the glassoFCM structure. For this reason, we proved that rearranging the input values and maintaining the target concept in the last position of the FCM weight matrix the values of the output concept were not changed significantly. Therefore, our findings could be effective to problems that expert knowledge and/or historical data could not determine with high assurance the directions of the edges among concepts of an FCM. As a consequence, this result has further strengthened our confidence to make meaningful decisions/predictions in hierarchal FCM structures, when the value of a single output (or receiver) concept is examined. Moreover, we hope that our research will serve as a base for future studies on cases that experts are not confident about causal-effect relationships among two concepts in decision-making problems.

REFERENCES

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

[2] B. Kosko, "Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence", Prentice-Hall. New York, 1992.

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

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

[5] E. I. Papageorgiou, A. Kontogianni, "Using fuzzy cognitive mapping in environmental decision making and management: a methodological primer and an application" in S. S. Young and S. E. Silvern, editors. International perspectives on global environmental change. InTech, Rijeka, Croatia, pp. 427-450, 2012.

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

[7] B. W. Brunton, and M. Beyeler, “Data-driven models in human neuroscience and neuroengineering”, Current Opinion in Neurobiology, 58, 21–29, 2019.

[8] M. E. J. Newman, “Networks: an introduction”; Oxford, UK: Oxford University Press, 2010.

[9] J. H. Friedman, T. Hastie, and R. Tibshirani, “ Sparse inverse covariance estimation with the graphical lasso”, Biostatistics, 9(3):432–441, 2008.

[10] G. Xirogiannis and M. Glykas, “Fuzzy Cognitive Maps in Business Analysis and Performance-Driven Change”, IEEE Transactions on Engineering Management, 51(3), 334–351, 2004.

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

[12] 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, Naples, Italy, July 9-12, 2017.

[13] Z. Dikopoulou, E. Papageorgiou, K. Vanhoof, “Retrieving sparser Fuzzy Cognitive Maps directly from large ordinal dataset using lasso graphical models and the max-threshold algorithm”, IEEE world congress on computational intelligence (WCCI), 19 - 24th July, 2020, Glasgow (UK), 2020. (accepted)

[14] P. Zhao, and B. Yu, “On model selection consistency of lasso”, The Journal of Machine Learning Research, 7:2541–2563, 2006.

[15] J. Chen and Z. Chen, “Extended bayesian information criteria for model selection with large model spaces”, Biometrika, 95 (3), 759–771, 2008.

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

[17] J. H. Friedman, T. Hastie, and R. Tibshirani “ glasso: Graphical lasso estimation of gaussian graphical models”, 2014. [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=glasso

[18] S. Epskamp, A. Cramer, L. Waldorp, V. D. Schmittmann, and D. Borsboom, “qgraph: Network visualizations of relationships in psychometric data”, Journal of Statistical Software, 48 (4), 1-18, 2012.

[19] H. M. Weiss, and K. L. Merlo, “Job Satisfaction”, International Encyclopedia of the Social & Behavioral Sciences, 833–838, 2015.

[20] E.I. Papageorgiou, “Fuzzy Cognitive Maps for Applied Sciences and Engineering From Fundamentals to Extensions and Learning Algorithms”, Intelligent Systems Reference Library, Vol. 54, 2014.

[21] M. Bougon, K. Weick, and D. Binkhorst, “Cognition in Organizations: An Analysis of the Utrecht Jazz Orchestra”, Administrative Science Quarterly, 22(4), 606, 1977.

[22] S. S. Shapiro and M. B. Wilk, “An analysis of variance test for normality (complete samples)”, Biometrika, 52 (3-4), pp. 591–611, 1965.

[23] R. A. Fisher, “The Correlation Between Relatives on the Supposition of Mendelian Inheritance”, Philosophical Transactions of the Royal Society of Edinburgh, 52, 399–433, 1918.

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

[25] W. Stach, W. Pedrycz, L. A. Kurgan, “Learning of fuzzy cognitive maps using density estimate”, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 42 (3) (2012) 900–912.

[26] A. Amirkhani, EI. Papageorgiou, A. Mohseni, MR. Mosavi, “A review of fuzzy cognitive maps in medicine: Taxonomy, methods, and applications”, Computer Methods and Programs in Biomedicine 142, (2017) 129-145.

[27] P. Muenning, “Decision Analytic Modeling”, International Encyclopedia of Public Health, 71–76, 2008.

[28] V.C.P. Knowland, H. Purser, M.S.C. Thomas, “Cross-Sectional Methodologies in Developmental Psychology”, International Encyclopedia of the Social & Behavioral Sciences, pp. 354-360, 2015.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] P. Zhang, “Model selection via multifold cv”, Ann. Statist., 21, (1993), 299-311.

[35] S. L. Lauritzen. “Graphical Models”, Springer Verlag, 1996.

[36] J. M. B. Haslbeck & , L. J Waldorp “mgm: Structure estimation for time-varying mixed graphical models in high-dimensional Data”, 2016.

[37] R. Mazumder, and T. Hastie, “The graphical lasso: New insights and alternatives”, Electron. J. Statist, 6, (2012), 2125-2149.

[38] S. Epskamp, “Regularized Gaussian Psychological Networks: Brief Report on the Performance of Extended BIC Model Selection”, 2016.

[39] S. Epskamp and E. I., Fried, “A Tutorial on Regularized Partial Correlation Networks”, Psychological Methods, 2007.

[40] K. Wu, J. Liu, Robust learning of large-scale fuzzy cognitive maps via the lasso from noisy time series, Knowl.-Based Syst. 113 (2016) 23–38.

[41] K. Wu and J. Liu, “Learning of sparse fuzzy cognitive maps using evolutionary algorithm with lasso initialization,” in Proc. Asia–Pac. Conf. Simulat. Evol. Learn., Shenzhen, China, Nov. 2017, pp. 385–396.

[42] Y. Chi and J. Liu, "Learning of Fuzzy Cognitive Maps With Varying Densities Using A Multiobjective Evolutionary Algorithm," in IEEE Transactions on Fuzzy Systems, vol. 24, no. 1, pp. 71-81, Feb. 2016.

[43] F. Shen, J. Liu, K. Wu, “Evolutionary multitasking fuzzy cognitive map learning”, Knowledge-Based Systems, 192, 2020.

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

[45] 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.

[46] S. Epskamp, D. Borsboom, E.I. Fried, “Estimating Psychological Networks and their Accuracy: A Tutorial Paper.” Behavior Research Methods, 2017.

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

[48] M. Kiani, J. Andreu-Perez, H. Hagras, E. I. Papageorgiou, M. Prasad and C. Lin, "Effective Brain Connectivity for fNIRS with Fuzzy Cognitive Maps in Neuroergonomics," in IEEE Transactions on Cognitive and Developmental Systems, 2019..