Information Sciences 486 (2019) 287–309

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Identification of structures and causation in flow graphs

Ning Yao, Duoqian Miao*

[a] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
[b] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Shanghai 201804, China

ARTICLE INFO

Article history:
Received 30 November 2017
Revised 16 October 2018
Accepted 13 February 2019
Available online 15 February 2019

Keywords:
Flow graph
Causal effect
Intervention
Rough sets
Equivalence class

ABSTRACT

Flow graphs in rough set theory exhibit intuitive and explicit formalization, straightforward computation, parallel processing and Markov property. This paper focuses on the extraction of flow graphs directly from data in the form of attribute-value tables as well as the identification of the causation between variables in such tables. Using the equivalence classes and the partitions derived directly from data tables, the variables having Markov property can be founded to form the structures that can be integrated into flow graphs. Then based on these structures, the causation hidden in data tables can be identified via the front-door criterion proposed by Pearl, also sometimes via the back-door criterion. The relation between the existence of the causation and the flow graph among variables has been established. According to the illustrations, the identification of the causation also relates to the selection of data sample, besides these flow graphs which are similar to the structures hidden in front-door criterion and part in back-door criterion.

1. Introduction

Using data in reasoning or as a departure point of reasoning has become a consensus among researchers. However, in reality, it is uncommon to get a perfect data sample and in most cases, the data collected are imperfect or incomplete. To understand and manipulate imperfect data or knowledge, rough set theory proposed by Zdzisław Pawlak has been shown to be a successful and intuitive attempt (e.g., [9,10,16,17,27,28,32,33]). In data analysis rough set theory does not need any preliminary or additional information about data such as probability in statistics, basic probability assignment in Dempster–Shafer theory and grade of membership in fuzzy set theory. The indiscernibility relation, especially equivalence relation derived from the attributes, is employed to define granular structure of data for the discovery of patterns in data. The search for data pattern begins with data tables, following with the characterization by approximations, decision rules or flow graphs. Graph representation can group facts into structures, by which the required information can be placed close to the variables involved in the task, and along the pathways, exhibited in the graph, many patterns of human reasoning can be explained.

In [21] Pawlak introduced flow graphs in rough set theory to characterize the relations between condition (input and hidden layers) and decision (output layer) granules as flow networks. In the graphs a flow of information is determined by certainty and coverage factors (i.e., the counterpart of conditional probabilities defined in rough set terms) and governed by the total probability theorem and Bayes' theorem. This class of flow networks centers on modelling the flow distribution in a network rather than the optimal flow analysis for flow networks introduced by Ford and Fulkerson. Flow graphs can be viewed as a special case of probabilistic graphical models, where attribute variables are modelled as nodes (or vertices)

* Corresponding author.
E-mail address: dqmiao@tongji.edu.cn (D. Miao).

https://doi.org/10.1016/j.ins.2019.02.037
0020-0255/

and connected by directed branches (or edges, links) without directed cycles and self-loops, that is, directed acyclic graphs having Markov property or alternatively a special case of chain graphs (i.e., a generalization of both Markov (undirected) and Bayesian (directed and acyclic) networks containing both undirected and directed branches) by omitting undirected branches. Every branch between nodes corresponds to a decision rule for condition node and decision node, denoted as an arrow and characterized by certainty and coverage factors.

Recent research on flow graphs has produced some interesting findings: for example, constructing a flow graph from a decision tree by removing the root and merging nodes labeled by the same attribute, provided the involved decision tree is given [24]; computing coefficients of flow graphs by exploiting the factorization to eliminate variables one by one in polynomial time instead of all at once in exponential time, provided the involved flow graphs are given [6]; representing and calculating coefficients (e.g., the certainty and coverage factors) in matrix form directly from an initial flow graph using a polynomial time with respect to the number of nodes, and integrating flow graphs from heterogeneous data sources to form a single flow graph, provided the involved flow graphs are given [8]. The literature on probabilistic graphical models shows us: the formalization of chain graph (i.e., Markovian and recursive) as well as the transformations of chain graph to Markovian graph with no arrows (undirected) via the existence of a strongly reducible dependence chain in chain graph and to recursive graph with solely arrows (oriented) by means of a complete ordering in chain graph which is locally strongly reducible and compatible to the ordering in recursive graph [19]; the Markov properties for chain graph [13]; qualitative abstractions of chain graph (e.g., the analysis of influences, synergies and sign propagation in chain graph) [18]; the use of Markov chain in analysis of behaviors for cognitive memory architectures and of control system [11,15]; the existence of an undirected graph and a directed acyclic graph (DAG) by examining the properties of the set of conditional independencies, such as symmetry, decomposition, intersection, strong union, transitivity and contraction, together with graphical rules and the faithfulness assumption between a graph and the set of conditional independencies in a distribution over the variables [26,31]; the construction of a directed acyclic graph given an ordering among variables related to fast closure of the set of conditional independence statements in an ordered triple form (i.e., containing only the maximal triples and no loss of information) [3]; the design of directed acyclic graphs for variable groups with a groupwise faithfulness between a group directed acyclic graph with the nodes of variable groups and the set of conditional independencies in the distribution over the groups [20]; incremental causal network with a combination of temporal precedence relationships and statistical dependency [1]; an extended Bayesian Networks for massive hierarchical data [2], and so forth. Research also has found that, DAG structures not only are primarily used in causal interpretation between variables as well as between actions and variables, based on the ability to represent and respond to external or spontaneous changes, but also can capture conditional independence judgments and express what we know or believe about the world (generally reflected by probabilistic relationships, such as marginal and conditional independencies) in a more natural and more reliable way, compared with probabilistic structures of associational knowledge [25,31].

Causal relationships describe objective physical constraints in our world and emphasize an understanding of durable relationships as well as transportability across situations [25]. They exhibit greater robustness to changes in the environment and in our knowledge about the environment than the corresponding probabilistic relationships which focus mainly on transitory relationships [25]. To evaluate the causation between variables, Pearl introduced the structural causal model framework to characterize counterfactuals and interventions from the observational (nonexperimental) data through two criteria: the back-door criterion and the front-door criterion, the former reflects the situation that the concomitant observations should be quite unaffected by the treatment, while the latter demonstrates the situation that the concomitants are affected by the treatment [12,14,25]. In the construction of causal explanations, the key is to predict the effect under hypothetical interventions, for which "*do*" operator and a belief change process ([5]) are used to encode interventions. Despite the importance of causation, the causal notions (e.g., intervention, counterfactual) in rough set theory remain unexplored. Concerning the back-door criterion, our strategy is to employ the lower approximation in rough set theory to construct the structures satisfying the back-door criterion for variables stored in data tables and then to estimate the causal effects, which has been detailed for a new topic and is beyond the scope of this paper. As to the front-door criterion, it is noticeable that flow graphs are quite similar to the structures embedded in the front-door criterion. Moreover, flow graphs are treated as a form of knowledge representation encoded in attribute-value tables, and reasoning based on flow graphs can be much more efficient, intuitive and interpretable than reasoning performed directly from data tables. Accordingly, the construction of rough set flow graphs and the estimate of the causal relationships in these graphs using the front-door criterion, form the core of this paper.

The purpose of the present paper is to build flow graphs directly from data, using the equivalence classes derived from the attributes in data tables in the context of rough sets, and then to identify the causation hidden in the established flow graphs, as depicted in the following diagram:

$$\begin{tikzcd}
Data\ table \arrow[r] & Equivalence\ classes \arrow[r] & Flow\ graph\ structures \arrow[r] \arrow[d, "Back-door(partially)"', "Front-door"] & Flow\ graphs \arrow[dl, "Front-door"', "Back-door(partially)"] \\
 & & Causal\ effects &
\end{tikzcd}$$

A simple sketch of the idea in this paper

Without any additional information (e.g., probabilities, decision tree, an ordering between attributes), we directly construct the structures, nested in flow graphs and satisfying the front-door criterion and part of the back-door criterion, from data through the notion of equivalence class determined by the attributes, since this notion is in fact the substance of rough sets. Using these structures we can exhibit the patterns hidden in data tables graphically and build the causal explanation in causal terms such as the interventions and counterfactuals. We first review the basic definitions involved in causation (e.g., the front-door criterion, probability of necessity, probability of sufficiency) and the construction of flow graphs (e.g., flow graph, certainty factor, equivalence class), together with some fundamental rules in probability theory (Section 2). In Section 3, the relations for Markov property, statistical independence and equivalence classes as well as the relation between equivalence classes and the front-door criterion are examined. Further we detail the flow graphs for variables, possessing Markov property within some constraints on equivalence classes determined by attribute variables, and the causal effect of interventions in these flow graphs. We also demonstrate the obtained theoretical results in the analysis of votes distribution and of driving a car in various driving conditions as well as the description of flu, including the hidden flow graphs and the interpretation and identification of causation for the attributes involved in the data tables (Section 4). Finally, the last section contains a short conclusion and plan for future research.

2. Preliminaries

In this section, we recall some basic knowledge on rough set theory and the causation related to flow graphs, which can be used to specify the judgement and construction of the flow graphs as well as the identification of causal effects in flow graphs.

2.1. Some basic definitions on rough set theory

In rough set theory, the pursuit of data patterns starts with data tables and the discovered knowledge can be expressed as the decision rules or as the flow graphs.

Definition 1 (Flow Graphs [23]). A flow graph is a directed acyclic finite graph $G=(N,\mathcal{B},\varphi)$, where N is a set of nodes, $\mathcal{B}\subseteq N\times N$ is a set of directed branches, $\varphi:\mathcal{B}\to R^{+}$ is a flow function, and R^{+} is the set of non-negative reals.

1. For any $w,\ y\in N$, if $(w,y)\in\mathcal{B}$ then w is an input of y and y is an output of w, $\varphi(w,\ y)$ is a throughflow from w to y and assume that $\varphi(w,\ y)\neq 0$. By $I(w)$ and $O(w)$ denote the set of all inputs of w and the set of all outputs of w respectively. Then the input and output of a flow graph G are defined as $I(G)=\{w\in N: I(w)=\emptyset\}$ and $O(G)=\{w\in N: O(w)=\emptyset\}$, $\varphi(G)$ is the throughflow of G and defined as

$$\varphi_{out\,flow}(G)=\sum_{w\in O(G)}\sum_{y\in I(w)}\varphi(y,w)=\varphi(G)=\sum_{w\in I(G)}\sum_{y\in O(w)}\varphi(w,y)=\varphi_{in\,flow}(G)$$

$$\varphi_{outflow}(w)=\sum_{y\in O(w)}\varphi(w,y)=\varphi(w)=\sum_{y\in I(w)}\varphi(y,w)=\varphi_{inflow}(w)\ \textit{for any internal node } w.$$

   Inputs and outputs of G are external nodes of G; other nodes are internal nodes of G.
2. Instead of $\varphi(w,\ y)$, we define a normalized flow function $\sigma:\mathcal{B}\to[0,1]$, i.e., $\sigma(w,y)=\frac{\varphi(w,y)}{\varphi(G)}$ for every $(w,y)\in\mathcal{B}$, which is also called the strength of $(w,\ y)$. The graph $G=(N,\mathcal{B},\sigma)$ is called a normalized flow graph. For every node w of a flow graph G, its normalized inflow and normalized outflow are given as

$$\begin{aligned}
\sigma_{outflow}(w) &= \sum_{y\in O(w)}\sigma(w,y)=\sum_{y\in O(w)}\frac{\varphi(w,y)}{\varphi(G)}\\
\sigma_{inflow}(w) &= \sum_{y\in I(w)}\sigma(y,w)=\sum_{y\in I(w)}\frac{\varphi(y,w)}{\varphi(G)}\\
\sigma_{inflow}(w) &= \sigma(w)=\sigma_{outflow}(w)\ \textit{when } w \textit{ is an internal node}
\end{aligned}$$

$$\sigma_{out\,flow}(G) = \sum_{w\in O(G)} \sum_{y\in I(w)} \frac{\varphi(y,w)}{\varphi(G)}$$

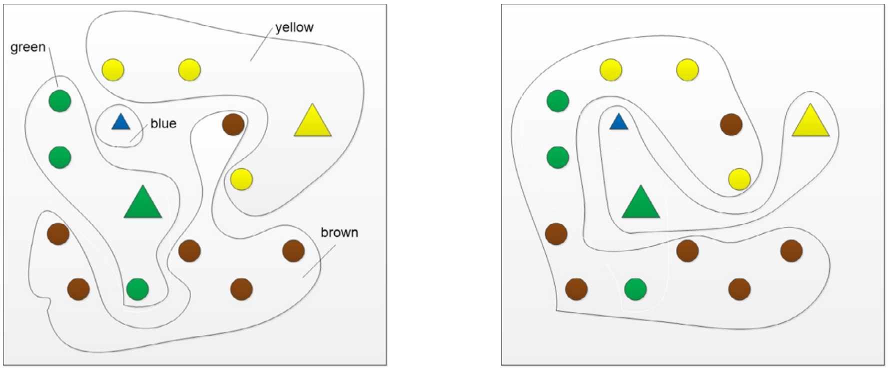
$$\sigma_{in\,flow}(G) = \sum_{w\in I(G)} \sum_{y\in O(w)} \frac{\varphi(w,y)}{\varphi(G)}$$

$$\sigma_{out\,flow}(G) = \sigma(G) = \sigma_{in\,flow}(G) = 1.$$

3. A (directed) path from w to y, $w \neq y$ in G is a sequence of nodes $w_1, \ldots, w_n$ such that $w_1 = w$, $w_n = y$ and $(w_i, w_{i+1}) \in \mathcal{B}$ for every i, $1 \leq i \leq n-1$. A path from w to y is denoted by $[w \ldots y]$. The set of all paths from w to y $(w \neq y)$ in G is denoted by $< w,\ y >$, called a connection from w to y in G. With every branch $(w,\ y)$ in G the certainty and coverage factors are associated, respectively defined as $cer(w,y) = \frac{\sigma(w,y)}{\sigma(w)}$ and $cov(w,y) = \frac{\sigma(w,y)}{\sigma(y)}$ under the condition that $\sigma(w) \neq 0$ and $\sigma(y) \neq 0$. The certainty and the coverage of the path $[w_1, \ldots, w_n]$ are defined by $cer[w_1, \ldots, w_n] = \prod_{i=1}^{n-1} cer(w_i, w_{i+1})$ and $cov[w_1, \ldots, w_n] = \prod_{i=1}^{n-1} cov(w_i, w_{i+1})$. For every connection $< w,\ y >$, the certainty is $cer < w, y >= \sum_{[w \ldots y] \in < w,y >} cer[w \ldots y]$ and the coverage is $cov < w, y >= \sum_{[w \ldots y] \in < w,y >} cov[w \ldots y]$.
4. For every $(w, y) \in \mathcal{B}$, the dependency factor of $(w,\ y)$, i.e., $\eta(w,\ y)$ is defined as $\eta(w,y) = \frac{cer(w,y) - \sigma(y)}{cer(w,y) + \sigma(y)} = \frac{cov(w,y) - \sigma(w)}{cov(w,y) + \sigma(w)}$. $\eta(w,y) = 0$ means w and y are independent on each other. $-1 < \eta(w,y) < 0$ represents w and y are negatively dependent. $0 < \eta(w,\ y) < 1$ implies w and y are positively dependent on each other.

Definition 2 (Equivalence class [22]). Let $U \neq \emptyset$ be a finite set of objects called the universe and R be an equivalence relation over U characterized by the attributes of the objects. Then the equivalence class $[x]_R = \{y \in U : xRy\}$, determined by the element $x \in U$, denotes a concept in R, and U/R denotes the family of all equivalence classes of R, which means knowledge associated with R (in short knowledge R) and forms a partition of U.

Consider an example illustrated by the following figures, i.e., a system containing a set of



(a) Blocks of Objects—Color (b) Blocks of Objects—Shape

colored circles and triangles with a feature set that contains two very obvious external attributes, namely color and shape. Using only one attribute the system can be partitioned into four blocks and each block posses the different colors such as green, yellow, brown and blue, or into two blocks and each has the same shape, namely, circle and triangle. By means of color and shape, the system is partitioned into six blocks and each has the same color and shape: green and circle, green and triangle, yellow and circle, yellow and triangle, brown and circle, blue and triangle.

Definition 3 (Lower and upper approximation [22]). Let X be a subset of the universe U. R-lower and R-upper approximations of X with respect to the equivalence relation R over U are defined by

$$\underline{R}(X) = \left\{x \in U : [x]_R \subseteq X\right\}, \quad \overline{R}(X) = \left\{x \in U : [x]_R \cap X \neq \emptyset\right\}.$$

R-boundary region of X is defined as $BN_R(X) = \overline{R}(X) - \underline{R}(X)$. A set is rough if it has nonempty boundary region; otherwise the set is crisp.

Consider the above illustration, the equivalence classes determined by color are $\{[x]_{green}, [x]_{yellow}, [x]_{brown}, [x]_{blue}\}$, similarly the equivalence classes derived from shape contains $\{[x]_{circle}, [x]_{triangle}\}$. It is easy to find that $[x]_{brown} \subseteq [x]_{circle}$ and $[x]_{blue} \subseteq [x]_{triangle}$, yielding *color*-lower approximations of $[x]_{circle}$ and $[x]_{triangle}$:

$$\left.\begin{array}{l}\underline{color}([x]_{circle}) = [x]_{brown} \\ \underline{color}([x]_{triangle}) = [x]_{blue}\end{array}\right\} \quad \underline{color}(shape) = [x]_{brown} \cup [x]_{blue}$$

here we write $\underline{\text{color}}(shape)$ instead of $\underline{\text{color}}([x]_{shape})$ if x is understood in the given system. In the same way *color*-upper approximations of $[x]_{circle}$ and $[x]_{triangle}$ are given as $\overline{color}([x]_{circle}) = [x]_{green} \cup [x]_{yellow} \cup [x]_{brown}$, $\overline{color}([x]_{triangle}) = [x]_{green} \cup [x]_{yellow} \cup [x]_{blue}$. The *color*-boundary regions for $[x]_{circle}$ and $[x]_{triangle}$ equal $BN_{color}([x]_{circle}) = [x]_{green} \cup [x]_{yellow}$, $BN_{color}([x]_{triangle}) = [x]_{green} \cup [x]_{yellow}$.

Definition 4 (Certainty factor [21])**.** Let $S = (U, C, D)$ be a decision table and U, C, D be sets of objects, condition and decision attributes respectively, with $C = \{c_1, \ldots, c_n\}$ and $D = \{d_1, \ldots, d_m\}$. For any $x \in U$, a sequence $[x]_{c_1}, \ldots, [x]_{c_n}$, $[x]_{d_1}, \ldots, [x]_{d_m}$ represents a decision rule induced by x and denoted as $[x]_{c_1}, \ldots, [x]_{c_n} \to [x]_{d_1}, \ldots, [x]_{d_m}$ (in short $C \to_x D$). A certainty factor of the decision rule, denoted as $cer_x(C, D)$, is defined by $cer_x(C, D) = \frac{|\,[x]_C \cap [x]_D\,|}{|\,[x]_C\,|}$, where $|X|$ denotes the cardinality (i.e., the number of elements) of X and this quantity can be interpreted statistically (in frequency) as a conditional probability that y belongs to $[x]_D$ given y belongs to $[x]_C$, symbolically $P_x(D|C)$. Instead of $C \to_x D$, $cer_x(C, D)$ and $P_x(D|C)$, we will write $C \to D$, $cer(C, D)$ and $P(D|C)$ if x is understood. In the same way a coverage factor of the decision rule, denoted $cov(C, D)$, is defined as $cov(C, D) = \frac{|\,[x]_C \cap [x]_D\,|}{|\,[x]_D\,|} = P(C|D)$, expressing the conditional probability of x to condition class $[x]_C$ given D. The unconditional probability that x in U satisfies condition attribute C, denoted $P(C)$, is defined as $P(C) = \frac{|\,[x]_C\,|}{|U|}$. The strength of the decision rule, denoted as $\sigma(C, D)$, is defined by $\sigma(C, D) = \frac{|\,[x]_C \cap [x]_D\,|}{|U|}$.

If $cer(C, D) = 1$ (alternatively $P(D|C) = 1$), then $C \to D$ will be called a certain decision rule in S; if $0 < cer(C, D) < 1$ (or $0 < P(D|C) < 1$), the decision rule will be referred to as an uncertain decision rule in S.

Remark 1. Throughout the paper the absence of an arrow between a pair of nodes reflects statistical independence and an arrow between a pair of nodes represents dependence for attribute variables in flow networks.

Again consider the above figure. We can calculate the certainty factor from *brown* to *circle* by $cer(brown, circle) = \frac{|[x]_{brown} \cap [x]_{circle}|}{|[x]_{brown}|} = \frac{|[x]_{brown}|}{|[x]_{brown}|} = 1$. Similarly there are $cer(blue, triangle) = \frac{|[x]_{blue} \cap [x]_{triangle}|}{|[x]_{blue}|} = 1$. This finding leads to another definition of approximations and boundary region of a set: $\underline{R}(X) = \{x \in U : cer(R, X) = 1\}$, $\overline{R}(X) = \{x \in U : 0 < cer(R, X) \leq 1\}$, $BN_R(X) = \{x \in U : 0 < cer(R, X) < 1\}$. By referring to this figure, there hold $\underline{color}(circle) = \{x \in U : cer(color, circle) = 1\} = [x]_{brown}$ and $\underline{color}(triangle) = \{x \in U : cer(color, triangle) = 1\} = [x]_{blue}$; $\overline{color}(circle) = \{x \in U : 0 < cer(color, circle) \leq 1\} = [x]_{brown} \cup [x]_{green} \cup [x]_{yellow}$ for $cer(green, circle) = \frac{3}{4} = cer(yellow, circle)$ and $cer(blue, circle) = 0$, $\overline{color}(triangle) = \{x \in U : 0 < cer(color, triangle) \leq 1\} = [x]_{blue} \cup [x]_{green} \cup [x]_{yellow}$ for $cer(green, triangle) = \frac{1}{4} = cer(yellow, triangle)$ and $cer(brown, triangle) = 0$; $BN_{color}(circle) = \{x \in U : 0 < cer(color, circle) < 1\} = [x]_{green} \cup [x]_{yellow}$ and $BN_{color}(triangle) = \{x \in U : 0 < cer(color, triangle) < 1\} = [x]_{green} \cup [x]_{yellow}$.

It is worth pointing out that the certainty factors from Definition 1 and Definition 4 are equivalent and can be treated as two different representations of the same decision rule based respectively on flow graph and data table, the same to coverage factor and strength of the decision rule. In this way the quantity $\varphi(x, y)$ in Definition 1 refers to the number of objects in the universe U satisfying the directed branch (x, y). Another key point is that the certainty factor and coverage factor of a path $[x_1, \ldots, x_n]$ have the Markov property according to Definition 1 and 4, which is also at the heart of extracting relevant flow graphs from data.

2.2. Some basic definitions on causation

The pioneering work on causation attributes to the elucidation of cause-effect relationships among variables or events using elementary mathematics, which was introduced by Pearl. The identification of causation begins with a directed acyclic graph (which is the same as the one used in the flow graph) and two graphical criteria are designed to the existence of causal effects: the front-door criterion and the back-door criterion. In this paper we mainly focus on the front-door criterion as a result of the similarity of the structures between this criterion and flow graph.

Definition 5 (Front-Door [25])**.** A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (W, Y) if:

(i) Z intercepts all directed paths from W to Y;
(ii) there is no back-door path from W to Z ; and
(iii) all back-door paths from Z to Y are blocked by W;

illustrated by the following diagram where V denotes the unobserved variable (i.e., the background or exogenous variable that is determined by factors outside the model, the links connecting the unobserved variables are designated by dashed arrows) and the variables W, Z, Y are assumed to be observed (i.e., the endogenous variables that are determined by variables in the model, the links connecting the observed variables are designated by solid arrows), v, w, y, z refer to the sets of variable values respectively for V, W, Y, Z. Two conditional independencies are assumed to be encoded with the absence of arrows in the graph: Z and V are conditionally independent given W (i.e., $P(v|z, w) = P(v|w)$, $P(z|v, w) = P(z|w)$); Y and W

are conditionally independent given Z and V (i.e., $P(y|z, v, w) = P(y|z, v),\ P(w|z, v, y) = P(w|z, v)$).

$$\begin{tikzcd} & V \arrow[dl, dashed] \arrow[dr, dashed] & \\ W \arrow[r] & Z \arrow[r] & Y \end{tikzcd}$$

A diagram representing the front-door criterion

The name "back-door" path from W to Z signifies the path between W and Z with an arrow into W, namely entering W through the back door. A "directed" path from W to Y means a sequence of branches such that every branch in this path is an arrow that points from the first to the second node of the pair. A path in a graph is a sequence of branches such that each branch starts with the node ending the preceding branch. By "blocking" or "intercepting" we mean stopping the flow of information (or of dependency) between the variables that are connected by paths.

Lemma 1 (Front-Door Adjustment [25]). *If Z satisfies the front-door criterion relative to (W, Y) and if $P(w, z) > 0$, then the causal effect of W on Y is identifiable and is given by the formula*

$$P(y|do(W = w_i)) = \sum_z P(z|w_i) \sum_w P(y|w, z)P(w) \tag{1}$$

where $w_i \in w$ denotes the specific value of W, i is a positive integer; $do(W = w_i)$ (or $do(w_i)$ for short) denotes an external intervention, i.e., setting $W = w_i$; $P(y|do(W = w_i))$ stands for the probability of achieving a yield level of $Y = y$, given that the treatment is set to level $W = w_i$ by external intervention.

Note that, in the front-door criterion, the set of variables Z is affected by W and measurements of Z can enable consistent estimation of the effect of interventions from non-experimental data. As to the case that Z is not affected by W, the back-door criterion ([25], Chapter 3) can be used for identifying the unbiased effect estimate. The front-door criterion studies the case that Z is located in the path directed along the arrows from W to Y, such as $W \to Z \to Y$; the back-door criterion concerns the case that Z is located in the back-door paths carrying the spurious associations from W to Y, such as $W \leftarrow Z \to Y$. Assume that Z, W, Y be three different variables in an acyclic graph, and Z is unaffected by W but may possibly affect Y. If both of $P(w|z) = P(w)$ and $P(y|z, w) = P(y|w)$ are violated, then W and Y are not stably unconfounded [25]. When W and Y are stably unconfounded, there is $P(y|do(w)) = P(y|w)$ plus the absence of a common ancestor of W and Y in the acyclic graph, which implies the satisfaction of either $P(w|z) = P(w)$ or $P(y|z, w) = P(y|w)$ [25]. This condition on the variable Z in an acyclic graph can be guaranteed via the back-door criterion and the adjustment formula $P(y|do(w)) = \sum_z P(y|w, z)P(z)$ [4,25].

Besides the mathematization of the effect of interventions, the counterfactual quantities, which are used for interpretation of causation, also have been mathematized. First comes the introduction of causal model involved in the definition of counterfactual quantities. A causal model consists of a set of background or exogenous variables (representing factors outside the model, which nevertheless affect relationships within the model), a set of endogenous variables (assumed to be observable, determined by variables in the model), and a set of functions that determine or simulate how values are assigned to each endogenous variable [4,25]. The diagram that captures the relationships among the variables in a causal model is called the causal graph of this model [4].

Definition 6 ([25]). Let W and Y be two binary variables in a causal model. Let w and y stand (respectively) for the propositions $W = \text{true}$ and $Y = \text{true}$, and let w' and y' denote their complements. $Y_{w'} = y'$ means that y would not have occurred in the absence of w, denoted as $y'_{w'}$ for short. y_w is the short for $Y_w = y$ representing that y would have occurred by setting w.

The probability that Y would be y' in the absence of event w, given that w and y did in fact occur, denoted as probability of necessity (PN), is defined as the expression $PN(w, y) = P(Y_{w'} = y'|W = w, Y = y)$.

The probability that setting w would produce y in a situation where w and y are in fact absent, denoted as probability of sufficiency (PS), is defined as $PS(w, y) = P(Y_w = y|W = w', Y = y')$.

Y is said to be monotonic relative to variable W **if and only if** $y'_w \wedge y_{w'} = \text{false}$. Monotonicity expresses the assumption that a change from $W = \text{false}$ to $W = \text{true}$ cannot, under any circumstance, make Y change from true to false.

Lemma 2 ([30]). *For the monotonic case that Y is monotonic relative to W, whenever the causal effect $P(y|do(w'))$ is identifiable, PN and PS are identifiable and given by*

$$PN = \frac{P(y) - P\big(y|do(w')\big)}{P(w, y)}$$

$$PS = \frac{P\big(y|do(w)\big) - P(y)}{P(w', y')}.$$

Further, using $P(y) = P(y|w)P(w) + P(y|w')(1 - P(w))$ (total probability formula), the equation PN can be rewritten as

$$PN = \frac{P(y|w) - P(y|w')}{P(y|w)} + \frac{P(y|w') - P(y|do(w'))}{P(w, y)}$$

For the general and nonmonotonic case that Y is nonmonotonic relative to W, whenever the causal effects $P(y|do(w))$, $P(y|do(w'))$ and $P(y'|do(w'))$ are identifiable, PN and PS are bounded with the following tight upper and lower bounds:

$$\max\left\{0, \frac{P(y) - P(y|do(w'))}{P(w, y)}\right\} \le PN \le \min\left\{1, \frac{P(y'|do(w')) - P(w', y')}{P(w, y)}\right\}$$

$$\max\left\{0, \frac{P(y|do(w)) - P(y)}{P(w', y')}\right\} \le PS \le \min\left\{1, \frac{P(y|do(w)) - P(w, y)}{P(w', y')}\right\}$$

Apart from these results on causality, we also need some basic results on probability theory used to characterize flow graph structures.

Lemma 3 ([25]). *Given an ordered set of n discrete random variables $A_1, A_2, \ldots, A_n$, then the probability of the joint variable $(A_1, A_2, \ldots, A_n)$ can be written as a product of n conditional probabilities, called the chain rule formula*

$$P(A_1, A_2, \ldots, A_n) = P(A_n|A_{n-1}, \ldots, A_2, A_1) \cdots P(A_2|A_1)P(A_1);$$

the marginal probability of (A_1, A_n) can be computed by marginalizing respectively over $A_2, \ldots, A_{n-1}$

$$P(A_1, A_n) = \sum_{A_2, \ldots, A_{n-1}} P(A_1, A_2, \ldots, A_{n-1}, A_n);$$

the probability of the conditional variable $(A_n|A_1)$ can be assessed from

$$P(A_n|A_1) = \sum_{A_2, \ldots, A_{n-1}} P(A_n|A_{n-1}, \ldots, A_2, A_1) \cdots P(A_2|A_1).$$

Further, if the sequence of $A_1, A_2, \ldots, A_n$ is said to be Markovian, namely the probability of moving to the next state depends only on the present state and not on the previous states

$$P(A_n|A_{n-1}, \ldots, A_2, A_1) = P(A_n|A_{n-1}),$$

then the conditional probability $P(A_n|A_1)$ and the joint probability $P(A_1, A_2, \ldots, A_n)$ can be written as

$$P(A_n|A_1) = \sum_{A_2, \ldots, A_{n-1}} P(A_n|A_{n-1}) \cdots P(A_2|A_1) = \sum_{A_2, \ldots, A_{n-1}} \prod_{i=1}^{n-1} P(A_{i+1}|A_i)$$

$$P(A_1, A_2, \ldots, A_n) = P(A_n|A_{n-1}) \cdots P(A_2|A_1)P(A_1) = \prod_{i=1}^{n-1} P(A_{i+1}|A_i)P(A_1).$$

3. Structures and interventions in flow graphs

In this section, we will deeply characterize the structures in flow graphs and the effect of interventions using the rough set tools and the graphical criteria depicted in above section. The diagram below shows the dependency between the Theorems (*Thm i*), Corollaries (*Cor i*) and Lemmas (Lem 1, 3) in this paper.

$$\begin{tikzcd}
Cor\ 1 & Thm\ 1 \arrow[l] \arrow[r, "Lem\ 1"] \arrow[dr] & Thm\ 2 \arrow[r] \arrow[d] & Cor\ 2 \arrow[dl, "Lem\ 3"] \\
 & & Thm\ 3 \arrow[dl] \arrow[dr] & \\
 & Cor\ 3 & & Cor\ 4 + Cor\ 5
\end{tikzcd}$$

A sketch of this section

Thm 1: the relation between equivalence classes and Markov property;
Cor 1: the relation between equivalence classes and conditional (marginal) independence (non-Markovian);
Thm 2: the relation between equivalence classes and front-door adjustment;
Cor 2: conditional probability and the effect of the intervention for the specific attribute values;

Thm 3: flow graphs and causal effects of the intervention for data table with three attributes;
Cor 3: the cases when equivalence class(es) for the attribute pairs is (or are) $\emptyset$;
Cor 4, 5: the cases when the number of the attributes are greater than 3.

At the heart of a path in a flow graph lies the Markov property for involved attributes. The first theorem is placed to characterize the Markov property using the rough set terms.

Theorem 1. *Let $S=(U,A)$ be an information system and U be a non-empty finite set of objects called the universe, A be an ordered set of attributes, i.e., $A=\{A_1,\dots,A_n\}$ respectively. Let $a_1,\dots,a_n$ refer to sets of attribute values respectively for variables $A_1,\dots,A_n$. Let U/A_i be the family of all equivalence classes determined by A_i, i.e., a partition of U, and $[x]_{A_i}$ be equivalence classes with respect to the attribute A_i containing element $x\in U$. By $|\cdot|$ we denote the cardinality of a set. Then for $P(A_1,\dots,A_{n-1})>0$, $P(A_n|A_1,\dots,A_{n-1})=P(A_n|A_{n-1})$ **if and only if** for every equivalence class $[x]_{a_{n-1}}$, every equivalence class $[x]_{a_1,\dots,a_{n-2},a_{n-1}}$, every equivalence class $[x]_{a_{n-1},a_n}$ and every equivalence class $[x]_{a_1,\dots,a_{n-2},a_{n-1},a_n}$, there exists some number $\gamma\geq 1$ such that $|[x]_{a_{n-1}}|=\gamma|[x]_{a_1,\dots,a_{n-2},a_{n-1}}|$ and $|[x]_{a_{n-1},a_n}|=\gamma|[x]_{a_1,\dots,a_{n-2},a_{n-1},a_n}|$.*

Proof. Given $P(A_n|A_1,\dots,A_{n-1})=P(A_n|A_{n-1})$, there is

$$\frac{|[x]_{a_1,\dots,a_n}|}{|[x]_{a_1,\dots,a_{n-1}}|}=\frac{|[x]_{a_{n-1},a_n}|}{|[x]_{a_{n-1}}|},\ [x]_{a_1,\dots,a_{n-1}}\neq\emptyset,\ [x]_{a_{n-1}}\neq\emptyset. \tag{2}$$

We write $[x]_{a_{n-1},a_n}=[x]_{a_{n-1}}\cap[x]_{a_n}$, $[x]_{a_1,\dots,a_{n-1}}=[x]_{a_1,\dots,a_{n-2}}\cap[x]_{a_{n-1}}$ and $[x]_{a_1,\dots,a_n}=[x]_{a_1,\dots,a_{n-2}}\cap[x]_{a_{n-1}}\cap[x]_{a_n}$. Therefore, $[x]_{a_1,\dots,a_{n-2}}\cap[x]_{a_{n-1}}\subseteq[x]_{a_{n-1}}$,

$[x]_{a_1,\dots,a_{n-2}}\cap[x]_{a_{n-1}}\cap[x]_{a_n}\subseteq[x]_{a_{n-1}}\cap[x]_{a_n}$, and further $|[x]_{a_1,\dots,a_{n-2}}\cap[x]_{a_{n-1}}|\leq|[x]_{a_{n-1}}|$, $|[x]_{a_1,\dots,a_{n-2}}\cap[x]_{a_{n-1}}\cap[x]_{a_n}|\leq|[x]_{a_{n-1}}\cap[x]_{a_n}|$. According to Eq. (2), there must be $|[x]_{a_{n-1}}|=\gamma|[x]_{a_1,\dots,a_{n-2}}\cap[x]_{a_{n-1}}|$ and $|[x]_{a_{n-1}}\cap[x]_{a_n}|=\gamma|[x]_{a_1,\dots,a_{n-2}}\cap[x]_{a_{n-1}}\cap[x]_{a_n}|$ for some number $\gamma\geq 1$.

Assume that for every equivalence class $[x]_{a_{n-1}}$, $[x]_{a_{n-1},a_n}$, $[x]_{a_1,\dots,a_{n-1}}$ and $[x]_{a_1,\dots,a_n}$, there are $|[x]_{a_{n-1}}|=\gamma|[x]_{a_1,\dots,a_{n-2},a_{n-1}}|$ and $|[x]_{a_{n-1},a_n}|=\gamma|[x]_{a_1,\dots,a_{n-2},a_{n-1},a_n}|$ for $\gamma\geq 1$. Since $[x]_{a_{n-1}}\neq\varnothing$ and $[x]_{a_1,\dots,a_{n-2},a_{n-1}}\neq\emptyset$, hence

$$\frac{|[x]_{a_{n-1},a_n}|}{|[x]_{a_{n-1}}|}=\frac{|[x]_{a_1,\dots,a_{n-2},a_{n-1},a_n}|}{|[x]_{a_1,\dots,a_{n-2},a_{n-1}}|},$$

which implies $P([x]_{a_n}|[x]_{a_{n-1}})=P([x]_{a_n}|[x]_{a_1,\dots,a_{n-2},a_{n-1}})$. Thus we get $P(A_n|A_1,\dots,A_{n-1})=P(A_n|A_{n-1})$. □

The intuition behind Theorem 1 is that, for the ordered triple (A_1, A_2, A_3) with $n=3$, the Markov property can be viewed as one of the properties for (A_1, A_2) to A_2 and (A_1, A_2, A_3) to (A_2, A_3), i.e., the ratios of $|[x]_{A_1,A_2}|$ to $|[x]_{A_2}|$ and $|[x]_{A_1,A_2,A_3}|$ to $|[x]_{A_2,A_3}|$ are the same, and denote this ratio by γ_1. In turn, for $n=4$ and (A_1, A_2, A_3, A_4), the validity of Markov property is to ensure that, the ratios of $|[x]_{A_1,A_2,A_3}|$ to $|[x]_{A_3}|$ and $|[x]_{A_1,A_2,A_3,A_4}|$ to $|[x]_{A_3,A_4}|$ are the same, and denote the ratio by γ_2. For $(A_1,\dots,A_n)$, the validity of Markov property is to ensure that, the ratios of $|[x]_{A_1,\dots,A_{n-1}}|$ to $|[x]_{A_{n-1}}|$ and $|[x]_{A_1,\dots,A_n}|$ to $|[x]_{A_{n-1},A_n}|$ are the same, and denote the ratio by γ_{n-2}. Therefore, the Markov property of the ordered set $(A_1,\dots,A_n)$ needs the existence of $n-2$ coefficients, i.e., $\gamma_1,\dots,\gamma_{n-2}$, which can be equal or unequal. In addition, for a set of attributes without any attribute orders, one can examine the coefficient γ directly from data, via the equivalence classes determined by the attributes, to capture the attributes that satisfy the Markov property. Relying on the relation of Markov property and equivalence classes above, we can easily make the judgement for all possible conditional independent and unconditional independent statements by putting some constraints on equivalence classes derived from attributes.

Corollary 1. *1. For $P(A_1,\dots,A_{n-1})>0$, $P(A_n|A_1,\dots,A_{n-1})=P(A_n|A_i)$ $(1\leq i<n-1)$ **if and only if** for every equivalence class with respect to A_i (i.e., $[x]_{a_i}$), every equivalence class with respect to $\{A_1,A_2,\dots,A_{n-1}\}$ (i.e., $[x]_{a_1,a_2,\dots,a_{n-1}}$), every equivalence class with respect to $\{A_i,A_n\}$ (i.e., $[x]_{a_i,a_n}$) and every equivalence class with respect to $\{A_1,A_2,\dots,A_{n-1},A_n\}$ (i.e., $[x]_{a_1,a_2,\dots,a_{n-1},a_n}$), there exists some number $\gamma\geq 1$ such that $|[x]_{a_i}|=\gamma|[x]_{a_1,a_2,\dots,a_{n-1}}|$ and $|[x]_{a_i,a_n}|=\gamma|[x]_{a_1,a_2,\dots,a_{n-1},a_n}|$.*

*2. For $P(A_{n-1},A_n)>0$, $P(A_n|A_{n-1})=P(A_n)$ **if and only if** for every equivalence class $[x]_{a_n}$, every equivalence class $[x]_{a_{n-1},a_n}$, every equivalence class $[x]_{a_{n-1}}$ and the universe U, there exists some number $\gamma\geq 1$ such that $|U|=\gamma|[x]_{a_{n-1}}|$ and $|[x]_{a_n}|=\gamma|[x]_{a_{n-1},a_n}|$.*

Once the Markov property between attributes for all possible attribute values exists, by the front-door adjustment in Lemma 1 and Definition 5, the effect of interventions between attributes can be estimated. This bridges the gap between interventions and equivalence classes from non-experimental data.

Theorem 2. *Let $S=(U,A)$ be an information system, U be the non-empty finite set of objects called the universe and $x\in U$, A be the set of attributes and W, Y, $Z\in A$ be three different attribute variables, respectively with the attribute values $w=\{w_i:1\leq i\leq n\}$, $y=\{y_j:1\leq j\leq m\}$, $z=\{z_k:1\leq k\leq q\}$ (here i, j, k, m, n, q are positive integers). Denote the equivalence class with respect to W having attribute value w_i by $[x]_{w_i}$, the equivalence class with respect to Y having attribute value y_j by $[x]_{y_j}$, the equivalence class with respect to Z having attribute value z_k by $[x]_{z_k}$. U/W, the partition of U determined by W, equals $\{[x]_{w_1},\dots,[x]_{w_i},\dots,[x]_{w_n}\}$, similarly $U/Y=\{[x]_{y_1},\dots,[x]_{y_j},\dots,[x]_{y_m}\}$ and $U/Z=\{[x]_{z_1},\dots,[x]_{z_k},\dots,[x]_{z_q}\}$. If Z satisfies the condition that, for every equivalence class $[x]_z$, $[x]_{z,y}$, $[x]_{z,w}$ and $[x]_{z,y,w}$, there exists some number $\gamma\geq 1$ such that $|[x]_z|=\gamma|[x]_{z,w}|$ and $|[x]_{z,y}|=\gamma|[x]_{z,y,w}|$, then the causal effect of W on Y is identifiable and is given by the front-door adjustment for Z.*

Proof. Using Theorem 1, we have $P(y|w, z) = P(y|z)$ and $P(w, z) > 0$, which satisfies the conditions of Lemma 1 and the front-door criterion of Definition 5 (e.g., the graph in Definition 5 by deleting the unobserved variable V). Hence, the effect $P(y|do(w))$ is estimated via the formula (1): $P(y|do(w_i)) = \sum_z P(z|w_i) \sum_w P(y|w, z)P(w)$. □

It is not uncommon that the Markov property is valid for some (not all) values of attributes, for example $P(y_j|w_i, z_t) \neq P(y_j|z_t)$ or $P(w_i, z_k) = 0$. The following details this case.

Corollary 2. *1. For $P(w_i, z_k) > 0$, $P(y_j|w_i, z_k) = P(y_j|z_k)$ **if and only if** for the single index i, j, k and equivalence classes $[x]_{z_k}$, $[x]_{z_k, y_j}$, $[x]_{z_k, w_i}$ and $[x]_{z_k, y_j, w_i}$, there exists some number $\gamma \geq 1$ such that $|[x]_{z_k}| = \gamma|[x]_{z_k, w_i}|$ and $|[x]_{z_k, y_j}| = \gamma|[x]_{z_k, y_j, w_i}|$. Moreover, given $P(w, z) > 0$ and only $P(y_j|w_i, z_k) = P(y_j|z_k)$, there is*

$$P(y_j|w_i) = P(y_j|z_k)P(z_k|w_i) + \sum_{t\in[1,q], t\neq k} P(y_j|w_i, z_t)P(z_t|w_i).$$

Further, the effect $P(y_j|do(w_i))$ is not identifiable using the front-door adjustment for Z in Lemma 1 unless $P(y_j|w, z) = P(y_j|z)$ for all possibilities of w and z.

*2. If $P(w_i, z_k) = 0$ for the single i, k (i.e., $[x]_{w_i, z_k} = \emptyset$) but $P(w_{e\neq i}, z) > 0$ and $P(w_i, z_{t\neq k}) > 0$ for all the rest of w and z (i.e., $[x]_{w_{e\neq i}, z} \neq \emptyset$ and $[x]_{w_i, z_{t\neq k}} \neq \emptyset$), then $P(y|w_{e\neq i}, z) = P(y|z)$ **if and only if** $|[x]_z| = \gamma_1|[x]_{z, w_{e\neq i}}|$ and $|[x]_{z,y}| = \gamma_1|[x]_{z, y, w_{e\neq i}}|$ ($\gamma_1 \geq 1$); $P(y|w_i, z_{t\neq k}) = P(y|z_{t\neq k})$ **if and only if** $|[x]_{z_{t\neq k}}| = \gamma_2|[x]_{z_{t\neq k}, w_i}|$ and $|[x]_{z_{t\neq k}, y}| = \gamma_2|[x]_{z_{t\neq k}, y, w_i}|$ ($\gamma_2 \geq 1$).*

Based on the above results concerning the Markov property and the measurement of interventions, we can construct all possible flow graphs for attributes and compute the causal effect under interventions.

Theorem 3. *Let $S = (U, A)$ be an information system, U be the non-empty finite set of objects called the universe and $x \in U$, A be the set of attributes and W, Y, $Z \in A$, respectively with the sets of attribute values $w = \{w_i : 1 \leq i \leq n\}$, $y = \{y_j : 1 \leq j \leq m\}$, $z = \{z_k : 1 \leq k \leq q\}$ (i, j, k, m, n, q are positive integers). $[x]_{w_i}$ denotes the equivalence class with respect to W having attribute value w_i, $[x]_{y_j}$ refers to the equivalence class with respect to Y having attribute value y_j, and $[x]_{z_k}$ indicates the equivalence class with respect to Z having attribute value z_k, together with $U/W = \{[x]_{w_1}, \ldots, [x]_{w_i}, \ldots, [x]_{w_n}\}$, $U/Y = \{[x]_{y_1}, \ldots, [x]_{y_j}, \ldots, [x]_{y_m}\}$ and $U/Z = \{[x]_{z_1}, \ldots, [x]_{z_k}, \ldots, [x]_{z_q}\}$.*

1. *If for every equivalence class $[x]_z$, $[x]_{z,y}$, $[x]_{z,w}$ and $[x]_{z,y,w}$, there exists some number $\gamma \geq 1$ such that $|[x]_z| = \gamma|[x]_{z,w}|$ and $|[x]_{z,y}| = \gamma|[x]_{z,y,w}|$, then the following equation holds*

$$P(y|w) = \sum_z P(y|z)P(z|w) \tag{3}$$

which implies flow graph for variables W, Z and Y (i.e., $w \to z \to y$), and

$$P(y|do(w_i)) = \sum_z P(z|w_i) \sum_w P(y|z, w)P(w). \tag{4}$$

2. *Assume that $[x]_{z_k, w_i} = \emptyset$, but the rest are non-empty and satisfy the condition that, for every equivalence class $[x]_z$, $[x]_{z, w_{e\neq i}}$, $[x]_{z,y}$ and $[x]_{z,y,w_{e\neq i}}$, there exist $\gamma_1 \geq 1$ and $\gamma_2 \geq 1$ such that $|[x]_z| = \gamma_1|[x]_{z,w_{e\neq i}}|$ and $|[x]_{z,y}| = \gamma_1|[x]_{z,y,w_{e\neq i}}|$; $|[x]_{z_{t\neq k}}| = \gamma_2|[x]_{z_{t\neq k}, w_i}|$ and $|[x]_{z_{t\neq k}, y}| = \gamma_2|[x]_{z_{t\neq k}, y, w_i}|$. Then the following equations hold*

$$P(y|w_i) = \sum_{t\neq k} P(y|z_{t\neq k})P(z_{t\neq k}|w_i) \tag{5}$$

$$P(y|w_{e\neq i}) = \sum_z P(y|z)P(z|w_e) \tag{6}$$

which entails the flow graphs for w, z and y (i.e., $w_i \to z_{t\neq k} \to y$ and $w_{e\neq i} \to z \to y$); moreover,

$$P(y|do(w_i)) = \sum_{z_{t\neq k}} P(z_{t\neq k}|w_i) \sum_w P(y|z_{t\neq k}, w)P(w) \tag{7}$$

but $P(y|do(w_{e\neq i}))$ is not identifiable via front-door adjustment for Z. Further, there is $P(y|do(z_{t\neq k})) = \sum_w P(y|w, z_{t\neq k})P(w)$ via the back-door adjustment for W with the structure $w \to z_{t\neq k} \to y$, but $P(y|do(z_k))$ is not identifiable via back-door adjustment for W.

Proof. Taking Theorem 1 and 2, we obtain equation (3) and (4). To prove (5) and (6), we write $P(z_k, w_i) = 0$, $P(w_{e\neq i}, z) > 0$ and $P(w_i, z_{t\neq k}) > 0$. Using Corollary 2, we get $P(y|w_{e\neq i}, z) = P(y|z)$, $P(y|w_i, z_{t\neq k}) = P(y|z_{t\neq k})$. Hence, according to Lemma 3, the following equations hold

$$\begin{aligned} P(y|w_i) &= P(y|w_i, z_k)P(z_k|w_i) + \sum_{z_{t\neq k}} P(y|z_{t\neq k})P(z_{t\neq k}|w_i) \\ &= \sum_{z_{t\neq k}} P(y|z_{t\neq k})P(z_{t\neq k}|w_i) \end{aligned}$$

$$\begin{aligned} P(y|w_{e\neq i}) &= \sum_z P(y|z, w_{e\neq i})P(z|w_{e\neq i}) \\ &= \sum_z P(y|z)P(z|w_{e\neq i}). \end{aligned}$$

Using Lemma 1 and taking Eq. (1), we find

$$\begin{aligned} &\sum_z P(z|w_i)\sum_w P(y|z,w)P(w) \\ &= P(z_k|w_i)\sum_w P(y|z_k,w)P(w) + \sum_{z_{t\neq k}} P(z_{t\neq k}|w_i)\sum_w P(y|z_{t\neq k},w)P(w) \\ &= \sum_{z_{t\neq k}} P(z_{t\neq k}|w_i)\sum_w P(y|z_{t\neq k},w)P(w) \\ &= \sum_{z_{t\neq k}} P(z_{t\neq k}|w_i)\sum_w P(y|z_{t\neq k})P(w) \\ &= \sum_{z_{t\neq k}} P(z_{t\neq k}|w_i)P(y|z_{t\neq k}) \\ &= P(y|do(w_i)) = P(y|w_i) \end{aligned}$$

which establishes Eq. (7). However,

$$\begin{aligned} &\sum_z P(z|w_{e\neq i})\sum_w P(y|z,w)P(w) \\ &= P(z_k|w_{e\neq i})\sum_w P(y|z_k,w)P(w) + \sum_{z_{t\neq k}} P(z_{t\neq k}|w_{e\neq i})\sum_w P(y|z_{t\neq k},w)P(w) \\ &= P(z_k|w_{e\neq i})\Big[P(y|z_k,w_i)P(w_i) + \sum_{w_{e\neq i}} P(y|z_k,w_{e\neq i})P(w_{e\neq i})\Big] \\ &\quad + \sum_{z_{t\neq k}} P(z_{t\neq k}|w_{e\neq i})\sum_w P(y|z_{t\neq k})P(w) \\ &= P(z_k|w_{e\neq i})\sum_{w_{e\neq i}} P(y|z_k,w_{e\neq i})P(w_{e\neq i}) + \sum_{z_{t\neq k}} P(z_{t\neq k}|w_{e\neq i})P(y|z_{t\neq k}) \\ &\neq P(y|do(w_{e\neq i})) \end{aligned}$$

which means that adjustment for Z would yield a biased result for the causal effect of $w_{e\neq i}$ on y. As to the causal effect of Z on Y, we find that $[x]_{w,z_{t\neq k}} \neq \emptyset$ and $P(y|w,z_{t\neq k}) = P(y|z_{t\neq k})$, hence

$$\sum_w P(y|w,z_{t\neq k})P(w) = \sum_w P(y|z_{t\neq k})P(w) = P(y|z_{t\neq k}) = P(y|do(z_{t\neq k})),$$

but due to $[x]_{w_i,z_k} = \emptyset$ and $P(y|w_e,z_k) = P(y|z_k)$ plus $\sum_w P(w) = P(w_i) + \sum_e P(w_e) = 1$ and $\sum_w P(w|z) = 1$, we get

$$\begin{aligned} &\sum_w P(y|w,z_k)P(w) = P(y|w_i,z_k)P(w_i) + \sum_{w_e} P(y|w_e,z_k)P(w_e) \\ &= \sum_{w_e} P(y|z_k)P(w_e) \neq P(y|do(z_k)). \end{aligned}$$ □

The extension to the case with more than three attributes as well as more empty equivalence classes, for the identification of flow graphs and the causal effects, is to be explored.

Corollary 3.

1. *Assume that* $[x]_{z_k,w_i} = \emptyset$ *and* $[x]_{z_k,w_j} = \emptyset$ *but the rest are non-empty and satisfy the condition that, for every equivalence class* $[x]_z$, $[x]_{z,y}$, $[x]_{z,y,w_{e\neq i,j}}$, $[x]_{z_{t\neq k},y,w_i}$ *and* $[x]_{z_{t\neq k},y,w_j}$, *there exist* $\gamma_1 \geq 1$, $\gamma_2 \geq 1$ *and* $\gamma_3 \geq 1$ *such that*

$$\begin{aligned} &|[x]_z| = \gamma_1|[x]_{z,w_{e\neq i,j}}| and |[x]_{z,y}| = \gamma_1|[x]_{z,y,w_{e\neq i,j}}|; \\ &|[x]_{z_{t\neq k}}| = \gamma_2|[x]_{z_{t\neq k},w_i}| and |[x]_{z_{t\neq k},y}| = \gamma_2|[x]_{z_{t\neq k},y,w_i}|; \\ &|[x]_{z_{t\neq k}}| = \gamma_3|[x]_{z_{t\neq k},w_j}| and |[x]_{z_{t\neq k},y}| = \gamma_3|[x]_{z_{t\neq k},y,w_j}|. \end{aligned}$$

Then there hold flow graphs $w_i \to z_{t\neq k} \to y$, $w_j \to z_{t\neq k} \to y$, *and* $w_e \to z \to y$. *Further,*

$$P(y|do(w_i)) = \sum_{z_{t\neq k}} P(z_{t\neq k}|w_i)\sum_w P(y|z_{t\neq k},w)P(w) = P(y|w_i),$$

$$P(y|do(w_j)) = \sum_{z_{t\neq k}} P(z_{t\neq k}|w_j) \sum_{w} P(y|z_{t\neq k}, w)P(w) = P(y|w_j),$$
$$P(y|do(w_e)) \neq \sum_{z} P(z|w_e) \sum_{w} P(y|z, w)P(w).$$

2. *Let $[x]_{z_k,w_i} = \emptyset$ and $[x]_{z_h,w_j} = \emptyset$ but the rest be non-empty, satisfying the condition that, for every equivalence class $[x]_z$, $[x]_{z,y}$, $[x]_{z,y,w_{e\neq i,j}}$, $[x]_{z_{t\neq k},y,w_i}$ and $[x]_{z_{t\neq h},y,w_j}$, there exist $\gamma_1 \geq 1$, $\gamma_2 \geq 1$ and $\gamma_3 \geq 1$ such that*

$$|[x]_z| = \gamma_1|[x]_{z,w_{e\neq i,j}}| and |[x]_{z,y}| = \gamma_1|[x]_{z,y,w_{e\neq i,j}}|;$$
$$|[x]_{z_{t\neq k}}| = \gamma_2|[x]_{z_{t\neq k},w_i}| and |[x]_{z_{t\neq k},y}| = \gamma_2|[x]_{z_{t\neq k},y,w_i}|;$$
$$|[x]_{z_{t\neq h}}| = \gamma_3|[x]_{z_{t\neq h},w_j}| and |[x]_{z_{t\neq h},y}| = \gamma_3|[x]_{z_{t\neq h},y,w_j}|.$$

Then there are flow graphs $w_i \to z_{t\neq k} \to y$, $w_j \to z_{t\neq h} \to y$, and $w_e \to z \to y$, and further

$$P(y|do(w_i)) \neq \sum_{z_{t\neq k}} P(z_{t\neq k}|w_i) \sum_{w} P(y|z_{t\neq k}, w)P(w) \neq P(y|w_i),$$
$$P(y|do(w_j)) \neq \sum_{z_{t\neq h}} P(z_{t\neq h}|w_j) \sum_{w} P(y|z_{t\neq h}, w)P(w) \neq P(y|w_j),$$
$$P(y|do(w_e)) \neq \sum_{z} P(z|w_e) \sum_{w} P(y|z, w)P(w).$$

3. *Assume W, Z_1, Z_2, $Y \in A$ and $w = \{w_i : 1 \leq i \leq n\}$, $y = \{y_j : 1 \leq j \leq m\}$, $z_1 = \{z_1^k : 1 \leq k \leq q\}$, $z_2 = \{z_2^h : 1 \leq h \leq r\}$ and h, i, j, k, m, n, q, r are positive integers. If there exist flow graphs $w \to z_1^{t\neq k} \to z_2 \to y$ and $w_{e\neq i} \to z_1^k \to z_2 \to y$ with $[x]_{w_i,z_1^k} = \emptyset$, then the following equation holds*

$$\begin{aligned} P(y|do(z_1^{t\neq k})) &= \sum_{z_2} P(z_2|z_1^{t\neq k}) \sum_{z_1} P(y|z_2, z_1)P(z_1) \\ &= \sum_{z_2} P(z_2|z_1^{t\neq k}) \sum_{z_1} P(y|z_2)P(z_1) \\ &= \sum_{z_2} P(z_2|z_1^{t\neq k})P(y|z_2) = P(y|z_1^{t\neq k}). \end{aligned}$$

If there exist flow graphs $w \to z_1^{t\neq k} \to z_2 \to y$ and $w_{e\neq i} \to z_1^k \to z_2^{c\neq h} \to y$ with $[x]_{w_i,z_1^k} = \emptyset = [x]_{z_1^k,z_2^h}$, then $\sum_{z_2} P(z_2|z_1^{t\neq k}) \sum_{z_1} P(y|z_2, z_1)P(z_1) \neq P(y|do(z_1^{t\neq k}))$ because of $P(z_2^h|z_1^{t\neq k})P(y|z_2^h, z_1^k)P(z_1^k) = 0$, however,

$$P(y|do(z_1^k)) = \sum_{z_2} P(z_2|z_1^k) \sum_{z_1} P(y|z_2, z_1)P(z_1) = P(y|z_1^k),$$

$$P(y|do(z_1^{t\neq k})) = \sum_{w} P(y|w, z_1^{t\neq k})P(w) = P(y|z_1^{t\neq k}) \quad \text{(by back-door criterion for } W).$$

Corollary 4. *Let an ordered set of attributes $\{W, Z_1, \ldots, Z_n, Y\} \subseteq A$ and w, z_1, $\ldots$, z_n, y denote sets of the attribute values respectively for $W, Z_1, \ldots, Z_n, Y$. Then 1. $P(z_1, \ldots, z_n, w) > 0$ and $P(y|z_1, \ldots, z_n, w) = P(y|z_n)$* ***if and only if*** *for every equivalence class $[x]_{z_n}$, $[x]_{z_n,y}$, $[x]_{z_n,z_1,\ldots,z_{n-1},w}$ and $[x]_{z_n,y,z_1,\ldots,z_{n-1},w}$, there exists $\gamma \geq 1$ such that $|[x]_{z_n}| = \gamma|[x]_{z_n,z_1,\ldots,z_{n-1},w}|$ and $|[x]_{z_n,y}| = \gamma|[x]_{z_n,y,z_1,\ldots,z_{n-1},w}|$. 2. Given $P(z_1, \ldots, z_n, w) > 0$ and $P(y|z_1, \ldots, z_n, w) = P(y|z_n)$, $\ldots$, $P(z_2|z_1, w) = P(z_2|z_1)$, there exists flow graph for $W, Z_1, \ldots, Z_n, Y$, i.e., $w \to z_1 \to \cdots \to z_n \to y$.*

Corollary 5. *Let $W, Z_1, \ldots, Z_n, Y \in A$ and $w, z_1, \ldots, z_n, y$ denote the sets of the attribute values respectively for $W, Z_1, \ldots, Z_n, Y$. Assume that there exists flow graph for $W, Z_1, \ldots, Z_n, Y$, i.e., $w \to z_1 \to \cdots \to z_n \to y$, then*

$$P(y|w) = \sum_{z_1,\ldots,z_n} P(y|z_n)P(z_n|z_{n-1}) \cdots P(z_1|w)$$
$$P(y|w) = \sum_{z_i, i\in[1,n]} P(y|z_i)P(z_i|w)$$
$$P(y|z_i) = \sum_{z_j, j\in(i,n]} P(y|z_j)P(z_j|z_i)$$
$$P(y|z_i, w) = \sum_{z_j, j\in(i,n]} P(y|z_j, z_i, w)P(z_j|z_i, w) = \sum_{z_j, j\in(i,n]} P(y|z_j)P(z_j|z_i) = P(y|z_i)$$

(alternatively, due to $\sum_{z_h} P(z_h|z_i, w) = 1$, *there is* $P(y|z_i, w) = \sum_{z_h, h\in[1,i)} P(y|z_h, z_i, w)P(z_h|z_i, w) = \sum_{z_h} P(y|z_i)P(z_h|z_i, w) = P(y|z_i)$.*) and further, let* $1 \le i < j \le n$, $z_i = \{z_i^t : 1 \le t \le N\}$ *and* $w = \{w_t : 1 \le t \le N\}$, *here i, j, n, t, N represent different positive integers,*

$$\begin{aligned}
P(y|do(z_i^t)) &= \sum_{z_j} P(z_j|z_i^t) \sum_{z_i} P(y|z_j, z_i)P(z_i) (\text{by front-door adjustment for} Z_j) \\
&= \sum_{z_j} P(z_j|z_i^t)P(y|z_j) = P(y|z_i^t) \\
P(y|do(z_i^t)) &= \sum_{w} P(y|z_i^t, w)P(w) (\text{using back-door adjustment for} W) \\
&= \sum_{w} P(y|z_i^t)P(w) = P(y|z_i^t)
\end{aligned}$$

$$\begin{aligned}
P(y|do(w_t)) &= \sum_{z_1,\ldots,z_n} P(z_1, \ldots, z_n|w_t) \sum_{w} P(y|z_1, \ldots, z_n, w)P(w) \\
&= \sum_{z_1,\ldots,z_n} P(z_n|z_{n-1}) \cdots P(z_1|w_t) \sum_{w} P(y|z_n)P(w) \\
&= \sum_{z_1,\ldots,z_n} P(z_n|z_{n-1}) \cdots P(z_1|w_t)P(y|z_n) = P(y|w_t).
\end{aligned}$$

Remark 2. It is worth noting that the causal effect of Z_i on Y is identifiable via both the front-door adjustment for $Z_j (i < j \le n)$ and the back-door adjustment for W or Z_k $(1 \le k < i)$. For example,

$$\begin{aligned}
P(y|do(z_i^t)) &= \sum_{z_k} P(y|z_i^t, z_k)P(z_k) \\
&= \sum_{z_k} P(y|z_i^t)P(z_k) = P(y|z_i^t).
\end{aligned}$$

The key aspect of this argument is that, the use of the front-door criterion (see figure (a)) requires $P(y|z_j, z_i) = P(y|z_j)$, i.e., Z_i is not associated with Y, conditional on Z_j, however the back-door criterion (see figures (b)-(c)) requires $P(y|z_i, w) = P(y|z_i)$ (i.e., W is not associated with Y, conditional on Z_i) or $P(y|z_i, z_k) = P(y|z_i)$ (i.e., Z_k is not associated with Y, conditional on Z_i).

$$Z_i \longrightarrow Z_j \longrightarrow Y$$

(a)

$$\begin{tikzcd}
 & W \arrow[dl] & \\
Z_i \arrow[rr] & & Y
\end{tikzcd}$$

(b)

$$\begin{tikzcd}
 & Z_k \arrow[dl] & \\
Z_i \arrow[rr] & & Y
\end{tikzcd}$$

(c)

(a) A diagram representing the use of front-door criterion for Z_j. (b) and (c) illustrating the use of back-door criterion for W and Z_k.

This finishes the theoretical analysis of the flow graph and the causation between attributes in the context of rough set theory, which provides a sound basis for identifying the structures and causal effects from data tables.

4. Flow graphs and the existing causation

Based on the above theoretical results, first we present a straightforward procedure that solves the extraction of a flow graph from data stored in the form of data tables through the partitions and the equivalence classes determined by the attributes; then we give a more sophisticated discussion on the causal relations between the attributes in the established flow graph via the front-door and back-door criteria for the construction of causation.

4.1. The procedure of the identification for flow graph and causation

The idea behind our procedure starts with the partitions of the universe from the attributes in data table. Suppose $S = (U, C, D)$ be a decision table, where U denotes the universe, C and D are respectively the sets of condition and decision attributes, and let $x \in U$.

Table 1
Voting Intentions.

Y_2	Y_3	Y_1			
		1	2	3	4
1	1	28	8	7	0
	2	153	114	53	14
	3	20	31	17	1
2	1	1	1	0	1
	2	165	86	54	6
	3	30	57	18	4

Table 2
Driving a car in various driving conditions.

Fact no.	Driving conditions			Consequence	Support
	weather (W)	road (R)	time (T)	accident (A)	
1	misty	icy	day	yes	80
2	foggy	icy	night	yes	140
3	misty	not icy	night	yes	40
4	sunny	icy	day	no	500
5	foggy	icy	night	no	20
6	misty	not icy	night	no	200

Table 3
Characterization of flu.

Patient	Headache (H)	Muscle-pain (M)	Temperature (T)	Flu (F)
p1	no	yes	high	yes
p2	yes	no	high	yes
p3	yes	yes	very high	yes
p4	no	yes	normal	no
p5	yes	no	high	no
p6	no	yes	very high	yes

Using Theorem 1, we find that for any three different variables *W, Z, Y*, denote the sets of values for these three variables respectively as *w, z, y*, if the number of the elements in $[x]_z$ (the equivalence class of *Z*) has the same nonzero multiple of the number of elements in $[x]_{z,\,w}$ as the multiple of the number of elements in $[x]_{y,z}$ to the number of elements in $[x]_{y,\,z,\,w}$, then $P(y|z, w) = P(y|z)$; if the number of the elements in *U* has the same nonzero multiple of the number of elements in $[x]_w$ as the multiple of the number of elements in $[x]_z$ to the number of elements in $[x]_{w,\,z}$, then $P(z|w) = P(z)$.

The flow graph "$w \rightarrow z \rightarrow y$", according to 1–3 of Definition 1 and 1 of Theorem 3, requires $P(y|w) = \sum_z P(y|z)P(z|w)$, where *w, y* and *z* induce three different partitions of the universe. In other words, for an individual equivalence class with respect to *w* and *y*, every element of this equivalence class must be outputted by *y* through *z*, namely the inflow (i.e., all elements of an individual $[x]_{wy}$) of the graph equals the outflow (i.e., all elements of an individual $[x]_{wy}$) of the graph. This requirement needs a condition to be satisfied, namely $P(y|z, w) = P(y|z)$, a condition that Markov properties exist for every element of an individual $[x]_{wy}$. If the Markov properties for part of the elements in an individual $[x]_{wy}$ hold, then the summation fails and the flow graph (*w, z, y*) does not exist. In this case we call the relevant structure (that satisfies Markov property and can be used for building the flow graph) *flow graph structure*. But it is worth noting that flow graph structure and flow graph fit the condition of the front-door criterion for the effect of *W* on *Y* and the condition of the back-door criterion for the effect of *Z* on *Y*.

To make the flow graph (structure) exist, we need to examine:

(1) the ratios of the number of elements (i.e., the objects in *U*) in equivalence classes determined by the attributes to the number of elements in equivalence classes from the combination of attributes, under the assumption that the partitions of *U* derived from the attributes as well as their combination are different, i.e., having the different combination of the elements in *U*, but each element has the same number of supports (here 'support' is used to represent the number of the analogous objects that satisfy the decision rule relative to each element), such as Table 1 by omitting the supports and Table 3;
(2) the ratios of the support for all elements in equivalence classes determined by the attributes to the support for all elements in equivalence classes from the combination of attributes, when the partitions determined by the attributes as well as their combination have the same combination of the elements in *U* but each element corresponds to different number of supports, such as Table 2;

(3) the ratios of the support for all elements in each equivalence class determined by the attributes to the support for all elements in equivalence classes from the combination of attributes, provided that the partitions of U from the attributes as well as their combinations are different and each object possess the different supports, such as Table 1.

To exploit the causation between the attributes in the established flow graphs, we adopt the quantities related to interventions and counterfactuals, and calculate the effect of the interventions, the probability of necessity and the probability of sufficiency via mainly the front-door adjustment and also sometimes via the back-door criterion.

Assume that W, Z, Y are three different variables, $w = \{w_i\}, z = \{z_j\}, y = \{y_k\}$ are the sets of values respectively for these three variables, i, j, k are positive integers. Regarding the causal effect of W on Y, if a mediate variable Z satisfies $P(y|w, z) = P(y|z)$, i.e., "$w \to z \to y$" for $P(w, z) > 0$, the effect $P(y|do(w))$ can be given by the front-door adjustment for Z; if Z meets $P(y|w_i, z_{t\neq k}) = P(y|z_{t\neq k})$ and $P(y|w_{e\neq i}, z) = P(y|z)$ for $P(w_i, z_k) = 0$ and the rest are greater than 0, then only $P(y|do(w_i))$ can be estimated by the front-door adjustment for Z, otherwise, $P(y|do(w))$ is not identifiable via the front-door criterion for Z. Moreover, to make $P(y|do(w))$ identifiable via the back-door criterion for a variable Z, in the simplest case we need the condition that either $P(y|w, z) = P(y|w)$, i.e., "$z \to w \to y$" or $P(w|z) = P(w)$, i.e., "$\begin{tikzcd} z \arrow[rr, bend left] & w \arrow[r] & y \end{tikzcd}$". By combining flow graph (structure) and front-door criterion, we can conclude that the front-door criterion requires that all elements of every equivalence classes $[x]_{w,y_k}$ satisfy the Markov property relative to the ordered triple (w, z, y_k); the flow graph demands that for an individual equivalence class $[x]_{w_i,y_k}$, every element satisfies the Markov property relative to the ordered triple (w_i, z, y_k); the flow graph structure requires part of the elements, for an individual equivalence class $[x]_{w_i,y_k}$, satisfy the Markov property relative to the ordered triple (w_i, z, y_k); the causation requires that all elements of every equivalence classes $[x]_{w,y_k}$ and $[x]_{w,y'_k}$ satisfy the Markov property relative to the ordered triples (w, z, y_k) and (w, z, y'_k).

4.2. Illustrations

Now we will illustrate the above presented ideas by means of several examples.

Example 1. Consider an example concerning votes distribution in Table 1 on the basis of sex (Y_2) and social class (Y_3) of voters between political parties (Y_1) from [22]. Y_1 includes four types, namely $1 =$ Conservatives, $2 =$ Labour, $3 =$ Liberal Democrat and $4 =$ Others. Y_2 has two values: $1 =$ male and $2 =$ female. Y_3 consists of three categories, i.e., $1 =$ high, $2 =$ middle and $3 =$ low. Suppose that sex and social class are condition attributes and political party is the decision attribute in the following decision table, where U consists of 22 decision rules and the

Decision table: Voting Intentions

U	Y_2	Y_3	Y_1	Support	U	Y_2	Y_3	Y_1	Support
1	1	1	1	28	12	2	1	1	1
2	1	1	2	8	13	2	1	2	1
3	1	1	3	7	14	2	1	4	1
4	1	2	1	153	15	2	2	1	165
5	1	2	2	114	16	2	2	2	86
6	1	2	3	53	17	2	2	3	54
7	1	2	4	14	18	2	2	4	6
8	1	3	1	20	19	2	3	1	30
9	1	3	2	31	20	2	3	2	57
10	1	3	3	17	21	2	3	3	18
11	1	3	4	1	22	2	3	4	4

support column denotes the number of voters for each decision rule. We want to find description of voting each political party in terms of condition attributes. Note that, when using the different supports to test the Markov property for Y_2, Y_3 and Y_1, there is no Markov property between variables. Consider the possibility of sample selection bias, here we assume each element has the same supports or omit the numbers in the support column when extracting the flow graph (structure) as well as the causal effects, the two rely strongly on the Markov property.

Step 1. Extracting relevant flow graphs from decision table

In this decision table about voting intentions, we have the universe U with 22 elements and its partitions determined by Y_2, Y_3 and Y_1 (here the subscript is the name of the equivalence class, the superscript represents the number of elements in the corresponding equivalence class, the notation a' stands for the negation of a):

$$U/Y_1 = \Big\{\{1, 4, 8, 12, 15, 19\}^6_{[x]_{Y_1=1}}, \{2, 5, 9, 13, 16, 20\}^6_{[x]_{Y_1=2}},$$
$$\{3, 6, 10, 17, 21\}^5_{[x]_{Y_1=3}}, \{7, 11, 14, 18, 22\}^5_{[x]_{Y_1=4}}\Big\}$$

$$U/Y_2 = \Big\{\{1, \dots, 11\}^{11}_{[x]_{Y_2=1}}, \{12, \dots, 22\}^{11}_{[x]_{Y_2=2}}\Big\}$$

$$U/Y_3 = \Big\{\{1,2,3,12,13,14\}^6_{[x]_{Y_3=1}}, \{4,5,6,7,15,16,17,18\}^8_{[x]_{Y_3=2}}, \{8,9,10,11,19,20,21,22\}^8_{[x]_{Y_3=3}}\Big\}$$

$$U/\{Y_2,Y_3\} = \Big\{\{1,2,3\}^3_{[x]_{Y_2=1,Y_3=1}}, \{4,5,6,7\}^4_{[x]_{Y_2=1,Y_3=2}}, \{8,9,10,11\}^4_{[x]_{Y_2=1,Y_3=3}}, \{12,13,14\}^3_{[x]_{Y_2=2,Y_3=1}}, \{15,16,17,18\}^4_{[x]_{Y_2=2,Y_3=2}}, \{19,20,21,22\}^4_{[x]_{Y_2=2,Y_3=3}}\Big\}$$

$$U/\{Y_1,Y_3\} = \Big\{\{1,12\}^2_{[x]_{Y_1=1,Y_3=1}}, \{4,15\}^2_{[x]_{Y_1=1,Y_3=2}}, \{8,19\}^2_{[x]_{Y_1=1,Y_3=3}}, \{2,13\}^2_{[x]_{Y_1=2,Y_3=1}}, \{5,16\}^2_{[x]_{Y_1=2,Y_3=2}}, \{9,20\}^2_{[x]_{Y_1=2,Y_3=3}}, \{3\}^1_{[x]_{Y_1=3,Y_3=1}}, \{6,17\}^2_{[x]_{Y_1=3,Y_3=2}}, \{10,21\}^2_{[x]_{Y_1=3,Y_3=3}}, \{14\}^1_{[x]_{Y_1=4,Y_3=1}}, \{7,18\}^2_{[x]_{Y_1=4,Y_3=2}}, \{11,22\}^2_{[x]_{Y_1=4,Y_3=3}}\Big\}$$

$$U/\{Y_1,Y_2\} = \Big\{\{1,4,8\}^3_{[x]_{Y_1=1,Y_2=1}}, \{12,15,19\}^3_{[x]_{Y_1=1,Y_2=2}}, \{2,5,9\}^3_{[x]_{Y_1=2,Y_2=1}}, \{13,16,20\}^3_{[x]_{Y_1=2,Y_2=2}}, \{3,6,10\}^3_{[x]_{Y_1=3,Y_2=1}}, \{17,21\}^2_{[x]_{Y_1=3,Y_2=2}}, \{7,11\}^2_{[x]_{Y_1=4,Y_2=1}}, \{14,18,22\}^3_{[x]_{Y_1=4,Y_2=2}}\Big\}$$

Thus by employing Definition 1, Theorem 1 and Theorem 3, we obtain

(a) $2\times|[x]_{Y_3=1,Y_2}| = |[x]_{Y_3=1}|$, $2\times|[x]_{Y_3=1,Y_1=1\ (\text{or}\ 2),Y_2}| = |[x]_{Y_3=1,Y_1=1\ (\text{or}\ 2)}|$, but, $|[x]_{Y_3=1,Y_1=3\ (\text{or}\ 4),Y_2}| = |[x]_{Y_3=1,Y_1=3\ (\text{or}\ 4)}|$, hence $P(Y_1=1\ (\text{or}\ 2)|Y_2,Y_3=1) = P(Y_1=1\ (\text{or}\ 2)|Y_3=1)$;
(b) $2\times|[x]_{Y_3=2\ (\text{or}\ 3),Y_2}| = |[x]_{Y_3=2\ (\text{or}\ 3)}|$, $2\times|[x]_{Y_3=2\ (\text{or}\ 3),Y_1,Y_2}| = |[x]_{Y_3=2\ (\text{or}\ 3),Y_1}|$, and therefore $P(Y_1|Y_2,Y_3=2\ (\text{or}\ 3)) = P(Y_1|Y_3=2\ (\text{or}\ 3))$;
(c) $2\times|[x]_{Y_3,Y_2}| = |[x]_{Y_3}|$, $2\times|[x]_{Y_3,Y_1\neq1\ (\text{or}\ 2),Y_2}| = |[x]_{Y_3,Y_1\neq1\ (\text{or}\ 2)}|$, $2\times|[x]_{Y_3=2\ (\text{or}\ 3),Y_2}| = |[x]_{Y_3=2\ (\text{or}\ 3)}|$, $2\times|[x]_{Y_3=2\ (\text{or}\ 3),Y_1\neq3\ (\text{or}\ 4),Y_2}| = |[x]_{Y_3=2\ (\text{or}\ 3),Y_1\neq3\ (\text{or}\ 4)}|$, hence $P(Y_1\neq1\ (\text{or}\ 2)|Y_2,Y_3) = P(Y_1\neq1\ (\text{or}\ 2)|Y_3)$ and $P(Y_1\neq3\ (\text{or}\ 4)|Y_2,Y_3=2\ (\text{or}\ 3)) = P(Y_1\neq3\ (\text{or}\ 4)|Y_3=2\ (\text{or}\ 3))$;
(d) $2\times|[x]_{y_2}| = |U|$, $2\times|[x]_{y_3,y_2}| = |[x]_{y_3}|$ and $2\times|[x]_{y'_3,y_2}| = |[x]_{y'_3}|$, hence $P(y_3|y_2) = P(y_3)$ and $P(y'_3|y_2) = P(y'_3)$;
(e) the structures for Y_1, Y_2 and Y_3 are given as the following figures (1)-(4) (here the superscripts or subscripts on the arrow represent the involved elements in U): according to figures (1) and (3), in this table, the **complete flow graph of** y_2, y_3 **and** $Y_1=1,2$ and the **complete flow graph of** y_2, y_3 **and**$Y_1\neq1\,\text{or}\,2$ exist, because the Markov properties of y_3 relative to the ordered triples $(Y_2,Y_3,Y_1=1,2)$ and $(Y_2,\ Y_3,\ Y_1\neq1\,\text{or}\,2)$ hold, which means every element of $[x]_{Y_2,Y_1=1,2}$ ($[x]_{Y_2,Y_1\neq1\ \text{or}\ 2}$) can be outputted by $Y_1=1,2$ ($Y_1\neq1\,\text{or}\,2$) via Y_3; but concerning figures (2) and (4), since the elements $3\in[x]_{Y_2=1,Y_1=3}$, $14\in[x]_{Y_2=2,Y_1=4}$, as well as the elements $\{1,2\}\subseteq[x]_{Y_2=1,Y_1\neq3}$, $\{1,2,3\}\subseteq[x]_{Y_2=1,Y_1\neq4}$, $\{12,13,14\}\subseteq[x]_{Y_2=2,Y_1\neq3}$, $\{12,13\}\subseteq[x]_{Y_2=2,Y_1\neq4}$, are not outputted by Y_1, in other words, these elements do not satisfy Markov properties for the ordered triples $(Y_2,Y_3=2,3,Y_1=3,4)$ and $(Y_2,Y_3=2,3,Y_1\neq3\ \text{or}\ 4)$. Therefore we only get **flow graph structure of** y_2, $Y_3=2,3$ **and** $Y_1=3,4$, **flow graph structure of** y_2, $Y_3=2,3$ **and**$Y_1\neq3\,\text{or}\,4$.

Step 2. Evaluating causal effects of variables within flow graph (structure)

The causal effect of Y_2 **on** Y_1 **via the front-door criterion for** Y_3.

By referring to **Flow graphs (1) and (3)**, based on 1 of heorem 3, we get the causal effect of Y_2 on $Y_1=1,2$ by the following equations

$$\begin{aligned} P(Y_1=Conservations|do(y_2)) &= \sum_{y_3}P(y_3|y_2)\sum_{y_2}P(Conservations|y_3,y_2)P(y_2) \\ &= \sum_{y_3}P(y_3|y_2)P(Conservations|y_3) \\ &= P(Conservations|y_2) \end{aligned}$$

$$y_2 \longrightarrow y_3 \longrightarrow Y_1 = 1\ (or\ 2)$$

(1) **Flow graph:** from Y_2 to $Y_1 = 1, 2$

$$\begin{tikzcd}
Y_2 = 1 \arrow[r, "6{,}7"] \arrow[dr, "10{,}11"] & Y_3 = 2 \arrow[r, "\mathbf{6{,}17}"] \arrow[dr, "7{,}18"] & Y_1 = 3 \\
Y_2 = 2 \arrow[r, "21{,}22"'] \arrow[ur, "17{,}18"'] & Y_3 = 3 \arrow[r, "11{,}22"'] \arrow[ur, "\mathbf{10{,}21}"'] & Y_1 = 4
\end{tikzcd}$$

(2) Flow graph structure: not including $3 \in [x]_{Y_2=1,Y_1=3}$, $14 \in [x]_{Y_2=2,Y_1=4}$

$$y_2 \longrightarrow y_3 \longrightarrow Y_1 \neq 1\ (or\ 2)$$

(3) **Flow graph:** from Y_2 to $Y_1 \neq 1\ (or\ 2)$

$$y_2 \longrightarrow Y_3 = 2, 3 \longrightarrow Y_1 \neq 3\ (or\ 4)$$

(4) Flow graph structure: not including $\{1, 2\} \subseteq [x]_{Y_2=1,Y_1\neq 3}$, $\{1, 2, 3\} \subseteq [x]_{Y_2=1,Y_1\neq 4}$, $\{12, 13, 14\} \subseteq [x]_{Y_2=2,Y_1\neq 3}$, $\{12, 13\} \subseteq [x]_{Y_2=2,Y_1\neq 4}$

$$\begin{aligned}
P(Y_1 = Labour|do(y_2)) &= \sum_{y_3} P(y_3|y_2) \sum_{y_2} P(Labour|y_3, y_2)P(y_2) \\
&= \sum_{y_3} P(y_3|y_2)P(Labour|y_3) = P(Labour|y_2)
\end{aligned}$$

$$\begin{aligned}
P(Y_1 \neq Labour|do(y_2)) &= \sum_{y_3} P(y_3|y_2) \sum_{y_2} P(Y_1 \neq Labour|y_3, y_2)P(y_2) \\
&= \sum_{y_3} P(y_3|y_2)P(Y_1 \neq Labour|y_3) = P(Y_1 \neq Labour|y_2)
\end{aligned}$$

$$\begin{aligned}
&P(Y_1 \neq Conservations|do(y_2)) \\
&= \sum_{y_3} P(y_3|y_2) \sum_{y_2} P(Y_1 \neq Conservations|y_3, y_2)P(y_2) \\
&= \sum_{y_3} P(y_3|y_2)P(Y_1 \neq Conservations|y_3) \\
&= P(Y_1 \neq Conservations|y_2).
\end{aligned}$$

However, as to the causal effect of Y_2 on $Y_1 = 3, 4$, we find in figure (2) that $P(Y_1 = Liberal\ Democrat|do(y_2))$ and $P(Y_1 = Others|do(y_2))$ are unidentifiable via the front-door adjustment for Y_3 because of $P(Y_1 = Liberal\ Democrat|y_2, Y_3 = high) \neq P(Y_1 = Liberal\ Democrat|Y_3 = high)$ and $P(Y_1 = Others|y_2, Y_3 = high) \neq P(Y_1 = Others|Y_3 = high)$. With the same reason in figure (4), the effects of Y_2 on $Y_3 \neq 3$ and $Y_3 \neq 4$, i.e.,

$P(Y_1 \neq Liberal\ Democrat|do(y_2))$ and $P(Y_1 \neq Others|do(y_2))$ can not be given via the front-door criterion for Y_3 due to $P(Y_1 \neq Others|y_2, Y_3 = high) \neq P(Y_1 \neq Others|Y_3 = high)$ and $P(Y_1 \neq Liberal\ Democrat|y_2, Y_3 = high) \neq P(Y_1 \neq Liberal\ Democrat|Y_3 = high)$.

Consider the causal explanations of Y_2 on Y_1, namely the necessity and sufficiency of causation, denoted respectively as "Probability of necessity"$PN(Y_2, Y_1)$ (i.e., how necessary Y_2 is for the production of Y_1) and "Probability of sufficiency" $PS(Y_2, Y_1)$ (i.e., how sufficient Y_2 is for the production of Y_1). More specifically, using Definition 6 and Lemma 2, one have $0 \leq PN(Y_2 = male, Y_1 = Conservations) \leq 1$, $0 \leq PS(Y_2 = male, Y_1 = Conservations) \leq \frac{3}{8}$;

$0 \leq PN(Y_2 = female, Y_1 = Conservations) \leq 1$, and

$0 \leq PS(Y_2 = female, Y_1 = Conservations) \leq \frac{3}{8}$;

$0 \leq PN(Y_2 = male, Y_1 = Labour) \leq 1$ and $0 \leq PS(Y_2 = male, Y_1 = Labour) \leq \frac{3}{8}$;

$0 \leq PN(Y_2 = female, Y_1 = Labour) \leq 1$ and $0 \leq PS(Y_2 = female, Y_1 = Labour) \leq \frac{3}{8}$.

Note that the probabilities PN and PS are bounded. As a result, according to the $PN > \frac{1}{2}$ criterion [30] and the fact that $0 \leq PS \leq \frac{1}{2}$ means that the probabilities of causation cannot be determined from statistical data [25], the conclusion is that there is no causation between sex and voting for the Conservations party and the Labour party using this data table. Referring to rough set theory, we have $\underline{male}(Conservations) = \emptyset = \underline{male}(Labour)$, $\underline{female}(Conservations) = \emptyset = \underline{female}(Labour)$, $\eta(male, Conservations) \approx -0.007$, $\eta(male, Labour) \approx 0.001$,

$\eta(female, Conservations) \approx 0.007$, $\eta(female, Labour) \approx -0.002$, as expected.

The causal effect of Y_3 **on** Y_1 **via the back-door criterion for** Y_2**.**

According to 2 of Theorem 3 as well as figures (1)-(4), plus $P(y_3|y_2) = P(y_3)$ and $P(y_3'|y_2) = P(y_3')$, the causal effect of Y_3 on Y_1 can be given via the back-door adjustment for Y_2, namely

$$P(y_1|do(y_3)) = \sum_{y_2} P(y_1|y_2, y_3)P(y_2) = P(y_1|y_3)$$
$$P(y_1'|do(y_3)) = \sum_{y_2} P(y_1'|y_2, y_3)P(y_2) = P(y_1'|y_3)$$
$$P(y_1'|do(y_3')) = \sum_{y_2} P(y_1'|y_2, y_3')P(y_2) = P(y_1'|y_3')$$
$$P(y_1|do(y_3')) = \sum_{y_2} P(y_1|y_2, y_3')P(y_2) = P(y_1|y_3')$$

thus yielding $\frac{1}{4} \leq PN(Y_3 = high, Y_1 = Conservatives) \leq 1$ and
$\frac{1}{9} \leq PS(Y_3 = high, Y_1 = Conservatives) \leq \frac{4}{9}$;
$0 \leq PN(Y_3 = middle, Y_1 = Conservatives) \leq 1$ and
$0 \leq PS(Y_3 = middle, Y_1 = Conservatives) \leq \frac{7}{20}$;
$0 \leq PN(Y_3 = low, Y_1 = Conservatives) \leq 1$ and
$0 \leq PS(Y_3 = low, Y_1 = Conservatives) \leq \frac{7}{20}$;
$\frac{1}{4} \leq PN(Y_3 = high, Y_1 = Labour) \leq 1$ and
$\frac{1}{9} \leq PS(Y_3 = high, Y_1 = Labour) \leq \frac{4}{9}$;
$0 \leq PN(Y_3 = middle, Y_1 = Labour) \leq 1$ and
$0 \leq PS(Y_3 = middle, Y_1 = Labour) \leq \frac{7}{20}$;
$0 \leq PN(Y_3 = low, Y_1 = Labour) \leq 1$ and
$0 \leq PS(Y_3 = low, Y_1 = Labour) \leq \frac{7}{20}$;
$0 \leq PN(Y_3 = high, Y_1 = Liberal\ Democrat) \leq 1$ and
$0 \leq PS(Y_3 = high, Y_1 = Liberal\ Democrat) \leq \frac{2}{9}$;
$0 \leq PN(Y_3 = high, Y_1 = Others) \leq 1$ and $0 \leq PS(Y_3 = high, Y_1 = Others) \leq \frac{2}{9}$;
$\frac{1}{7} \leq PN(Y_3 = middle, Y_1 = Liberal\ Democrat) \leq 1$ and
$\frac{1}{22} \leq PS(Y_3 = middle, Y_1 = Liberal\ Democrat) \leq \frac{7}{22}$;
$\frac{1}{7} \leq PN(Y_3 = middle, Y_1 = Others) \leq 1$ and $\frac{1}{22} \leq PS(Y_3 = middle, Y_1 = Others) \leq \frac{7}{22}$;
$\frac{1}{7} \leq PN(Y_3 = low, Y_1 = Liberal\ Democrat) \leq 1$ and
$\frac{1}{22} \leq PS(Y_3 = low, Y_1 = Liberal\ Democrat) \leq \frac{7}{22}$;
$\frac{1}{7} \leq PN(Y_3 = low, Y_1 = Others) \leq 1$ and $\frac{1}{22} \leq PS(Y_3 = low, Y_1 = Others) \leq \frac{7}{22}$.

By combining the $PN > \frac{1}{2}$ criterion and the fact that $0 \leq PS \leq \frac{1}{2}$ means that the probabilities of causation cannot be determined from statistical data, we obtain the chance that high social class was necessary for the vote of the Conservations party and the Labour party is bounded on $[\frac{1}{4}, 1]$; the chance that middle social class and low social class were necessary for the vote of the Liberal Democrat party and the other party is bounded on $[\frac{1}{7}, 1]$; but the sufficient causation of the social classes on voting for political parties, as well as the necessary causation of high social class on voting for the Liberal Democrat party and the other party and of middle social class and low social class on voting for the Conservations party and the Labour party, cannot be determined from this table.

In rough set theory, there are $\underline{Y_2, Y_3}(Y_1) = \emptyset = \underline{Y_2}(Y_1) = \underline{Y_3}(Y_1)$, $\underline{high}(Conservatives) = \emptyset = \underline{high}(Labour)$, $\eta(high, Conservatives) \approx 0.159$, $\eta(high, Labour) \approx -0.271$;

$\underline{middle}(Liberal\ Democrat) = \emptyset = \underline{middle}(Others)$, $\eta(middle, Liberal\ Democrat) \approx -0.017$, $\eta(middle, Others) \approx 0.018$; $\underline{low}(Liberal\ Democrat) = \emptyset = \underline{low}(Others)$,

$\eta(low, Liberal Democrat) \approx 0.068$, $\eta(low, Others) \approx -0.033$, reflecting that, the dependency between high social class and the vote of the Conservations party and the Labour party is more apparent, compared with the dependency between middle social class and the vote of the Liberal Democrat party and the other party as well as between low social class and the vote of the Liberal Democrat party and the other party, according to the statistical data table.

Remark 3. It is necessary to emphasize that we place a constraint in the beginning, i.e., omitting the sample selection bias or the different supports. As exhibited in Example 1, we can construct the flow graphs hidden in data table and measure the causal effects of Y_2 on Y_1 and of Y_3 on Y_1 via the front-door and back-door criteria without considering the supports. This may also imply that there exists the bias towards sample selection in this table.

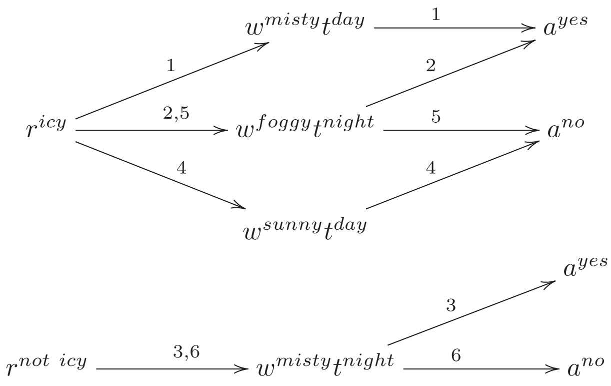


Fig. 1. From “road” to “accident”.

Example 2. Consider another decision table from [21] of driving a car in various driving conditions (e.g., weather, road and time) plus 980 analogous cases concerning information whether an accident has occurred or not. Take driving conditions as condition attributes and accident as decision attribute. The support column denotes the number of analogous cases for each fact. The partitions for 980 cases by employing the attributes are given as follows:

$$
\begin{aligned}
U/W &= \{\{1,3,6\}^{320}_{[x]_{W=misty}}, \{2,5\}^{160}_{[x]_{W=foggy}}, \{4\}^{500}_{[x]_{W=sunny}}\}\\
U/R &= \{\{1,2,4,5\}^{740}_{[x]_{R=icy}}, \{3,6\}^{240}_{[x]_{R=noticy}}\}\\
U/T &= \{\{1,4\}^{580}_{[x]_{T=day}}, \{2,3,5,6\}^{400}_{[x]_{T=night}}\}\\
U/A &= \{\{1,2,3\}^{260}_{[x]_{A=yes}}, \{4,5,6\}^{720}_{[x]_{A=no}}\}\\
U/\{W,R,T\} &= \{\{1\}^{80}_{[x]_{misty,icy,day}}, \{3,6\}^{240}_{[x]_{misty,noticy,night}},\\
&\quad \{2,5\}^{160}_{[x]_{foggy,icy,night}}, \{4\}^{500}_{[x]_{sunny,icy,day}}\}\\
U/\{W,R\} &= \{\{1\}^{80}_{[x]_{misty,icy}}, \{3,6\}^{240}_{[x]_{misty,noticy}}, \{2,5\}^{160}_{[x]_{foggy,icy}}, \{4\}^{500}_{[x]_{sunny,icy}}\}\\
U/\{W,T\} &= \{\{1\}^{80}_{[x]_{misty,day}}, \{3,6\}^{240}_{[x]_{misty,night}}, \{2,5\}^{160}_{[x]_{foggy,night}}, \{4\}^{500}_{[x]_{sunny,day}}\}\\
U/\{R,T\} &= \{\{1,4\}^{580}_{[x]_{icy,day}}, \{3,6\}^{240}_{[x]_{noticy,night}}, \{2,5\}^{160}_{[x]_{icy,night}}\}
\end{aligned}
$$

Step 1. Extracting relevant flow graphs from decision table

Obviously, there exist $U/\{W,R,T,A\} = U/\{W,R,A\} = U/\{W,T,A\}$ but $[x]_{misty,icy,night}$, $[x]_{misty,not\,icy,day}$, $[x]_{foggy,icy,day}$, $[x]_{sunny,icy,night}$ are empty, and $P(W,\ R|T) \neq P(W,\ R)$, $P(W,\ T|R) \neq P(W,\ T)$ because of $|[x]_t| \neq |U|$ and $|[x]_r| \neq |U|$ by Corollary 1. Using Corollary 2 we have $|[x]_{misty,icy,day}| = |[x]_{misty,icy}| = |[x]_{misty,day}|$ and $|[x]_{misty,icy,day,a}| = |[x]_{misty,icy,a}| = |[x]_{misty,day,a}|$, hence $P(a|misty,icy,day) = P(a|misty,icy) = P(a|misty,day)$, i.e.,

$$t^{day} \longrightarrow w^{misty}r^{icy} \longrightarrow a^{yes} \qquad r^{icy} \longrightarrow w^{misty}t^{day} \longrightarrow a^{yes};$$

likewise, $P(a|misty, not\ icy, night) = P(a|misty, not\ icy) = P(a|misty, night)$, i.e.,

$$t^{night} \longrightarrow w^{misty}r^{noticy} \longrightarrow a \qquad r^{noticy} \longrightarrow w^{misty}t^{night} \longrightarrow a;$$

$P(a|foggy,icy,night) = P(a|foggy,icy) = P(a|foggy,night)$, i.e.,

$$t^{night} \longrightarrow w^{foggy}r^{icy} \longrightarrow a \qquad r^{icy} \longrightarrow w^{foggy}t^{night} \longrightarrow a;$$

$P(a|sunny,icy,day) = P(a|sunny,icy) = P(a|sunny,day)$, i.e.,

$$t^{day} \longrightarrow w^{sunny}r^{icy} \longrightarrow a^{no} \qquad r^{icy} \longrightarrow w^{sunny}t^{day} \longrightarrow a^{no}.$$

According to these structures, we can construct the following flow graphs (see Flow Graph 1 and Flow Graph 2) related to the compound attributes (here the superscripts or subscripts on the arrow represent the involved elements in U, i.e., the fact number involved):

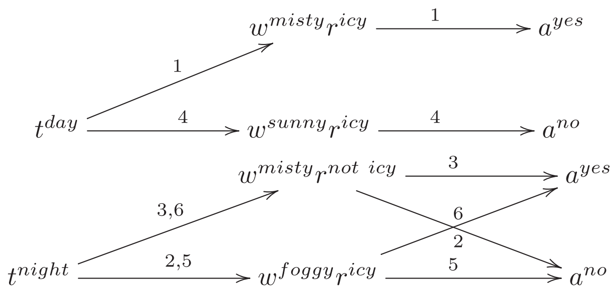


Fig. 2. From “time” to “accident”.

Regarding $P(a|w, r, t)$, it is easy to find that $P(a|w, r, t) \neq P(a|t)$ while $P(a|foggy, icy, night) = P(a|foggy)$, $P(a|sunny, icy, day) = P(a|sunny)$ and $P(a|misty, not\ icy, night) = P(a|not\ icy)$.

Concerning the relations of R, W and A, we prove that $P(a|r, misty) \neq P(a|misty)$ while $P(a|foggy, icy) = P(a|foggy)$ and $P(a|sunny, icy) = P(a|sunny)$, i.e.,

$$r^{icy} \longrightarrow w^{foggy,sunny} \longrightarrow a$$

$P(a|w, icy) \neq P(a|icy)$ but $P(a|noticy, misty) = P(a|noticy)$, i.e.,

$$w^{misty} \longrightarrow r^{noticy} \longrightarrow a.$$

In terms of the relations of R, T and A, we have $P(a|night, r) \neq P(a|night)$ while $P(a|day, icy) = P(a|day)$, i.e., $r^{icy} \longrightarrow t^{day} \longrightarrow a$;

$P(a|icy, t) \neq P(a|icy)$ but $P(a|not\ icy, night) = P(a|not\ icy)$, i.e.,

$$t^{night} \longrightarrow r^{noticy} \longrightarrow a.$$

As to the relations of T, W and A, there exist $P(a|w, day) \neq P(a|day)$ and $P(a|w, night) \neq P(a|night)$, but $P(a|foggy, night) = P(a|foggy)$ and $P(a|sunny, day) = P(a|sunny)$, namely

$$t^{night} \longrightarrow w^{foggy} \longrightarrow a \qquad t^{day} \longrightarrow w^{sunny} \longrightarrow a.$$

As to the relations of W, R and T, we find that P(w|t, icy)\neq P(w|icy) but

$$\begin{aligned}
P(w|night, noticy) &= P(w|noticy), i.e., \quad t^{night} \longrightarrow r^{noticy} \longrightarrow w;\\
P(sunny|r, night) &= P(sunny|night) and P(w|icy, day) = P(w|day), i.e.,\\
&\quad r \longrightarrow t^{night} \longrightarrow w^{sunny} \qquad r^{icy} \longrightarrow t^{day} \longrightarrow w;\\
P(r|foggy, night) &= P(r|foggy) and P(r|sunny, day) = P(r|sunny), i.e.,\\
&\quad t^{night} \longrightarrow w^{foggy} \longrightarrow r \qquad t^{day} \longrightarrow w^{sunny} \longrightarrow r;\\
P(noticy|day) &= P(noticy|day, misty) = P(noticy|day, sunny), i.e.,\\
&\quad w^{misty} \longrightarrow t^{day} \longrightarrow r^{noticy} \quad w^{sunny} \longrightarrow t^{day} \longrightarrow r^{not\ icy};\\
P(t|foggy, icy) &= P(t|foggy) and P(t|sunny, icy) = P(t|sunny), namely\\
&\quad r^{icy} \longrightarrow w^{foggy} \longrightarrow t \qquad r^{icy} \longrightarrow w^{sunny} \longrightarrow t;\\
P(r|foggy, night) &= P(r|foggy) and P(r|sunny, day) = P(r|sunny), namely\\
&\quad t^{night} \longrightarrow w^{foggy} \longrightarrow r \qquad t^{day} \longrightarrow w^{sunny} \longrightarrow r.
\end{aligned}$$

Consider the relations or structures of R, T, W and A on the basis of the above established flow graph structures. By merging the established graphical structures, we have the following two flow graphs (see Flow graph 3 and Flow graph 4):

Apart from the above two flow graphs, regarding the structure of $W = misty$ and A, we get the following graph (see Graphical model): where only exists the **flow graph** $(misty, not\ icy, a = no)$ because of $[x]_{misty,a=no} = \{6\}$; the rest of $(misty, r, a = yes)$ or $(misty, day, not\ icy, a = yes)$ cannot be taken as the flow graph on account of $P(a|icy, day, misty) \neq P(a|day) \neq P(a|icy) \neq P(a|misty)$ (i.e., the Markov property of icy relative to $(misty, day, icy, a)$ is invalid), in other words, $\{1\} = [x]_{misty,a=yes}$ is not outputted by A through R or T.

Step 2. Evaluating causal effects of variables within flow graph (structure)

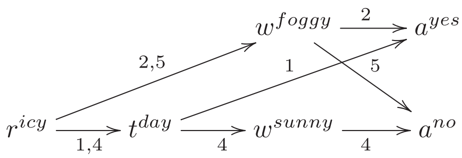


Fig. 3. Plotting from "icy road" to the occurrence of "accident".

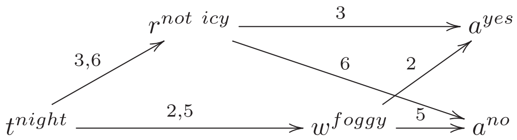
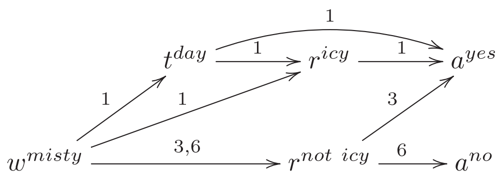


Fig. 4. Depicting from "night time" to the occurrence of "accident".

Based on the above graphical structures for variables *W, R, T* and *A*, we can learn that: for *W* and *A*,

(a) in terms of the structures in the **graphical model from "misty weather" to "accident"**, the causal effect $P(a|do(w))$ is not identifiable via the front-door adjustment for *R*, with the reason that there does not hold $w \to r \to a$, although there is $w^{misty} \to r^{not\,icy} \to a$. In other words, for all values of (w, r), $P(w, r)$ is not always greater than 0 (e.g., $P(W = foggy, R = not\ \ icy) = 0 = P(W = sunny, R = not\ \ icy)$); further for $P(W = foggy, R = not\ \ icy) = 0$ there is $P(a|foggy,\ icy) \neq P(a|icy)$, and $P(a|sunny,\ icy) \neq P(a|icy)$ for $P(W = sunny, R = not\ \ icy) = 0$.

(b) according to Flow Graph 4, $P(a|do(w))$ is also not identifiable via the back-door adjustment for *T*. The reason is, there exists $P(W = sunny, T = night) = 0 = P(W = foggy, T = day)$ and, although there is $P(W = misty, T) > 0$, there holds $P(a|misty,\ t) \neq P(a|misty)$. $P(a|do(w))$ is not given by back-door criterion for *R* with the same reason, that is, $P(W = foggy, R = not\ \ icy) = 0 = P(W = sunny, R = not\ \ icy)$ and $P(a|misty,\ r) \neq P(a|misty)$ for $P(W = misty, R) > 0$.



Graphical model: showing from "misty weather" to the occurrence of "accident"

for *T* and *A*,

(a) by means of Flow Graphs 2 and 4, the causal effect $P(a|do(t))$ is not identifiable via the front-door adjustment for *R* and *W*, with the reason that $P(T = day, R = not\ \ icy) = 0 = P(T = night, W = sunny) = P(T = day, W = foggy)$ and further, $P(a|day,\ icy) \neq P(a|icy)$ for $P(T = day, R = not\ \ icy) = 0$, $P(a|night,\ sunny) \neq P(a|sunny)$ for $P(T = night, W = sunny) = 0$, and $P(a|day,\ foggy) \neq P(a|foggy)$ for $P(T = day, W = foggy) = 0$.

(b) according to the structure $(R = icy, T = day, A)$ of Flow Graph 3, $P(a|do(t))$ is not given by back-door criterion for *R* with the reason of $P(T = day, R = not\ \ icy) = 0$ and $P(a|night,\ r) \neq P(a|night)$ for $P(T = night, R) > 0$.

for *R* and *A*,

(a) referring to Flow Graphs 1 and 3, the causal effect $P(a|do(r))$ is not identifiable via the front-door adjustment for *T*, with the reason that $P(T = day, R = not\ \ icy) = 0$ and further, $P(a|night,\ icy) \neq P(a|night)$ for $P(T = night, R = icy) > 0$. $P(a|do(r))$ is not identifiable via the front-door adjustment for *W* due to $P(W = foggy, T = not\ \ icy) = 0 = P(W = sunny, T = not\ \ icy)$ and $P(a|misty,\ icy) \neq P(a|misty)$ for $P(W = misty, T = icy) > 0$.

(b) according to Flow Graph 4, $P(a|do(r))$ is not identifiable by back-door criterion for *T* with the reason of $P(T = day, R = not\ \ icy) = 0$ and $P(a|t,\ icy) \neq P(a|icy)$ for $P(T, R = icy) > 0$.

In the context of rough set theory, there are $\underline{W,R,T}(A) = \{1,4\} = \underline{W,R}(A) = \underline{W,T}(A)$, $\underline{R,T}(A) = \emptyset$, reflecting that the significance of attribute *W* on *A* is greater than those of attributes *R* on *A* and *T* on *A*; the dependency of $(w = misty, a = yes)$, i.e., $\eta(w = misty, a = yes) = \frac{cer(w=misty, a=yes) - \sigma(a=yes)}{cer(w=misty, a=yes) + \sigma(a=yes)} = \frac{P(a=yes|w=misty) - P(a=yes)}{P(a=yes|w=misty) + P(a=yes)} \approx \frac{0.375 - 0.265}{0.375 + 0.265} \approx 0.172$, meaning that $w = misty$ and $a = yes$ are positively dependent on each other.

Example 3. Consider a simple example of the characterization of flu (this is also used in [29]). Take Headache, Muscle-pain and Temperature as condition attributes and Flu as the decision attribute. We want to find description of Flu in terms of condition attributes. The universe of patients *U* is $U = \{p_1, \ldots, p_6\}$. *H* represents the syndrome of headache using h_1 and

h_0 for the presence and absence of this syndrome, M represents having ($M = m_1$) and not having ($M = m_0$) muscle pain, T represents the temperature with the value of very high (t_1), high (t_2) and normal (t_3), F represents the outcome of flu using f_1 and f_0 for the occurrence and non-occurrence of flu.

$U/\{T\} = \{\{p1, p2, p5\}, \{p3, p6\}, \{p4\}\} = \{[x]_{t_2}, [x]_{t_1}, [x]_{t_3}\}$,
$U/\{H\} = \{\{p1, p4, p6\}, \{p2, p3, p5\}\} = \{[x]_{h_0}, [x]_{h_1}\}$,
$U/\{M\} = \{\{p1, p3, p4, p6\}, \{p2, p5\}\} = \{[x]_{m_1}, [x]_{m_0}\}$,
$U/\{F\} = \{\{p1, p2, p3, p6\}, \{p4, p5\}\} = \{[x]_{f_1}, [x]_{f_0}\}$,
$U/\{H, M\} = \{\{p1, p4, p6\}, \{p2, p5\}, \{p3\}\} = \{[x]_{h_0,m_1}, [x]_{h_1,m_0}, [x]_{h_1,m_1}\}$,
$U/\{H, T\} = \{\{p4\}, \{p1\}, \{p6\}, \{p2, p5\}, \{p3\}\}$
$= \{[x]_{h_0,t_3}, [x]_{h_0,t_2}, [x]_{h_0,t_1}, [x]_{h_1,t_2}, [x]_{h_1,t_1}\}$,
$U/\{M, T\} = \{\{p4\}, \{p1\}, \{p3, p6\}, \{p2, p5\}\} = \{[x]_{m_1,t_3}, [x]_{m_1,t_2}, [x]_{m_1,t_1}, [x]_{m_0,t_2}\}$,
$U/\{H, M, T\} = \{\{p4\}, \{p1\}, \{p6\}, \{p2, p5\}, \{p3\}\}$
$= \{[x]_{h_0,m_1,t_3}, [x]_{h_0,m_1,t_2}, [x]_{h_0,m_1,t_1}, [x]_{h_1,m_0,t_2}, [x]_{h_1,m_1,t_1}\}$.

Step 1. Extracting relevant flow graphs from decision table

It is easy to find that $[x]_{h_1,m_0,t_2} = \{p2, p5\} = [x]_{m_0}$ and $[x]_{h_0,m_1,t_3} = \{p4\} = [x]_{t_3}$, yielding $P(F|h_1, m_0, t_2) = P(F|m_0)$ and $P(F|h_0, m_1, t_3) = P(F|t_3)$;
$[x]_{h_0,m_1,t_1} = \{p6\} = [x]_{h_0,m_1,t_1,f_1}$ and $[x]_{t_1} = \{p3, p6\} = [x]_{t_1,f_1}$, which means
$P(f_1|h_0, m_1, t_1) = P(f_1|t_1)$, likewise $P(f_1|h_1, m_1, t_1) = P(f_1|t_1)$. In the same way, we have $P(F|h_1, m_0) = P(F|m_0)$, $P(F|h_0, t_3) = P(F|t_3)$, $P(F|h_0, t_1) = P(F|t_1)$,
$P(F|h_1, t_1) = P(F|t_1)$, $P(F|m_1, t_1) = P(F|t_1)$, $P(F|m_1, t_3) = P(F|t_3)$, thus obtaining the following graphical structures (see figures (1)-(3), here the superscripts or subscripts on the arrow represent the involved patients). From these figures, it is easy to find that **flow graphs**

$$\begin{tikzcd}
h_1 \arrow[r, "p2{,}p5"] & m_0 \arrow[r, "p2"] \arrow[dr, "p5"] & f_1 \\
 & & f_0
\end{tikzcd}$$

(1) Flow graph structure for (H, M, F)

$$\begin{tikzcd}
h_1 \arrow[r, "p3"] & t_1 \arrow[r, "p3{,}p6"] & f_1 \\
h_0 \arrow[ur, "p6"] \arrow[r, "p4"] & t_3 \arrow[r, "p4"] & f_0
\end{tikzcd}$$

(2) Flow graph structure for (H, T, F)

$$\begin{tikzcd}
m_1 \arrow[r, "p3{,}p6"] \arrow[dr, "p4"] & t_1 \arrow[r, "p3{,}p6"] & f_1 \\
 & t_3 \arrow[r, "p4"] & f_0
\end{tikzcd}$$

(3) Flow graph structure for (M, T, F)

of (h_1, m_0, f_0), (h_0, t_3, f_0) and (m_1, t_3, f_0) exist, because of $\{p5\} = [x]_{h_1,f_0}$ and $\{p4\} = [x]_{h_0,f_0} = [x]_{m_1,f_0}$. But unfortunately, the complete flow graphs do not exist for the remaining structures in figures (1)-(3). The reason is that: for figure (1), $p3 \in [x]_{h_1,f_1}$ is not outputted by F through M, due to the nonexistence of the Markov property between m_1 and {H, F}, namely $p3$ does not satisfy the Markov property relative to the ordered triple (H, M, F) ; for figure (2), $p2 \in [x]_{h_1,f_1}$ and $p1 \in [x]_{h_0,f_1}$ are not outputted by F through T, because the Markov property between t_2 and {H, F} does not hold, that is, $p1$ and $p2$ do not satisfy the Markov property relative to (H, T, F); for figure (3), $p1 \in [x]_{m_1,f_1}$ is not outputted by F through T, due to the invalidity of the Markov property between t_2 and {M, F}, i.e., $p1$ does not satisfy the Markov property relative to (M, T, F).

Step 2. Evaluating causal effects of variables within flow graph (structure)

According to the structures given in Step 1 as well as Theorem 3 and Corollary 3, regarding the ordered triples (h, m, f) and (h, t, f) from figures (1)-(2), the effect of F under the intervention $H = h$, i.e., $P(f|do(h))$ cannot be estimated via the front-door adjustment for M due to $\Sigma_H P(F|m_1, H)P(H) \neq P(F|m_1, H)$ (the Markov property $P(F|m_1, H) = P(F|m_1)$ is invalid) and $[x]_{m_0,h_0} = \emptyset$ (alternatively $\Sigma_H P(F|m_0, H)P(H) \neq P(F|m_0, H)$, the Markov property $P(F|m_0, h_0) = P(F|m_0)$ is invalid), likewise unidentifiable via the front-door adjustment for T because of $\Sigma_H P(F|t_2, H)P(H) \neq P(F|t_2, H)$ (the Markov property $P(F|t_2, H) = P(F|t_2)$ is invalid) and $[x]_{t_3,h_1} = \emptyset$.

Regarding the ordered triple (m, t, f) from figure (3), the causal effect of M on F, i.e., $P(f|do(m))$ is unidentifiable via the front-door adjustment for T because of
$\Sigma_M P(F|t_2, M)P(M) \neq P(F|t_2, M)$ (the Markov property $P(F|t_2, M) = P(F|t_2)$ is invalid) and $[x]_{t_3,m_0} = \emptyset = [x]_{t_1,m_0}$.

As to the causal effect of T on F, i.e., $P(f|do(t))$ is unidentifiable via the back-door adjustment for H and M from figures (2)-(3). The reason is that $P(t_2) \neq P(t_2|h)$, $P(t_3) \neq P(t_3|h)$, $P(f|t_2) \neq P(f|t_2, h)$, $P(f|t_3) \neq P(f|t_3, h)$; $P(t) \neq P(t|m)$, $[x]_{t_1,m_0} = \emptyset = [x]_{t_3,m_0}$, $P(f_1|t_1) \neq P(f_1|t_1, m_0)$, $P(f|t_2) \neq P(f|t_2, m)$. Although there holds $P(t|h) = P(t)$ and $P(f|t, h) = P(f|t)$ for setting $t_0 = t_2 + t_3$, the distribution of F and T has been changed and thus the case of $t = \{t_1, t_0\}$ is not considered.

In the context of rough set theory, there are $\underline{H, M, T}(F) = \{p1, p3, p4, p6\}$, $\underline{H, M}(F) = \{p3\}$, $\underline{H, T}(F) = \{p1, p3, p4, p6\} = \underline{M, T}(F)$, reflecting that the significance of attribute T on F is greater than those of attributes H on F and M on F; the dependency of (t_1, f_1), i.e., $\eta(t_1, f_1) = \frac{cer(t_1, f_1) - \sigma(f_1)}{cer(t_1, f_1) + \sigma(f_1)} = \frac{P(f_1|t_1) - P(f_1)}{P(f_1|t_1) + P(f_1)} = \frac{1 - \frac{2}{3}}{1 + \frac{2}{3}} = \frac{1}{5}$, meaning that t_1 and f_1 are positively dependent on each other.

The established flow graphs are totally different from flow graphs constructed in [29]: the former allows that some relationship between h_1 and f_1 can be calculated directly by the relationships between h_1 and m_0 and between m_0 and f_1; the latter does not, for example, the Markov property does not hold, which implies the invalidity of the front-door criterion for the latter flow graph.

In summary, with regard to developing flow graphs from data, it has been found that flow graph is a directed, acyclic, finite graph and it does not always exist for decision tables, considering that flow graph requires the Markov property for every element in an individual equivalence class induced by the attributes and the attribute values. The front-door criterion needs Markov property for every element in all equivalence classes, a more demanding condition than that of the flow graph. The designing of flow graphs can be achieved by means of the partitions and the equivalence classes induced from the attribute variables, more specifically, the ratio of cardinality of the equivalence classes derived from the different attributes to cardinality of the equivalence classes derived from the different combination of attributes.

5. Conclusion

In this paper, we mostly work with the structures and the causal relationships hidden in flow graphs in the context of rough set theory with the aid of the graphical criteria established in causality literature. By referring to the Pawlak's definition of flow graph, we can extract flow graphs from data within the constraints on the equivalence classes derived from the attributes in data tables. The front-door criterion can be directly applied to the established flow graphs and also some of the attribute variables in flow graphs satisfy the back-door criterion. This is not valid in the flow graphs constructed in [29]. The construction in this paper begins directly with the data table without the requirement of external information such as decision tree, and portrays the conditional independence statements using the properties derived from the attribute variables in a more specific and vivid way compared to the methods of developing a directed acyclic graph or a chain graph in [3,26,31], besides the Markov property for variables.

This is the first attempt to combine flow graph in rough set theory with the tools of identifying the causation in causality literature. The study of flow graph corresponding to the front-door criterion and of the structures satisfying the back-door criterion (the absence of flow graphs) form the general strategies for identifying causal effects from observational variables. According to the illustrations, it has been found that the estimate of causal effects from observations also relates to sample selection bias as well as the more data sample (e.g., new attributes or new attribute values), and if this happens, the alternative is to estimate the causal effects directly from experimental data, although it is sometimes very difficult to do the experiments in real life. For future research, the relationship between the established flow graphs, the causation of variables and attribute reduction ([34,35]) is worth giving close attention, as well as the precise picture of the computational complexity for the proposed procedure of building flow graphs and causation hidden in the graphs. Research in this paper also has shown that we can decompose the established flow graphs into new flow graphs with part of the attribute variables in original flow graphs. When a new attribute is observed, how the flow graphs change depends on the conditional independence of the new attribute and the existed attributes. In addition, how to describe the notion of intervention in rough set terms and how to understand and analyze human behaviors ([7]) using flow graphs are interesting topics. These also attracts our attention.

Acknowledgments

The authors are very grateful to all of the editors and reviewers for making constructive and valuable comments and suggestions on the draft. This research was supported in parts by the National Key R&D Program of China (213), National Natural Science Foundation of China (Grant Numbers 61673301, 61273304, 61573255, 61573259, 61673299, 61603278), Major Project of Ministry of Public Security (20170004), and Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant Number 20130072130004).

References

[1] S. Acharya, B.S. Lee, Incremental causal network construction over event streams, Inf. Sci. (Ny) 261 (2014) 32–51.
[2] K. Aljadda, M. Korayem, C. Ortiz, et al., Mining massive hierarchical data using a scalable probabilistic graphical model, Inf. Sci. (Ny) 425 (2018) 62–75.
[3] M. Baioletti, G. Busanello, B. Vantaggi, Acyclic directed graphs representing independence models, Int. J. Approx. Reas. 52 (2011) 2–18.
[4] E. Bareinboim, J. Pearl, Causal Inference from Big Data: Theoretical Foundations and the Data-fusion Problem (No. TR-R-450), in: Department of Computer Science, University of California, Proceedings of the National Academy of Sciences, Los Angeles, CA, 2015. Forthcoming.
[5] S. Benferhat, Interventions and belief change in possibilistic graphical models, Artif. Intell. 174 (2010) 177–189.
[6] C.J. Butz, W. Yan, B. Yang, An efficient algorithm for inference in rough set flow graphs, in: J.F. Peters, A. Skowron (Eds.), Transactions on Rough Sets V, Springer-Verlag, Berlin Heidelberg, 2006, pp. 102–122.
[7] L. Cao, In-depth behavior understanding and use: the behavior informatics approach, Inf. Sci. (Ny) 180 (2010) 3067–3085.
[8] D. Chitcharoen, P. Pattaraintakorn, Novel matrix forms of rough set flow graphs with applications to data integration, Comput. Math. Appl. 60 (2010) 2880–2897.
[9] G. Chiaselotti, T. Gentile, F. Infusino, Knowledge pairing systems in granular computing, Knowl. Based Syst. 124 (2017) 144–163.
[10] K.J. Cios, W. Pedrycz, R.W. Swiniarski, Rough sets, in: In: Data Mining Methods for Knowledge Discovery, Springer, US, 1998, pp. 27–71.
[11] A. Cuenca, U. Ojha, J. Salt, M.Y. Chow, A non-uniform multi-rate control strategy for a markov chain-driven networked control system, Inf. Sci. (Ny) 321 (2015) 31–47.
[12] T. Eiter, T. Lukasiewicz, Complexity results for structure-based causality, Artif. Intell. 142 (2002) 53–89.
[13] M. Frydenberg, The chain graph markov property, Scand. J. Stat. 17 (4) (1990) 333–353.

[14] D. Galles, J. Pearl, Axioms of causal relevance, Artif. Intell. 97 (1997) 9–43.
[15] J.S. Gero, W. Peng, Understanding behaviors of a constructive memory agent: a markov chain analysis, Knowl. Based Syst. 22 (2009) 610–621.
[16] J.W. Grzymala-Busse, Knowledge acquisition under uncertainty–a rough set approach, J. Intell. Rob. Syst. 1 (1988) 3–16.
[17] J. Hu, T. Li, H. Wang, H. Fujita, Hierarchical cluster ensemble model based on knowledge granulation, Knowl. Based. Syst. 91 (2016) 179–188.
[18] M. Lappenschaar, A. Hommersom, P.J.F. Lucas, Qualitative chain graphs and their application, Int. J. Approx. Reason. 55 (4) (2014) 957–976.
[19] S.L. Lauritzen, N. Wermuth, Graphical models for associations between variables, some of which are qualitative and some quantitative, Ann. Stat. 17 (1) (1989) 31–57.
[20] P. Parviainen, S. Kaski, Learning structures of bayesian networks for variable groups, Int. J. Approx. Reason. 88 (2017) 110–127.
[21] Z. Pawlak, In pursuit of patterns in data reasoning from data–the rough set way, in: J.J.A.e. al. (Ed.), Rough sets and current trends in computing, Springer-Verlag, Berlin Heidelberg, 2002, pp. 1–9.
[22] Z. Pawlak, Rough sets, decision algorithms and bayes' theorem, Eur. J. Oper. Res. 136 (2002) 181–189.
[23] Z. Pawlak, Flow graphs and data mining, in: J.F. Peters, A. Skowron (Eds.), Transactions on Rough Sets III, Springer-Verlag, Berlin Heidelberg, 2005, pp. 1–36.
[24] Z. Pawlak, Decision trees and flow graphs, in: S.G.e. al. (Ed.), Rough Sets and Current Trends in Computing, Springer-Verlag, Berlin Heidelberg, 2006, pp. 1–11.
[25] J. Pearl, Causality: Models, Reasoning, and Inference, second ed., Cambridge University Press, New York, 2009.
[26] J. Pearl, A. Paz, Graphoids: A graph-based logic for reasoning about relevance relations, in: B.D.B.e. al. (Ed.), Advances in Artificial Intelligence-II, North Holland, Amsterdam, 1986, pp. 357–363.
[27] W. Pedrycz, Granular Computing : An Introduction, in: In: Future directions for intelligent systems and information sciences, Physica-Verlag HD, 2000, pp. 309–328.
[28] W. Pedrycz, A. Skowron, Fuzzy and rough sets, in: In: Handbook of data mining and knowledge discovery, Oxford University Press, Inc., New York, NY, USA, 2002, pp. 680–689.
[29] J. Sun, H. Liu, H. Zhang, An extension of Pawlak's flow graphs, in: G.W.e. al. (Ed.), Rough Sets and Knowledge Technology, Springer-Verlag, Berlin Heidelberg, 2006, pp. 191–199.
[30] J. Tian, J. Pearl, Probabilities of causation: bounds and identification, Ann. Math. Artif. Intell. 28 (2000) 287–313.
[31] T. Verma, J. Pearl, An algorithm for deciding if a set of observed independencies has a causal explanation, in: In: Eighth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers Inc, 1992, pp. 323–330.
[32] N. Yao, D.Q. Miao, Z.F. Zhang, G.M. Lang, Probabilistic estimation for generalized rough modus ponens and rough modus tollens, in: F.V.e. al. (Ed.), International Joint Conference on Rough Sets, Springer, Santiago, Chile, 2016, pp. 166–176.
[33] Y.Y. Yao, The two sides of the theory of rough sets, Knowl. Based Syst. 80 (2015) 67–77.
[34] Y.Y. Yao, Y. Zhao, Attribute reduction in decision-theoretic rough set models, Inf. Sci. (Ny) 178 (2008) 3356–3373.
[35] X.Y. Zhang, D.Q. Miao, Reduction target structure-based hierarchical attribute reduction for two-category decision-theoretic rough sets, Inf. Sci. (Ny) 277 (2014) 755–776.