Electronic Journal of Statistics
Vol. 18 (2024) 4382–4406
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2299

Path-dependent parametric decompositions in Ising models

Monia Lupparelli

Department of Statistics, Computer Science, Applications, University of Florence
PO BOX 50134, Florence, Italy,
e-mail: monia.lupparelli@unifi.it

Giovanni M. Marchetti

Department of Statistics, Computer Science, Applications, University of Florence
PO BOX 50134, Florence, Italy,
e-mail: giovanni.marchetti@unifi.it

Abstract: The analysis of paths in undirected graph models can be used to quantify the relevance of the strength of association in multiple paths connecting a pair of vertices of the graph. Some results are available in multivariate Gaussian settings as the covariance of two variables can be decomposed into the sum of measures related to paths joining the variables of the underlying graph. This paper studies the analysis of paths in undirected graph models for binary data, with special focus on Ising models, where the propagation of the variable status through multiple paths joining a pair of vertices is an aspect of interest. A novel logistic regression approach for baseline events in multi-way tables is proposed to show that a parameter of pairwise association can be computed by the sum of components related to paths. These components are based on products of odds ratios which are typically used to measure the dependence represented by the edges in Ising models. Specifically, two parametric decompositions are developed to gain insight on a twofold aspect of interest: the relevance of the multivariate dependence within each path connecting a pair of vertices and the interaction between the multivariate dependence in each path and in the rest of the graph. The results are illustrated through an application to cyber-security risk assessment in industrial networks.

Keywords and phrases: Active path, collpased table, dependence ratio, mean parameter decomposition, odds ratio, undirected graph, Yule's measure.

Received October 2022.

1. Introduction

Graphical Markov models are extensively used in contexts of high complexity to reveal independence structures in multivariate distributions; see Wermuth and Cox (2015) and Maathuis et al. (2019) for recent literature reviews and applications in several scientific fields. Undirected graph models (Lauritzen, 1996) represent a relevant exception where missing edges between vertices of the graph correspond to conditional independencies for the variables associated to the vertex set. Non-missing edges are also relevant for the specification of graph attributes as paths or cliques. The set of paths joining two vertices is of special interest since it gives a picture of the relationship between the related pair of variables and provides insight to interpret the multivariate dependence of variables in the path setting (Cox and Wermuth, 2008).

In graphical modeling, path analysis has a long tradition and dates back to Wright (1921) who studied causal interpretation of directed paths in directed acyclic graph models. The use of path analysis in Gaussian undirected graph models has been introduced only more recently by Jones and West (2005) to quantify the dependence relationships in multiple paths. They propose a decomposition of the covariance for a pair of variables into the sum of path weights which are function of the partial correlation coefficients related to the edge set of each path. A broad analysis and interpretation of this result have been later provided by Roverato and Castelo (2020) who derive a generalized decomposition of a measure of association, named the inflated correlation coefficient.

This paper discusses the use of path analysis in undirected graph models for discrete distributions, which, to the best of our knowledge, is not explored in the literature. Specifically, an approach for the analysis of paths is proposed in Ising models (Ising, 1925) for binary data to study the intensity of the propagation of the variable status in multiple paths (Marsman et al., 2018). The Ising model, originally introduced to represent the joint distribution of magnetic solid materials in a lattice (Besag, 1974), is an undirected graph model for atomic variables where the probability of becoming "active" or "non-active" for each variable depends only on the status of its neighbours (Ravikumar, Wainwright and Lafferty, 2010). Regardless of the applications in statistical physics, these models are increasingly used to study the co-occurrence of patterns of homologous variables as symptoms, psychological traits or disorders (Kruis and Maris, 2021). They are also employed for the analysis of catastrophic or cybersecurity risks (Denuit and Robert, 2022), as well as to model multivariate binary autoregressive time series in finance (Campajiola et al., 2021). Paths identify special patterns of variables in a graph and define a multiple way for the transition of the variable status between two vertices. The proposed approach to path analysis provides a decomposition of a parameter of marginal association between two variables into the sum of interpretable measures which can be used to compare the intensity of the transition of the variable status across different paths.

The pairwise marginal probability is a mean parameter for the Ising model and can be considered as a measure of marginal association whose strength is determined with respect to the departure from the independence condition, according to the literature on the dependence ratio parameter (Ekholm, Smith and McDonald, 1995). Two parametric decompositions for the probability of the co-occurrence of a couple of variables are derived based on the path setting joining the related vertices. A novel logistic regression approach for baseline events in multi-way tables is developed and the properties are formally derived to prove that the proposed decompositions, based on path dependence measures, are function of the odds ratio parameters which are typically used to detect conditional associations in binary undirected graph models (Whittaker, 1990). In turn, these decompositions are alternative path-dependent mappings between the mean and the canonical parameter in Ising models. The first decomposition provides insight to quantify the relevance of the multivariate dependence in each path and to interpret the propagation of the variable status in multiple paths. The second decomposition is complemented by the inclusion of a parameter inspired on Yule's measure of association (Yule, 1900) with the aim to shed light on the relationship between the dependence structure of the variables inside and outside the path. An application of these results is discussed for the analysis of cyber-security risk in an industrial network system where the propagation of a cyber attack can run through the path setting of the underlying graph.

2. Preliminaries

2.1. Background and notation

Given a finite set V, let $X_V = \{X_j\}_{j\in V}$ be a binary random vector with levels $i \in \mathcal{I}_V$, where $\mathcal{I}_V = \{0,1\}^{|V|}$ is a $2^{|V|}$ probability table. Any variable X_j is said to be *active* when it takes level 1 and *non-active* otherwise. Let $\pi = (\pi_D)_{D\subseteq V}$ be the probability parameter for the joint distribution of X_V, where the generic element

$$\pi_D = P(X_D = 1_D, X_{V\setminus D} = 0_{V\setminus D}), \qquad D \subseteq V$$

is the probability of the event associated to the cell i_D of the table $\mathcal{I}_V$; 1_D defines a vector of 1s of size $|D|$ and $0_{V\setminus D}$ is defined accordingly. Then, π_D is the probability that the random vector X_D is active and the rest of variables $X_{V\setminus D}$ is non-active. If the vector X_V follows a multivariate Bernoulli distribution $P(X_V;\pi)$, let $\log\psi = (\log\psi_D)_{D\subseteq V}$ be the log-linear parameter vector which is the canonical parameter in the exponential family theory; see Barndorff-Nielsen (1978). The mapping between the probability and canonical parameter is based on the Möbius transformation (Lauritzen, 1996) that is a log-linear transformation of the probabilities $\pi_D \in \pi$,

$$\log\psi_D = \sum_{D'\subseteq D} (-1)^{|D\setminus D'|} \log\pi_{D'}, \qquad \log\pi_D = \sum_{D'\subseteq D} \log\psi_{D'}, \qquad D\subseteq V. \quad (1)$$

Example 1 *Given two binary variables $\{X_1, X_2\}$, let $\pi = (\pi_\emptyset, \pi_1, \pi_2, \pi_{12})^T$ be the set of joint probabilities. The vector $\log\psi = (\log\psi_\emptyset, \log\psi_1, \log\psi_2, \log\psi_{12})^T$ of canonical parameters is obtained as a suitable log-linear combination of joint probabilities:*

$$\log\psi_\emptyset = \log\pi_\emptyset, \qquad \log\psi_1 = \log\frac{\pi_1}{\pi_\emptyset},$$
$$\log\psi_2 = \log\frac{\pi_2}{\pi_\emptyset}, \qquad \log\psi_{12} = \log\frac{\pi_{11}\times\pi_\emptyset}{\pi_1\times\pi_2}.$$

The non-active status of the variables is used as baseline event $\{X_1 = 0, X_2 = 0\}$ to compute the log-linear combinations.

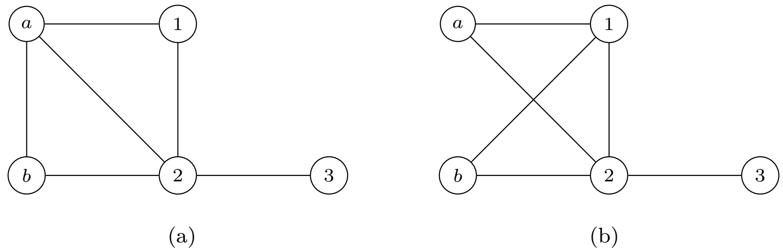

FIG 1. *Ising graphical models for the random vector $X_V = \{X_a, X_b, X_1, X_2, X_3\}$; variables X_a, X_b represent the pair of interest for the analysis of paths.*

For any $j \in V$, the parameter ψ_j is the odds for the variable X_j, for any pair $j, k \in V$, the parameter ψ_{jk} is the conditional odds ratio for the variables X_j, X_k, both computed in the full table $\mathcal{I}_V$. The odds ratio measures the strength of association between X_j and X_k given a fixed value, the baseline non-active level $X_{V\setminus jk} = 0_{V\setminus jk}$, of the remaining set of variables. The odds ratio is often interpreted on the log-scale because $\log \psi_{jk} > 0$ denotes a positive association between X_j and X_k as well as a negative association is given by $\log \psi_{jk} < 0$. Parameters ψ_D, with $D \subseteq V$ and $|D| > 2$, are higher-order measures of association based on combination of odds ratios. Independence models can be specified by imposing zero constraints on log-linear parameters (Whittaker, 1990), specifically, for any pair $j, k \in V$,

$$X_j \perp\!\!\!\perp X_k | X_{V\setminus jk} \quad \text{if and only if} \quad \log \psi_D = 0, \quad j, k \in D \subseteq V.$$

Given the product $\Psi = \prod_{D\subseteq V} \psi_D$, let Ψ_D be any sub-product including only terms ψ_E with $E \subseteq D$, for any $D \subseteq V$. From equation (1), $\pi_D = \Psi_D$ is the probability that only variables in X_D are active. The mean parameter of the joint distribution of X_V is defined by the vector $\mu = (\mu_D)_{D\subseteq V}$ where

$$\mu_D = P(X_D = 1_D), \qquad D \subseteq V,$$

is the probability that the event i_D in the marginal table $\mathcal{I}_D = \{0,1\}^{|D|}$ occurs, that is, the probability that variables in X_D are active regardless of the status of other variables. For any pair D, E of disjoint subsets of V, the conditional probability

$$\pi_{D|E} = P(X_D = 1_D, X_{V\setminus D\cup E} = 0_{V\setminus D\cup E} | X_E = 1_E), \qquad D, E \subseteq V,\ D \cap E = \emptyset,$$

denotes the probability of the event i_D in the slice of the table $\mathcal{I}_V$ defined by $i_E \in \mathcal{I}_E$. Then, $\pi_{D|E}$ is the probability that only variables in X_D are active given that variables in X_E are active as well.

The random vector X_V is associated with a given undirected graph $\mathcal{G} = (V, \mathcal{E})$ where V is a set of vertices/nodes and $\mathcal{E}$ is a set of pairs of vertices; if the couple $\{j, k\} \in \mathcal{E}$, the vertices j and k of the graph are joined by an undirected edge; see Figure 1. A *path* δ is defined by an ordered sequence $(\delta_1, \delta_2, \ldots, \delta_m)$ of distinct vertices, with $\delta_1, \delta_2, \dots, \delta_m \in V$, such that any couple of adjacent vertices in the sequence is joined by an edge in $\mathcal{G}$. Repetition of vertices is not allowed in this definition of paths. Let Δ_{ab} be the set of all paths with endpoints $a, b \in V$ where the generic element of the set is a path $\delta = (\delta_1, \dots, \delta_m)$ with $\{\delta_j, \delta_{j+1}\} \in \mathcal{E}$, for any couple of adjacent vertices. The set of paths joining a and b in the graph of Figure 1(a) is $\{(a, b), (a, 1, 2, b), (a, 2, b)\}$. An *active path* occurs when all variables along the path are active.

2.2. Ising model

The Ising model is a special type of independence model for the binary vector X_V over an undirected graph $\mathcal{G}$.

Definition 1 *Given a random vector X_V of binary variables associated to an undirected graph $\mathcal{G} = (V, \mathcal{E})$, the Ising model is the family of Ising probability distributions $P_{\mathcal{G}}(X_V; \psi)$ where $\log(\psi_{jk}) = 0$ for every pair $\{j, k\} \notin \mathcal{E}$. The probability parameters are*

$$\pi_D \propto \left\{ \prod_{j \in D} \psi_j^{x_j} \times \prod_{j,k \in D : \{j,k\} \in \mathcal{E}} \psi_{jk}^{x_j x_k} \right\}, \qquad x_j, x_k \in \{0, 1\}, \quad D \subseteq V. \tag{2}$$

The Ising model was originally introduced by Ising (1925) to model solid magnetic materials where each variable represents an atom having a spin in one of two states, active or non-active. It is a special case of hierarchical log-linear model for binary random vectors where $\log(\psi_{jk}) = 0$ if $\{j, k\} \notin \mathcal{E}$ and $\log \psi_D = 0$ for any $|D| > 2$ with $D \subseteq V$. Ising models represent a class of graphical independence models where, for any pair $\{j, k\} \notin \mathcal{E}$, $\log \psi_{jk} = 0$ and then $X_j \perp\!\!\!\perp X_k | X_{V \setminus jk}$.

Example 2 *The Ising model in Figure 1(a) is defined by the independence statements $X_3 \perp\!\!\!\perp X_1 | \{X_2, X_a, X_b\}$, $X_3 \perp\!\!\!\perp X_a | \{X_1, X_2, X_b\}$, $X_3 \perp\!\!\!\perp X_b | \{X_1, X_2, X_a\}$ and $X_1 \perp\!\!\!\perp X_b | \{X_a, X_2, X_3\}$. The model parameters are the odds $\psi_a, \psi_b, \psi_1, \psi_2, \psi_3$ and the odds ratios $\psi_{ab}, \psi_{a1}, \psi_{a2}$, $\psi_{b2}, \psi_{12}, \psi_{23}$ for the probability table $\mathcal{I}_V$.*

Based on the ferromagnetism principle in statistical physics, magnets have the general tendency to be aligned as well as a set of atomic *homologous variables* (e.g., symptoms, risk factors) may tend to have a synchronized behavior of their status. The tendency for pattern of variables to be aligned depends on the *intensity of the propagation* or of the transition of their status in the graph which is measured by positive edge parameters $\log \psi_{jk}$. Nowadays, there is an increasing use of Ising models to study the co-occurrence of the active status for patterns of variables; see Marsman et al. (2018) for an interesting interpretation of Ising models in network psychometrics.

A path represents a special pattern of variables in a graph since it shows a specific track for the status propagation. More generally, the set of all paths joining a pair of vertices $a, b \in V$ represents the multiple propagation track between variables X_a and X_b. Studying the intensity of the transition across active paths provides insight on the co-occurrence of X_a and X_b and on the simultaneous behavior of variables along the paths.

This paper studies path-based mappings between the mean parameter μ and the canonical parameter $\log(\psi)$ to derive a decomposition of the marginal probability μ_{ab} of X_a and X_b to be active. These decompositions are based on sums of comparable measures directly related to the strength of association and to the intensity of the status transition in multiple paths joining the vertices $a, b \in V$.

2.3. A preliminary result

The mapping between the mean parameter μ_{ab} and the canonical parameter vector $\log(\psi)$ is analytically available since for any $a, b \in V$,

$$\mu_{ab} = \sum_{E \subseteq V \setminus ab} \pi_{abE} = \sum_{E \subseteq V \setminus ab} \Bigg(\prod_{F \subseteq abE} \psi_F \Bigg), \tag{3}$$

where abE is used as shorthand notation for the set $a \cup b \cup E$. This equation does not directly provide insights in terms of path analysis since the contribution of each path $\delta \in \Delta_{ab}$ to compute μ_{ab} is intrinsically embedded in equation (3). The following lemma, instead, derives a decomposition of μ_{ab} which depends on the graphical path setting.

Lemma 1 *Let $P_{\mathcal{G}}(X_V; \psi)$ be an Ising model. The marginal probability μ_{ab} is*

$$\mu_{ab} = \frac{\sum_{\delta \in \Delta_{ab}} \pi_{\delta ab}}{\sum_{\delta \in \Delta_{ab}} \pi_{\delta | ab}}, \tag{4}$$

where Δ_{ab} is the set of paths joining the pair $a, b \in V$ of vertices.

Proof. From the definition of conditional probability we have

$$\mu_{ab} = \frac{\pi_{\delta ab}}{\pi_{\delta | ab}}, \qquad \delta \in \Delta_{ab}. \tag{5}$$

The result follows by averaging equation (5) across all paths $\delta \in \Delta_{ab}$ including the normalized weights

$$\frac{\pi_{\delta | ab}}{\sum_{\delta \in \Delta_{ab}} \pi_{\delta | ab}}, \qquad \delta \in \Delta_{ab}. \tag{6}$$

□

The marginal probability μ_{ab} is decomposed into the sum of the joint probabilities $\pi_{\delta ab}$ that the path δ and the vertices a, b are active and the rest of the network is non-active, normalized with the sum of the conditional probabilities $\pi_{\delta | ab}$ that, given the vertices a, b active, only the path δ is active, with $\delta \in \Delta_{ab}$.

Equation (4) does not completely satisfy the purpose of this work. The probability $\pi_{\delta ab}$ cannot be uniquely assigned to an active path $\delta \in \Delta_{ab}$. For instance, in the graph in Figure 1(b), paths $(a, 1, 2, b)$ and $(a, 2, 1, b)$ have the same probability π_{12ab} to be active, however these paths invoke different pairwise associations and different multivariate dependence structures. The interpretation of the weights in equation (6) is interesting because they enable the comparison of the probabilities of active path events. However, the conditional probability $\pi_{\delta|ab}$ is a non-trivial function of the parameter ψ used to model pairwise associations detected by edges in Ising models. A suitable decomposition of both probabilities $\pi_{\delta ab}$ and $\pi_{\delta|ab}$ involved in equation (4) is required to compare the intensity of the propagation status for the variables X_a, X_b across different paths.

3. Logistic regression parameters for baseline events in multi-way tables

This section proposes a logistic regression approach for modeling the probability of a baseline event $\{X_\delta = 1_\delta, X_{\setminus\delta} = 0_{\setminus\delta} | X_a = 1, X_b = 1\}$ in the conditional distribution of $X_{V\setminus ab}|\{X_a, X_b\}$, where $X_{\setminus\delta}$ is adopted as shorthand notation of $X_{V\setminus a\delta b}$, for any $\delta \in \Delta_{ab}$. This baseline event represents the active status of the variables along the path δ, the non-active status for the remaining set of variables, given that variables X_a and X_b are active. The main intent is proving that the conditional probability $\pi_{\delta|ab}$ is a function of the parameter ψ, for any $\delta \in \Delta_{ab}$. First we define a suitable transformation of the variables.

Definition 2 *Given the set X_V of the variables associated to an undirected graph $\mathcal{G} = (V, \mathcal{E})$, for any path $\delta \in \Delta_{ab}$ with $a, b \in V$, let Y_δ be a binary variable which takes level 1 if the event associated to the baseline cell $i_\delta \in \mathcal{I}_{V\setminus ab}$ realizes, and 0 otherwise.*

For any $\delta \in \Delta_{ab}$, consider the partition $\{X_a, X_b, X_\delta, X_{\setminus\delta}\}$ of the random vector X_V. The random variable Y_δ derives from a dichotomization of the marginal table $\mathcal{I}_{V\setminus ab}$ with respect to the cell associated to the event $\{X_\delta = 1_\delta, X_{\setminus\delta} = 0_{\setminus\delta}\}$. Then, Y_δ takes level 1 if only path δ is active in the network structure, regardless of the status of the extremes a, b; the level 0 occurs when path δ is not fully active, regardless of the rest of the network, for any $\delta \in \Delta_{ab}$:

$$Y_\delta = 1 \quad \text{if } \{X_\delta = 1_\delta, X_{\setminus\delta} = 0_{\setminus\delta}\}, \quad \text{and} \quad Y_\delta = 0 \quad \text{otherwise.}$$

Any path $\delta \in \Delta_{ab}$ induces a set of three binary variables $W_K = (Y_\delta, X_a, X_b)$ indexed by $K = \{\delta, a, b\}$ and related to a 2^3 probability table $\mathcal{I}_K$ following a trivariate Bernoulli distribution with probability parameter $\pi^\delta = (\pi^\delta_E)_{E\subseteq K}$ with elements

$$\pi^\delta_E = \pi_E, \qquad \text{if} \qquad \delta \in E \subseteq K, \tag{7}$$

$$\pi^\delta_E = \sum_{F\subseteq V\setminus ab\delta} \pi_{EF}, \qquad \text{if} \qquad \delta \notin E \subseteq K, \tag{8}$$

where E denotes a generic subset of $\{a, b, \delta\}$. When E includes the path set δ, the probability π^δ_E of the induced random vector W_K coincides with the joint probability π_E of the random vector X_V. The mapping $\pi \mapsto \pi^\delta$ is based on a linear surjective function defined by equations (7) and (8). The corresponding log-linear parameter is $\log\theta^\delta = (\log\theta^\delta_E)_{E\subseteq K}$ with generic element

$$\log\theta^\delta_E = \sum_{E'\subseteq E} (-1)^{|E\setminus E'|}\log\pi^\delta_{E'}, \qquad E\subseteq K. \tag{9}$$

The vector $\log\theta^\delta$ is the canonical parameter of the probability table $\mathcal{I}_K$ obtained by collapsing cells of $\mathcal{I}_V$. The following example gives some insight on the probability and on the log-linear parameter of the distribution of W_K which is, in general, not an Ising model.

Example 3 *Consider the path $(a, 2, b)$ in Figure 1(b). It induces the table $\mathcal{I}_K$ associated to three binary variables $\{Y_{(2)}, X_a, X_b\}$ where $\delta = (2)$. The variable $Y_{(2)}$ takes level 1 if the baseline event $\{X_2 = 1, X_1 = 0, X_3 = 0\}$ occurs and level 0 otherwise. The probability parameter $\pi^{(2)}$ is*

$$\pi^{(2)} = (\pi^{(2)}_\emptyset,\ \pi^{(2)}_a,\ \pi^{(2)}_b,\ \pi^{(2)}_2,\ \pi^{(2)}_{ab},\ \pi^{(2)}_{a2},\ \pi^{(2)}_{b2},\ \pi^{(2)}_{ab2})^T$$

where $\pi^{(2)}_2 = \pi_2$, $\pi^{(2)}_{2a} = \pi_{2a}$, $\pi^{(2)}_{2b} = \pi_{2b}$, $\pi^{(2)}_{2ab} = \pi_{2ab}$ according to equation (7); other elements are obtained as sums of joint probabilities of X_V across the events associated to $Y_{(2)} = 0$, for instance, following equation (8),

$$\pi^{(2)}_a = \pi_a + \pi_{a1} + \pi_{a3} + \pi_{a13} + \pi_{a12} + \pi_{a13} + \pi_{a123}.$$

The log-linear parameter is

$$\theta^{(2)} = (\theta^{(2)}_\emptyset,\ \theta^{(2)}_a,\ \theta^{(2)}_b,\ \theta^{(2)}_2,\ \theta^{(2)}_{ab},\ \theta^{(2)}_{a2},\ \theta^{(2)}_{b2},\ \theta^{(2)}_{a2b})^T$$

where the generic element is an odds-based measure of association, e.g.,

$$\theta^{(2)}_\emptyset = P(X_a = 0, X_b = 0, Y_{(2)} = 0), \quad \theta^{(2)}_a = \frac{P(X_a = 1, X_b = 0, Y_{(2)} = 0)}{P(X_a = 0, X_b = 0, Y_{(2)} = 0)},$$

$$\theta^{(2)}_{ab} = \frac{P(X_a = 1, X_b = 1, Y_{(2)} = 0)P(X_a = 0, X_b = 0, Y_{(2)} = 0)}{P(X_a = 1, X_b = 0, Y_{(2)} = 0)P(X_a = 0, X_b = 1, Y_{(2)} = 0)},$$

$$\theta^{(2)}_{2ab} = \frac{P(X_a = 1, X_b = 1, Y_{(2)} = 1)P(X_a = 0, X_b = 0, Y_{(2)} = 1)}{P(X_a = 1, X_b = 0, Y_{(2)} = 1)P(X_a = 0, X_b = 1, Y_{(2)} = 1)} \times \frac{1}{\theta^{(2)}_{ab}}.$$

To study the relationship between θ^δ and ψ parameters, let us consider the logit of Y_δ given $\{X_a, X_b\}$, for any $\delta \in \Delta_{ab}$. The logistic regression parameters are, for any $\delta \in \Delta_{ab}$,

$$\log\frac{P(Y_\delta = 1|X_a = 1, X_b = 1)}{1 - P(Y_\delta = 1|X_a = 1, X_b = 1)} = \beta_\delta + \beta_{\delta a}X_a + \beta_{\delta b}X_b + \beta_{\delta ab}X_aX_b. \tag{10}$$

The following theorem proves that logistic regression coefficients in equation (10) formally depend on parameter ψ.

Theorem 1 *Let $P_{\mathcal{G}}(X_V;\psi)$ be an Ising model. For any $\delta \in \Delta_{ab}$ with $a,b \in V$, consider the set $\{Y_\delta, X_a, X_b\}$ of binary variables and the logistic regression parameters in equation (10) for the conditional distribution of $Y_\delta|\{X_a,X_b\}$. The following equivalences hold:*

$$\beta_\delta = \log\left(\frac{\prod_{j,k\in\delta}\psi_{jk}}{\theta^\delta_\emptyset}\right), \tag{11}$$

$$\beta_{\delta a} = \log\left(\frac{\prod_{j\in\delta}\psi_{aj}}{\theta^\delta_a}\right), \tag{12}$$

$$\beta_{\delta b} = \log\left(\frac{\prod_{j\in\delta}\psi_{bj}}{\theta^\delta_b}\right), \tag{13}$$

$$\beta_{\delta ab} = \log\left(\frac{\psi_{ab}}{\theta^\delta_{ab}}\right). \tag{14}$$

Proof. The proof is based on the fact that logistic regression coefficients are equal to log-linear interaction terms in the joint distribution of the explanatory and the response variables (Bishop, Fienberg and Holland, 1975). Given $\beta_\delta = \log\theta^\delta_\delta = \log\pi^\delta_\delta/\pi^\delta_\emptyset$, we obtain

$$\theta^\delta_\delta = \frac{\pi_\delta}{\pi^\delta_\emptyset} = \frac{\prod_{j,k\in\delta}\psi_{jk}}{\theta^\delta_\emptyset};$$

since $\pi^\delta_\delta = \pi_\delta$ from equation (7), π_δ is a product of exponential canonical interaction terms in ψ and $\pi^\delta_\emptyset = \theta^\delta_\emptyset$ by definition. Given $\beta_{\delta a} = \log\theta^\delta_{\delta a} = \log(\pi^\delta_{\delta a}\times\pi^\delta_\emptyset)/(\pi^\delta_a\times\pi^\delta_\delta)$, we obtain

$$\theta^\delta_{\delta a} = \frac{\pi_{\delta a}\times\pi^\delta_\emptyset}{\pi_\delta\times\pi^\delta_a} = \frac{\prod_{j,k\in\delta\cup a}\psi_{jk}}{\prod_{j\in a}\psi_j}\times\frac{1}{\theta^\delta_a}$$

since $\pi^\delta_{\delta a} = \pi_{\delta a}$ form equation (7) and $\pi^\delta_a = \theta^\delta_a$ by definition. The same argument is used to prove $\beta_{\delta b} = \log\theta^\delta_{\delta b} = \log\prod_{j\in\delta}\psi_{bj}/\theta^\delta_b$ since $\theta^\delta_{\delta b} = (\pi^\delta_{\delta b}\times\pi^\delta_\emptyset)/(\pi^\delta_b\times\pi^\delta_\delta)$. For the interaction term $\beta_{\delta ab} = \theta^\delta_{\delta ab}$, we have

$$\theta^\delta_{\delta ab} = \frac{\pi^\delta_{\delta ab}\times\pi^\delta_\delta}{\pi^\delta_{\delta a}\times\pi^\delta_{\delta b}}\times\frac{1}{\theta^\delta_{ab}} = \frac{\pi_{\delta ab}\times\pi_\delta}{\pi_{\delta a}\times\pi_{\delta b}}\times\frac{1}{\theta^\delta_{ab}}.$$

The result follows since $\psi_{ab} = \frac{\pi_{\delta ab}\times\pi_\delta}{\pi_{\delta a}\times\pi_{\delta b}}$. □

The following example is illustrative of the result of Theorem 1.

Example 4 *(follows Example 3). Let us consider the logistic regression parameters of $Y_{(2)}|\{X_a,X_b\}$, where $\pi^{(2)}_{2|ab} = \pi_{2|ab}$:*

$$\log\frac{\pi_{2|ab}}{1-\pi_{2|ab}} = \beta_2 + \beta_{2a}X_a + \beta_{2b}X_b + \beta_{2ab}X_aX_b.$$

From the result of Theorem 1, we have

$$\log \frac{\pi_{2|ab}}{1-\pi_{2|ab}} = \log \frac{\psi_\emptyset \psi_2}{\theta_\emptyset^{(2)}} + \log \frac{\psi_a \psi_{a2}}{\theta_a^{(2)}} + \log \frac{\psi_b \psi_{b2}}{\theta_b^{(2)}} + \log \frac{\psi_{ab}}{\theta_{ab}^{(2)}}.$$

Under the undirected graph model in Figure 1(b), $\psi_{ab} = 1$, then the logistic regression model simplifies to

$$\log \frac{\pi_{2|ab}}{1-\pi_{2|ab}} = \log \frac{\psi_\emptyset \psi_2}{\theta_\emptyset^{(2)}} + \log \frac{\psi_a \psi_{a2}}{\theta_a^{(2)}} + \log \frac{\psi_b \psi_{b2}}{\theta_b^{(2)}} - \log \theta_{ab}^{(2)}.$$

Theorem 1 is employed as a technical result in the next section to derive a path-based decomposition. However, the logistic regression model for baseline events is essentially an univariate logistic regression where the response variable is obtained by collapsing a multivariate response vector with respect to a specific event of interest. It can be used to simplify a multivariate regression framework when the interest is only in modeling the probability of a baseline configuration of response variables.

4. Path-based decomposition of the mean parameter

A path-based decomposition in Ising models is derived for the mean parameter μ_{ab}. The result is discussed through an example.

4.1. The main result

The following theorem provides a path-based decomposition of the marginal probability μ_{ab} by using the result of Theorem 1 to compute the conditional probability $\pi_{\delta|ab}$ in Lemma 1 for any path $\delta \in \Delta_{ab}$.

Theorem 2 *Let $P_{\mathcal{G}}(X_V;\psi)$ be an Ising model. Given the set Δ_{ab} of paths joining the pair $a, b \in V$ of vertices,*

$$\mu_{ab} = \sum_{\delta \in \Delta_{ab}} \omega_\delta \frac{\Psi_{\delta ab \setminus \omega_\delta}}{1^T \pi(\Delta_{ab})}, \tag{15}$$

where

$$\omega_\delta = \psi_{a\delta_1} \psi_{\delta_1 \delta_2} \dots \psi_{\delta_m b}, \qquad \delta \in \Delta_{ab}, \tag{16}$$

$\Psi_{\delta ab \setminus \omega_\delta}$ is the sub-product of parameters ψ_{jk}, $j, k \in \delta ab$ with entries omitted in ω_δ, and $\pi(\Delta_{ab})$ is a column vector with generic element

$$\frac{\Psi_{\delta ab}}{\Psi_{\delta ab} + \Theta_{ab}^\delta}, \qquad \delta \in \Delta_{ab},$$

with $\Theta_{ab}^\delta = \prod_{E \subseteq ab} \theta_E^\delta$.

Proof. We preliminary need to prove that, given the Ising model $P_{\mathcal{G}}(X_V;\psi)$, for any $\delta \in \Delta_{ab}$,

$$\pi_{\delta|ab} = \frac{\Psi_{\delta ab}}{\Psi_{\delta ab} + \Theta^{\delta}_{ab}}. \tag{17}$$

Firstly, consider that, for any $\delta \in \Delta_{ab}$, $Y_\delta|\{X_a = 1, X_b = 1\}$ is a binary variables which follows a Bernoulli distribution with probability parameter $\pi_{\delta|ab}$ as $P(Y_\delta = 1|X_a = 1, X_b = 1) = P(X_\delta = 1, X_{\setminus\delta} = 0|X_a = 1, X_b = 1)$ for Definition 2. Consider the logit of Y_δ given $\{X_a, X_b\}$ to compute $\pi_{\delta|ab}$:

$$\pi_{\delta|ab} = \frac{\exp(\beta_\delta + \beta_{\delta a} + \beta_{\delta b} + \beta_{\delta ab})}{1 + \exp(\beta_\delta + \beta_{\delta a} + \beta_{\delta b} + \beta_{\delta ab})} = \frac{\theta^{\delta}_{\delta}\theta^{\delta}_{\delta a}\theta^{\delta}_{\delta b}\theta^{\delta}_{\delta ab}}{1 + \theta^{\delta}_{\delta}\theta^{\delta}_{\delta a}\theta^{\delta}_{\delta b}\theta^{\delta}_{\delta ab}}.$$

Then, the result follows from Theorem 1 since

$$\theta^{\delta}_{\delta}\theta^{\delta}_{\delta a}\theta^{\delta}_{\delta b}\theta^{\delta}_{\delta ab} = \frac{\Psi_{\delta ab}}{\Theta^{\delta}_{ab}}$$

and, by using simple algebra,

$$\pi_{\delta|ab} = \frac{\Psi_{\delta ab}}{\Theta^{\delta}_{ab}} \times \frac{\Theta^{\delta}_{ab}}{\Psi^{\delta}_{ab} + \Theta_{\delta ab}} = \frac{\Psi^{\delta}_{ab}}{\Psi^{\delta}_{ab} + \Theta_{\delta ab}}.$$

Considering that $1^T\pi(\Delta_{ab}) = \sum_{\delta\in\Delta_{ab}} \pi_{\delta|ab}$, and given that $\pi_{\delta ab} = \omega_\delta(\Psi_{\delta ab\setminus\omega_\delta})$ for the Möbius inversion in equation (1), the result follows since equation (15) is an application of the equation (4) of Lemma 1. □

We prove that the parametric decomposition provided by Theorem 2 is an explicit function of parameter ψ and it may be used to define a path-based mapping between mean and canonical parameter for Ising models.

Corollary 1 *In an Ising model $P_{\mathcal{G}}(X_V;\psi)$, for any $\delta \in \Delta_{ab}$, with $a, b \in V$,*

$$\Theta^{\delta}_{ab} = \sum_{E\subseteq V\setminus ab: E\neq\delta} \Psi_{abE}, \qquad \delta \in \Delta_{ab}. \tag{18}$$

Proof. Given an undirected graph $\mathcal{G}$ associated to a multivariate binary random vector $X_V = \{X_a, X_b, X_\delta, X_{\setminus\delta}\}$, for any $\delta \in \Delta_{ab}$,

$$\Theta^{\delta}_{ab} = \pi^{\delta}_{ab} = P(Y_\delta = 0, X_a = 1, X_b = 1)$$

from the inverse mapping of equation (9) and, specifically, from equation (8),

$$\Theta^{\delta}_{ab} = \sum_{E\subseteq V\setminus ab: E\neq\delta} \pi_{abE}$$

where, for any $E \subseteq V \setminus ab : E \neq \delta$, $\pi_{abE} = \Psi_{abE}$. □

For any $\delta \in \Delta_{ab}$, the decomposition in equation (15) includes the measure ω_δ in equation (16) which is an interpretable parameter of multivariate dependence able to quantify the intensity of the propagation of the variable status since it is a product of the odds ratios associated to the edge set of the path. The term $\Psi_{\delta ab \setminus \omega_\delta}$ represents the residual strength of association to compute the joint probability of the active path event, but it does not directly influence the transition of the variable status along the path. Finally, $1^T \pi(\Delta_{ab})$ is a normalizing measure given by the sum over all $\delta \in \Delta_{ab}$ of the probability $\pi_{\delta|ab}$ that X_a and X_b are active only because of variables of a specific path δ are active.

The parameter μ_{ab} is dependent on the variable coding by definition, then a large measure ω_δ of the propagation status results when positive associations $\log \psi_{\delta_j \delta_k}$ run along the path and, more generally, when the strength of positive association prevails over the negative one. This feature is coherent with the research question addressed by Ising models for ferromagnetism and generalized to contexts when the interest is modeling the joint occurrence of patterns of variables (Marsman et al., 2018). An alternative mean parameter-based measure is discussed in the next section to replace μ_{ab} as outcome of interest in the path-based decomposition, however it is well-established that mean parameterizations for binary variables are code dependent; see Ekholm, Smith and McDonald (1995). The marginal odds ratio could be also considered as measure of marginal association, nevertheless this is not a representative mean parameter and the mapping between the marginal and the joint odds ratio may rise some challenges (Didelez, Kreiner and Keiding, 2010; Stanghellini and Doretti, 2019).

4.2. An example

Consider the set of Reinis data from a prospective study of coronary heart disease carried out in Czechoslovakia in 1981 and discussed in Reinis et al. (1981). The data set includes six risk factors observed on a sample of 1841 car-workers: A: systolic blood pressure less than 140mm (yes, no); B: ratio of beta to alpha litoproteins less than 3 (yes, no); C: smoker (no, yes); D: work mentally strenuous (no, yes); E: work physically strenuous (no, yes) and F: familiarity with heart coronary disease (no, yes). The variables have been suitably relabelled to be homologous so that any path is active when risk factors along the path jointly occur. These data are also analyzed in Edwards (2000) who selects a log-

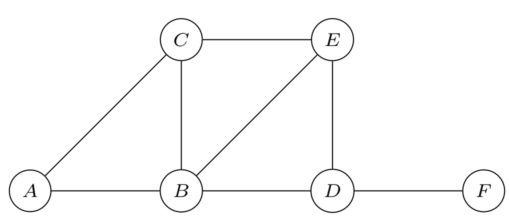

Fig 2. *Ising model for Reinis data*

Table 1
The analysis of paths for Reinis data

Path	size	$\hat{\omega_\delta}$	$\hat{se}(\hat{\omega_\delta})$	$\hat{\psi}_{\delta_1\delta_2}$	$\hat{\psi}_{\delta_2\delta_3}$	$\hat{\psi}_{\delta_3\delta_4}$	$\hat{\mathcal{Y}^*}_\delta$
(1) (B, A, C, E)	4	3.568	0.710	1.460	1.424	1.716	0.42
(2) (B, C, E)	3	1.055	0.349	0.615	1.716	-	0.25
(3) (B, D, E)	3	0.047	0.274	0.765	0.062	-	0.06
(4) (B, E)	2	1.339	0.370	1.339	-	-	0.27

linear graphical model with cliques ABC, BCE, CDE and DF with deviance equal to 51.36 and 46 degrees of freedoms. Since all the three-order log-linear interactions are not significant, an Ising model is considered and the model represented in Figure 2 is selected using a stepwise procedure. The model has a deviance 55.03 with 49 degree of freedoms and all significant pairwise log-linear associations. The association between blood pressure, smoking and lipoproteins (cholesterol) are unsurprisingly positive, as well as most of the associations between other risk factors. As noted in Edwards (2000), physical work E can influence both directly and indirectly the cholesterol level B; the estimated odds ratio is $\hat{\psi}_{BE} = 1.339$ which shows a positive association between the two risk factors; the marginal odds ratio under the selected model is still greater than one, i.e. 1.47, but the marginal association, unlike the conditional one, is poorly significant. The proposed path-based decomposition may provide insight in the role of different active paths of risk factors on the joint occurrence of the factors B and E under the focus, with estimated probability $\hat{\mu}_{BE} = 0.237$. The set Δ_{BE} includes a direct path between B and E and other three paths, that, when active, define different risk factor profiles of car workers: (B, A, C, E), (B, C, E) and (B, D, E).

The estimates of the path measure ω_δ are collected in Table 1 with their estimated standard errors computed using the delta method. Path (1) shows the highest propagation intensity of the active status and defines the profile of workers which also have high blood pressure and are smokers. For path (3) the estimate of the active status propagation is lower than 1 mainly because the pairwise association between D and E is negative showing that car workers are rarely both physically and mentally stressed. Following the Jones and West (2005) interpretation, estimates for ω_δ lower than 1 identify 'moderating' profiles which reduce or mitigate the probability of having risk factors B and E; on the other hand, estimates higher than 1 identify 'mediating' profiles which tend to increase the co-occurrence of B and E. We consider that the higher estimate of ω_δ, the higher the propagation intensity of the active status between risk factors. These terms are not referred to causal effects but to association relationships, since in principle path-based decompositions can be applied regardless of the type of variables and the type of graph. Path analysis in undirected graph models is useful because it does not require to include information about the ordering of the variable set. It is based only on the graph skeleton and, for this reason, consistent with each variable ordering.

5. An alternative parametric decomposition

The above decomposition provides a tool to assess and compare the intensity of propagation in active paths quantified through the strength of association between variables. However, it does not consider the multivariate dependence between the variables along the path and the variables outside the path, which so far, have been assumed to be in a non-active status. This section proposes an alternative decomposition with the intention of shedding light on how the propagation intensity along the path can be influenced by status changes in the rest of the graph. The possibility of applying this decomposition to a measure of marginal association between two variables rather than to marginal probability is also discussed.

5.1. Yule's measure for path-based analysis

For a pair of binary variables X_1 and X_2, Yule's measure (Yule, 1900) is a nonlinear one-to-one function of the odds ratio ψ_{12}:

$$\mathcal{Y}_{12} = \frac{\sqrt{\psi_{12}} - 1}{\sqrt{\psi_{12}} + 1}. \tag{19}$$

Since $0 < \psi_{12} < \infty$, we have that $-1 \leq \mathcal{Y}_{12} \leq 1$. Specifically, $\psi_{12} = 1$ if and only if $X_1 \perp\!\!\!\perp X_2$ and if and only if $\mathcal{Y}_{12} = 0$, also $\mathcal{Y}_{12} > 0$ if $\psi_{12} > 1$ and $\mathcal{Y}_{12} < 0$ if $0 < \psi_{12} < 1$. A suitable two-way table is defined to apply Yule's measure for our purposes.

For any path, two binary variables are defined in order to collapse the full table $\mathcal{I}_V$ into a two-way table providing information on the status of the variables both along and outside the path. Any $\delta \in \Delta_{ab}$ induces a 2×2 probability table for the set $\{Z_\delta, Z_{\bar{\delta}}\}$ of binary variables where Z_δ and $Z_{\bar{\delta}}$ are defined considering a partition of the variable set X_V into $\{X_{\delta ab}\}$ and $\{X_{V \setminus \delta ab}\}$, respectively. Z_δ takes level 1 if variables along the path (a, δ, b) are active and $Z_{\bar{\delta}}$ takes level 1 if variables out of the path are active. Then,

$$Z_\delta = 1 \quad \text{if } X_{\delta ab} = 1_{\delta ab}, \qquad \text{and} \qquad Z_\delta = 0 \quad \text{otherwise},$$

$$Z_{\bar{\delta}} = 1 \quad \text{if } X_{V \setminus \delta ab} = 1_{V \setminus \delta ab}, \quad \text{and} \qquad Z_{\bar{\delta}} = 0 \quad \text{otherwise}.$$

The random vector $(Z_\delta, Z_{\bar{\delta}})^T$ follows a bivariate Bernoulli distribution where the mean parameter denoted by $\tilde{\mu}^\delta$ is a sub-vector of mean parameter μ of X_V, so that

$$\tilde{\mu}^\delta = (\mu_\emptyset, \quad \mu_{\delta ab}, \quad \mu_{V \setminus \delta ab}, \quad \mu_V)^T, \qquad \delta \in \Delta_{ab}, \tag{20}$$

where $\mu_\emptyset = 1$. The probability parameter $\tilde{\pi}^\delta$ and the log-linear parameter $\log \tilde{\theta}^\delta$ of $(Z_\delta, Z_{\bar{\delta}})^T$ can be analytically computed using the inverse transformation, respectively,

$$\begin{aligned} \tilde{\pi}^\delta &= M \tilde{\mu}^\delta, \\ \log \tilde{\theta}^\delta &= M^T \log \tilde{\pi}^\delta, \end{aligned}$$

where M is a 4×4 Möbius matrix

$$M = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

For any $\delta \in \Delta_{ab}$, Yule's measure applied to the odds ratio $\tilde{\theta}_{\delta\bar{\delta}} \in \tilde{\theta}^{\delta}$ provides a measure of the strength of the interaction between the multivariate dependence of variables inside and outside the path. This measure is comparable across the set Δ_{ab} of all paths. Then, for any $\delta \in \Delta_{ab}$, let us consider

$$\mathcal{Y}_{\delta} = \frac{\sqrt{\tilde{\theta}_{\delta\bar{\delta}}} - 1}{\sqrt{\tilde{\theta}_{\delta\bar{\delta}}} + 1}, \qquad \delta \in \Delta_{ab}, \tag{21}$$

where $\tilde{\theta}_{\delta\bar{\delta}} = (\tilde{\pi}^{\delta}_{\emptyset} \times \tilde{\pi}^{\delta}_{\delta\bar{\delta}})/(\tilde{\pi}^{\delta}_{\delta} \times \tilde{\pi}^{\delta}_{\bar{\delta}})$. Parameter $\mathcal{Y}_{\delta}$ measures the association between the status of the two sub-networks induced by any $\delta \in \Delta_{ab}$. A value of $\mathcal{Y}_{\delta}$ close to zero can be interpreted as a poor dependence between the two sub-network models because the probability of having an active path δ barely depends on the active or non-active status of the rest of the network. An alternative decomposition for μ_{ab} is proposed which is complemented with additional information provided by the measure $\mathcal{Y}_{\delta}$, for each path $\delta \in \Delta_{ab}$.

Theorem 3 *Consider an Ising model $P_{\mathcal{G}}(X_V; \psi)$. Given the set Δ_{ab} of paths joining the pair $a, b \in V$ of vertices,*

$$\mu_{ab} = \sum_{\delta \in \Delta_{ab}} (\Psi_{\delta} + \Theta^{\delta}_{ab}) \mathcal{Y}^{*}_{\delta}, \tag{22}$$

where, for any $\delta \in \Delta_{ab}$,

$$\mathcal{Y}^{*}_{\delta} = \frac{| \mathcal{Y}_{\delta} |}{\sum_{\delta \in \Delta_{ab}} | \mathcal{Y}_{\delta} |}. \tag{23}$$

Proof. Firstly, we need to prove that

$$\mu_{ab} = \Psi_{\delta} + \Theta^{\delta}_{ab}. \tag{24}$$

Considering the probability table for $\{Y_{\delta}, X_a, X_b\}$, we compute $\mu_{ab} = \pi^{\delta}_{\delta ab} + \pi^{\delta}_{ab}$ marginalizing over Y_{δ}. The equation (24) is verified since $\pi^{\delta}_{\delta ab} = \pi_{\delta ab} = \Psi_{\delta}$ from equation (7) and $\pi^{\delta}_{ab} = \Theta^{\delta}_{ab}$ from the inverse mapping in equation (9). Then, the result follows because μ_{ab} is averaged across all paths in Δ_{ab} using the normalized Yule's measure $\mathcal{Y}^{*}_{\delta}$ for any $\delta \in \Delta_{ab}$. □

For each path $\delta \in \Delta_{ab}$, the result in equation (22) includes the probability Ψ_{δ} for the path to be fully active when the rest of the network is non-active and the complementary probability Θ^{δ}_{ab} that the path is not fully active, weighted with a normalized value $\mathcal{Y}^*_\delta$ of Yule's measure which quantifies the intensity of the interaction between the two sub-networks induced by the path δ.

Theorem 3 can be easily generalized to compute the dependence ratio τ_{ab} which is a marginal parameter of association. Given the pair X_a, X_b of variables, the dependence ratio is

$$\tau_{ab} = \frac{\mu_{ab}}{\mu_a \mu_b}, \qquad a, b \in V, \tag{25}$$

where $\mu_a = P(X_a = 1)$ and $\mu_b = P(X_b = 1)$. The parameter τ takes value 1 if and only if $X_a \perp\!\!\!\perp X_b$; see Ekholm, Smith and McDonald (1995).

Corollary 2 *In an Ising model $P_\mathcal{G}(X_V;\psi)$, given the set Δ_{ab} of paths joining the pair $a, b \in V$ of vertices,*

$$\tau_{ab} = \frac{\sum_{\delta\in\Delta_{ab}}(\Psi_\delta + \Theta^\delta_{ab})\mathcal{Y}^*_\delta}{(\Psi_a + \Theta^\emptyset_a)(\Psi_b + \Theta^\emptyset_b)}, \tag{26}$$

Proof. The result follows since μ_j, for $j \in V$, can be obtained by applying the logistic regression approach for the baseline event $\{X_j = 1, X_{V\setminus j} = 0\}$, so that $\mu_j = (\Psi_j + \Theta^\emptyset_j)$. □

A decomposition of a marginal measure of association rather than of a marginal probability might be useful in case of effect reversal (Cox and Wermuth, 2003) to explore whether the graphical path structure may explain the different direction between marginal and conditional pairwise associations, or even when the variables of interest are not adjacent in the graph, and so conditionally independent, but marginally associated; see Section 5.3 for an example.

5.2. The case of rare baseline events

In case of large tables, the baseline event $\{X_\delta = 1_\delta, X_{\setminus\delta} = 0_{\setminus\delta}\}$ and Y_δ, defined accordingly in Section 3, could be reasonably rare for each $\delta \in \Delta$. Then, the decompositions provided by Theorems 2 and 3 can be approximated using a log-linear rather than logistic regression parameters (Bishop, Fienberg and Holland, 1975). If the outcome Y_δ is rare, we have

$$\log P(Y_\delta = 1 | X_a = 1, X_b = 1) \approx \log\theta^\delta_\delta + \log\theta^\delta_{\delta a} + \log\theta^\delta_{\delta b} + \log\theta^\delta_{\delta ab}, \quad \delta \in \Delta_{ab},$$

where $\approx$ stands for approximately equal. The conditional probability $\pi_{\delta|ab}$ can be approximated as

$$\pi_{\delta|ab} \approx \frac{\Psi_{\delta ab}}{\Theta^\delta_{ab}} \tag{27}$$

and this simplifies the parametric decomposition in Theorem 2, since

$$1^T\pi(\Delta_{ab}) \approx \sum_{\delta\in\Delta_{ab}} \frac{\Psi_{\delta ab}}{\Theta^\delta_{ab}}.$$

We also derive an approximation of the result of Theorem 3.

Theorem 4 *Consider an Ising model $P_{\mathcal{G}}(X_V;\psi)$. Given the set Δ_{ab} of paths joining the pair $a, b \in V$ of vertices, let $P(Y_\delta = 1, X_a = 1, X_b = 1) \approx 0$ for any $\delta \in \Delta_{ab}$. Then,*

$$\mu_{ab} \approx \sum_{\delta \in \Delta_{ab}} \Theta^{\delta}_{ab} \mathcal{Y}^*_\delta. \tag{28}$$

Proof. For any $\delta \in \Delta_{ab}$, by definition we have $\mu_{ab} = P(Y_\delta = 1, X_a = 1, X_b = 1) + P(Y_\delta = 0, X_a = 1, X_b = 1)$. If $P(Y_\delta = 1, X_a = 1, X_b = 1) \approx 0$, then $\mu_{ab} \approx P(Y_\delta = 0, X_a = 1, X_b = 1) = \Theta^{\delta}_{ab}$ from Corollary 1. The result follows since $\mu_{ab} \approx \frac{1}{|\Delta_{ab}|} \sum_{\delta \in \Delta_{ab}} \Theta^{\delta}_{ab}$ and the same average can be computed including normalized weights $\mid \mathcal{Y}_\delta \mid / (\sum_{\delta \in \Delta_{ab}} \mid \mathcal{Y}_\delta \mid)$ for any path $\delta \in \Delta_{ab}$. □

The previous result can be generalized in terms of dependence ratio.

Corollary 3 *Consider an Ising model $P_{\mathcal{G}}(X_V;\psi)$. Given the set Δ_{ab} of paths joining the pair $a, b \in V$ of vertices, let $P(Y_\delta = 1, X_a = 1, X_b = 1) \approx 0$ for any $\delta \in \Delta_{ab}$. Then,*

$$\tau_{ab} \approx \frac{\sum_{\delta \in \Delta_{ab}} \Theta^{\delta}_{ab} \mathcal{Y}^*_\delta}{\Theta^{\emptyset}_a \Theta^{\emptyset}_b}. \tag{29}$$

Proof. The result direclty follows from Theorem 4 since $\mu_a \approx \Theta^{\emptyset}_a$ and $\mu_b \approx \Theta^{\emptyset}_b$. □

5.3. An example

The proposed measure of Yule is used to extend the Reinis data example in Section 4.2. The last column of Table 1 shows the estimates of $\mathcal{Y}^*_\delta$ for any $\delta \in \Delta_{BE}$. The highest estimate is for the sub-networks induced by path (1); this estimate becomes lower for sub-networks induced by paths (2) and (4), instead it is close to zero when path (3) is considered. The association between variables D and E has an important role to interpret the results since the absolute value of $\log \hat{\psi}_{DE}$ is at least five times larger than the estimates of other log odds ratios. When this edge is not involved in the path structure, the association between the sub-networks is strong and the co-occurrence of the related path is affected by the active or non-active status of the rest of the network, as the case of paths (1), (2), (4). Conversely, when the edge joining D and E belongs to the path, the probability of this path to be active is less influenced by the status of the rest of the network.

The dependence ratio $\hat{\tau}_{BE} = 1.11$ under the selected model is weakly greater than 1 and the chi-square test for the 2×2 marginal table of B and E supports the hypothesys that these variables are marginally independent. On the other hand, the conditional odds-ratio computed in the model shows a significant positive association between the cholesterol level and the physical stress given the other risk factors. In this case, the use of a measure of association is more informative than the marginal probability because this is an example of Simpson's paradox (Simpson, 1951), where the selected model suggests that variables B and E are conditionally associated, but marginally independent. The use of a

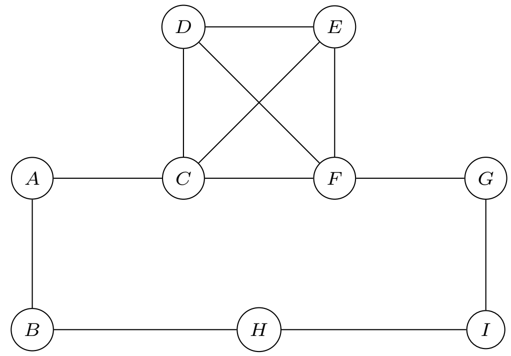

Fig 3. *Ising model in a cyber-security case-study*

measure of association is also useful to disclose collapsibility conditions and of effect reversal which arise, respectively, when marginal and conditional association are comparable and when marginal and conditional associations have a different direction.

6. Application

Protecting industrial networks from sophisticated and modern cyber attacks is a serious issue and cyber-security risk assessment represents an active area of research in several fields. Inspired by the recent stream of the literature on attack graphs (Lallie, Debattista and Bal, 2020), graphical models have proven to be a valuable tool for providing information on the propagation of cyber attacks in a network, in particular on the risk propagation in multiple paths. Despite Bayesian networks (also known as directed acyclic graphs) are often used, in some contexts the attack propagation does not have a unique direction, so an undirected graph model might be preferable. Denuit and Robert (2022) recently discuss the use of Ising models in actuarial sciences and, more specifically, for cyber-security analysis where the risk rises from the interconnections and interdependencies of a network system. This section shows the potential use of path analysis in Ising models for industrial networks with the focus on cyber-security in operational technology (OT) systems.

6.1. Case study

Let us consider the undirected graph in Figure 3 as an illustrative case study. The vertices of the graph represent items of an industrial network, as servers, firewalls, computers, devises, able to monitoring the industrial equipments, assets and processes. The structure of the graph is a priori defined depending on the connections/interactions between items and a basic assumption in OT systems is that a cyber attack propagation can run only through the paths of

Table 2
Path-based analysis of attack propagation risk in a cyber-security case-study

Path	size	ω_δ	$\psi_{\delta_1\delta_2}$	$\psi_{\delta_2\delta_3}$	$\psi_{\delta_3\delta_4}$	$\psi_{\delta_4\delta_5}$	$\psi_{\delta_5\delta_6}$	$\pi^*_{\delta\|BF}$	$\mathcal{Y}^*_\delta$
(1) (B, A, C, D, E, F)	6	9.03	1.22	2.23	1.49	1.35	1.65	0.33	0.14
(2) (B, A, C, D, F)	5	4.95	1.22	2.23	1.49	1.22	-	0.07	0.21
(3) (B, A, C, E, D, F)	6	8.16	1.22	2.23	1.82	1.35	1.22	0.33	0.14
(4) (B, A, C, E, F)	5	8.17	1.22	2.23	1.82	1.65	-	0.12	0.20
(5) (B, A, C, F)	4	6.69	1.22	2.23	2.46	-	-	0.04	0.22
(6) (B, H, I, G, F)	5	24.59	2.23	2.46	2.23	2.01	-	0.11	0.09

the graph. The variables associated to items are binary and take level 1 if the item receives a cyber attack, and 0 otherwise. For any pair $\{j, k\} \notin \mathcal{E}$ of not adjacent vertices, we have that the probability of X_j to be attacked/non-attacked is independent of the probability of X_k to be attacked/non-attacked, given the status of their neighbors. An Ising model is assumed and the following matrix includes the log-linear parameters $\log \psi$ which have been assigned on the basis of subject-matter considerations supported by the expertise of collaborators working in cyber-security field:

$$\begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \\ I \end{array} \begin{pmatrix} 0.1 & 0.2 & 0.8 & 0 & 0 & 0 & 0 & 0 & 0 \\ . & 0.3 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0 \\ . & . & -0.2 & 0.4 & 0.6 & 0.9 & 0 & 0 & 0 \\ . & . & . & 0.1 & 0.3 & 0.2 & 0 & 0 & 0 \\ . & . & . & . & 0.1 & 0.5 & 0 & 0 & 0 \\ . & . & . & . & . & 0.2 & 0.7 & 0 & 0 \\ . & . & . & . & . & . & -0.6 & 0 & 0.8 \\ . & . & . & . & . & . & . & 0.1 & 0.9 \\ . & . & . & . & . & . & . & . & 0.2 \end{pmatrix}:$$

the elements in the main diagonal represent the propensity of each item to be attacked depending on its vulnerabilities; the off-diagonal terms are related to the graph edges and reveal the risk of a local attack propagation which depends on the intensity and on the type of interaction between the items, on their own vulnerabilities and of their neighbours.

Suppose that item B received an attack which spread to item F. We are interested in measuring the propagation risk of this attack over all paths starting from B to F. The same reasoning holds for an attack going from F to B. The path-based decomposition of the probability μ_{BF} provides insights to assess and to mitigate, at the same time, the risk of the attack propagation; under this Ising model, $\mu_{BF} = 0.64$.

Table 2 includes the list of six paths running between vertices B and F with the related risk measures ω_δ of attack propagation computed using the decomposition in Theorem 2. Path (6) has the highest risk since it includes edges with high scores showing great interconnections between vertices. The table includes the conditional probability $\pi^*_{\delta|BF}$, normalized over all the possible paths as in equation (6), that, since B, F have beed attacked, the attack propagation occurred through path $\delta \in \Delta_{BF}$; paths (1) and (3) show the highest probability.

Notice that paths (1) and (3), sharing the same vertices with different orderings, have the same probability $\pi^*_{\delta|BF}$ but different propagation risks ω_δ; conversely, paths (3) and (4) have very similar risks, since they show similar intensity of interconnection along the path, but different probabilities. A naive risk assessment for the network system focused on the pair of items B and F can be obtained by averaging the propagation risk ω_δ weigthed with the probabilities $\pi^*_{\delta|BF}$ across all paths, i.e,

$$R_{BF} = \sum_{\delta \in \Delta_{BF}} \omega_\delta \times \pi^*_{\delta|BF} = 9.95.$$

The last column of the table includes the normalized measure of Yule which reflects the sensitivity of each path with respect to the network structure external to the path. Path (6) with the highest propagation risk shows the lowest sensitivity. Interestingly, in this example where all associations are positive (with $\psi_{jk} > 1$), Yule's measure seems to be inversely related with the magnitude of the propagation risk ω_δ. So, the probability that the attack occurs through paths with high propagation risk, as path (6), poorly depends on the active or non-active status of the rest of the network; conversely, the probability that the attack realizes through paths with low propagation risk, as paths (2) or (5), is more affected by the status of the rest of the network. Intuitively, this reverse relation may be explained by the fact that all associations are quite strong and positive, then, once an attack is running along a path, the greater its propagation intensity, the less impact the status of the rest of the network has on this propagation.

We also explore the effect of risk mitigation in the network structure. Suppose that we are able to protect both item B and the interconnection $H - I$ such that the parameters become $\log \psi_B = -0.1$ and $\log \psi_{HI} = 0.3$, respectively. Then, the propagation risks for paths (1)-(5) are unchanged since the edge $H - I$ appears only in the last path where the risk reduces from 24.54 to 13.46. Similarly, the normalized probabilities of paths (1)-(5) slightly change, whereas the conditional probability that path (6) is active goes from 11% to 6%. The naive risk assessment becomes $R_{BF} = 8.49$. Another possible action could be devoted to protecting the interconnection represented by the edge $A - C$ which is crucial for five paths. If the log-linear parameter becomes $\log \psi_{AC} = 0.4$, then the propagation risks ω_δ for paths (1)-(5) reduce to 6.04, 3.32, 5.47, 5.47 and 4.48, respectively. The related conditional probabilities slightly decreases with the exception of the last one that increases from 11% to 16%. The overall risk is almost the same, i.e. $R_{BF} = 8.43$. In this setting, the mitigation of the pairwise parameter provides a reduction of the risk propagation ω_δ; however this action does not reduce the overall risk which also depends on main effect parameters revealing the propensity of each item to be attacked/non-attacked. The R code to implement and reproduce the proposed results is made available at https://github.com/StaThin/Ising_paths.

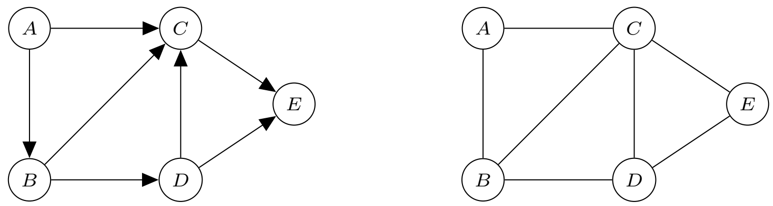

FIG 4. *A Bayesian network and an undirected graph with the same skeleton.*

6.2. On path-based decompositions in Bayesian networks

The choice of graph to represent industrial networks necessarily depends on the case study. When an attack can only propagate in one direction, the use of a Bayesian network can be preferred to that of an undirected graph (Kaynar, 2016). In the literature, there are various metrics to define the riskiness of an attack when the industrial system is represented through a graphical Markov model, specifically a Bayesian network. A review on this would also require introducing technical aspects on directed acyclic graphs, on their Markov properties and on the related independence model, which is beyond the scope of this article. However, we consider it useful to mention a classical approach directly comparable with the topic treated here and to discuss the possible generalization of the methods we propose to the case of directed networks.

Based the approach discussed in Dacier, Deswarte and Kaâniche (1996), the riskiness of an attack path is quantified inversely to the Mean Time To Failure (MTTF) founded on the principle that the higher the MTTF, the greater the security. Let us consider the following graph representing the simple case where there is only one direct path from the first vertex that received the attack to the fifth target vertex:

$$1 \xrightarrow{\pi_{2|1}} 2 \xrightarrow{\pi_{3|2}} 3 \xrightarrow{\pi_{4|2}} 4 \xrightarrow{\pi_{5|4}} 5$$

In this example, MTTF is computed by $\sum_{i=1}^{5}(1/\lambda_i)$ where each λ_i is the success rate of the attack obtained as a function of the mean sojourn time in item i weighted with the conditional probability $\pi_{i|i-1}$ of transition from item $i-1$ to item i, with $\pi_{i|i-1} = \pi_i$ when $i = 1$. These conditional probabilities are the parameters associated to the edges in a Bayesian network, e.g. $\pi_{2|1}$ is the probability to transition from node 1 to node 2 computed by ignoring the status of the rest of the network; see Dacier, Deswarte and Kaâniche (1996) and references therein. Intuitively, the MTTF and the related system security are positively related to the path length. Given a path length, the MTTF is positively related to the mean sojourn time and inversely related to the transition probability. The same approach could be also applied when the network is represented by an undirected Ising model without imposing a specific attack direction. Given a physical criterion to compute the mean sojourn time in item i, the edge parameters ψ_{ij} can be used as weights to compute the MTTF, since they represent the intensity of the status propagation.

Following a similar argument, we conjecture that the path-based decomposition proposed in this paper could be reformulated and applied to the case of Bayesian networks with appropriate differences in terms of the interpretation of the model parameters. In the undirected case, edge parameters of interest are function of conditional probabilities related to the distribution of the full network. The directed case, on the other hand, implies an ordering of vertices, so edge parameters are usually computed in marginal distributions involving the variable associated with the receiving vertex of an arrow and the variables related to the preceding vertices from which the attack may have originated. An important aspect is the definition of edge parameters of interest so that they are comparable across different paths.

Let us consider the Bayesian network in Figure 4 to represent a system where each cyber attack can only propagate along the directed paths of the graph, and the corresponding undirected graph with the same skeleton where any direction is allowed. The two graphs define the same connections but the propagation of the attack is quite different because oriented networks do not consider all propagation paths. In fact, the main difference between the two graphical choices is given by the path sets joining a pair for vertices. For instance, let us compare the paths from item B to item C, under the Bayesian and the undirected network, respectively,

$$
\begin{aligned}
\Delta_{BC}^{\rightarrow} &= \{\{B,C\},\{B,D,C\}\} \\
\Delta_{BC} &= \{\{B,C\},\{B,D,C\},\{B,D,E,C\},\{B,A,C\}\}.
\end{aligned}
$$

These sets are different, specifically $\Delta_{BC}^{\rightarrow}$ is strictly included in Δ_{BC}. It is worth remarking that, in this example, the two graphs are Markov equivalent (Lauritzen, 1996), that is they define the same independence model. However, as discussed earlier, the parameters associated to the edges of the graphs and their interpretation are quite different. Consequently, the risk assessment might be also different. Then, the graphical representation, from our perspective, is mainly a model choice that must faithfully reproduce the case study and allow for the appropriate assignment of edge parameters which, in general, is based on expert knowledge (Xie et al.).

6.3. Discussion

Cyber-security risk assessment rises serious challenges for future research developments involving also computer science expertises. The industrial network is typically huge, not sparse and implementing efficient algorithms to derive all attack paths is a fundamental and preliminary aspect of any risk assessment. Most of the time, considering all possible paths is not feasible and the risk analysis is limited to a reduced sets of short paths that are more likely to occur. Any methodology for risk assessment needs to be scalable and able to handle very large tables. From this side, the proposed decompositions directly work on the rectangular log-linear parameter space, rather than on the probability space, and this represents a technical gain since log-linear parameters are variation independent. A relevant issue consists in the assignment of model parameters that should be dynamically inferred from the activity of the network which rapidly and continuously changes over the time, as well as the vulnerability of vertices and of their interactions. In particular, the learning of model parameters should be performed both from historical attack data and from online data (Huang et al., 2017) where temporal dependence cannot be a priori neglected.

7. Conclusion

Path analysis in undirected graph models is useful to have a full description of the relationship between two variables and to compare the strength of association in multiple paths. It provides an interesting decomposition of the marginal association into measures of multivariate dependence that are consistent with any ordering of the variables. Nevertheless, any path-based methodology needs to be tailored to the nature of the variables and the interpretation of the measures related to paths becomes crucial to gain insight on the problem at hand.

The mean parameter decompositions developed in this paper aim to answer research questions typically rising in models for binary data, in particular in Ising models, and also to provide a first approach for the analysis of paths in non-Gaussian settings. These results could be generalized to non-Ising binary models and to models for categorical data where the complexity of the parameter space increases as well as the interpretation of the measures related to paths.

In principle, these results could be also extended to non-graphical models to study the strength of association in patterns of variables of interest. In the example on the risk factors illustrated in Section 4.2, given a log-linear model for the joint probability table (not necessarily a graphical model), we might be interested in measures that quantify the propagation intensity of the active status in patterns of factors, regardless of any transition ordering induced by a path. This generalization becomes not relevant when the research question is related to the graph structure of an independence model, as in the cyber-security application illustrated in Section 6, where the propagation of a cyber attack in OT systems can only run through the paths of the industrial network.

Acknowledgments

We gratefully acknowledge Alberto Roverato for helpful discussions on path analysis, Massimiliano Latini and Andrea Lazzerini for the support in the application to cyber-security. We also thank the Reviewers for the careful reading of the manuscript and for the insightful and constructive comments.

Funding

The first author was partially supported by the Italian Ministry of University and Research (MUR), Department of Excellence project 2023-2027 ReDS'Rethinking Data Science' - Department of Statistics, Computer Science, Applications - University of Florence, the European Union - NextGenerationEU - National Recovery and Resilience Plan, Mission 4 Component 2 - Investment 1.5 - THE - Tuscany Health Ecosystem - ECS00000017 - CUP B83C22003920001, and the MUR-PRIN grant 2022 SMNNKY, CUP B53D23009470006, funded by the European Union Next Generation EU, Mission 4, Component 2.

References

BARNDORFF-NIELSEN, O. (1978). *Information and exponential families in statistical theory*. Wiley, New York. MR3221776

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B.* **36** 192-236. MR373208

BISHOP, Y. M., FIENBERG, S. E. and HOLLAND, P. W. (1975). *Discrete multivariate analysis: theory and practice*. MIT Press, Cambridge. MR2344876

CAMPAJIOLA, C., LILLO, F., MAZZARISI, P. and TANTARI, D. (2021). On the equivalence between the kinetic Ising model and discrete autoregressive processes. *Journal of Statistical Mechanics: Theory and Experiments* **3** Paper No. 033412, 13. MR4270427

COX, D. R. and WERMUTH, N. (2003). A general condition for avoiding effect reversal after marginalization. *Journal of the Royal Statistical Society, Series B* **65** 937-941. MR2017879

COX, D. R. and WERMUTH, N. (2008). Distortion of effects caused by indirect confounding. *Biometrika* **95** 17–33. MR2409712

DACIER, M., DESWARTE, Y. and KAÂNICHE, M. (1996). *Models and tools for quantitative assessment of operational security* In *Information Systems Security: Facing the information society of the 21st century* 177–186. Springer US, Boston, MA. `https://doi.org/10.1007/978-1-5041-2919-0_15`

DENUIT, M. and ROBERT, C. Y. (2022). Conditional mean risk sharing in the individual model with graphical dependencies. *Annals of Actuarial Science* **16** 183–209.

DIDELEZ, V., KREINER, S. and KEIDING, N. (2010). Graphical models for inference under outcome-dependent sampling. *Statistical Science* **25** 368–387. MR2791673

EDWARDS, D. (2000). *Introduction to graphical modelling, 2nd edition*. Springer-Verlag, New York. MR1880319

EKHOLM, A., SMITH, P. W. F. and MCDONALD, J. W. (1995). Marginal regression analysis of a multivariate binary response. *Biometrika* **82** 847–854. MR1380819

HUANG, K., ZHOU, C., TIAN, Y. C., TU, W. and PENG, Y. (2017). Application of Bayesian network to data-driven cyber-security risk assessment in SCADA networks. In *27th International Telecommunication Networks and Applications Conference (ITNAC)* 1–6. IEEE.

ISING, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift fur Physik A Hadrons and Nuclei* **31** 253–258.

Jones, B. and West, M. (2005). Covariance decomposition in undirected Gaussian graphical models. *Biometrika* **92** 779–786. MR2234185

Kaynar, K. (2016). A taxonomy for attack graph generation and usage in network security. *Journal of Information Security and Applications* **29** 27–56.

Kruis, J. and Maris, G. (2021). Three representations of the Ising model. *Scientific Report* **6** Article No. 34175.

Lallie, H. S., Debattista, K. and Bal, J. (2020). A review of attack graph and attack tree visual syntax in cyber security. *Computer Science Review* **35** 1–41. MR4051972

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press, Oxford. MR1419991

Maathuis, M., Drton, M., Lauritzen, S. and Wainwright, M. (2019). *Handbook of graphical models*. Chapman & Hall/CRC Press, Boca Raton FL. MR3889064

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., van der Maas, H. L. J. and Maris, G. (2018). An Introduction to network psychometrics: relating Ising network odels to Item Response Theory models. *Multivariate Behavioral Research* **53** 15-35.

Ravikumar, P., Wainwright, M. and Lafferty, J. (2010). High-dimensional Ising model selection using l_1-regularized logistic regression. *Annals of Statistics* **38** 1287-1319. MR2662343

Reinis, Z., Pokorny, J., Bazika, V., Tiserova, J., Gorican, K., Horakova, D., Stuchlikova, E., Havranek, T. and Hrabovsky, F. (1981). Prognosticky vyznam rizikoveh profilu v prevenci ischemike choroby srdce. *Bratrisl Lek Listy* **76** 137–150.

Roverato, A. and Castelo, R. (2020). Path weights in concentration graphs. *Biometrika* **107** 705-722. MR4138985

Simpson, E. H. (1951). The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society: Series B* **13** 238–241. MR51472

Stanghellini, E. and Doretti, M. (2019). On marginal and conditional parameters in logistic regression models. *Biometrika* **106** 732-739. MR3992403

Wermuth, N. and Cox, D. R. (2015). Graphical Markov models: overview. *International Encyclopedia of the Social and Behavioral Sciences, 2nd edition* **10** 341-350.

Whittaker, J. (1990). *Graphical models in applied multivariate analysis*. John Wiley & Sons, Chichester. MR1112133

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research* **20** 557–585.

Xie, P., Hi, J. H., Ou, X., Liu, P. and Levy, R. Using Bayesian network for cyber-security analysis. In *IEEE/IFIP International Conference on Dependable System Networks & (DSN)* 211–220.

Yule, G. U. (1900). On the association of attributes in statistics: with illustrations from the material of the Childhood Society, &c. *Philosophical Transactions of the Royal Society of London, Series A* **194** 257-319.