Social Indicators Research (2026) 182:40
https://doi.org/10.1007/s11205-026-03844-6

ORIGINAL RESEARCH

Graphical Ising Models for Missing Data Patterns Detection in Sustainability Surveys

Andrea Mecca[1,2] · **Anna Gottard**[1] · **Francesca Gagliardi**[2]

Received: 16 November 2025 / Accepted: 24 March 2026 / Published online: 10 April 2026

Abstract
Ensuring sustainability in the agri-food sector requires comprehensive data analysis. This study examines missing data patterns in a large-scale survey of Italian agri-food companies within the Italian National Research Center for Technology in Agriculture (Agritech), focusing on sustainability variables. The underlying idea is that failure to report a value for these variables indicates low attention to sustainability. We employ graphical Ising models to infer the conditional independence structure among missingness indicators and fully observed farm characteristics, which are modeled as binary variables. The graph structure is selected through node-wise logistic regressions with variable selection based on a backward stepwise procedure guided by the Bayesian Information Criterion (BIC). This approach enables the recovery of sparse and interpretable graphs while controlling for model complexity. We are not interested in causal relationships between missingness indicators and fully observed variables; rather, our focus is on the dependence structure among these variables. The method is applied to the Agritech data, yielding both a national graph and macro-regional graphs, as well as differential networks that highlight structural differences between each macro-region and the national graph. These results provide new insights into systematic patterns of missing data, offering a rigorous framework for improving data quality in terms of completeness and reliability.

Keywords Agri-food sustainability · Graph learning · Ising graphical models · Missing data

1 Introduction

The sustainability of agri-food systems is increasingly assessed through composite sets of social, environmental, and economic indicators. International policy frameworks, most prominently the European Green Deal (European Commission, 2019) and the UN 2030

This paper was presented at the SIS 2025 Conference Statistics for Innovation. The original title of the paper presented at the SIS 2025 Conference Statistics for Innovation by the author Andrea Mecca was "Using Graphical Models for Missing Data Patterns Detection in Sustainability Surveys".

Extended author information available on the last page of the article

Agenda (United Nations, 2015), have accelerated the diffusion of indicator-based assessment and disclosure along agri-food value chains. Within this policy framework, the Italian Agritech Center (PNRR, 2022-2026) coordinates universities and research institutes through a hub-and-spoke structure to align measurement practices with European and international standards and to develop an integrated platform for data collection and analysis. Our study is embedded in these activities, contributing an indicator-quality perspective focused on the structure of non-response in sustainability surveys.

Sustainability assessment is best framed within the debate on how indicator systems can capture the multidimensional nature of the phenomenon. A consistent stream of research argues that the environmental, economic and social pillars should be assessed jointly and complemented by governance, traceability, and disclosure mechanisms. These elements shape the credibility of the resulting measures (Grimm, 2025; Liu et al., 2024). Indicator sets that make interdependencies explicit improve comparability between heterogeneous firms, while also remaining sensitive to managerial logics and territorial constraints (Gallo et al., 2021; Poponi et al., 2022). Traceability systems, in particular, operate simultaneously as verification channels and as observable proxies for process control. Their prominence has been documented for both companies and consumers (Betti et al., 2024; Tessitore et al., 2022). From this point of view, sustainability cannot be collapsed into a single bundle of technical indicators. Rather, it calls for a comprehensive framework in which dimensions are connected and assessed jointly (Betti et al., 2025). In addition to multidimensionality, heterogeneity is a defining characteristic of the agri–food sector. Production chains differ in biological cycles, quality conventions, and market structures, all of which shape the type and salience of sustainability indicators (Banterle et al., 2006; Mania et al., 2018). Territorial variations are equally consequential: infrastructural endowments, regulatory environments, and market access alter the feasibility, costs, and verification of sustainable practices. Treating territorial differentiation as analytically primary reduces the risk of conflating structurally distinct contexts and provides a substantive rationale for selecting separate network representations at the macro-regional level (Betti et al., 2024).

To investigate these issues, a national survey of Italian agri-food enterprises was conducted in 2024 using CATI/CAWI interviews with stratified sampling across Italian regions, following established approaches in the sustainability literature (Do Canto et al., 2021; Grimm et al., 2016; León Bravo et al., 2021). The questionnaire covered more than 120 items, namely demographic and structural items and a broad set of sustainability practices that combine social, environmental, biodiversity, and economic domains. A common issue in such surveys, which gather extensive quantitative data, is the presence of missing responses. We are not interested in the non-response mechanism, nor do we assume any specific missingness mechanism affecting the values of variables with missing data. Instead, we focus solely on the presence or absence of missing records within the subset of selected variables. Early diagnostics showed heterogeneous missingness: resource-intensive items displayed much higher non-response than structural descriptors, motivating a closer, system-level look at how missingness co-varies across items. Rather than treating missing data as a limitation to be repaired ex post, we take a different perspective in which systematic missingness is itself substantive, potentially reflecting salience, feasibility, or sensitivity of sustainability items. Accordingly, our goal is to treat item non-responses as an informative signal and to examine whether systematic missingness in specific sustainability aspects

helps identify groups of similar firms with different propensities to engage with, or report on, sustainability dimensions. Our proposal relates to the framework of missingness graph models (Mohan & Pearl, 2021), which assume a known causal graph, i.e., a known causal relationship between observed variables and missingness indicators. However, unlike causal uses of such graphs, we do not impose any a priori structure or attempt to identify causal effects. Instead, we aim to infer from data the conditional independence structure among missingness indicators and fully observed firm characteristics, using it as an indication of a firm's attention to sustainability.

We consider the class of the Ising graphical model (Ising, 1925), an undirected graphical model for binary variables. This model was originally introduced by Ising (1925) to describe interactions among states in ferromagnetic solid materials. Here, we are on the special case of the Ising model called *graphical Ising model* discussed by Ravikumar et al. (2010), which is a parsimonious model for sets of binary variables. Exploiting the equivalence of the Ising graphical model to node-wise logistic regressions (Ravikumar et al., 2010), on the basis of a simulation study conducted to compare widely adopted estimation strategies, we select the graph structure via a BIC-guided backward stepwise selection. This approach allows us to analyze how the missingness indicators related to sustainability are associated with firm characteristics. In addition, we can provide the same analysis for different populations. Interestingly, we select (i) a national graphical model summarizing the structure of non-response across the full sample and (ii) macro-regional networks to capture territorial differentiation. A significant by-product of this analysis is given by the differential graphs, able to highlight the difference between the local and the national graph, in terms of added or missing edges. This multi-group design recognizes the heterogeneity of Italian agri-food systems and the territorial embeddedness of sustainability practices and reporting frictions.

The proposed methodology's contribution to social indicator research is threefold.

1. Indicator fragility diagnostics. Graphs reveal tightly connected clusters among operational items (e.g., energy and water use, revenues), suggesting shared reporting frictions; by contrast, several compliance/organizational items form localized neighborhoods, and some social indicators (e.g., women in management) appear structurally isolated.
2. Territory aware risk mapping. Macro-regional graphs and their differentials uncover shifts in dependencies, e.g. stronger social and environmental coupling in some areas, or narrower neighborhoods for revenue reporting in others, informing targeted data-quality interventions.
3. A replicable workflow. The proposed method provides a transparent and reproducible diagnosis of missingness dependencies. Assuming that non-response signals lower sustainability attention, it identifies systematic patterns across missing indicators and firm attributes, informing survey design and data-quality assessment. These elements align with recent discussions on multidimensional assessment, traceability, and governance in agri-food sustainability (Betti et al., 2024; Grimm, 2025; Liu et al., 2024).

The remainder of the paper is organized as follows. Section 2 describes the survey data and methodological approach. Section 3 reports national and macro-regional results, including differential networks. Section 4 concludes with some final remarks.

2 Dataset and Methods

2.1 Data

In this study, we analyze data from the survey conducted in 2024 within the Agritech Center. The original dataset covers 3, 002 agri-food enterprises and includes more than 120 structured items covering demographic, economic, environmental, and organizational dimensions. We focus on the core section of the questionnaire, which is concentrated on sustainability-related practices and structural characteristics of enterprises. Before defining the analytical sample, we carried out an exploratory assessment of missing data. The results, reported in Table 1, highlight the heterogeneity of non-response across items: environmen-

Table 1 Distribution of missing values for the core variables and the original data from the Agritech questionnaire

Variable description	Missing (N)	Missing (%)
Water consumption	1752	58.36
Gas consumption	871	29.01
Cultivated land area	843	28.08
Diesel consumption	842	28.05
Electricity consumption	770	25.65
Annual revenue	703	23.42
Uncultivated land area	659	21.95
Renewable energy produced	654	21.79
Employees from EU countries	642	21.39
Farms providing employee benefits	563	18.75
Farms employing people with disabilities	563	18.75
Forested land area	551	18.35
Farms receiving subsidies	357	11.89
Certification costs	332	11.06
Employees from not EU countries	261	8.69
Employees under 25 years old	237	7.89
Women in managerial positions	235	7.83
Farms with product certifications	210	7.00
Farms using fertilizers	210	7.00
Farms supplying school canteen	210	7.00
Farms with rainwater collection systems	210	7.00
Farms with organic matter reintroduction programs	210	7.00
Farms with hedgerows delimiting fields	210	7.00
Farms reporting soil organic matter percentage	210	7.00
Farms with ponds or similar water bodies	210	7.00
Seasonal employees	178	5.93
Female employees	162	5.40
Male employees	162	5.40
Total number of employees	130	4.33
Soil organic matter percentage	82	2.73
Average age of machinery	26	0.87
Number of agricultural machines	7	0.23

tal and energy-related questions exhibit the highest proportions of missing values, whereas structural variables show almost complete coverage.

Within the set of variables subject to missingness, we consider a subset of variables that might be informative for sustainability. For each of these variables, we construct a binary missing indicator (see subsequent (10)) to represent the presence or absence of a response. The selected variables describe environmental, economic, social, and biodiversity domains. They include one land-use indicator (cultivated land area) and four measures of resource consumption within the atmospheric domain: electricity, gas, diesel, and water. Additional indicators refer to biodiversity and social aspects, namely the presence of rainwater collection systems and the share of women in managerial positions. Finally, two economic variables, annual revenue and the presence of product certifications, are also considered.

The set of fully observed variables includes binary indicators describing key structural and organizational characteristics of the firms. They identify whether the respondent is the owner, the firm belongs to a consortium, operates as a society or an enterprise, is family-owned, plays a production role in the agri-food chain, has invested in renewable energy plants, and is specialized in the wine sector.

For the analytical sample, we first exclude enterprises with excessive item non-response; among the remaining cases, we further remove those that did not report the year of foundation, yielding a final sample of $n = 2689$ units, divided geographically into the macro regions North ($n_N = 1019$), Center ($n_C = 801$) and South and Islands ($n_S = 869$). The year of foundation is dichotomized at the 85-th percentile of its distribution (2015) and retained as a fully observed structural variable. Table 2 summarizes the final set of variables used in the empirical analyzes.

Table 2 Final set of variables included in the empirical analyzes

Dimension	Variable	Label
Variables subject to missingness		
Soil	Cultivated land area	M1
Atmosphere	Annual electricity consumption	M2
	Annual gas consumption	M3
	Annual diesel consumption	M4
	Annual water consumption	M5
Biodiversity & Society	Rainwater collection systems	M6
	Women in managerial positions	M7
Economy	Annual revenue	M8
	Product certification	M9
Fully observed variables		
Firm characteristics	Owner as respondent	Y1
	Consortia membership	Y2
	Farm type (Society/enterprise)	Y3
	Family business	Y4
	Production role	Y5
	Investment in renewable power plants	Y6
	Operates in wine sector	Y7
	Year of foundation (dichotomized at 2015)	Y8

2.2 Methods

Graphical models are a class of multivariate models that study the conditional independence structure among variables (Lauritzen, 1996) and represent these independencies via a graph $\mathcal{G}$. Let $\mathrm{X}_V = (X_1, \ldots, X_p)$ denote p random variables, and let $\mathcal{G} = (V, E)$ be a graph where each node $j \in V$ corresponds to a random variable X_j. Graphs may be undirected, directed, or mixed, according to the permitted edge type. For this study, we adopt undirected graphical models, also called Markov Random fields or Networks. Consequently, E contains only undirected edges and the variables are considered on equal footing, with no a priori causal ordering. Conditional independence corresponds to the absence of an edge between two nodes, and it is encoded by the Global, the Local, and the Pairwise Markov properties. For instance, the Pairwise Markov property states that if $(j, k) \notin E$, then $X_j \perp\!\!\!\perp X_k \mid \mathrm{X}_{V\setminus\{j,k\}}$. Here $\mathrm{X}_{V\setminus\{j,k\}}$ is the set of variables in V excluding X_j and X_k. Now, let $\mathrm{ne}(j) = \{k \in V : (j, k) \in E\}$ denote the set of neighbors of a node j. The Local Markov property states that each variable is independent of all the other variables, conditional on its neighbors, that is $X_j \perp\!\!\!\perp \mathrm{X}_{V\setminus(\{j\}\cup\mathrm{ne}(j))} \mid \mathrm{X}_{\mathrm{ne}(j)}$. We refer to Lauritzen (1996) for further details on undirected graphical models.

For categorical variables, a model for the joint distribution of X_V is given by a hierarchical log-linear model (Darroch et al., 1980). Denoting as $\mathcal{I}$ the p-dimensional contingency table, with cells $\iota = (\iota_1, \ldots, \iota_p)$ and probabilities $p(\iota)$, expected cell counts $\mu(\iota) = np(\iota)$ satisfy

$$\log \mu(\iota) = \sum_{a \subseteq \{1,\ldots,p\}} \lambda^a(\iota_a), \tag{1}$$

where $\lambda^a(\iota_a)$ are interaction parameters of order $|a|$. In this setting, the conditional independence $X_j \perp\!\!\!\perp X_k \mid \mathrm{X}_{V\setminus\{j,k\}}$ holds whenever all interaction terms involving both j and k vanish. The saturated log-linear model corresponds to the complete graph, including interactions of all orders up to p. Although conceptually appealing, this parameterization quickly becomes infeasible in practice. The number of cells in $\mathcal{I}$ grows exponentially with p, and even in the binary case, the saturated log-linear model requires 2^p parameters. As p increases, the contingency table becomes extremely sparse, leading to unstable or undefined maximum likelihood estimates (Banzato et al., 2025). This curse of dimensionality severely limits the applicability of general log-linear models in high-dimensional survey data. To circumvent these difficulties, one may restrict attention to models that retain only first and second-order interactions. A natural choice in the binary case is the Ising graphical model (Ising, 1925; Ravikumar et al., 2010). Let $X_j \in \{0, 1\} \quad \forall j = 1, \ldots, p$, the joint distribution is specified as

$$P(\mathrm{X}_V) \propto \exp\left(\sum_{j=1}^{p} \theta_j X_j + \sum_{j<k} \theta_{jk} X_j X_k\right), \tag{2}$$

where X_j is a variable in X_V, θ_j its node-specific parameter and $\theta_{jk} = 0$ if and only if $X_j \perp\!\!\!\perp X_k \mid \mathrm{X}_{V\setminus\{j,k\}}$. This model can be seen as a special case of log-linear models where only the two-way interaction terms are retained. As a result, the Ising graphical model requires at most only $\binom{p}{2}$ interaction parameters, providing a parsimonious but interpretable representation of the dependence structure. However, direct likelihood-based

estimation is computationally infeasible. The likelihood function is known only up to a normalizing constant that involves a summation over all possible configurations of X_V and is therefore analytically intractable (Ravikumar et al., 2010).

To learn a sparse graph, we consider a node-wise regression approach to detect the neighbor set of each variable (Ravikumar et al., 2010). Under sparsity of the graph and positivity of the distribution assumptions, their estimation procedure consistently recovers each variable neighborhood via the Local Markov Property. According to this procedure, for each binary random variable $X_j \in \mathrm{X}_V$ a logistic regression model is fitted treating X_j as the response variable and all other variables $X_{V\setminus\{j\}}$ as predictors

$$\mathrm{logit}(P(X_{ji} = 1|\mathrm{X}_{V\setminus\{j\},i})) = \beta_{j0} + \beta_{j1}X_{1i} \dots \beta_{j(j-1)}X_{(j-1)i} + \beta_{j(j+1)}X_{(j+1)i} \dots \beta_{jp}X_{pi} \quad (3)$$

where $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \dots, \beta_{j(j-1)}, \beta_{j(j+1)}, \dots, \beta_{jp})^\top$ is the vector of the regression coefficients for the response X_j, $j = 1, \dots, p$. The effectiveness of the node-wise selection approach lies in the fact that $\theta_{jk} = 0$ iff $\beta_{jk} = 0$. Then, the conditional independence structure encoded by the Ising graphical model can be estimated by exploiting its equivalence to a system of p logistic regressions (Ravikumar et al., 2010) of the kind shown in Eq. 3.

Several approaches have been proposed to learn the graph structure of an Ising model. The most widely adopted methods are the eLasso (Van Borkulo et al., 2014), node-wise ℓ_1 -regularized logistic regression with model selection based on a K-fold cross-validation (Hastie et al., 2009), node-wise stepwise backward logistic regression with model selection based on the Bayesian Information Criterion (BIC), and node-wise logistic regression with stability selection (Meinshausen & Bühlmann, 2010; Shah & Samworth, 2012).

In the eLasso (Van Borkulo et al., 2014), the estimation of the neighborhood set of a node j is based on computing an ℓ_1-regularized logistic regression of X_j on the rest of the variables $\mathrm{X}_{V\setminus\{j\}}$. The optimization procedure (Ravikumar et al., 2010) is performed sequentially for each variable, and the regularized regression problem takes the form of the following convex program

$$\widehat{\boldsymbol{\beta_j}} = \underset{\boldsymbol{\beta}_j \in \mathbb{R}^{p-1}}{\mathrm{argmin}} \{\ell(\boldsymbol{\beta}_j; \mathrm{X}_{V\setminus\{j\}}) + \lambda\|\boldsymbol{\beta}_j\|_1\}, \quad (4)$$

where $\lambda \in \mathbb{R}^+$ is a regularization parameter to be selected, and $\ell(\boldsymbol{\beta}_j; \mathrm{X}_{V\setminus\{j\}})$ is the rescaled negative log-likelihood of the logistic regression model in (3). In the eLasso estimation procedure, the regularization parameter, determining the best set of neighbors, can be chosen via the Extended Bayesian Information Criterion (EBIC) (Chen & Chen, 2008; Foygel et al., 2010)

$$\mathrm{EBIC}_\gamma = -2\,\ell(\widehat{\boldsymbol{\beta_j}}) + |E|\log(n) + 2\,\gamma\,|E|\,\log(p) \quad (5)$$

where $\ell(\widehat{\boldsymbol{\beta}_j})$ denotes the maximized log-likelihood of a candidate model with parameter estimate $\widehat{\boldsymbol{\beta}_j}$, $|E|$ is the number of corresponding retained edges, and $\gamma \in [0, 1]$ is a further parameter related to the assumed sparsity level. The final estimated edge set $\widehat{E}$ is determined by applying the AND rule, i.e. an edge between nodes X_j and X_k is included in $\widehat{E}$ if and only if both the regression coefficients β_{jk} and β_{kj} are nonzero (Van Borkulo et al., 2014), that is

$$(j, k) \in \widehat{E} \quad \text{iff} \quad j \in \widehat{\mathrm{ne}}(k) \quad \text{AND} \quad k \in \widehat{\mathrm{ne}}(j) \quad (6)$$

where $\widehat{\text{ne}}(k)$ and $\widehat{\text{ne}}(j)$ are the estimated neighborhood set of nodes k and j respectively. Instead of relying on the EBIC for selecting the regularization parameter, a widely adopted and theoretically grounded approach is represented by a K-fold cross-validation. The optimal λ can be determined using the one-standard-error rule (1-SE rule) (Hastie et al., 2009), which selects the most parsimonious value of λ within one standard error of the minimum cross-validated error. The 1-SE rule takes a conservative stance, favoring sparser and more stable solutions while maintaining comparable predictive adequacy.

As an alternative to ℓ_1-regularized estimation, model selection can be performed using a stepwise backward procedure guided by the BIC. For each node, the algorithm starts from the saturated model including all possible predictors and proceeds iteratively by removing one variable at a time whenever its exclusion leads to an improvement in the BIC value. The elimination continues until no further improvement in the information criterion is achieved. Since the estimation is performed separately for each node, the resulting neighborhoods are not necessarily symmetric. To obtain the final estimated undirected graph, the individual estimates are combined using the AND rule in (6).

Despite the use of regularization and model selection criteria, node-wise estimation may yield unstable graphs, as small data perturbations may alter the set of selected edges. To improve the robustness and reliability of the graph estimation procedure, Meinshausen and Bühlmann (2010) propose, for Gaussian graphical models, a node-wise lasso procedure with a stability selection step, which can be easily adapted to the case of graphical Ising models. This method combines ℓ_1-regularized logistic regressions with subsampling to enhance the stability of the estimated graphical structure, controlling for the expected number of falsely selected variables, i.e., the per-family error rate (PFER). Now, let $\widehat{\text{ne}}_A^{\lambda}(j)$ denote the set of selected neighbors of a given node obtained from a lasso regression with regularization parameter $\lambda \in \mathbb{R}^+$ on a random subsample $A \subset \{1, \ldots, n\}$ of size $\lfloor n/2 \rfloor$ drawn without replacement, where $\lfloor \cdot \rfloor$ denotes the floor operator, which returns the largest integer less than or equal to the quantity in brackets. Data are repeatedly perturbed through subsampling, and only variables with a high selection probability will be retained. Let $\widehat{\Pi}_{\lambda}^{j}(K)$, $j = 1, \ldots, p$, be the probability distribution induced by subsampling that a subset of variables $K \subseteq \mathrm{X}_{V \setminus \{j\}}$ is included in the selected set $\widehat{\text{ne}}_A^{\lambda}(j)$. Then, for a set Λ of values for the regularization parameter λ, the set of stable neighbors is defined as

$$\widehat{\text{ne}}^{\text{stable}}(j) = \{K : \max_{\lambda \in \Lambda} \widehat{\Pi}_{\lambda}^{j}(K) \geq \tau\}, \tag{7}$$

where $\tau \in (0, 1)$ is cut-off parameter related to stability. Under certain assumptions, Meinshausen and Bühlmann (2010) derived the upper bound of the per-family error rate

$$\text{PFER} \leq \frac{1}{2\tau - 1} \cdot \frac{q_{\Lambda}^2}{p}, \tag{8}$$

where q_{Λ} is the expected number of selected variables across regularization levels. The node-wise lasso with stability selection proceeds by repeatedly drawing random subsamples without replacement, applying the node-wise lasso to each subsample, recording the frequency with which each potential edge is selected, and retaining only those edges with selection probability greater than τ. The AND rule in (6) is then applied across nodes.

Shah and Samworth (2012) extends the stability selection procedure by Meinshausen and Bühlmann (2010), proposing a complementary pairs stability selection (CPSS) procedure, that modifies the subsampling scheme to form complementary pairs of disjoint subsamples. Specifically, let $\{(A_{2b-1}, A_{2b}) : b = 1, \ldots, B\}$ denote B complementary pairs of disjoint subsamples of size $\lfloor n/2 \rfloor$. The node-wise lasso is applied independently to each subsample. Denoting with $\widehat{\Pi}^j_B(X_k)$ the proportion of times the variable X_k is selected for the response X_j across the $2B$ samples, the estimated neighbor set is

$$\widehat{\text{ne}}^{CPSS}(j) := \{X_k : \widehat{\Pi}^j_B(X_k) \geq \tau\}, \tag{9}$$

where $\tau \in [0, 1]$, again, a cut-off parameter related to stability. Shah and Samworth (2012) proved that for $\tau \in (\frac{1}{2}, 1]$ the PFER is also bounded with no need of Meinshausen and Bühlmann (2010) assumptions.

Brusco et al. Brusco et al. (2023), in their simulation study, showed that the stepwise node-wise regression combined with the BIC attains accuracy comparable to that obtained with the eLasso using BIC or EBIC selection criteria, while exhibiting higher specificity and slightly lower sensitivity for smaller samples ($n = 250, 500$), and being computationally more efficient for moderate to large sample sizes ($n \geq 1000$). To complement these findings, we performed an *ad hoc* simulation study to compare the alternative estimation procedures in a setting similar to ours. Since the underlying graph in our application is expected to be relatively dense, we designed the comparison to include scenarios with different levels of sparsity while keeping the same sample size ($n = 2689$) and the same number of binary variables ($p = 17$). Simulation results, reported extensively in Appendix A, indicate that the backward stepwise node-wise logistic regression guided by the BIC consistently achieves high accuracy and strong specificity across all simulated scenarios. In scenarios with lower and moderate sparsity, alternative methods occasionally exhibited higher sensitivity, whereas the stepwise BIC approach provides the highest accuracy among the methods. Based on the simulation results, the backward stepwise BIC-guided node-wise estimation was selected for the empirical analysis, as it ensures stable and interpretable structure learning given the sample size and dimensionality considered in this study.

3 Results

Since the underlying network is not expected to be highly sparse, the backward stepwise node-wise logistic regression guided by the BIC is employed, following the simulation evidence in Appendix A. We consider the data matrix $\mathrm{X}_V = \{\mathrm{Y}_1, \ldots, \mathrm{Y}_8, \mathrm{M}_1, \ldots, \mathrm{M}_9\}$, where $\mathrm{Y} = \{\mathrm{Y}_1, \ldots, \mathrm{Y}_8\}$ is the set of fully observed variables and $\mathrm{M} = \{\mathrm{M}_1, \ldots, \mathrm{M}_9\}$ is the set of binary missing indicators (see Table 2). Each binary missing indicator $\mathrm{M}_j = (M_{ji})_{i=1}^n$ is specified as

$$M_{ji} = \begin{cases} 1, & \text{if } X_j \text{ is missing for unit } i \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

for all $j = \{1, \ldots, 9\} \in \mathrm{M}$ and $i = 1, \ldots, n$. In what follows, we report the graphs selected for the Agritech dataset at both the national and macro-regional levels. To make comparisons

across macro-regions easier, we use the same layout for all graphs. Appendix B provides alternative visualizations with an *ad hoc* node arrangement that makes the relevant conditional independencies more apparent. The same appendix also lists all local and selected global conditional independence statements implied by the graphs, and the stability checks.

To facilitate reading of the selected graphs, we briefly recall the three Markov properties. The pairwise property states that two variables whose nodes are not connected are conditionally independent given all remaining variables. The local property states that each variable is independent of all non-neighbors given its neighbors. The global property asserts that if a subset of nodes S separates the subsets of nodes A and B (i.e., every path from a node A to a node in B needs to pass through a node in S), then the variables in A are independent of the variables in B given those in S. These conditional independence statements are encoded by the graph and hold simultaneously. They pertain to the joint distribution and to every marginal and conditional distribution implied by the model.

3.1 National Level Graphical Model

We first select the graph structure at the national level (Fig. 1, selected adjacency matrix is reported in Appendix B). Three features stand out. First, the missingness indicator for women in managerial positions (M7) appears as a singleton in the national graph, suggesting that non-response on this item follows a distinct mechanism that is not aligned with the other missingness patterns captured by the model. Second, two compliance-related items, rainwater collection systems (M6) and product certification (M9), form a self-contained pair that is disconnected from the rest of the network. This indicates a circumscribed compliance-related non-response mechanism that is largely unrelated to both firm characteristics and other sustainability items. Notably, this separation persists across all macro-regional graphs (Section 3.2). Third, the missingness indicators related to the atmospheric/technical domain, electricity (M2), gas (M3), diesel (M4), and water consumption (M5), form a densely connected block, and annual revenue (M8) is closely tied to this same block. This configuration suggests that non-response for these items tends to co-occur, consistent with a shared reporting burden for technical, infrastructural, and economic quantities. Links

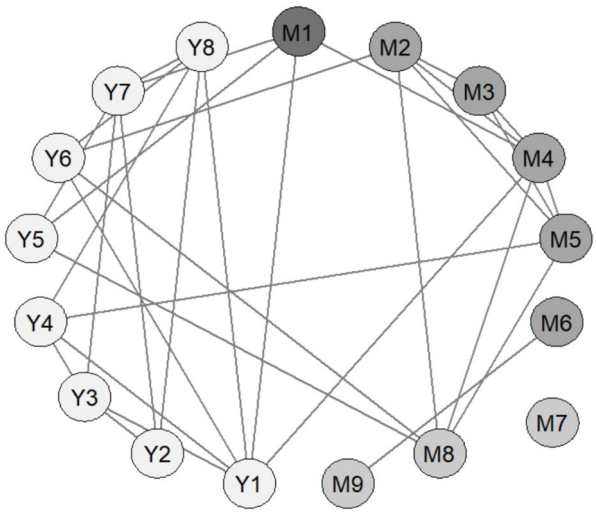


Fig. 1 Selected graphical model at national level

between this technical block and fully observed firm characteristics are concentrated on a small set of bridging variables, such as respondent role and a few structural firm attributes, rather than being broadly distributed across all firm descriptors.

Overall, Figure 1 supports the view that item non-response is structured and domain-specific. A cohesive technical/economic missingness component, a narrow compliance component, and an idiosyncratic mechanism for the social indicator M7. This modular structure suggests that multiple non-response processes coexist in the survey, rather than a single uniform missingness mechanism.

Summarizing, the most interesting conditional independencies implied by the selected graphs are the following.

- $M7 \perp\!\!\!\perp \mathrm{X}_{V\setminus\{M7\}}$
- $(M6, M9) \perp\!\!\!\perp \mathrm{X}_{V\setminus\{M6,M9\}}$
- $(M2, M3, M4, M5) \perp\!\!\!\perp (Y2, Y3, Y5, Y7, Y8) \mid (M1, M8, Y1, Y4, Y6)$
- $(M2, M4, M5, M8) \perp\!\!\!\perp (Y2, Y3, Y7, Y8) \mid (M1, M3, Y1, Y4, Y5, Y6)$
- $M1 \perp\!\!\!\perp (M2, M3, M5, M6, M7, M8, M9, Y2, Y3, Y4, Y6, Y8) \mid (M4, Y1, Y5, Y7)$
- $(M1, M4, M8) \perp\!\!\!\perp (Y2, Y3, Y4, Y8) \mid (M2, M3, M5, Y1, Y5, Y6, Y7)$

See Appendix B, Table 4) for other conditional independence statements.

3.2 Macro-Regional Level Graphical Models

In this section, we discuss the selected graphs at macro-regional level. For completeness, we also report the differential network of the macro-regional specific graphs and the national graph. In these differential networks, only the differences in the graph structure are reported to highlight macro-regional specific behavior in terms of conditional dependencies and independencies.

We replicate the analysis at the macro-regional level (North, Center, South and Islands). Figure 2 reports the selected graphs and the corresponding differential networks relative to the national graph. Appendix B reports adjacency matrices and the implied conditional independencies (Tables 5, 6 and 7). Below, we emphasize what is stable across territories and what changes.

Two patterns are remarkably stable. Across all macro-regions, the compliance pair (M6, M9) remains separated from the rest of the variables, and the atmospheric/technical missingness indicators (M2–M5) remain tightly interconnected, indicating persistent co-missingness for technical items. Territorial differences mainly concern where the social indicator M7 attaches, how strongly annual revenue missingness (M8) is embedded in the technical block, and which firm characteristics serve as bridges between missingness and observed firm attributes.

3.2.1 North Macro-Region

The selected network for the North, depicted in Fig. 2 (a1), is sparser than the national graph, with a clearer separation between the technical missingness block and the fully observed firm characteristics (see the differential network in Fig. 2 (b1)). The compliance pair (M6, M9) remains isolated.

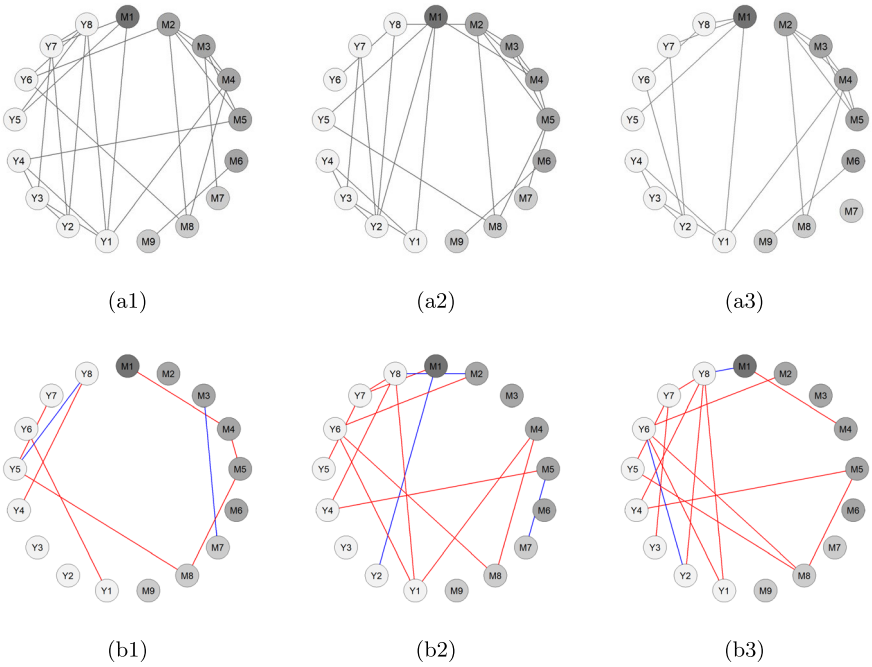


Fig. 2 Selected graphical models for (a1) North, (a2) Center, and (a3) South and Islands macro-regions. Differential networks for the (b1) North, (b2) Center, and (b3) South and Islands macro-regions. In each differential network, blue edges indicate links present in the macro-regional graph but absent at the national level, whereas red edges indicate links absent in the macro-regional graph but present in the national graph

A notable macro-regional feature concerns the social indicator, M7, that is no longer isolated and connects to missingness in gas consumption (M3). This suggests that social non-response in the North is more closely aligned with the technical reporting dimension than at the national level. Once energy-related items and ownership structure are taken into account, the non-response on the environmental and social indicators (M3, M5, M7) reflects a self-contained reporting mechanism. Annual revenue missingness (M8) shows a more restricted neighborhood than in the national graph, consistent with a more narrowly defined economic reporting mechanism in the North. Overall, the differential network (Fig. 2 (b1)) indicates that several links involving organizational variables, for instance, the firm's production role, weaken or disappear, while the technical block remains the dominant structure governing non-response.

We report here in detail some conditional independence statements and refer to Appendix B Table 5.

- $M7 \perp\!\!\!\perp \mathrm{X}_{V\setminus\{M7,M3\}} \mid M3$
- $(M6, M9) \perp\!\!\!\perp \mathrm{X}_{V\setminus\{M6,M9\}}$
- $(M2, M3, M4, M5, M7, M8) \perp\!\!\!\perp (Y2, Y3, Y5, Y7, Y8, M1) \mid (Y1, Y4, Y6)$
- $M1 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M8, M9, Y2, Y3, Y4, Y6, Y8) \mid (Y1, Y5, Y7)$
- $M8 \perp\!\!\!\perp (M1, M3, M5, M6, M7, M9, Y1, Y2, Y3, Y4, Y5, Y7, Y8) \mid (M2, M4, Y6)$
- $(M3, M5, M7) \perp\!\!\!\perp (M1, M6, M8, M9, Y1, Y2, Y3, Y5, Y6, Y7, Y8) \mid (M2, M4, Y4)$

3.2.2 Center Macro-Region

Regarding the macro-region Center, the selected network, as shown in Fig. 2 (a2), confirms several conditional independencies that are present at the national level, with an isolated compliance pair (M6, M9) and a strongly interconnected atmospheric/technical block. However, the Center exhibits a stronger coupling between social and environmental missingness: the social indicator M7 attaches to the technical domain via water consumption missingness (M5). In addition, annual revenue missingness (M8) is linked to a small set of technical and organizational variables, suggesting a clearer bridge between economic reporting difficulties and specific aspects of firm organization. Compared with the national graph, the respondent role shows a reduced centrality in shaping missingness patterns, while new links emerge among the missingness indicators themselves, pointing to a more internally structured environmental–social reporting mechanism in the Center.

Interestingly, missingness in women in managerial positions (M7), gas (M3), and water (M5) are jointly independent of the whole set of firm characteristics once electricity (M2), diesel (M4), and annual revenue (M8) are controlled, indicating an autonomous environmental–social response mechanism. Missingness in cultivated land area (M1), together with electricity (M2) and diesel (M4), appears to be jointly independent of certain firms characteristics one conditioning on their neighbors set, suggesting that land-related reporting is primarily explained by energy and size effects. Finally, firm age (Y8) and investment in renewable power plants (Y6) are jointly independent from the rest of the variables once electricity (M2) and consortia membership (Y2) are accounted for, highlighting that technological modernization and organizational maturity represent distinct aspects, weakly connected to the main non-response mechanisms.

Annual revenue (M8) depends on electricity (M2), water (M5), and production role (Y5), reflecting a more restricted neighborhood set than in the national graph. This suggests that for the Center, economic reporting difficulties may be more closely related to technical rather than structural aspects. Compared to the national graph (see Fig. 2 (b2)), the Center exhibits a denser set of connections between social and environmental missingness indicators, alongside a weaker role of the respondent type (Y1), year of foundation (Y8), and investment in renewable power plants (Y6) in shaping organizational dependencies. Overall, the differential graph highlights new edges between social and environmental non-response indicators, as well as the reduced role of firm characteristics in shaping dependencies.

In summary, in the central macro-region, social and environmental missingness indicators show a stronger interdependence. The atmospheric domain indicators structure remains densely connected, while M8 is linked to M2, M5, and Y5, indicating a bridge between technical reporting and firm characteristics. Respondent role (Y1), a strong hub at the national level, is here connected only to the firm structure (Y3, Y4) and soil-related indicator (M1). Accordingly, the key conditional independence statements are as follows.

- $M7 \perp\!\!\!\perp \mathrm{X}_{V\setminus\{M7,M5\}} \mid M5$
- $(M6, M9) \perp\!\!\!\perp \mathrm{X}_{V\setminus\{M6,M9\}}$
- $(M2, M3, M4, M5, M7, M8) \perp\!\!\!\perp (Y1, Y2, Y3, Y4, Y5, Y6, Y7) \mid (M1, Y5, Y8)$
- $(M3, M5, M7) \perp\!\!\!\perp \mathrm{Y} \mid (M2, M4, M8)$
- $(M1, M2, M4) \perp\!\!\!\perp (Y3, Y4, Y6, Y7) \mid (M3, M5, M8, Y1, Y2, Y5, Y8)$
- $(Y6, Y8) \perp\!\!\!\perp \mathrm{X}_{V\setminus\{Y6,Y8,M2,Y2\}} \mid (M2, Y2)$

– $M8 \perp\!\!\!\perp (M1, M3, M4, M6, M7, M9, Y1, Y2, Y3, Y4, Y5, Y7, Y8) \mid (M2, M5, Y6)$

3.2.3 South and Islands Macro-Region

The selected graph for the South and Islands macro-region, shown in Fig. 2 (a3), emphasizes a more modular structure. The social missingness indicator M7 again appears as an isolated node, and the compliance pair (M6, M9) remains disjoint from the rest of the system. The atmospheric/technical missingness block remains cohesive, but it becomes more sharply separated from firm characteristics, with respondent role acting as the main bridge between technical missingness and the remainder of the graph. Missingness in cultivated land area (M1) is primarily associated with firm attributes in this macro-region, rather than being embedded in the technical missingness block.

The structure of conditional independencies among electricity (M2), gas (M3), diesel (M4), and water (M5) consumption remains the same, while annual revenue (M8) presents a smaller neighborhood set. Interestingly, the set of variables formed by (M2, M3, M4, M5, M8) appears to be jointly independent of the rest of the variables once conditioning on the respondent role (Y1). This suggests the importance of the respondent type in this macro-region.

In summary, for the South and Islands, the selected network reveals clearer separation between domains, with a territorial configuration in which technical, compliance, and organizational components operate through more distinct, weakly connected non-response mechanisms (see Fig. 2 (b3)). The key conditional independence statements are as follows. See Appendix B Table 7.

– $M7 \perp\!\!\!\perp \mathrm{X}_{V \setminus \{M7\}}$
– $(M6, M9) \perp\!\!\!\perp \mathrm{X}_{V \setminus \{M6, M9\}}$
– $(M2, M3, M4, M5, M8) \perp\!\!\!\perp (M1, M6, M7, M9, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (Y1)$
– $M1 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M8, M9, Y2, Y3, Y4, Y6) \mid (Y1, Y5, Y7, Y8)$

4 Conclusions

Motivated by the idea that systematic non-response reflects structured signals rather than mere noise, this study analyzed the missingness mechanism of a large-scale survey conducted in 2024 within the Agritech National research center. Our focus on sustainability supposes that non-responses to specific questions can be driven by different attitudes towards sustainability. We treat item non-responses as binary indicators and model their conditional independencies, along with fully observed firm attributes, using undirected (non-causal) Ising graphical models. We view non-response patterns as an indirect signal of firm-level propensity toward sustainability.

Unlike approaches that assume a known missingness graph or aim for causal identification (Mohan & Pearl, 2021), we adopt an undirected Ising graphical model and infer the graph from the data to explore conditional associations, and assess how the missingness mechanism is related to attitudes towards sustainability. Estimation and graph selection use node-wise logistic regressions with BIC-guided backward selection to learn the graph structure, without assuming sparsity. The approach is grounded in recent contributions to graph structural learning (Maathuis et al., 2018; Ravikumar et al., 2010; Roverato, 2017; Van Borkulo et al., 2014) and is validated through a Monte Carlo study.

We first estimate our model at the national level and then replicate the analysis considering three Italian macro-regions: North, Center, and South and Islands. At all levels, missingness in atmospheric indicators (electricity, gas, diesel, and water consumption) turned out to be strongly interconnected. Interestingly, missingness indicators in compliance variables, such as rainwater collection and product certification, are jointly independent of the other variables. Non-response indicator on the presence of women in managerial roles appears in general as a singleton, that is, it is both marginally and conditionally independent of all the other variables. Consequently, allowing for macro-regional specificities, three distinct patterns seem to link sustainability and missingness indicators. The first concerns atmospheric/environmental sustainability (such as power and water consumption), which is linked to power and water consumption and relates to economic return, and is connected to some firm characteristics. The second pattern concerns active sustainability behaviors, such as awareness of a firm's certifications. The third pattern pertains to social sustainability and attitudes toward gender equity in managerial roles. These last two patterns appear to be disjoint from each other and from the other characteristics. These three distinct non-response patterns, environmental, behavioral, and social, suggest that sustainability is a multifaceted phenomenon, characterized by different association structures with firm characteristics.

In our analysis, we estimate the graph structure for each of the three macro-regions independently, refraining from imposing any structural similarity across macro-regions. When the identification of a shared or partially shared structure is of primary interest, the three graphs can instead be estimated jointly using multiple-graph models that allow for borrowing of information across groups. For Bayesian formulations of joint graph learning, see Avalos-Pacheco et al. (2025).

Importantly for a correct interpretation, we reiterate that our approach is intended to characterize patterns of reporting behavior related to the sustainability dimensions covered by the questionnaire. This should not be read as a claim about the underlying missingness generating mechanism for the values of variables with missing data. Accordingly, the proposed approach is not intended to identify a missingness mechanism in the formal sense (MCAR/MAR/MNAR) or to guide for multiple imputation, which concerns the missing values themselves rather than the missingness indicators.

Appendix A Simulation Study

The simulation study was conducted to identify the most appropriate selection procedure for the empirical analysis, in a controlled setting calibrated to the empirical problem size ($n = 2689$, $p = 17$). Three sparsity scenarios were considered, defined relative to the reference edge probability $\pi = 2/(p-1) = 0.125$: high $\pi = 0.0625$, moderate $\pi = 0.125$, and low $\pi = 0.1875$ sparsity levels. For each sparsity scenario, a single undirected ground-truth random graph was generated using the `R` package `huge`, which allows the generation of data from a multivariate normal distribution $\mathcal{N}_p(0, \boldsymbol{\Sigma})$ with different graph structures. The resulting ground-truth graphs corresponding to the three sparsity levels are shown in Fig. 3.

Denoting by $\boldsymbol{\Sigma}_{gt}$ the variance–covariance matrix associated with the ground-truth graph, we then performed 100 Monte Carlo replications in which data were sampled from a multivariate normal distribution $\mathcal{N}_p(0, \boldsymbol{\Sigma}_{gt})$ and subsequently dichotomized by thresholding each variable at zero to obtain binary observations.

For each replication, the network structure was selected using the alternative estimation strategies described in Section 2.2: *i)* node-wise ℓ_1-regularized logistic regression with model selection based on the EBIC (IsingFit & EBIC); *ii)* node-wise ℓ_1-regularized logistic regression with the regularization parameter λ selected by 10-fold cross-validation using the 1-SE rule (Lasso & CV); *iii)* node-wise ℓ_1-regularized logistic regression combined with stability selection (Lasso & SS); *iv)* node-wise ℓ_1-regularized logistic regression combined with complementary pairs stability selection (Lasso & CPSS); and *v)* node-wise stepwise backward logistic regression with model selection based on the BIC (Stepwise & BIC). Specifically, the first strategy was implemented via the `R` package `IsingFit` with the tuning parameter set to $\gamma = 0.25$, as recommended by Van Borkulo et al. (2014). Strategies *iii)* and *iv)* were implemented through the `R` package `stabs` with parameter choices following the theoretical frameworks proposed by Meinshausen and Bühlmann (2010) and Shah and Samworth (2012) respectively. Subsampling was performed without replacement on half-size samples $\lfloor n/2 \rfloor$ with $B = 100$ resampling replicates under the subsampling scheme of Meinshausen and Bühlmann (2010) and $B = 50$ under the CPSS variant. Following the recommendations of these authors, the selection probability threshold was fixed at $\tau = 0.75$, the per-family error rate was controlled at $PFER_{max} = 1$ providing a conservative upper bound on the expected number of falsely selected edges.

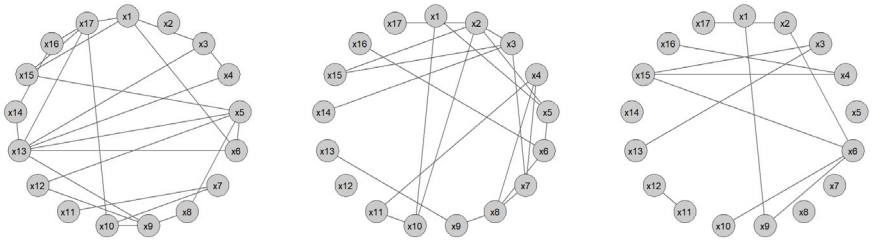
Performance is summarized with standard metrics: sensitivity, specificity, and accuracy calculated as

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{12}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{13}$$

where TP, FP, TN, and FN denote, respectively, the number of true positives, that is, the number of true arcs correctly selected as present, false positives, true negatives, and false negatives.



(a) low (b) moderate (c) high

Fig. 3 True graph structures used in the simulation study under the three sparsity levels: (**a**) low, (**b**) moderate, and (**c**) high sparsity. Each graph represents the ground-truth conditional independence structure corresponding to the specified edge probability

Across the three sparsity scenarios, as reported in Table 3, all selection strategies exhibited satisfactory overall performance, with accuracy levels remaining high throughout. However, distinct patterns emerged across the methods, reflecting the inherent trade-off between sparsity control and edge recovery. In the low-sparsity scenario, Stepwise & BIC, Lasso & CV, and IsingFit & EBIC methods achieved high accuracy and perfect sensitivity, albeit at the cost of slightly reduced specificity. In contrast, stability-based procedures (Stability & SS and Stability & CPSS) produced more conservative graphs, characterized by nearly perfect specificity but markedly lower sensitivity, consistent with the theoretical rationale of stability selection, which prioritizes controlling false positives over detecting weaker connections. As sparsity increased, moving to the moderate and high sparsity scenarios, differences among methods became less pronounced, and all approaches achieved very high accuracy and specificity. In the high-sparsity setting, the Stability & CPSS method proves to be the best in terms of accuracy, while Lasso & CV and Stepwise & BIC performed comparably, providing accurate and stable estimates of the underlying structure. Taking into account these results, and since the empirical phenomenon under investigation is not expected to be excessively sparse, the Stepwise & BIC selection approach was selected for the empirical analysis.

Table 3 Performance of alternative graph selection strategies across the three sparsity scenarios. For each metric, mean values across 100 Monte Carlo replications are reported with standard deviations in parentheses

Sparsity	Method	Accuracy	Sensitivity	Specificity
low	IsingFit & EBIC	0.877 (0.028)	1.000 (0.000)	0.850 (0.034)
	Lasso & CV	0.912 (0.021)	1.000 (0.000)	0.893 (0.026)
	Stability & SS	0.854 (0.008)	0.206 (0.044)	**1.000** (0.000)
	Stability & CPSS	0.891 (0.011)	0.408 (0.059)	0.999 (0.001)
	Stepwise & BIC	**0.963** (0.015)	1.000 (0.000)	0.954 (0.019)
moderate	IsingFit & EBIC	0.930 (0.025)	1.000 (0.000)	0.918 (0.030)
	Lasso & CV	0.960 (0.017)	1.000 (0.000)	0.953 (0.020)
	Stability & SS	0.904 (0.011)	0.347 (0.072)	**1.000** (0.000)
	Stability & CPSS	0.963 (0.005)	0.746 (0.035)	**1.000** (0.001)
	Stepwise & BIC	**0.972** (0.012)	1.000 (0.000)	0.967 (0.014)
high	IsingFit & EBIC	0.956 (0.014)	1.000 (0.000)	0.952 (0.016)
	Lasso & CV	0.978 (0.012)	1.000 (0.000)	0.976 (0.014)
	Stability & SS	0.960 (0.005)	0.510 (0.064)	**1.000** (0.000)
	Stability & CPSS	**0.982** (0.007)	0.797 (0.086)	0.999 (0.003)
	Stepwise & BIC	0.979 (0.010)	1.000 (0.000)	0.977 (0.011)

Appendix B Supplementary Results

In this section, we provide supplementary results complementing Section 3. For each selected graph, we report an alternative visualization of the selected network structure, the corresponding adjacency matrix, and tables summarizing conditional independencies according to the Local Markov property, along with a selection of the most relevant ones according to the Global Markov property. The corresponding selected adjacency matrices are denoted as $\widehat{A}_{NL}$ for the national level, $\widehat{A}_C$ for the Center, $\widehat{A}_N$ for the North, and $\widehat{A}_{SI}$ for the South and Islands macro-regional levels. Moreover, to assess the stability of the selected graphs, we conducted a stability check. Specifically, for each territorial disaggregation, we generated 1000 bootstrap resamples and re-estimated the graph using the nodewise stepwise procedure. Table 8 reports the bootstrap mean accuracy, sensitivity, and specificity, with standard deviations in parentheses, across the territorial levels.

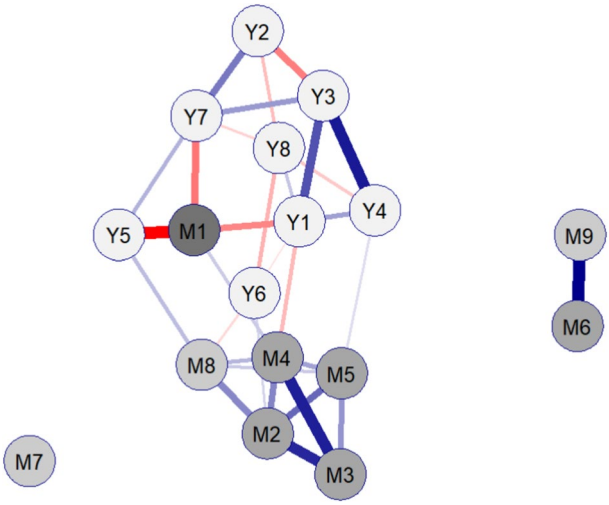


Fig. 4 Different representation of the national level selected graph structure reported in Fig. 1. Blue edges denote positive associations, red edges denote negative associations. Edges' thickness corresponds to the strength of the associations

$$\widehat{\mathbf{A}}_{NL} = \begin{array}{c} \\ M1 \\ M2 \\ M3 \\ M4 \\ M5 \\ M6 \\ M7 \\ M8 \\ M9 \\ Y1 \\ Y2 \\ Y3 \\ Y4 \\ Y5 \\ Y6 \\ Y7 \\ Y8 \end{array} \begin{array}{c} \begin{array}{cccccccccccccccc} M1 & M2 & M3 & M4 & M5 & M6 & M7 & M8 & M9 & Y1 & Y2 & Y3 & Y4 & Y5 & Y6 & Y7 & Y8 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

Fig. 5 Adjacency matrix selected for the graph at the national level

Table 4 Conditional independencies for the national-level selected graph by Global and Local Markov Property. For the Global Markov Property we only report some relevant conditional independencies

Markov Property	Conditional independence statement
Global	$(M1, M4, M8) \perp\!\!\!\perp (Y2, Y3, Y4, Y8) \mid (M2, M3, M5, Y1, Y5, Y6, Y7)$
	$(M2, M4, M5, M8) \perp\!\!\!\perp (Y2, Y3, Y7, Y8) \mid (M1, M3, Y1, Y4, Y5, Y6)$
	$(M2, M3, M4, M5) \perp\!\!\!\perp (Y2, Y3, Y5, Y7, Y8) \mid (M1, M8, Y1, Y4, Y6)$
	$(M6, M9) \perp\!\!\!\perp \mathrm{X}_{V \setminus (M6, M9)}$
	$(M6, M7) \perp\!\!\!\perp \mathrm{X}_{V \setminus \{M6, M7, M9\}} \mid (M9)$
Local	$M1 \perp\!\!\!\perp (M2, M3, M5, M6, M7, M8, M9, Y2, Y3, Y4, Y6, Y8) \mid (M4, Y1, Y5, Y7)$
	$M2 \perp\!\!\!\perp (M1, M6, M7, M9, Y1, Y2, Y3, Y4, Y5, Y7, Y8) \mid (M3, M4, M5, M8, Y6)$
	$M3 \perp\!\!\!\perp (M1, M6, M7, M8, M9, Y1, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (M2, M4, M5)$
	$M4 \perp\!\!\!\perp (M6, M7, M9, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (M1, M2, M3, M5, M8, Y1)$
	$M5 \perp\!\!\!\perp (M1, M6, M7, M9, Y1, Y2, Y3, Y5, Y6, Y7, Y8) \mid (M2, M3, M4, M8, Y4)$
	$M6 \perp\!\!\!\perp \mathrm{X}_{V \setminus \{M6, M9\}} \mid (M9)$
	$M7 \perp\!\!\!\perp \mathrm{X}_{V \setminus \{M7\}}$
	$M8 \perp\!\!\!\perp (M1, M3, M6, M7, M9, Y1, Y2, Y3, Y4, Y7, Y8) \mid (M2, M4, M5, Y5, Y6)$
	$M9 \perp\!\!\!\perp \mathrm{X}_{V \setminus \{M9, M6\}} \mid (M6)$
	$Y1 \perp\!\!\!\perp (M2, M3, M5, M6, M7, M8, M9, Y2, Y5, Y7) \mid (M1, M4, Y3, Y4, Y6, Y8)$
	$Y2 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y1, Y4, Y5, Y6) \mid (Y3, Y7, Y8)$
	$Y3 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y5, Y6, Y8) \mid (Y1, Y2, Y4, Y7)$
	$Y4 \perp\!\!\!\perp (M1, M2, M3, M4, M6, M7, M8, M9, Y2, Y5, Y6, Y7) \mid (M5, Y1, Y3, Y8)$
	$Y5 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M9, Y1, Y2, Y3, Y4, Y6, Y8) \mid (M1, M8, Y7)$
	$Y6 \perp\!\!\!\perp (M1, M3, M4, M5, M6, M7, M9, Y2, Y3, Y4, Y5, Y7) \mid (M2, M8, Y1, Y8)$
	$Y7 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M8, M9, Y1, Y4, Y6) \mid (M1, Y2, Y3, Y5, Y8)$
	$Y8 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y3, Y5) \mid (Y1, Y2, Y4, Y6, Y7)$

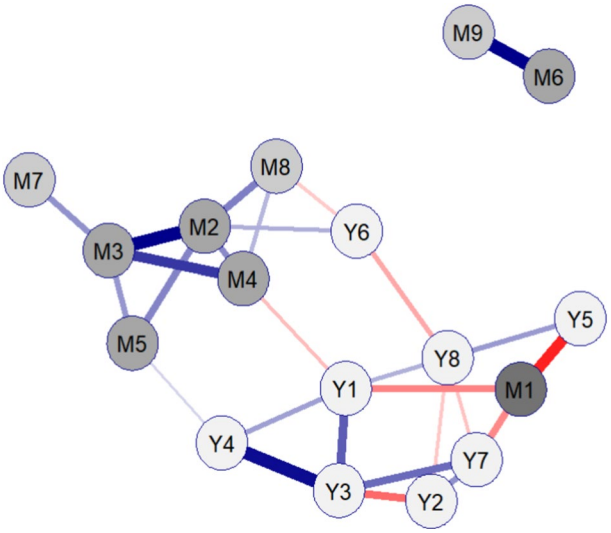


Fig. 6 Different representation of the selected graph structure for the North macro-region reported in Fig. 2 (a1). Blue edges denote positive associations, red edges denote negative associations. Edges' thickness corresponds to the strength of the associations

$$\widehat{\mathbf{A}}_N = \begin{array}{c} \\ M1 \\ M2 \\ M3 \\ M4 \\ M5 \\ M6 \\ M7 \\ M8 \\ M9 \\ Y1 \\ Y2 \\ Y3 \\ Y4 \\ Y5 \\ Y6 \\ Y7 \\ Y8 \end{array} \begin{array}{c} \begin{array}{cccccccccccccccc} M1 & M2 & M3 & M4 & M5 & M6 & M7 & M8 & M9 & Y1 & Y2 & Y3 & Y4 & Y5 & Y6 & Y7 & Y8 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

Fig. 7 Adjacency matrix selected for the graph of the North macro-region

Table 5 Conditional independencies of the selected graph for the North macro-region by Global and Local Markov Property. For the Global Markov Property we only report some relevant conditional independencies

Markov Property	Conditional independence statement
Global	$(M6, M9) \perp\!\!\!\perp \mathrm{X}_V$
	$(M2, M3, M4, M5, M7, M8) \perp\!\!\!\perp (Y2, Y3, Y5, Y7, Y8, M1) \mid (Y1, Y4, Y6)$
	$(M2, M4, M8) \perp\!\!\!\perp (Y2, Y3, Y5, Y7, Y8, M1) \mid (Y1, Y4, Y6)$
	$(M3, M5, M7) \perp\!\!\!\perp (M1, M6, M8, M9, Y1, Y2, Y3, Y5, Y6, Y7, Y8) \mid (M2, M4, Y4)$
	$(M2, M4, M8) \perp\!\!\!\perp (M1, M6, M7, M9, Y2, Y3, Y4, Y5, Y7, Y8) \mid (M3, M5, Y1, Y6)$
Local	$M1 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M8, M9, Y2, Y3, Y4, Y6, Y8) \mid (Y1, Y5, Y7)$
	$M2 \perp\!\!\!\perp (M1, M6, M7, M9, Y1, Y2, Y3, Y4, Y5, Y7, Y8) \mid (M3, M4, M5, M8, Y6)$
	$M3 \perp\!\!\!\perp (M1, M6, M8, M9, Y1, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (M2, M4, M5, M7)$
	$M4 \perp\!\!\!\perp (M1, M5, M6, M7, M9, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (M2, M3, M8, Y1)$
	$M5 \perp\!\!\!\perp (M1, M4, M6, M7, M8, M9, Y1, Y2, Y3, Y5, Y6, Y7, Y8) \mid (M2, M3, Y4)$
	$M6 \perp\!\!\!\perp X_{V \setminus \{M6, M9\}} \mid (M9)$
	$M7 \perp\!\!\!\perp X_{V \setminus \{M7, M3\}} \mid (M3)$
	$M8 \perp\!\!\!\perp (M1, M3, M5, M6, M7, M9, Y1, Y2, Y3, Y4, Y5, Y7, Y8) \mid (M2, M4, Y6)$
	$M9 \perp\!\!\!\perp X_{V \setminus \{M9, M6\}} \mid (M6)$
	$Y1 \perp\!\!\!\perp (M2, M3, M5, M6, M7, M8, M9, Y2, Y5, Y6, Y7) \mid (M1, M4, Y3, Y4, Y8)$
	$Y2 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y1, Y4, Y5, Y6) \mid (Y3, Y7, Y8)$
	$Y3 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y5, Y6, Y8) \mid (Y1, Y2, Y4, Y7)$
	$Y4 \perp\!\!\!\perp (M1, M2, M3, M4, M6, M7, M8, M9, Y2, Y5, Y6, Y7, Y8) \mid (M5, Y1, Y3)$
	$Y5 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M8, M9, Y1, Y2, Y3, Y4, Y6, Y7) \mid (M1, Y8)$
	$Y6 \perp\!\!\!\perp (M1, M3, M4, M5, M6, M7, M9, Y1, Y2, Y3, Y4, Y5, Y7) \mid (M2, M8, Y8)$
	$Y7 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M8, M9, Y1, Y4, Y5, Y6) \mid (M1, Y2, Y3, Y8)$
	$Y8 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y3, Y4) \mid (Y1, Y2, Y5, Y6, Y7)$

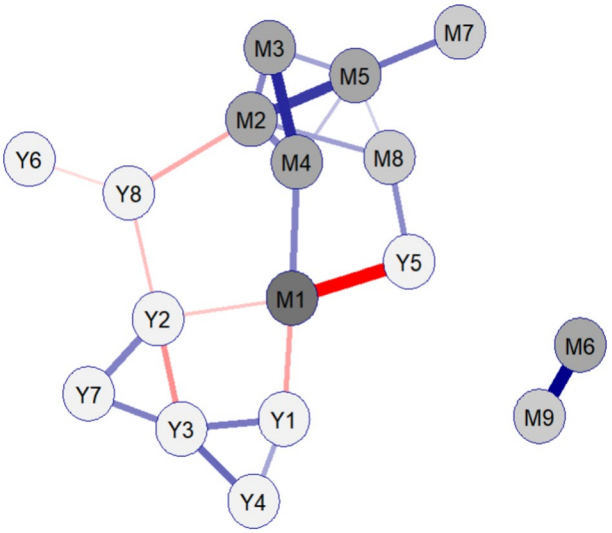


Fig. 8 Different representation of the selected graph structure for the Center macro-region reported in Fig. 2 (a2). Blue edges denote positive associations, red edges denote negative associations. Edges' thickness corresponds to the strength of the associations

$$\widehat{\mathbf{A}}_C = \begin{array}{c} \\ M1 \\ M2 \\ M3 \\ M4 \\ M5 \\ M6 \\ M7 \\ M8 \\ M9 \\ Y1 \\ Y2 \\ Y3 \\ Y4 \\ Y5 \\ Y6 \\ Y7 \\ Y8 \end{array} \begin{array}{c} \begin{array}{cccccccccccccccc} M1 & M2 & M3 & M4 & M5 & M6 & M7 & M8 & M9 & Y1 & Y2 & Y3 & Y4 & Y5 & Y6 & Y7 & Y8 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

Fig. 9 Adjacency matrix selected for the graph of the Center macro-region

Table 6 Conditional independencies of the selected graph for the Center macro-region by Global and Local Markov Property. For the Global Markov Property we only report some relevant conditional independencies

Markov Property	Conditional independence statement
Global	$(M2, M3, M4, M5, M7, M8) \perp\!\!\!\perp (Y1, Y2, Y3, Y4, Y5, Y6, Y7) \mid (M1, Y5, Y8)$
	$(M6, M9) \perp\!\!\!\perp X_{V\setminus\{M6,M9\}}$
	$(M3, M5, M7) \perp\!\!\!\perp \mathrm{Y} \mid (M2, M4, M8)$
	$(M1, M2, M4) \perp\!\!\!\perp (Y3, Y4, Y6, Y7) \mid (M3, M5, M8, Y1, Y2, Y5, Y8)$
	$(Y6, Y8) \perp\!\!\!\perp \mathrm{X}_{V\setminus\{Y6,Y8,M2,Y2\}} \mid (M2, Y2)$
Local	$M1 \perp\!\!\!\perp (M2, M3, M5, M6, M7, M8, M9, Y3, Y4, Y6, Y7, Y8) \mid (M4, Y1, Y2, Y5)$
	$M2 \perp\!\!\!\perp (M1, M6, M7, M9, Y1, Y3, Y4, Y5, Y6, Y7) \mid (M3, M4, M5, M8, Y8)$
	$M3 \perp\!\!\!\perp (M1, M6, M7, M8, M9, Y1, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (M2, M4, M5)$
	$M4 \perp\!\!\!\perp (M6, M7, M8, M9, Y1, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (M1, M2, M3, M5)$
	$M5 \perp\!\!\!\perp (M1, M6, M9, Y1, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (M2, M3, M4, M7, M8)$
	$M6 \perp\!\!\!\perp \mathrm{X}_{V\setminus\{M6,M9\}} \mid (M9)$
	$M7 \perp\!\!\!\perp \mathrm{X}_{V\setminus\{M7,M5\}} \mid (M5)$
	$M8 \perp\!\!\!\perp (M1, M3, M4, M6, M7, M9, Y1, Y2, Y3, Y4, Y5, Y7, Y8) \mid (M2, M5, Y6)$
	$M9 \perp\!\!\!\perp \mathrm{X}_{V\setminus\{M9,M6\}} \mid (M6)$
	$Y1 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M8, M9, Y2, Y5, Y6, Y7, Y8) \mid (M1, Y3, Y4)$
	$Y2 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M8, M9, Y1, Y3, Y4, Y5, Y6) \mid (M1, Y7, Y8)$
	$Y3 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y5, Y6, Y8) \mid (Y1, Y2, Y4, Y7)$
	$Y4 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y2, Y5, Y6, Y7, Y8) \mid (Y1, Y3)$
	$Y5 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M9, Y1, Y2, Y3, Y4, Y6, Y7, Y8) \mid (M1, M8)$
	$Y6 \perp\!\!\!\perp \mathrm{X}_{V\setminus\{Y6,Y8\}} \mid (Y8)$
	$Y7 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y1, Y4, Y5, Y6, Y8) \mid (Y2, Y3)$
	$Y8 \perp\!\!\!\perp (M1, M3, M4, M5, M6, M7, M8, M9, Y4, Y5, Y6, Y7) \mid (M2, Y2)$

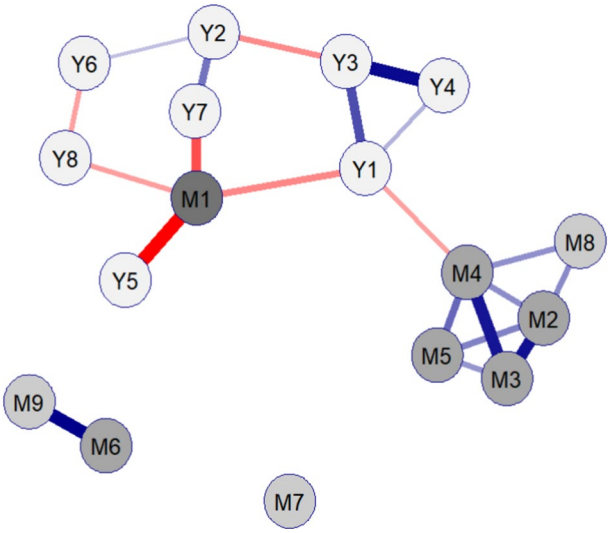


Fig. 10 Different representation of the selected graph structure for the South and Islands macro-region reported in Fig. 2 (a3). Blue edges denote positive associations, red edges denote negative associations. Edges' thickness corresponds to the strength of the associations

$$\widehat{\mathrm{A}}_{SI} = \begin{array}{c} \\ M1 \\ M2 \\ M3 \\ M4 \\ M5 \\ M6 \\ M7 \\ M8 \\ M9 \\ Y1 \\ Y2 \\ Y3 \\ Y4 \\ Y5 \\ Y6 \\ Y7 \\ Y8 \end{array} \begin{array}{c} \begin{array}{cccccccccccccccc} M1 & M2 & M3 & M4 & M5 & M6 & M7 & M8 & M9 & Y1 & Y2 & Y3 & Y4 & Y5 & Y6 & Y7 & Y8 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

Fig. 11 Adjacency matrix selected for the graph of the South and Islands macro-region

Table 7 Conditional independencies of the selected graph for the South and Islands macro-region by Global and Local Markov Property. For the Global Markov Property we only report some relevant conditional independencies.

Markov Property	Conditional independence statement
Global	$(M6, M9) \perp\!\!\!\perp \mathrm{X}_{V\setminus\{M6,M9\}}$
	$(M6, M7) \perp\!\!\!\perp \mathrm{X}_{V\setminus\{M6,M7,M9\}} \mid (M9)$
	$(M2, M3, M4, M5, M8) \perp\!\!\!\perp (M1, M6, M7, M9, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (Y1)$
Local	$M1 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M8, M9, Y2, Y3, Y4, Y6) \mid (Y1, Y5, Y7, Y8)$
	$M2 \perp\!\!\!\perp (M1, M6, M7, M9, Y1, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (M3, M4, M5, M8)$
	$M3 \perp\!\!\!\perp (M1, M6, M7, M8, M9, Y1, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (M2, M4, M5)$
	$M4 \perp\!\!\!\perp (M1, M6, M7, M9, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (M2, M3, M5, M8, Y1)$
	$M5 \perp\!\!\!\perp (M1, M6, M7, M8, M9, Y1, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (M2, M3, M4)$
	$M6 \perp\!\!\!\perp X_{V\setminus\{M6,M9\}} \mid (M9)$
	$M7 \perp\!\!\!\perp X_{V\setminus\{M7\}}$
	$M8 \perp\!\!\!\perp (M1, M3, M5, M6, M7, M9, Y1, Y2, Y3, Y4, Y5, Y6, Y7, Y8) \mid (M2, M4)$
	$M9 \perp\!\!\!\perp X_{V\setminus\{M9,M6\}} \mid (M6)$
	$Y1 \perp\!\!\!\perp (M2, M3, M5, M6, M7, M8, M9, Y2, Y5, Y6, Y7, Y8) \mid (M1, M4, Y3, Y4)$
	$Y2 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y1, Y3, Y5, Y6) \mid (Y4, Y7, Y8)$
	$Y3 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y5, Y6, Y7, Y8) \mid (Y1, Y2, Y4)$
	$Y4 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y2, Y5, Y6, Y7, Y8) \mid (Y1, Y3)$
	$Y5 \perp\!\!\!\perp X_{V\setminus\{Y5,M1\}} \mid (M1)$
	$Y6 \perp\!\!\!\perp (M1, M2, M3, M4, M5, M6, M7, M8, M9, Y1, Y2, Y4, Y5, Y7) \mid (Y3, Y8)$
	$Y7 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M8, M9, Y1, Y3, Y4, Y5, Y6) \mid (M1, Y2)$
	$Y8 \perp\!\!\!\perp (M2, M3, M4, M5, M6, M7, M8, M9, Y1, Y2, Y3, Y4, Y5, Y7) \mid (M1, Y6)$

Table 8 Stability of the selected graph across the four territorial disaggregation. For each metric, mean values across 1000 bootstrap resamples are reported with standard deviations in parentheses.

Territorial level	Accuracy	Sensitivity	Specificity
National	0.93 (0.02)	0.82 (0.07)	0.95 (0.02)
North	0.92 (0.02)	0.86 (0.06)	0.94 (0.02)
Center	0.93 (0.02)	0.87 (0.05)	0.95 (0.02)
South & Islands	0.93 (0.02)	0.91 (0.06)	0.94 (0.02)

Acknowledgements This study was partly conducted in the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)–MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4—D.D. 1032 17/06/2022, CN00000022). Additional financial support was provided by the MUR-PRIN grant 2022 SMNNKY, CUP B53D23009470006 Unione europea- Next Generation EU, Missione 4 Componente 2 - CUP B53D23009470006. This manuscript reflects solely the authors' views and opinions, not the positions of the European Union or the European Commission.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement.

Data Availability The data generated by the survey research and analyzed during the current study are available on reasonable request from the corresponding author.

Declarations

Conflicts of Interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Avalos-Pacheco, A., Lazzerini, A., Lupparelli, M., et al. (2025). Bayesian inference of multiple ising models for heterogeneous public opinion survey networks. *Journal of the Royal Statistical Society: Series C: Applied Statistics, 74*(5), 1395–1426. https://doi.org/10.1093/jrsssc/qlaf028

Banterle, A., Stranieri, S., & Baldi, L. (2006). Voluntary traceability and transaction costs: an empirical analysis in the italian meat processing supply chain. In *99th Seminar of the European Association of Agricultural Economists, Bonn, Germany*. https://doi.org/10.22004/ag.econ.7722

Banzato, E., Roverato, A., Buja, A., et al. (2025). Understanding multimorbidity: insights with graphical models. *BMC Medical Research Methodology,25*(84). https://doi.org/10.1186/s12874-025-02536-y

Betti, G., Evangelista, D., Gagliardi, F., et al. (2024). Towards integrating information systems of statistical indicators on traceability, quality and safety of italian agrifood systems for citizens, institutions and policy-makers. *Sustainability, 16*(15), 6330. https://doi.org/10.3390/su16156330

Betti, G., Gagliardi, F., Mecca, A., et al. (2025). Sustainability and circularity of the agri-food systems: How to measure it? a first attempt on the italian system. *Sustainability, 17*(7), 3169. https://doi.org/10.3390/su17073169

Brusco, M. J., Steinley, D., & Watts, A. L. (2023). A comparison of logistic regression methods for ising model estimation. *Behavior Research Methods, 55*, 3566–3584. https://doi.org/10.3758/s13428-022-01976-4

Chen, J., & Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika, 95*(3), 759–771. https://doi.org/10.1093/biomet/asn034

Darroch, J. N., Lauritzen, S. L., & Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Annals of Statistics, 8*(3), 522–539. https://doi.org/10.1214/aos/1176345006

Do Canto, N., Bossle, M., Vieira, L., et al. (2021). Supply chain collaboration for sustainability: a qualitative investigation of food supply chains in brazil. *Management of Environmental Quality: An International Journal, 32*(6), 1210–1232. https://doi.org/10.1108/MEQ-12-2019-0275

European Commission (2019). The european green deal (com(2019) 640 final). European Commission.

Foygel, R., & Drton, M. (2010). Extended bayesian information criteria for gaussian graphical models. In J. Lafferty, C. Williams, J. Shawe-Taylor, et al. (Eds.) *Advances in Neural Information Processing Systems,* vol 23. Curran Associates Inc. https://proceedings.neurips.cc/paper_files/paper/2010/file/072b030ba126b2f4b2374f342be9ed44-Paper.pdf.

Gallo, A., Accorsi, R., Goh, A., et al. (2021). A traceability-support system to control safety and sustainability indicators in food distribution. *Food Control, 124*, Article 107866. https://doi.org/10.1016/j.foodcont.2021.107866

Grimm, C. (2025). Unraveling circular conundrums with a cheeky twist: Proposal for a new way of measuring circular economy efforts at the product level within procurement-to-waste system boundaries–a case study from the airline industry. *Sustainability, 17*(3), 807. https://doi.org/10.3390/su17030807

Grimm, J., Hofstetter, J., & Sarkis, J. (2016). Exploring sub-suppliers' compliance with corporate sustainability standards. *Journal of Cleaner Production, 112*, 1971–1984. https://doi.org/10.1016/j.jclepro.2014.11.036

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edn. Springer Series in Statistics, Springer, New York. https://doi.org/10.1007/978-0-387-84858-7

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik, 31*(1), 253–258. https://doi.org/10.1007/BF02980577

Lauritzen, S. L. (1996). *Graphical Models*. Oxford Statistical Science Series: Oxford University Press, Oxford.

León Bravo, V., Moretto, A., & Caniato, F. (2021). A roadmap for sustainability assessment in the food supply chain. *British Food Journal, 123*(13), 199–220. https://doi.org/10.1108/BFJ-04-2020-0293

Liu, S., Xie, Y., & Liang, W. (2024). Optimisation of the circular economy based on the resource circulation equation. *Sustainability, 16*(15), 6514. https://doi.org/10.3390/su16156514

Maathuis, M. H., Drton, M., Lauritzen, S. L., et al. (2018). *Handbook of Graphical Models*. Boca Raton: Chapman and Hall/CRC.

Mania, I., Delgado, A.M., Barone, C., et al. (2018). The extra tool–a practical example of extended food traceability for cheese productions. In *Traceability in the Dairy Industry in Europe*. Springer, Cham, pp. 29–66. https://doi.org/10.1007/978-3-030-00446-0_3

Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72*(4), 417–473. https://doi.org/10.1111/j.1467-9868.2010.00740.x

Mohan, K., & Pearl, J. (2021). Graphical models for processing missing data. *Journal of the American Statistical Association, 116*(534), 1023–1037. https://doi.org/10.1080/01621459.2021.1874961

Poponi, S., Arcese, G., Pacchera, F., et al. (2022). Evaluating the transition to the circular economy in the agri-food sector: Selection of indicators. *Resources, Conservation and Recycling, 176*, Article 105916. https://doi.org/10.1016/j.resconrec.2021.105916

Ravikumar, P., Wainwright, M., & Lafferty, J. (2010). High-dimensional ising model selection using regularized logistic regression. *Annals of Statistics, 38*(3), 1287–1319. https://doi.org/10.1214/09-AOS691

Roverato, A. (2017). Graphical Models for Categorical Data. *SemStat Elements, Cambridge University Press, Cambridge,*. https://doi.org/10.1017/9781108277495

Shah, R. D., & Samworth, R. J. (2012). Variable selection with error control: Another look at stability selection. *Journal of the Royal Statistical Society, Series B: Statistical Methodology, 75*(1), 55–80. https://doi.org/10.1111/j.1467-9868.2011.01034.x

Tessitore, S., Iraldo, F., Apicella, A., et al. (2022). Food traceability as driver for the competitiveness in italian food service companies. *Journal of Foodservice Business Research, 25*(1), 57–84. https://doi.org/10.1080/15378020.2021.1918536

United Nations (2015). Transforming our world: The 2030 agenda for sustainable development. resolution a/res/70/1. United Nations.

Van Borkulo, C. D., Borsboom, D., Epskamp, S., et al. (2014). A new method for constructing networks from binary data. *Scientific Reports, 4*(1), 5918. https://doi.org/10.1038/srep05918

Authors and Affiliations

Andrea Mecca[1,2] · Anna Gottard[1] · Francesca Gagliardi[2]

✉ Andrea Mecca
andrea.mecca@unifi.it

✉ Anna Gottard
anna.gottard@unifi.it

✉ Francesca Gagliardi
gagliardi10@unisi.it

[1] University of Florence, Florence, Italy

[2] University of Siena, Siena, Italy