Dynamic Uncertain Causality Graph for Knowledge Representation and Probabilistic Reasoning: Directed Cyclic Graph and Joint Probability Distribution

Qin Zhang

***Abstract*— Probabilistic graphical models (PGMs) such as Bayesian network (BN) have been widely applied in uncertain causality representation and probabilistic reasoning. Dynamic uncertain causality graph (DUCG) is a newly presented model of PGMs, which can be applied to fault diagnosis of large and complex industrial systems, disease diagnosis, and so on. The basic methodology of DUCG has been previously presented, in which only the directed acyclic graph (DAG) was addressed. However, the mathematical meaning of DUCG was not discussed. In this paper, the DUCG with directed cyclic graphs (DCGs) is addressed. In contrast, BN does not allow DCGs, as otherwise the conditional independence will not be satisfied. The inference algorithm for the DUCG with DCGs is presented, which not only extends the capabilities of DUCG from DAGs to DCGs but also enables users to decompose a large and complex DUCG into a set of small, simple sub-DUCGs, so that a large and complex knowledge base can be easily constructed, understood, and maintained. The basic mathematical definition of a complete DUCG with or without DCGs is proved to be a joint probability distribution (JPD) over a set of random variables. The incomplete DUCG as a part of a complete DUCG may represent a part of JPD. Examples are provided to illustrate the methodology.**

***Index Terms*— Causality, complex system, intelligent system, probabilistic reasoning, uncertainty.**

I. Introduction

IN RECENT decades, probabilistic graphical models (PGMs) have been widely applied in uncertain knowledge representation and probabilistic reasoning [1]–[5]. One of the well-known PGMs is Bayesian network (BN) [6]–[14]. As a new member of PGMs, dynamic uncertain causality graph (DUCG) was recently presented [15]–[19]. The basic idea of DUCG is to introduce virtual random events linking parent variables and child variables to represent the uncertain causalities between them. The causality uncertainties are quantified as probabilities of these virtual random events.

In [15], basic concepts and algorithms of DUCG in terms of events are presented, in which it is pointed out that the compact representation of uncertain causalities is desired in practice, because the huge number of parameters in conditional probability tables (CPTs) of BN is not easy to specify, in particular when the statistical data are insufficient. Moreover, although the existing compact representation models (e.g., noisy-OR [6], context-specific independence [20], dynamic causality diagram [21], and so on) are applicable to single-valued cases, they are questionable to be applied to multivalued cases, because an imposed normalization formula has to be applied to keep the normalization over the child variable states [15]. Here, the so-called single-valued means that only the causes of a single true state of a child variable can be specified, while the false state can be treated only as the complement of the true state. This case is modeled as S-DUCG in DUCG. The so-called multivalued means that the causes of all states of a child variable can be specified separately. This case is modeled as M-DUCG in DUCG. The free combination of S-DUCG and M-DUCG is DUCG. No imposed normalization is needed.

In [16], the statistic base of DUCG is addressed. It is pointed out that the parameters of DUCG can be either learned from data or specified by domain experts directly based on their knowledge. Moreover, the representation and inference algorithm of DUCG are extended from the event level to the variable/matrix level. A real application of DUCG, i.e., the online fault diagnoses for a generator system of a real nuclear power plant, is presented, which involves 633 variables and 2952 causal links, and the most exact inferences can be done within 1 s with a laptop computer. Continuous variables, uncertain evidence, and failure forecast are addressed in [17]. The dynamic fault diagnoses are addressed in [18]. An application of DUCG in diagnosing causes of vertigo is addressed in [19]. The diagnostic accuracy rate is 88.3%. In comparison, BN achieves 71.7% as cited in [19].

However, until now, only DUCG without directed cyclic graphs (DCGs) was addressed. It is well known that BN is defined as a factorized joint probability distribution (JPD) over a set of random variables, and, therefore, is defined on only directed acyclic graphs (DAGs), because otherwise the Markov condition or the conditional statistical independence will not be satisfied.

In practice, feedback among variables is frequently encountered, which may cause DCGs. In dynamic cases, an earlier state of a variable may be a descendant of the same

Manuscript received May 17, 2014; revised December 2, 2014 and February 1, 2015; accepted February 5, 2015. Date of publication March 12, 2015; date of current version June 16, 2015. This work was supported by the National Natural Science Foundation of China under Grant 61273330.

The author is with the Institute of Nuclear and New Energy Technology, Tsinghua University, Beijing 100084, China (e-mail: zhangqin@buaa.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2015.2402162

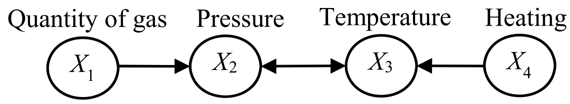

Fig. 1. DCG for a gas container system.

variable later. To avoid DCGs in BN, the involved variables are put into different time slices, and the graph is still DAG. This type of BN is called dynamic BN (DBN), which includes hidden Markov model and Kalman filter model as its special cases [22]. Similar models include recurrent neural networks and recurrent BNs [23], [24], where the recurrent BN can be viewed as a folded DBN. Such dynamic cases in DUCG are addressed in [18], but not in this paper. To clarify terminology, it should be emphasized that DCG exists only in static cases. Time is not involved, because otherwise, DCGs can be broken by putting causes into earlier time slices.

In many cases, DCG is not avoidable. In other cases, a DCG is a combination of a set of DAGs. The two types of DCG are illustrated as follows.

A. First Type of DCG

A typical example of this type of DCG, mentioned in [16], is given in Fig. 1: Some chemical reactions occur in a gas container at a certain pressure (X_2) and temperature (X_3). Meanwhile, abnormal temperature can be caused by overheating or less heating (X_4), and abnormal pressure can be caused by abnormal quantity of gas (X_1) pumped into or pumped out of the container. The query may be $\Pr\{X_{ij}|E\} = ?$, in which and in DUCG, the second subscript indexes the state of a variable, E denotes the evidence and can be $E = X_{2,1}$ (pressure is low), $E = X_{2,2}X_{3,2}$ (pressure and temperature are high), and so on. Note that a state of a variable is actually an event.

In Fig. 1, there is a bidirectional arc between X_2 and X_3, which means that they are cause/parent and consequence/child of each other. The bidirectional arc represents two directed arcs: 1) $X_2 \rightarrow X_3$ and 2) $X_2 \leftarrow X_3$. Obviously, this is a DCG. This is because of the physical nature of X_2 and X_3. The physics law says $X_3/X_2 =$ constant, which means that X_2 and X_3 cause each other exactly simultaneously. Time is not involved in this physics law, and there is no way to put them into different time slices. In other words, DBN is not applicable. It should be pointed out that X_2 and X_3 can cause each other in only one direction in one case, although each direction is possible, which means that X_2 and X_3 may increase or decrease simultaneously with a same proportion, but not increase or decrease repeatedly. This nature generally applies to all DCGs.

This example was presented in [16, Fig. 1], but was revised as a DAG as shown in [16, Fig. 2], i.e., $X_2 \rightarrow X_3$ was removed, so as to meet the requirement of DAG. However, this revision does not reflect the real world. In this paper, the case without revision will be addressed.

In practice, even though a consequence is after a cause, they may be observed simultaneously. That is, when we observe the cause and consequence, they may appear in the same time slice, because the time interval of observations is larger than the time interval between the cause and consequence. For example, suppose two gas containers are connected through a pipe and each has a measured pressure. If one container has a leakage, its pressure will decrease, and then the other container's pressure will decrease immediately after the former. This is because pressure propagation through the pipe is very fast (at sound speed). In real applications, we may measure the two pressures once per second. However, this interval is still too large to detect which decrease is earlier. Therefore, what we know is only that the two pressures can be the cause of each other simultaneously, which means a DCG.

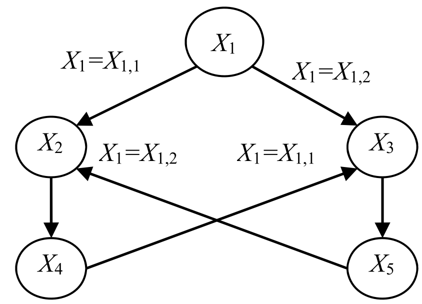

Fig. 2. Cyclic CBN for the hurricane example.

B. Second Type of DCG

Many other DCG examples can be found in [23]–[35], in which an interesting example shown in Fig. 2 is provided in [29]: a hurricane is going to strike two cities, Alphatown and Betaville, but it is not known which city will be hit first. The amount of damage in each city depends on the level of preparations made in each city. Also, the level of preparations in the second city depends on the amount of damage in the first city. The causalities can be represented by a cyclic contingent BN (CBN) [29], as shown in Fig. 2.

In this example, all variables are binary and the definitions of the variable states are listed below.

1) $X_{1,1} \equiv$ {Hit Alphatown first}, $X_{1,2} \equiv$ {Hit Betaville first}.
2) $X_{2,1} \equiv$ {Alphatown's ordinary preparation}, $X_{2,2} \equiv$ {Alphatown's special preparation}.
3) $X_{3,1} \equiv$ {Betaville's ordinary preparation}, $X_{3,2} \equiv$ {Betaville's special preparation}.
4) $X_{4,1} \equiv$ {Alphatown's mild damage}, $X_{4,2} \equiv$ {Alphatown's severe damage}.
5) $X_{5,1} \equiv$ {Betaville's mild damage}, $X_{5,2} \equiv$ {Betaville's severe damage}.

The queries may be $\Pr\{X_{4k}\} = ?$ and $\Pr\{X_{4k}|X_{5j}\} = ?$. It is seen that $X_2 \rightarrow X_4 \rightarrow X_3 \rightarrow X_5 \rightarrow X_2$ forms a DCG.

However, detailed analysis shows that the DCGs in Fig. 2 can be decomposed as two DAGs, as shown in Fig. 3.

The problem is that people may not be able to identify this decomposition. We should find a way to solve such DCGs without decomposition.

In this paper, the methodology to solve DCGs in terms of DUCG is presented. However, before that, we should clarify

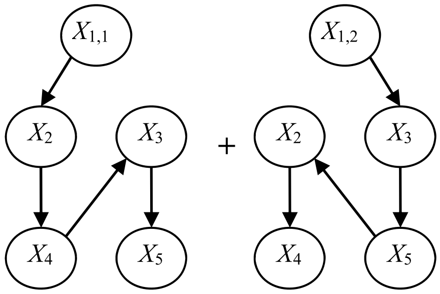

Fig. 3. Decomposed cyclic CBN for the hurricane example.

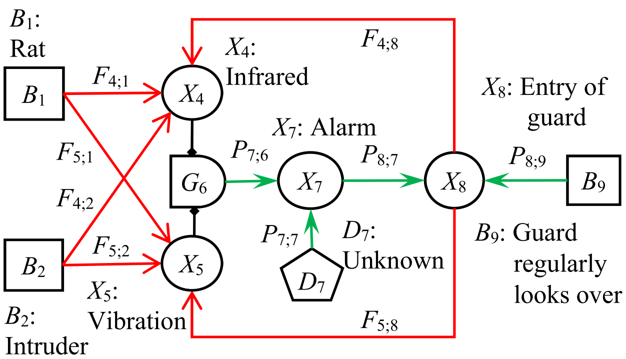

Fig. 4. DUCG with DCGs for an alarm system-detecting intruder.

j	G_{6j}
1	$(X_{4,2}+X_{4,3})X_{5,2}$
2	$X_{4,3}X_{5,3}$
3	Remnant, i.e. $\bar{G}_{6,1}\bar{G}_{6,2} = X_{4,1} \cup X_{5,1} + X_{4,2}X_{5,3}$

Fig. 5. LGS$_6$ in Fig. 4.

the mathematical meaning of DCGs. As well known, DAG is defined as a JPD over a set of random variables. What is the mathematical meaning of DCG? In this paper, it is shown that DUCG with or without DCGs can represent a JPD over a set of variables.

In Section II, DUCG is briefly introduced. In Section III, the mathematical definition of DUCG with or without DCGs is discussed. In Section IV, after a brief introduction on how to deal with DCGs in S-DUCG, which is an extension to DCD presented in [21], the algorithm to calculate the JPD over a set of random variables in both S-DUCG and M-DUCG with DCGs is presented. In Section V, the methodology of dealing with logic gates involved in DCGs is presented. In Section VI, the inference algorithms of S-DUCG and M-DUCG with DCGs are presented and illustrated with examples. Finally, the conclusion is drawn in Section VII. In this paper, it is assumed that the parameters of DUCG have been either obtained from data learning or directly specified by domain experts.

II. Brief Introduction to DUCG

An example of DUCG with DCGs is shown in Figs. 4 and 5. This example is an extension of the example shown in [15, Fig. 33].

The definitions of variable states are similar to those in [15] and are listed below.

1) $B_{1,1} \equiv$ {Rat appears}, $B_{1,2} \equiv$ {No rat}.
2) $B_{2,1} \equiv$ {Intruder appears}, $B_{2,2} \equiv$ {No intruder}.
3) $X_{4,1} \equiv$ {No infrared}, $X_{4,2} \equiv$ {Slight infrared}, $X_{4,3} \equiv$ {Strong infrared}.
4) $X_{5,1} \equiv$ {No vibration}, $X_{5,2} \equiv$ {Slight vibration}, $X_{5,3} \equiv$ {Strong vibration}.
5) $X_{7,1} \equiv$ {Alarm on}, $X_{7,2} \equiv$ {Alarm off}.
6) $D_7 \equiv$ {Unknown cause invoking alarm}.
7) $X_{8,1} \equiv$ {Guard enters}, $X_{8,2} \equiv$ {Guard does not enter}.
8) $B_{9,1} \equiv$ {Guard regularly enters to look over without closing alarm}, $B_{9,2} \equiv \bar{B}_{9,1}$.

The logic gate specification for G_6 (LGS$_6$) is shown in Fig. 5, which is same as [15, Fig. 31].

In this example, the infrared sensor (X_4) and vibration sensor (X_5) send signals to invoke the alarm (X_7) according to the logic specified in the logic gate G_6; some unknown cause (D_7) may also invoke the alarm; the infrared and vibration signals may be caused by the entry of rat (B_1), intruder (B_2), or the guard (X_8); the guard entry can be caused by either the alarm (X_7) or the regular looking over while the guard forgets to close the alarm system (B_9). The query may be $\Pr\{B_{2,1}|X_{7,1}\} =?$, where $X_{7,1}$ represents the event that the alarm is ON and $B_{2,1}$ represents the intruder entry.

In DUCG, B-type variable or event drawn as square represents root cause/parent without input, X-type variable or event drawn as circle represents consequence/child and can also be a cause/parent, G-type variable or event drawn as gate is introduced in DUCG to represent the logic relationship among its input and output variables and/or events, and must be a cause/parent, and D-type variable or event drawn as pentagon represents the default or unknown cause/parent of an X-type variable/event.

Different shapes of a directed arc in DUCG have different meanings. Suppose X_n is a child variable and V_i, $V \in \{X, B, G, D\}$, are its parent variables. Then, $\longrightarrow$ denotes the correlated causal link specified in the CPT between X_n and all V_i. $\longrightarrow$ denotes the linkage event matrix $P_{n;i}$ between X_n and V_i in the single valued cases (modeled as S-DUCG), where the member event $P_{nk;ij}$ in $P_{n;i}$ is a virtual random event introduced in DUCG to represent the uncertain causal mechanism that V_{ij} causes X_{nk}. $\longrightarrow$ denotes the weighted functional event matrix $F_{n;i} \equiv (r_{n;i}/r_n)A_{n;i}$ between X_n and V_i in the multivalued cases (modeled as M-DUCG), where the member event $A_{nk;ij}$ in $A_{n;i}$ is a virtual random event introduced in DUCG to represent the uncertain causal mechanism that V_{ij} causes X_{nk}. $r_{n;i} > 0$ quantifies the uncertain causal relationship intensity between V_i and X_n. If the causality between V_i and X_n does not exist, $r_{n;i} = 0$ and the directed arc does not exist, otherwise $r_{n;i} > 0$ quantifies the relative intensity. $r_n \equiv \sum_i r_{n;i}$, and, therefore, $(r_{n;i}/r_n)$ is a weighting factor of this causality. $\dashrightarrow$ and $\dashrightarrow$ denote the conditional $P_{n;i}$ and $F_{n;i}$, respectively, with condition $Z_{n;i}$, where $Z_{n;i}$ can be any event observable. The subscripts of directed arcs follow the order. 1) child indices; and 2) parent indices.

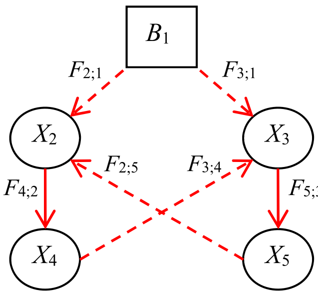

Fig. 6. DUCG with DCGs for the hurricane example.

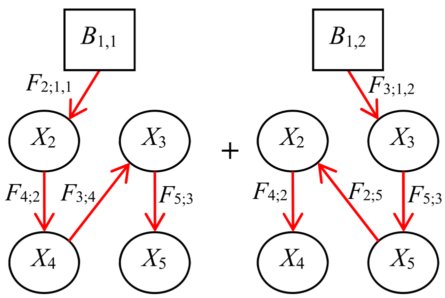

Fig. 7. Combination of a set of DUCGs without DCG for Fig. 6.

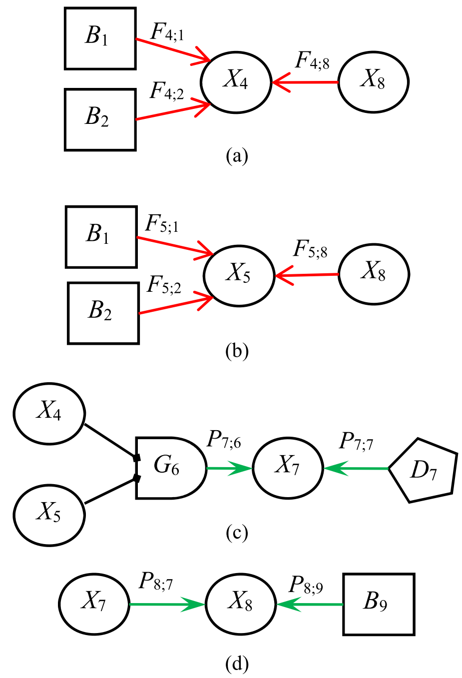

Fig. 8. Modules composing the DUCG in Fig. 4. (a) Module for X_4. (b) Module for X_5. (c) Module for X_7. (d) Module for X_8.

In terms of DUCG, Figs. 2 and 3 can be redrawn as Figs. 6 and 7, respectively, in which B_1 is defined the same as X_1 in Fig. 2.

In Fig. 6, the red dashed directed arcs represent $F_{2;5}$ and $F_{3;1}$ conditional on $Z_{2;5} = Z_{3;1} = B_{1,2}$, and $F_{2;1}$ and $F_{3;4}$ conditional on $Z_{2;1} = Z_{3;4} = B_{1,1}$. Given $B_{1,1}$ and $B_{1,2}$, Fig. 6 becomes Fig. 7.

Note that the $\{F\text{-}, P\text{-}\}$-type texts in Figs. 6 and 7 are not necessary and are usually ignored, because they can be easily recognized according to the shapes and colors of directed arcs. However, in this paper, they remain in all DUCG graphs, so that the readers not familiar with DUCG can understand the meanings of $\{F\text{-}, P\text{-}\}$-type texts easily.

In practice, people tend to represent the causal knowledge locally within a module composed of a child variable and its parent variables, without thinking over the whole system. The DUCG in Fig. 4 can be decomposed as four small and simple modules, as shown in Fig. 8. Or, in reversion, Fig. 4 is constructed by simply connecting the four modules in Fig. 8 through fusing same variables. Where Fig. 8(a) and (b) is in M-DUCG model and Fig. 8(c) and (d) is in S-DUCG model.

It is seen that a module is actually a sub-DUCG composed of only one child variable and its parent variables. For a large and complex system, such as a nuclear power plant, the decomposed construction of knowledge base is even necessary, because otherwise the knowledge base is too large and complex to be constructed and understood. This has been demonstrated in the generator system example presented in [16].

It is easy to understand that the decomposed construction of DUCG may involve DCGs. For the example shown in Fig. 8, the four modules/sub-DUCGs do not have DCGs themselves. However, once they are connected together, DCGs appear as shown in Fig. 4. Therefore, it is not easy for BN to apply the decomposition method to construct a large and complex knowledge base, because BN does not allow DCGs.

In S-DUCG model, the expanding algorithm is given as

$$X_n = \bigcup_i P_{n;i} V_i, \quad V \in \{B, X, G, D\}. \tag{1}$$

In (1), for simplicity, X_{nk} and $P_{nk;i}$ are simplified as X_n and $P_{n;i}$, because only the true/abnormal state k of X_n is involved in S-DUCG. When only one state of V_i is involved, the state of V_i is also ignored, and V_i and $P_{n;i}$ become events. For the example of Fig. 8(d), suppose X_7 and B_9 have only one state involved, according to (1), we have

$$X_8 = P_{8;7} X_7 \cup P_{8;9} B_9. \tag{2}$$

All $\{X\text{-}, P\text{-}, B\text{-}\}$-type notations in (2) represent events.

Equation (1) can be applied repeatedly until only $\{P\text{-}, B\text{-}\}$-type events are in the expanded event expression. The inference can be made according to

$$\Pr\{H_{kj} | E\} = \Pr\{H_{kj} E\} / \Pr\{E\} \tag{3}$$

where H_{kj} denotes the concerned hypothesis composed of $\{B\text{-}, X\text{-}, D\text{-}, G\text{-}, P\text{-}\}$-type events, and E denotes the observed evidence composed of $\{X\text{-}, B\text{-}\}$-type events. It is obvious that logic operation must be applied to expand E and $H_{kj} E$ based on (1), to get the disjoint event expressions for calculating probabilities, and then to calculate (3). More details of S-DUCG can be found in [15], [16], and [21].

In M-DUCG model, similar to (1) and in terms of variables/matrices

$$X_n = \sum_i F_{n;i} V_i = \sum_i \left(\frac{r_{n;i}}{r_n}\right) A_{n;i} V_i, \quad V \in \{B, X, G, D\}. \tag{4}$$

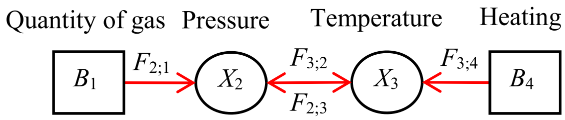

Fig. 9. DUCG with DCG for a gas container system.

In terms of events

$$X_{nk} = \sum_{i,j} F_{nk;ij} V_{ij} = \sum_{i,j} (r_{n;i}/r_n) A_{nk;ij} V_{ij} = \sum_{i} (r_{n;i}/r_n) A_{nk;i} V_i. \quad (5)$$

Equations (4) and (5) can be repeatedly applied to expand a variable/event as in the form of sum-of-products composed of only $\{A\text{-}, B\text{-}\}$-type variable/events associated with r-type parameters.

Equation (3) also applies to M-DUCG, but H_{kj} is composed of $\{B\text{-}, X\text{-}, D\text{-}, G\text{-}, A\text{-}\}$-type events. The logic operation has to be applied too. But different from S-DUCG, the weighting factor operation is involved. Because of this, some new algorithm called weighted set theory is presented in [16].

In DUCG, the upper case letters represent variables/matrices. The lower case letters represent the corresponding probabilities. For example, $v_{nk} \equiv \Pr\{V_{nk}\}$, $V \in \{B, X, G\}$ and $v \in \{x, b, g\}$, and $f_{nk;ij} \equiv \Pr\{F_{nk;ij}\} = (r_{n;i}/r_n)a_{nk;ij}$, $a_{nk;ij} \equiv \Pr\{A_{nk;ij}\}$, $p_{nk;ij} \equiv \Pr\{P_{nk;ij}\}$. Note $\Pr\{D_n\} \equiv 1$, because D_n represents an inevitable event. The $\{b\text{-}, a\text{-}, p\text{-}, r\text{-}\}$-type parameters are specified by users during the construction of DUCG based on either data learning or domain knowledge [16].

Thus, corresponding to (4) and (5), we have

$$x_n = \sum_i f_{n;i} v_i = \sum_i (r_{n;i}/r_n) a_{n;i} v_i \quad (6)$$

$$x_{nk} = \sum_{i,j} f_{nk;ij} v_{ij} = \sum_{i,j} (r_{n;i}/r_n) a_{nk;ij} v_{ij} = \sum_i (r_{n;i}/r_n) a_{nk;i} v_i. \quad (7)$$

In terms of DUCG, Fig. 1 can be redrawn as Fig. 9, which will be discussed in detail in the rest of this paper to illustrate the DUCG methodology dealing with DCGs.

DUCG can bring people with many benefits. First, the uncertain causalities among real variables can be compactly represented with exponentially reduced number of parameters compared with conditional probabilities in CPTs of BN. Second, the logics among parent variables hidden in CPTs of BN are explicitly represented, so that they are easier to be specified based on domain expert's knowledge when historical data are insufficient. Third, the probabilistic inference can be more efficient due to the exponentially reduced scale of problem conditional on the observed evidence. Fourth, the knowledge base can be incomplete (only those concerned need to be specified), while exact inference concerned can still be made. Fifth, low requirement to the accuracy of probability parameters, which means that DUCG is robust. Finally, the inference results can be explained graphically to users, so that they know not only the inference results, but also why the results are such. In the logic operation, the complex probability computation resulting from correlated variables and/or events is simplified as independent variable and/or event probability computation. Sometimes, only the logic operation is needed to get the final results, even without knowing the probability parameters [15], [16].

III. JPD of DUCG With or Without DCGs

It is noted that DCG does not satisfy the Markov condition. In other words, given the states of parent variables, a child variable and its descendants are not independent of its ancestors, because the child may be its own ancestor. Therefore, DCG cannot be simply interpreted as a JPD over a set of random variables. Then, what does DCG mean? So far, Spirtes [25] and Pearl and Dechter [26] prove that the D-separation can be applied in discrete DCG cases. Many studies [21], [27]–[35] just present practical DCG examples without detailed mathematical discussions, or present the exact inference algorithm applicable in only single-valued cases [21], or present approximate inference algorithms [28], [29]. Heckerman *et al.* [31] define a dependence network (DN) allowing DCGs, in which the Gibbs sampling method is used to get the samples of DN, and then the samples are used to represent a JPD over random variables, and to make probabilistic inference. However, DN is viewed as does not always define a unique joint distribution [36]. In summary, in the sense of uncertain causalities quantified by probabilities, the mathematical meaning of DCGs is still unclear.

However, if all $\Pr\{V_{1j_1}V_{2j_2}, \ldots, V_{nj_n}\}$ can be exactly calculated according to the algorithm defined in a PGM, regardless of whether DCGs are involved or not, where $V_{1j_1}V_{2j_2}, \ldots, V_{nj_n}$ represents a state combination of all concerned random variables included in the PGM, and if $\sum_{j_1,j_2,\ldots,j_n} \Pr\{V_{1j_1}V_{2j_2} \ldots V_{nj_n}\} = 1$, according to the definition of JPD, we can conclude that this PGM represents a JPD over $\{V_{1j_1}, V_{2j_2}, \ldots, V_{nj_n}\}$. For the example of Fig. 9, if all $\Pr\{X_{2k}X_{3g}\}$ or $\Pr\{B_{1j}X_{2k}X_{3g}B_{4y}\}$ can be exactly calculated and if $\sum_{k,g} \Pr\{X_{2k}X_{3g}\} = 1$ or $\sum_{j,k,g,y} \Pr\{B_{1j}X_{2k}X_{3g}B_{4y}\} = 1$, the JPD over $\{X_2, X_3\}$ or $\{B_1, X_2, X_3, X_4\}$ is actually given. Therefore, whether a DUCG with or without DCGs represents a JPD over a set of random variables depends on the DUCG algorithm of calculating $\Pr\{V_{1j_1}V_{2j_2} \ldots V_{nj_n}\}$ exactly while satisfying $\sum_{j_1,j_2,\ldots,j_n} \Pr\{V_{1j_1}V_{2j_2} \ldots V_{nj_n}\} = 1$.

In fact, the inference algorithm of DUCG without DCGs has been presented in [15] and [16], which includes the algorithm to calculate $\Pr\{V_{1j_1}V_{2j_2} \ldots V_{nj_n}\}$. In Section IV of this paper, the algorithm to calculate the JPD of a DUCG with DCGs will be presented. These DUCG algorithms ensure that when a DUCG is complete (all $\{a\text{-},b\text{-}\}$-type parameters are given, $\sum_k a_{nk;ij} = 1$ and $\sum_j b_{ij} = 1$), all $\Pr\{V_{1j_1}V_{2j_2} \ldots V_{nj_n}\}$ can be exactly calculated, while satisfying $\sum_{j_1,j_2 \ldots j_n} \Pr\{V_{1j_1}V_{2j_2} \ldots V_{nj_n}\} = 1$, regardless of whether DCGs are involved or not.

It is noted that the completeness of DUCG is an important condition for interpreting a DUCG as a JPD over a set of random variables, because otherwise, not all $\Pr\{V_{1j_1}V_{2j_2} \ldots V_{nj_n}\}$ can be exactly calculated.

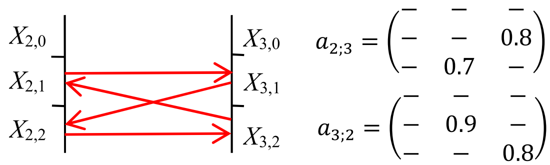

Fig. 10. Incomplete DUCG with DCG for a feedback control system.

Incomplete DUCG means that only the concerned parameters are given, but not all. Fig. 10 is an example of incomplete DUCG, in which the parameters not concerned are given as "−." The meaning of Fig. 10 can be illustrated with a real case of nuclear power plant: Given a certain electricity load of the plant, the low reactor power $X_{2,1}$ causes the low steam temperature $X_{3,1}$, the low steam temperature $X_{3,1}$ causes the high reactor power $X_{2,2}$, the high reactor power $X_{2,2}$ causes the high steam temperature $X_{3,2}$, the high steam temperature $X_{3,2}$ causes the low reactor power $X_{2,1}$, and so on. Only $a_{2,1;3,2}$, $a_{2,2;3,1}$, $a_{3,1;2,2}$, and $a_{3,2;2,1}$ are given, and other parameters are not given, because they are not in question. The parameter values quantify the uncertainties of the causalities between corresponding parent events and child events, mainly because the causalities have uncertain delay. We are not sure whether the effect or consequence will appear immediately after the cause or not. Of course, we can use the function $a_{nk;ij}(t)$ to represent the probability change with respect to time t, as being presented in [21]. But doing this is too trivial and may not be necessary in practice. If $a_{nk;ij}(t)$ is used, we have to keep $\sum_k a_{nk;ij}(t) \leq 1$ (the looser constraint shown in [15, eq. (33)]). However, in this paper, we consider only the case of $a_{nk;ij}$ without time t.

In fact, the only difference between a complete and an incomplete DUCG is that some $\{a$-, b-$\}$-type parameters in the incomplete DUCG are not specified, resulting in that not all $\Pr\{V_{1j_1}V_{2j_2}\dots V_{nj_n}\}$ can be exactly calculated. If we complete these parameters, this incomplete DUCG becomes a complete DUCG. Since an incomplete DUCG is a part of the complete DUCG, an incomplete DUCG can represent a part of a JPD. It should be mentioned that Theorem 1 of DUCG proved in [15] (i.e., $\sum_k X_{nk} = \sum_k (r_{n;i}/r_n) \sum_{i,j} A_{nk;ij} V_{ij} = 1$, where 1 denotes universal set) ensures that causality chains in DUCG are self-relied, and, therefore, some exact inferences concerned can be made based on incomplete DUCG.

Note that any DCG included in a DUCG must have at least one root cause as the DCG's input/ancestor. Otherwise, the DCG cannot be expanded as the expression in the form of sum-of-products composed of only $\{B$-, D-, P-, A-$\}$-type events/variables along with weighting factors $(r_{n;i}/r_n)$ associated with $A_{n;i}$, which means that the expansion result will be null set 0, because no root cause means no source of the DCG. For the example shown in Fig. 9, there must be B_1 or B_4 as a root cause of X_2 or X_3. Otherwise, the DCG between X_2 and X_3 is practically meaningless. For the example shown in Fig. 10, there must be at least one B-type event to cause the vibrational change of the reactor power (X_2) and the steam temperature (X_3). Otherwise, the vibrational change is practically meaningless. In general, a DUCG with DCGs has to satisfy the following assumption.

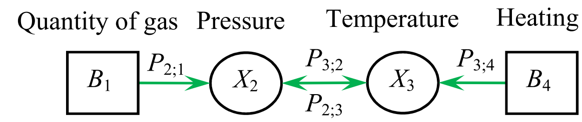

Fig. 11. S-DUCG of the gas container system.

Assumption 3: There must be at least one root cause as an ancestor of any DCG in a DUCG.

For convenience, the assumptions applied in DUCG are indexed sequentially in the serial papers of DUCG. Obviously, only $\{B$-, D-$\}$-type events are such root causes.

In summary, a complete DUCG with or without DCGs together with the algorithm to calculate all $\Pr\{V_{1j_1}V_{2j_2}\dots V_{nj_n}\}$ exactly, while satisfying $\sum_{j_1,j_2,\dots,j_n} \Pr\{V_{1j_1}V_{2j_2}\dots V_{nj_n}\} = 1$, defines a JPD over a set of random variables. Such algorithm includes how to break DCGs, which is presented and illustrated in the following sections. A DUCG with DCGs may be a combination of a set of exclusive and exhaustive DAGs and each DAG defines a JPD. Therefore, such a DUCG with DCGs defines a JPD. An incomplete DUCG represents a part of a JPD represented by its corresponding complete DUCG. The DUCG inference can be made based on a complete or incomplete JPD, which is basically to calculate the posterior probability of a hypothesis H_{kj} concerned conditional on evidence E, as shown in (3).

IV. Algorithm to Calculate the JPD of the DUCG With DCG

In [21], the exact algorithm to deal with DCGs in S-DUCG is initially presented, in which the following assumption is applied.

Assumption 4: Any state of a variable cannot be the cause of any state of the same variable at a same time.

At a same time, a variable must be in a same state. As mentioned earlier, a state of a variable is an event. It is obvious that an event cannot be a cause and consequence of itself at a same time. This is because of the law of contradiction. Therefore, this assumption is universally valid.

A. S-DUCG Model

To illustrate the methodology of S-DUCG with DCGs, we change Fig. 9 as Fig. 11.

As defined in S-DUCG, only the causalities among true/abnormal states are represented. Therefore, for simplicity, the states of variables are ignored, because all of them (X_n, B_i, and $P_{n;i}$) deal with only true/abnormal states. In Fig. 11, B_1 denotes abnormal quantity of gas, X_2 denotes abnormal pressure, X_3 denotes abnormal temperature, and B_4 denotes abnormal heating.

According to the event expanding algorithm defined in [15], which is illustrated in (1) and (2), and based on Fig. 11, we have

$$\begin{aligned} X_2 &= P_{2;1}B_1 \cup P_{2;3}X_3 = P_{2;1}B_1 \cup P_{2;3}(P_{3;4}B_4 \cup P_{3;2}X_2) \\ &= P_{2;1}B_1 \cup P_{2;3}P_{3;4}B_4 \end{aligned} \quad (8)$$

$$\begin{aligned} X_3 &= P_{3;4}B_4 \cup P_{3;2}X_2 = P_{3;4}B_4 \cup P_{3;2}(P_{2;1}B_1 \cup P_{2;3}X_3) \\ &= P_{3;4}B_4 \cup P_{3;2}P_{2;1}B_1. \end{aligned} \quad (9)$$

Based on Assumption 4, the repeated events on the right side of above equations are treated as null set 0.

Suppose $b_1 \equiv \Pr\{B_1\} = 0.1$, $b_4 \equiv \Pr\{B_4\} = 0.2$, $p_{2;1} \equiv \Pr\{P_{2;1}\} = 0.8$, $p_{2;3} \equiv \Pr\{P_{2;3}\} = 0.9$, $p_{3;2} \equiv \Pr\{P_{3;2}\} = 0.8$, $p_{3;4} \equiv \Pr\{P_{3;4}\} = 0.7$. To calculate the probabilities, we may apply (10) and (11) cited in [21] to get the disjoint cut set expressions

$$\begin{aligned} &C_1 \cup C_2 \cup \ldots \cup C_n \\ &\quad = C_1 + \bar{C}_1 C_2 + \bar{C}_1 \bar{C}_2 C_3 + \cdots + \bar{C}_1 \bar{C}_2 \ldots \bar{C}_{n-1} C_n \quad (10) \\ &\bar{C} = \overline{V_1 V_2 \ldots V_m} \\ &\quad = \bar{V}_1 + V_1 \bar{V}_2 + V_1 V_2 \bar{V}_3 + \cdots + V_1 \ldots V_{m-1} \bar{V}_m \quad (11) \end{aligned}$$

where $+$ denotes XOR, $C = V_1 V_2 \ldots V_m$ and $V \in \{P, B\}$. Thus, (8) and (9) can be further expanded and calculated as

$$\begin{aligned} \Pr\{X_2\} &= \Pr\{P_{2;1} B_1 \cup P_{2;3} P_{3;4} B_4\} \\ &= \Pr\{P_{2;1} B_1 + \overline{P_{2;1} B_1} P_{2;3} P_{3;4} B_4\} \\ &= \Pr\{P_{2;1} B_1 + (\overline{P}_{2;1} + P_{2;1} \overline{B}_1) P_{2;3} P_{3;4} B_4\} \\ &= p_{2;1} b_1 + (\overline{p}_{2;1} + p_{2;1} \overline{b}_1) p_{2;3} p_{3;4} b_4 \\ &= 0.8 \times 0.1 + (0.2 + 0.8 \times 0.9) 0.9 \times 0.7 \times 0.2 \\ &= 0.19592 \quad (12) \\ \Pr\{X_3\} &= \Pr\{P_{3;4} B_4 \cup P_{3;2} P_{2;1} B_1\} \\ &= \Pr\{P_{3;4} B_4 + (\overline{P}_{3;4} + P_{3;4} \overline{B}_4) P_{3;2} P_{2;1} B_1\} \\ &= 0.19504. \quad (13) \end{aligned}$$

Through combining the event expressions in (12) and (13), we have

$$\begin{aligned} &\Pr\{X_2 X_3\} \\ &\quad = \Pr\left\{ \begin{array}{l} (P_{2;1} B_1 + (\overline{P}_{2;1} + P_{2;1} \overline{B}_1) P_{2;3} P_{3;4} B_4) \\ \cdot (P_{3;4} B_4 + (\overline{P}_{3;4} + P_{3;4} \overline{B}_4) P_{3;2} P_{2;1} B_1) \end{array} \right\} \\ &\quad = \Pr\left\{ \begin{array}{l} P_{2;1} B_1 P_{3;4} B_4 + (\overline{P}_{3;4} + P_{3;4} \overline{B}_4) P_{3;2} P_{2;1} B_1 \\ + (\overline{P}_{2;1} + P_{2;1} \overline{B}_1) P_{2;3} P_{3;4} B_4 \end{array} \right\} \\ &\quad = p_{2;1} b_1 p_{3;4} b_4 + (\overline{p}_{3;4} + p_{3;4} \overline{b}_4) p_{3;2} p_{2;1} b_1 \\ &\qquad + (\overline{p}_{2;1} + p_{2;1} \overline{b}_1) p_{2;3} p_{3;4} b_4 \\ &\quad = 0.8 \times 0.1 \times 0.7 \times 0.2 + (0.3 + 0.7 \times 0.8) 0.8 \\ &\qquad \times 0.8 \times 0.1 + (0.2 + 0.8 \times 0.9) 0.9 \times 0.7 \times 0.2 \\ &\quad = 0.18216. \quad (14) \end{aligned}$$

Meanwhile, according to S-DUCG model and probability theory, we have

$$\begin{aligned} &\Pr\{\overline{X}_2\} \\ &\quad = \Pr\{\overline{P_{2;1} B_1 \cup P_{2;3} P_{3;4} B_4}\} \\ &\quad = \Pr\{\overline{P_{2;1} B_1} \cdot \overline{P_{2;3} P_{3;4} B_4}\} \\ &\quad = \Pr\{(\overline{P}_{2;1} + P_{2;1} \overline{B}_1)(\overline{P}_{2;3} + P_{2;3} \overline{P}_{3;4} + P_{2;3} P_{3;4} \overline{B}_4)\} \\ &\quad = (\overline{p}_{2;1} + p_{2;1} \overline{b}_1)(\overline{p}_{2;3} + p_{2;3} \overline{p}_{3;4} + p_{2;3} p_{3;4} \overline{b}_4) \\ &\quad = (0.2 + 0.8 \times 0.9)(0.1 + 0.9 \times 0.3 + 0.9 \times 0.7 \times 0.8) \\ &\quad = 0.80408. \quad (15) \end{aligned}$$

Similarly, we have

$$\begin{aligned} \Pr\{\overline{X}_3\} &= 0.80496 \quad (16) \\ \Pr\{\overline{X}_2 X_3\} &= 0.01288 \quad (17) \\ \Pr\{X_2 \overline{X}_3\} &= 0.01376 \quad (18) \\ \Pr\{\overline{X}_2 \overline{X}_3\} &= 0.7912. \quad (19) \end{aligned}$$

It is seen that

$$\Pr\{X_2 X_3\} + \Pr\{\overline{X}_2 X_3\} + \Pr\{X_2 \overline{X}_3\} + \Pr\{\overline{X}_2 \overline{X}_3\} = 1. \quad (20)$$

Moreover, we see

$$\begin{aligned} \Pr\{X_2\} + \Pr\{\overline{X}_2\} &= 0.19592 + 0.80408 = 1 \\ \Pr\{X_3\} + \Pr\{\overline{X}_3\} &= 0.19504 + 0.80496 = 1 \\ \Pr\{X_2 X_3\} + \Pr\{\overline{X}_2 X_3\} &= 0.18216 + 0.01288 \\ &= 0.19504 = \Pr\{X_3\} \\ \Pr\{X_2 X_3\} + \Pr\{X_2 \overline{X}_3\} &= 0.18216 + 0.01376 \\ &= 0.19592 = \Pr\{X_2\}. \quad (21) \end{aligned}$$

Obviously, the S-DUCG with DCG shown in Fig. 11 represents the JPD over $\{X_2, X_3\}$. It is easy to understand that this S-DUCG also represents the JPD over $\{B_1, X_2, X_3, B_4\}$, which can be obtained by multiplying the logic expressions in (16)–(19) with $B_1 B_4$, $\overline{B}_1 B_4$, $B_1 \overline{B}_4$, and $\overline{B}_1 \overline{B}_4$, respectively. For example

$$\begin{aligned} &\Pr\{X_2 X_3 B_1 \overline{B}_4\} \\ &\quad = \Pr\left\{ \begin{pmatrix} P_{2;1} B_1 P_{3;4} B_4 + (\overline{P}_{3;4} + P_{3;4} \overline{B}_4) P_{3;2} P_{2;1} B_1 \\ + (\overline{P}_{2;1} + P_{2;1} \overline{B}_1) P_{2;3} P_{3;4} B_4 \end{pmatrix} B_1 \overline{B}_4 \right\} \\ &\quad = \Pr\{P_{3;2} P_{2;1} B_1 \overline{B}_4\} = p_{3;2} p_{2;1} b_1 \overline{b}_4 \\ &\quad = 0.8 \times 0.8 \times 0.1 \times 0.8 = 0.0512. \quad (22) \end{aligned}$$

It is easy to prove that such calculated probabilities sum up to 1, which is necessary for them to be a JPD over $\{X_2, X_3\}$ or $\{B_1, X_2, X_3, B_4\}$. This property of S-DUCG can be generally proved as follows.

Proof: Let $E_\mu = V_{1j_1} V_{2j_2} \ldots V_{nj_n}$, where μ denotes a set of state subscripts $j_1, j_2, \ldots, j_n$. According to the S-DUCG algorithm presented in [15] and [16], and this paper, through expanding E_μ we can get the expression of E_μ as in the form of sum-of-products composed of only $\{P\text{-}, B\text{-}\}$-type events, regardless of whether DCGs are involved or not. According to the ordinary set theory, through summing up all states of all $\{P\text{-}, B\text{-}\}$-type variables, the sum-of-products becomes the universal set 1. This is because $\{P\text{-}, B\text{-}\}$-type events are just ordinary events satisfying $P + \bar{P} = 1$ and $\sum_{j_i} B_{ij_i} = 1$. In terms of equation, we have

$$\begin{aligned} &\sum_{j_1, j_2, \ldots, j_n} \Pr\{V_{1j_1} V_{2j_2} \ldots V_{nj_n}\} \\ &\quad = \sum_\mu \Pr\{E_\mu\} = \Pr\left\{ \sum_\mu E_\mu \right\} = \Pr\left\{ \sum_\mu \bigcup_\theta \hat{P}_\theta \hat{B}_\theta \right\} \\ &\quad = \Pr\left\{ \bigcup_\theta \sum_{\mu_\theta} \hat{P}_\theta \hat{B}_\theta \right\} = \Pr\left\{ \bigcup_\theta 1 \right\} = \Pr\{1\} = 1. \quad (23) \end{aligned}$$

In which θ indexes the products in the sum-of-products, $\hat{P}_\theta$ denotes all the P-type events in product θ, $\hat{B}_\theta$ denotes all the B-type events in product θ, $\hat{P}_\theta \hat{B}_\theta$ is the product indexed by θ, and μ_θ denotes all the state subscripts of the $\{B\text{-}, P\text{-}\}$-type events involved in product θ. ■

B. M-DUCG Model

To break DCGs, Assumption 4 can also be applied in multivalued cases. For the example shown in Fig. 9, in terms of matrices, $F_{2;3}F_{3;4}B_4$ can be a cause of X_2. But $F_{2;3}F_{3;2}X_2$ cannot be a cause of X_2, because the cause and consequence involve the same variable X_2 at a same time. $F_{2;1}B_1$ is a cause of X_2. It is easy to understand that only $F_{2;3}F_{3;4}B_4$ and $F_{2;1}B_1$ are the causes causing X_2. Each of them is an independent causality chain. According to M-DUCG model presented in [15], they are in a simple summation relationship.

In the case of S-DUCG, a cycle can be broken at a repeated variable during the logic expanding by simply treating the repeated variable as null 0 like what has been done in (8) and (9). But in the case of M-DUCG, the situation is not so simple, because the elimination of a repeated variable X_i will cause the change of the weighting factors $(r_{n;i'}/r_n)$, $i' \neq i$, i.e., the causal relationship intensity $r_{n;i}$ will be eliminated from r_n. For the example of Fig. 9

$$\begin{aligned} X_2 &= (F_{2;1}B_1 + F_{2;3}X_3) \\ &= (F_{2;1}B_1 + F_{2;3}(F_{3;4}B_4 + F_{3;2}X_2)) \\ &= F_{2;1}B_1 + F_{2;3}F_{3;4}B_4 + F_{2;3}F_{3;2}X_2. \end{aligned} \tag{24}$$

The last product $F_{2;3}F_{3;2}X_2$ forms a cycle $X_2 \leftarrow X_3 \leftarrow X_2$ and should be treated as null by applying Assumption 4, which means that only $F_{3;4}B_4$ in $(F_{3;4}B_4 + F_{3;2}X_2)$ just after the second $=$ in (24) remains. As $F_{3;2}X_2$ is eliminated, r_3 is changed from $r_3 = r_{3;4} + r_{3;2}$ to $r_{3;4}^{\{2\}} = r_{3;4}$, where the superscript $\{2\}$ indicates that X_2, as a parent of X_3, is eliminated. To denote this change, the remaining $F_{3;4}B_4$ is denoted as $F_{3;4}^{\{2\}}B_4$. In more detail

$$F_{3;4}^{\{2\}} = \left(r_{3;4}/r_3^{\{2\}}\right)A_{3;4} = (r_{3;4}/r_{3;4})A_{3;4} = A_{3;4}.$$

In contrast

$$F_{3;4} = (r_{3;4}/r_3)A_{3;4} = (r_{3;4}/(r_{3;4} + r_{3;2}))A_{3;4}.$$

Consequently, (24) becomes

$$\begin{aligned} X_2 &= F_{2;1}B_1 + F_{2;3}\left(F_{3;4}^{\{2\}}B_4\right) \\ &= (r_{2;1}/r_2)A_{2;1}B_1 + (r_{2;3}/r_2)A_{2;3}\left(\left(r_{3;4}/r_3^{\{2\}}\right)A_{3;4}B_4\right) \\ &= (r_{2;1}/r_2)A_{2;1}B_1 + (r_{2;3}/r_2)A_{2;3}((r_{3;4}/r_{3;4})A_{3;4}B_4) \\ &= (r_{2;1}/r_2)A_{2;1}B_1 + (r_{2;3}/r_2)A_{2;3}A_{3;4}B_4. \end{aligned} \tag{25}$$

Similarly, we have

$$\begin{aligned} X_3 &= F_{3;4}B_4 + F_{3;2}X_2 \\ &= (r_{3;4}/r_3)A_{3;4}B_4 + (r_{3;2}/r_3)A_{3;2}A_{2;1}B_1. \end{aligned} \tag{26}$$

Equations (25) and (26) express the possible but different weighted causality chains causing X_2 and X_3, respectively, and can be drawn as Figs. 12 and 13, respectively, in which $F_{n;i}$ are replaced by $A_{n;i}$, because the weighting factors equal 1. Note that all weighting factors in (25) and (26) sum up to 1, respectively.

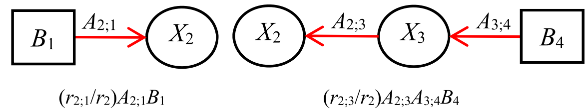

Fig. 12. All possible but different causality chains causing X_2.

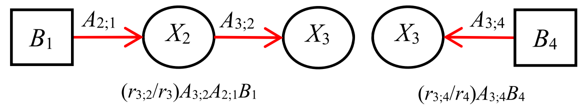

Fig. 13. All possible but different causality chains causing X_3.

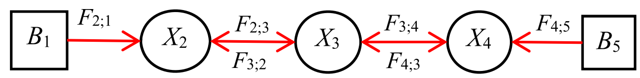

Fig. 14. Illustration for parent elimination in different causality chains.

It is important to note that a pair of brackets ( ) is used to include the expansion of a child variable. In a same ( ), any elimination of parent variable V_i, $V \in \{X, B, G, D\}$, will result in the elimination of $r_{n;i}$ from r_n in and only in this ( ), so that the sum of weighting factors in this ( ) is always 1.

In general, $F_{n;i}^{S_h}$ and $r_n^{S_h}$ denote the case where the parent variables indexed in S_h are eliminated from the parent variables of X_n. In (25), $S_h = \{2\}$. In (26), $S_h = \{3\}$. In some cases, different causality chain may require different elimination of parent variables for a same child variable. For the example shown in Fig. 14, we have

$$\begin{aligned} X_2 &= F_{2;1}B_1 + F_{2;3}\left(F_{3;4}^{\{2\}}\left(F_{4;5}^{\{3\}}B_5\right)\right) \\ &= (r_{2;1}/r_2)A_{2;1}B_1 + (r_{2;3}/r_2)A_{2;3}\left(r_{3;4}/r_3^{\{2\}}\right) \\ &\quad \cdot A_{3;4}\left(r_{4;5}/r_4^{\{3\}}\right)A_{4;5}B_5 \\ &= (r_{2;1}/r_2)A_{2;1}B_1 + (r_{2;3}/r_2)A_{2;3}(r_{3;4}/r_{3;4}) \\ &\quad \cdot A_{3;4}(r_{4;5}/r_{4;5})A_{4;5}B_5 \\ &= (r_{2;1}/r_2)A_{2;1}B_1 + (r_{2;3}/r_2)A_{2;3}A_{3;4}A_{4;5}B_5 \end{aligned} \tag{27}$$

$$\begin{aligned} X_4 &= F_{4;5}B_5 + F_{4;3}\left(F_{3;2}^{\{4\}}\left(F_{2;1}^{\{3\}}B_1\right)\right) \\ &= (r_{4;5}/r_4)A_{4;5}B_5 + (r_{4;3}/r_4)A_{4;3}\left(r_{3;2}/r_3^{\{4\}}\right) \\ &\quad \cdot A_{3;2}\left(r_{2;1}/r_2^{\{3\}}\right)A_{2,1}B_1 \\ &= (r_{4;5}/r_4)A_{4;5}B_5 + (r_{4;3}/r_4)A_{4;3}(r_{3;2}/r_{3;2}) \\ &\quad \cdot A_{3;2}(r_{2;1}/r_{2;1})A_{2,1}B_1 \\ &= (r_{4;5}/r_4)A_{4;5}B_5 + (r_{4;3}/r_4)A_{4;3}A_{3;2}A_{2,1}B_1. \end{aligned} \tag{28}$$

In which $r_2^{\{3\}} = r_{2;1}$, $r_3^{\{2\}} = r_{3;4}$, $r_3^{\{4\}} = r_{3;2}$ and $r_4^{\{3\}} = r_{4;5}$. It is seen that in different causality chain, different parent variable of X_3 is eliminated, e.g., $r_3^{\{2\}}$ and $r_3^{\{4\}}$, where $S_h = \{2\}$ and $S_h = \{4\}$, respectively. Similarly, the products in (27) and (28) represent all possible but different weighted causality chains causing X_2 and X_4, respectively.

In general, any event can be expanded as in the form of weighted sum-of-products composed of $\{B\text{-}, D\text{-}, A\text{-}\}$-type events and/or variables along with weighting factors

$(r_{n;i}/r_n)$ as illustrated above, regardless of whether the expansion involves DCGs or not. Then, we may multiply these events together when needed. For the example of Fig. 9, according to (25) and (26), we have

$$\begin{aligned} X_{2k}X_{3j} &= ((r_{2;1}/r_2)A_{2k;1}B_1 + (r_{2;3}/r_2)A_{2k;3j}A_{3j;4}B_4) \\ &\quad \cdot((r_{3;4}/r_3)A_{3j;4}B_4 + (r_{3;2}/r_3)A_{3j;2k}A_{2k;1}B_1) \\ &= (r_{2;1}/r_2)(r_{3;4}/r_3)A_{2k;1}B_1A_{3j;4}B_4 \\ &\quad + (r_{2;1}/r_2)(r_{3;2}/r_3)A_{3j;2k}A_{2k;1}B_1 \\ &\quad + (r_{2;3}/r_2)(r_{3;4}/r_3)A_{2k;3j}A_{3j;4}B_4 \\ &\quad + (r_{2;3}/r_2)(r_{3;2}/r_3)A_{2k;3j}A_{3j;4}B_4A_{3j;2k}A_{2k;1}B_1. \end{aligned} \tag{29}$$

In which the last product can be absorbed by all the first three products. Here, the last product is called the absorbed product and the first three products are called the absorbing products. According to Assumption 2 presented in [16], the absorbed event should be equally absorbed by all the three absorbing events, i.e., the last product is divided as three equal parts: $(1/3)(r_{2;3}/r_2)(r_{3;2}/r_3)A_{2k;3j}A_{3j;4}B_4A_{3j;2k}A_{2k;1}B_1$. Then, (29) becomes

$$\begin{aligned} &X_{2k}X_{3j} \\ &\quad = (r_{2;1}/r_2)(r_{3;4}/r_3)A_{2k;1}B_1A_{3j;4}B_4 \\ &\qquad + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3)A_{2k;3j}A_{3j;4}B_4A_{3j;2k}A_{2k;1}B_1 \\ &\qquad + (r_{2;1}/r_2)(r_{3;2}/r_3)A_{3j;2k}A_{2k;1}B_1 \\ &\qquad + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3)A_{2k;3j}A_{3j;4}B_4A_{3j;2k}A_{2k;1}B_1 \\ &\qquad + (r_{2;3}/r_2)(r_{3;4}/r_3)A_{2k;3j}A_{3j;4}B_4 \\ &\qquad + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3)A_{2k;3j}A_{3j;4}B_4A_{3j;2k}A_{2k;1}B_1 \\ &\quad = \left((r_{2;1}/r_2)(r_{3;4}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3)\right) A_{2k;1}B_1 \\ &\qquad \cdot A_{3j;4}B_4 \\ &\qquad + \left((r_{2;1}/r_2)(r_{3;2}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3)\right) \\ &\qquad \cdot A_{3j;2k}A_{2k;1}B_1 \\ &\qquad + \left((r_{2;3}/r_2)(r_{3;4}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3)\right) \\ &\qquad \cdot A_{2k;3j}A_{3j;4}B_4. \end{aligned} \tag{30}$$

According to the methodology of M-DUCG presented in [15] and illustrated in (4)–(7), through changing the upper case letters in (30) to lower case letters, we get

$$\begin{aligned} &\Pr\{X_{2k}X_{3j}\} \\ &\quad = ((r_{2;1}/r_2)(r_{3;4}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3)) \\ &\qquad \times a_{2k;1}b_1a_{3j;4}b_4 \\ &\qquad +((r_{2;1}/r_2)(r_{3;2}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))a_{3j;2k} \\ &\qquad \times a_{2k;1}b_1 \\ &\qquad +((r_{2;3}/r_2)(r_{3;4}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))a_{2k;3j} \\ &\qquad \times a_{3j;4}b_4. \end{aligned} \tag{31}$$

The final three weighted products represent all possible but different weighted causality chains causing $X_{2k}X_{3j}$. It is easy to see that the three weighting factors in (30) and (31) sum up to 1 by noting $r_{2;1} + r_{2;3} = r_2$ and $r_{3;2} + r_{3;4} = r_3$.

Suppose the parameters in this example are $r_{2;1} = 0.5$, $r_{2;3} = 1$, $r_{3;2} = r_{3;4} = 1$, $r_2 = r_{2;1} + r_{2;3} = 0.5 + 1 = 1.5$, $r_3 = r_{3;2} + r_{3;4} = 1 + 1 = 2$,

$$b_1 = (0.8 \quad 0.1 \quad 0.1)^T, \quad b_4 = (0.7 \quad 0.1 \quad 0.2)^T$$

$$a_{2;1} = \begin{pmatrix} 1 & 0.2 & 0.2 \\ 0 & 0.8 & 0 \\ 0 & 0 & 0.8 \end{pmatrix}, \quad a_{2;3} = \begin{pmatrix} 1 & 0.1 & 0.1 \\ 0 & 0.9 & 0 \\ 0 & 0 & 0.9 \end{pmatrix}$$

$$a_{3;2} = \begin{pmatrix} 1 & 0.2 & 0.2 \\ 0 & 0.8 & 0 \\ 0 & 0 & 0.8 \end{pmatrix}, \quad a_{3;4} = \begin{pmatrix} 1 & 0.2 & 0.3 \\ 0 & 0.8 & 0 \\ 0 & 0 & 0.7 \end{pmatrix}.$$

We calculate the JPD over $\{X_2, X_3\}$, as shown in Table I.

TABLE I

JPD OVER $\{X_2, X_3\}$ IN FIG. 9

	$X_{2,0}$	$X_{2,1}$	$X_{2,2}$	$\sum$
$X_{3,0}$	0.7620	0.02178	0.02178	0.8056
$X_{3,1}$	0.0222	0.05156	0.00178	0.0755
$X_{3,2}$	0.0389	0.00311	0.07689	0.1189
$\sum$	0.8231	0.07645	0.10045	1

For example

$$\begin{aligned} &\Pr\{X_{2,2}X_{3,2}\} \\ &\quad = ((r_{2;1}/r_2)(r_{3;4}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))a_{2,2;1}b_1 \\ &\qquad \times a_{3,2;4}b_4 \\ &\qquad + ((r_{2;1}/r_2)(r_{3;2}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))a_{3,2;2,2} \\ &\qquad \times a_{2,2;1}b_1 \\ &\qquad + ((r_{2;3}/r_2)(r_{3;4}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))a_{2,2;3,2} \\ &\qquad \times a_{3,2;4}b_4 \\ &\quad = 0.07689. \end{aligned} \tag{32}$$

From Table I, we know

$$\begin{aligned} x_{2,0} &\equiv \Pr\{X_{2,0}\} = 0.8231 \\ x_{2,1} &\equiv \Pr\{X_{2,1}\} = 0.07645 \\ x_{2,2} &\equiv \Pr\{X_{2,2}\} = 0.10045 \\ x_{3,0} &\equiv \Pr\{X_{3,0}\} = 0.8056 \\ x_{3,1} &\equiv \Pr\{X_{3,1}\} = 0.0755 \\ x_{3,2} &\equiv \Pr\{X_{3,2}\} = 0.1189. \end{aligned} \tag{33}$$

Based on (30), it is easy to calculate the JPD over $\{B_1, X_2, X_3, B_4\}$. For example (see Appendix I for details)

$$\Pr\{B_{1,2}X_{2,2}X_{3,2}B_{4,2}\} = 0.01227. \tag{34}$$

As described and illustrated above, it is obvious that for any complete M-DUCG with or without DCGs, the probability of any state combination $\Pr\{V_{1j_1}V_{2j_2}\dots V_{nj_n}\}$ can be exactly calculated. We can prove that they exactly sum up to 1.

Proof: Let $E_\mu = V_{1j_1}V_{2j_2}\dots V_{nj_n}$, where μ denotes a set of state subscripts $j_1, j_2, \dots, j_n$. According to the M-DUCG algorithm presented in [15] and [16], and this paper, through expanding E_μ, we can get the expression of E_μ as in the form of sum-of-products composed of only $\{A\text{-}, B\text{-}\}$-type events associated with r-type weighting factors, regardless of whether

DCGs are involved or not. That is, we always have

$$\sum_{j_1,j_2,\ldots,j_n} \Pr\{V_{1j_1}V_{2j_2}\ldots V_{nj_n}\} = \sum_{\mu}\Pr\{E_\mu\} = \Pr\left\{\sum_{\mu}E_\mu\right\} = \Pr\left\{\sum_{\mu}\sum_{\theta}\hat{r}_\theta\hat{A}_\theta\hat{B}_\theta\right\} = \Pr\left\{\sum_{\theta}\sum_{\mu_\theta}\hat{r}_\theta\hat{A}_\theta\hat{B}_\theta\right\} = \Pr\left\{\sum_{\theta}\hat{r}_\theta\sum_{\mu_\theta}\hat{A}_\theta\hat{B}_\theta\right\}. \quad (35)$$

In which θ indexes the products in the sum-of-products, $\hat{r}_\theta$ denotes all the r-type weighting factors in product θ and is not related to the state subscripts μ_θ, $\hat{A}_\theta$ denotes all the A-type events in product θ, $\hat{B}_\theta$ denotes all the B-type events in product θ, and $\hat{A}_\theta\hat{B}_\theta$ is the event product indexed by θ associated with $\hat{r}_\theta$. Included in product $\hat{A}_\theta\hat{B}_\theta$, there must be causality chain(s) $\breve{A}_{\theta_i}B_{ij_i}$ with B_{ij_i} as the root cause, in which θ_i indexes all A-type events involved in this causality chain. Suppose $\tilde{A}_\theta\tilde{B}_\theta$ is the rest of $\hat{A}_\theta\hat{B}_\theta$. Based on (35), we have

$$\sum_{j_1,j_2,\ldots,j_n} \Pr\{V_{1j_1}V_{2j_2}\ldots V_{nj_n}\} = \Pr\left\{\sum_{\theta}\hat{r}_\theta\sum_{\mu_\theta}\tilde{A}_\theta\tilde{B}_\theta\breve{A}_{\theta_i}B_{ij_i}\right\} = \Pr\left\{\sum_{\theta}\hat{r}_\theta\sum_{\mu_\theta}\tilde{A}_\theta\tilde{B}_\theta\sum_{j_i}B_{ij_i}\sum_{\mu_\theta - j_i}\breve{A}_{\theta_i}\right\} = \Pr\left\{\sum_{\theta}\hat{r}_\theta\sum_{\mu_\theta}\tilde{A}_\theta\tilde{B}_\theta\right\}. \quad (36)$$

Through repeating this process, we have

$$\sum_{j_1,j_2,\ldots,j_n} \Pr\{V_{1j_1}V_{2j_2}\ldots V_{nj_n}\} = \Pr\left\{\sum_{\theta-\mu_\theta}\hat{r}_\theta\right\} = \sum_{\theta-\mu_\theta}\hat{r}_\theta = 1. \quad (37)$$

In which the condition of a complete M-DUCG, i.e., $\sum_{k_n} A_{nk_n;ij_i} = 1$ and $\sum_{j_i} B_{ij_i} = 1$ corresponding to $\sum_{k_n} a_{nk_n;ij_i} = 1$ and $\sum_{j_i} b_{ij_j} = 1$, respectively, is used. ■

The mixture of S-DUCG and M-DUCG (i.e., DUCG) with or without DUCs is too trivial to be discussed in this paper. However, since both S-DUCG and M-DUCG with or without DCGs represent a JPD, it is reasonable to expect that a DUCG also represents a JPD.

In summary, whether a PGM such as BN and DUCG represents a JPD depends on two factors: 1) graphical representation $\mathcal{R}$ and 2) algorithm $\mathcal{A}$ to calculate JPD. In BN, $\mathcal{R}$ must be a DAG graph; $\mathcal{A}$ is the ordinary probability theory. In DUCG, $\mathcal{R}$ is a different graph with or without DCGs; $\mathcal{A}$ is similar but a bit different from the ordinary probability theory. Compared with BN, in the case of S-DUCG, Assumption 4 is added; in the case of M-DUCG, the so-called weighted set theory presented in [16] is added in addition to that Assumption 4 is added.

V. Dealing With Logic Gates Involved in DCGs

In Section IV, we have discussed the general algorithm of DUCG with or without DCGs to calculate JPD in terms of both events and variables/matrices. But the logic gates in DUCG are not involved. In DCG cases, to break logic cycles through applying Assumption 4, we may need to eliminate the input variables or events of a logic gate. Consider the example shown in Fig. 4, in which LGS_6 is specified in Fig. 5. When we expand X_4 in Fig. 4, we encounter the elimination of the input variable of G_6:

$$X_4 = F_{4;1}B_1 + F_{4;2}B_2 + F_{4;8}(F_{8;9}B_9 + F_{8;7}(F_{7;7}D_7 + F_{7;6}G_6\{X_4, X_5\{F_{5;1}B_1 + F_{5;2}B_2 + F_{5;8}X_8\}\})). \quad (38)$$

In which $G_i\{V_h, V_g, \ldots\}$ denotes that G_i is a function of variables V_h, V_g, and so on and $X_n\{\}$ denotes that the content in $\{\}$ is the expansion of X_n.

It is noted that in (38), X_4 and X_8 on the right side are the repeated variables and should be treated as null according to Assumption 4. Through doing this, we have

$$X_4 = F_{4;1}B_1 + F_{4;2}B_2 + F_{4;8}\big(F_{8;9}B_9 + F_{8;7}\big(F_{7;7}D_7 + F_{7;6}G_6 \big\{X_4 = 0, X_5\big\{F^{\{8\}}_{5;1}B_1 + F^{\{8\}}_{5;2}B_2\big\}\big\}\big)\big). \quad (39)$$

Based on Fig. 5, $X_4 = 0$ results in $G_{6,1} = 0$ and $G_{6,2} = 0$, and then G_6 should be eliminated from the parent variables of X_7, because $G_{6,3}$ does not have any output to X_7. Thus, we have

$$X_4 = F_{4;1}B_1 + F_{4;2}B_2 + F_{4;8}\big(F_{8;9}B_9 + F_{8;7}\big(F^{\{6\}}_{7;7}D_7\big)\big) = F_{4;1}B_1 + F_{4;2}B_2 + F_{4;8}F_{8;9}B_9 + F_{4;8}F_{8;7}F^{\{6\}}_{7;7}D_7. \quad (40)$$

In which all the four possible but different weighted causality chains causing X_4 are found.

In general, if V_i is an input variable of G_n and is treated as a null variable due to any reason, LGS_n must be reformed as follows.

1) Get the most simplified LGS_n.
2) Treat the states of V_i in the most simplified LGS_n as null events and calculate the new expressions from the most simplified LGS_n.
3) If the new expression is null, this expression and its corresponding state of G_n (e.g., G_{nk}) are eliminated from the new LGS_n. Meanwhile, $F_{g;nk}$ is also eliminated from $F_{g;n}$.
4) If all new expressions are null, this logic gate is eliminated and the F-type variables from this gate are also eliminated.

The so-called most simplified LGS includes two factors: 1) the state simple and 2) the product simple.

The state simple means that different states of G_n have different functions to its child variables, so that they have to be separate. Of course, the domain expert is the right person to decide how to define different states of G_n. Besides the domain expert, one can easily recognize the same function of different states of G_n: if $a_{gh;nk} = a_{gh;nk'}$ for all g and h where $k \neq k'$, then G_{nk} and $G_{nk'}$ should be combined as a single state, while the corresponding expressions should also be combined.

The state simple is to ensure that all states of G_n are necessary. For example, suppose $V_{3,1} + V_{3,2} = 1$.

Then, $G_{nk} = V_{1,1}V_{2,1}V_{3,1}$ and $G_{nk'} = V_{1,1}V_{2,1}V_{3,2}$ are not state simple given $a_{gh;nk} = a_{gh;nk'}$, because G_{nk} and $G_{nk'}$ can be combined as $G_{nk} = V_{1,1}V_{2,1}$ without any different influence to G_n's child variables. The difference is that if $V_{1,1}V_{2,1}$ is divided as two states: 1) $V_{1,1}V_{2,1}V_{3,1}$ and 2) $V_{1,1}V_{2,1}V_{3,2}$, when V_3 is eliminated (treated as a null variable), $V_{1,1}V_{2,1}V_{3,1} = V_{1,1}V_{2,1}V_{3,2} = 0$. But if not being divided, $V_{1,1}V_{2,1}$ remains.

The product simple means that different products in one expression include and only include the necessary events, so that the null event associated with the null variable will only eliminate the products including this null event. For example, suppose $V_{3,1} + V_{3,2} = 1$ and $G_{nk} = V_{1,1}V_{2,1}V_{3,1} + V_{1,1}V_{2,1}V_{3,2}$. It is obvious that products $V_{1,1}V_{2,1}V_{3,1}$ and $V_{1,1}V_{2,1}V_{3,2}$ are not product simple, because $V_{1,1}V_{2,1}V_{3,1} + V_{1,1}V_{2,1}V_{3,2} = V_{1,1}V_{2,1}$ which means that $V_{3,1}$ and $V_{3,2}$ are not necessary to the two products, respectively. If V_3 is treated as a null variable, $V_{1,1}V_{2,1}V_{3,1} + V_{1,1}V_{2,1}V_{3,2} = 0$, because $V_{3,1}$ and $V_{3,2}$ are null events. But $V_{1,1}V_{2,1} \neq 0$. Therefore, all products in an expression must be product simple. To achieve product simple, the original expression of a state of a logic gate given by domain experts should be complemented twice. This method is the well-known De Morgan's laws. For the example above

$$\begin{aligned}
&G_{nk}\\
&= V_{1,1}V_{2,1}V_{3,1} + V_{1,1}V_{2,1}V_{3,2}\\
&= \overline{\overline{V_{1,1}V_{2,1}V_{3,1} + V_{1,1}V_{2,1}V_{3,2}}}\\
&= \overline{(\bar{V}_{1,1} + \bar{V}_{2,1} + \bar{V}_{3,1})(\bar{V}_{1,1} + \bar{V}_{2,1} + \bar{V}_{3,2})}\\
&= \overline{\begin{pmatrix} \bar{V}_{1,1} + \bar{V}_{1,1}\bar{V}_{2,1} + \bar{V}_{1,1}\bar{V}_{3,2} + \bar{V}_{2,1}\bar{V}_{1,1} + \bar{V}_{2,1} + \bar{V}_{2,1}\bar{V}_{3,2}\\ +\bar{V}_{3,1}\bar{V}_{1,1} + \bar{V}_{3,1}\bar{V}_{2,1} + \bar{V}_{3,1}\bar{V}_{3,2} \end{pmatrix}}\\
&= \overline{\bar{V}_{1,1} + \bar{V}_{2,1} + \bar{V}_{3,1}\bar{V}_{3,2}}\\
&= V_{1,1}V_{2,1}(V_{3,1} + V_{3,2})\\
&= V_{1,1}V_{2,1}. \qquad (41)
\end{aligned}$$

In summary, state simple deals with different states of a logic gate, and product simple deals with the products within a state of a logic gate. Based on the most simplified LGS_n, the eliminated variable is treated as a null 0, and all the products including the event associated with this null variable become 0. This method can be applied repeatedly for the case in which more than one input variables of G_n are eliminated.

VI. Inference Algorithm and Examples

Equation (3) gives the general inference algorithm of DUCG with or without DCGs. Section V has presented the algorithm to calculate a JPD over a set of variables. Based on which, we can calculate $H_{kj}E$ and E, respectively.

A. Container Example

1) S-DUCG Case: As shown in Fig. 11, the parameters have been given in Section IV. Suppose $E = X_2X_3$ and $H_1 = B_1$, where the state subscripts are omitted for simplicity. In (14), we have calculated $\Pr\{E\} = \Pr\{X_2X_3\} = 0.18216$. Based on the logic expression in (14), we have (see Appendix II for details of (42). The details of (43) are similar)

$$\Pr\{B_1X_2X_3\} = 0.06876 \qquad (42)$$

$$\Pr\{B_4X_2X_3\} = 0.13096. \qquad (43)$$

According to (14), (42)–(43), we have

$$\Pr\{B_1|X_2X_3\} = \frac{\Pr\{B_1X_2X_3\}}{\Pr\{X_2X_3\}} = \frac{0.06876}{0.18216} = 0.3775 \qquad (44)$$

$$\Pr\{B_4|X_2X_3\} = \frac{\Pr\{B_4X_2X_3\}}{\Pr\{X_2X_3\}} = \frac{0.14632}{0.18216} = 0.7189. \qquad (45)$$

It is easy to understand that B_4 is $0.7189/0.3775 = 1.9$ times more likely to be true than B_1, given X_2 and X_3, because $b_4 = 0.2$ is twice higher than $b_1 = 0.1$ while $p_{3;4} = 0.7$ and $p_{2;3} = 0.9$ are a bit different from $p_{2;1} = 0.8$ and $p_{3;2} = 0.8$, respectively.

2) M-DUCG Case: As shown in Fig. 9, the parameters have been given in IV. Suppose $E = X_{2,2}X_{3,2}$, $H_{1,2} = B_{1,2}$ and $H_{4,2} = B_{4,2}$. In (32), we have calculated $\Pr\{E\} = \Pr\{X_{2,2}X_{3,2}\} = 0.07689$. Based on the logic expression in (32), we have (see Appendix III for details)

$$\Pr\{B_{1,2}X_{2,2}X_{3,2}\} = 0.02649 \qquad (46)$$

$$\Pr\{B_{4,2}X_{2,2}X_{3,2}\} = 0.06267. \qquad (47)$$

According to (32), (46)–(47), we have

$$\Pr\{B_{1,2}|X_{2,2}X_{3,2}\} = \frac{\Pr\{B_{1,2}X_{2,2}X_{3,2}\}}{\Pr\{X_{2,2}X_{3,2}\}} = \frac{0.02649}{0.07689} = 0.346. \qquad (48)$$

This result is larger than $\Pr\{B_{1,2}|X_{2,2}X_{3,2}\} = 0.138$ in [16], because in [16], $F_{3;2}$ is removed. The larger result is reasonable because in Fig. 9, $B_{1,2}$ along can cause $X_{2,2}X_{3,2}$, while $B_{1,2}$ along cannot cause $X_{2,2}X_{3,2}$, given that $F_{3;2}$ is removed.

Similarly, we have

$$\Pr\{B_{4,2}|X_{2,2}X_{3,2}\} = \frac{\Pr\{B_{4,2}X_{2,2}X_{3,2}\}}{\Pr\{X_{2,2}X_{3,2}\}} = \frac{0.06267}{0.07689} = 0.815. \qquad (49)$$

This result is different from $\Pr\{B_{4,2}|X_{2,2}X_{3,2}\} = 1$ in [16] in which $F_{3;2}$ is removed. This is also reasonable because $B_{4,2}$ in Fig. 9 is no longer necessary to cause $X_{2,2}X_{3,2}$.

It should be noted that if $\{b$-, a-$\}$-type parameters are given as

$$b_1 = \begin{pmatrix} - & - & 0.1 \end{pmatrix}^T, \quad b_4 = \begin{pmatrix} - & - & 0.2 \end{pmatrix}^T$$

$$a_{2;1} = \begin{pmatrix} - & - & - \\ - & - & - \\ 0 & 0 & 0.8 \end{pmatrix}, \quad a_{2;3} = \begin{pmatrix} - & - & - \\ - & - & - \\ 0 & 0 & 0.9 \end{pmatrix}$$

$$a_{3;2} = \begin{pmatrix} 1 & 0.1 & 0.2 \\ 0 & 0.9 & 0 \\ 0 & 0 & 0.8 \end{pmatrix}, \quad a_{3;4} = \begin{pmatrix} - & - & - \\ - & - & - \\ 0 & 0 & 0.7 \end{pmatrix}.$$

The results of (46)–(49) remain unchanged. This is because the property of [15, Th. 1], which ensures that the causality chains in M-DUCG are self-relied. Therefore, in the case of incomplete DUCG with DCGs, we can still get correct inference results.

B. Hurricane Example

In Fig. 6, suppose the parameters are as follows:

$$b_1 = \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix},\quad a_{2;1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},\quad a_{4;2} = \begin{pmatrix} 0.4 & 0.9 \\ 0.6 & 0.1 \end{pmatrix}$$
$$a_{3;4} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\quad a_{5;3} = \begin{pmatrix} 0.3 & 0.8 \\ 0.7 & 0.2 \end{pmatrix},\quad a_{3;1} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$
$$a_{2;5} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$
$$r_{2;1} = r_{3;1} = r_{4;2} = r_{5;3} = r_{2;5} = r_{3;4} = 1.$$

In what follows, two methods will be applied to perform the inference: 1) to treat the case as the weighted sum of DAGs, as shown in Fig. 7 and 2) to treat the case as M-DUCG with DCGs, as shown in Fig. 6. We will see that the results are exactly same. The later shows a general way of DUCG to deal with DCGs that are actually the sum of a set of DAGs.

1) Method of the Weighted Sum of DAGs: Based on Fig. 7, all $(r_{n;i}/r_n) = 1$ due to that one child variable has only one parent variable. According to (5), we have

$$\begin{aligned} X_{4k} &= F_{4k;2}F_{2;1,1}B_{1,1} + F_{4k;2}F_{2;5}F_{5;3}F_{3;1,2}B_{1,2} \\ &= A_{4k;2}A_{2;1,1}B_{1,1} + A_{4k;2}A_{2;5}A_{5;3}A_{3;1,2}B_{1,2} \end{aligned} \quad (50)$$
$$\begin{aligned} X_{5j} &= F_{5j;3}F_{3;4}F_{4;2}F_{2;1,1}B_{1,1} + F_{5j;3}F_{3;1,2}B_{1,2} \\ &= A_{5j;3}A_{3;4}A_{4;2}A_{2;1,1}B_{1,1} + A_{5j;3}A_{3;1,2}B_{1,2}. \end{aligned} \quad (51)$$

And then

$$\begin{aligned} X_{4k}X_{5j} &= (A_{4k;2}A_{2;1,1}B_{1,1} + A_{4k;2}A_{2;5}A_{5;3}A_{3;1,2}B_{1,2}) \\ &\quad \cdot(A_{5j;3}A_{3;4}A_{4;2}A_{2;1,1}B_{1,1} + A_{5j;3}A_{3;1,2}B_{1,2}) \\ &= A_{4k;2}A_{2;1,1}B_{1,1}A_{5j;3}A_{3;4}A_{4;2}A_{2;1,1}B_{1,1} \\ &\quad + A_{4k;2}A_{2;1,1}B_{1,1}A_{5j;3}A_{3;1,2}B_{1,2} \\ &\quad + A_{4k;2}A_{2;5}A_{5;3}A_{3;1,2}B_{1,2}A_{5j;3}A_{3;4}A_{4;2}A_{2;1,1}B_{1,1} \\ &\quad + A_{4k;2}A_{2;5}A_{5;3}A_{3;1,2}B_{1,2}A_{5j;3}A_{3;1,2}B_{1,2} \\ &= A_{5j;3}A_{3;4}A_{4k;2}A_{2;1,1}B_{1,1} + 0 + 0 \\ &\quad + A_{4k;2}A_{2;5j}A_{5j;3}A_{3;1,2}B_{1,2} \\ &= A_{5j;3}A_{3;4k}A_{4k;2}A_{2;1,1}B_{1,1} \\ &\quad + A_{4k;2}A_{2;5j}A_{5j;3}A_{3;1,2}B_{1,2}. \end{aligned} \quad (52)$$

In this paper, operator $\cdot$ also denotes AND.

In the expansion of (52), Rules/Corollaries 11–14 presented in [16] are used, e.g., $X_5X_{5j} = X_{5j}$ results in $A_{2;5}A_{5j;3} = A_{2;5j}A_{5j;3}$, $B_{1,1}B_{1,1} = B_{1,1}$, $B_{1,1}B_{1,2} = 0$, and $(A_{2;1,1}B_{1,1})^2 = A_{2;1,1}B_{1,1}$, and so on.

In (50)–(52), each event and/or variable product is a causality chain and all products represent all possible but different causality chains. These causality chains are composed of only independent $\{A\text{-}, B\text{-}\}$-type events and/or variables and are in a simple weighted summation relationship. For example, (50) reveals that X_{4k} can be caused by either $A_{4k;2}A_{2;1,1}B_{1,1}$ or $A_{4k;2}A_{2;5}A_{5;3}A_{3;1,2}B_{1,2}$. Since there is no repeated event and variable in anyone of causality chains, and the $\{A\text{-}, B\text{-}\}$-type event and variables are independent of each other, according to M-DUCG model presented in [15] and [16], we have

$$\Pr\{X_{4k}\} = a_{4k;2}a_{2;1,1}b_{1,1} + a_{4k;2}a_{2;5}a_{5;3}a_{3;1,2}b_{1,2} \quad (53)$$
$$\Pr\{X_{5j}\} = a_{5j;3}a_{3;4}a_{4;2}a_{2;1,1}b_{1,1} + a_{5j;3}a_{3;1,2}b_{1,2} \quad (54)$$
$$\begin{aligned} \Pr\{X_{4k}X_{5j}\} &= a_{5j;3}a_{3;4k}a_{4k;2}a_{2;1,1}b_{1,1} \\ &\quad + a_{4k;2}a_{2;5j}a_{5j;3}a_{3;1,2}b_{1,2}. \end{aligned} \quad (55)$$

Then, we get

$$\begin{aligned} &\Pr\{X_{4k}|X_{5j}\} \\ &= \frac{\Pr\{X_{4k}X_{5j}\}}{\Pr\{X_{5j}\}} \\ &= \frac{a_{5j;3}a_{3;4k}a_{4k;2}a_{2;1,1}b_{1,1} + a_{4k;2}a_{2;5j}a_{5j;3}a_{3;1,2}b_{1,2}}{a_{5j;3}a_{3;4}a_{4;2}a_{2;1,1}b_{1,1} + a_{5j;3}a_{3;1,2}b_{1,2}}. \end{aligned} \quad (56)$$

According to (50), we have

$$\begin{aligned} \Pr\{X_{4,1}\} &= a_{4,1;2}a_{2;1,1}b_{1,1} + a_{4,1;2}a_{2;5}a_{5;3}a_{3;1,2}b_{1,2} \\ &= \begin{pmatrix} 0.4 & 0.9 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} 0.4 \\ &\quad + \begin{pmatrix} 0.4 & 0.9 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0.3 & 0.8 \\ 0.7 & 0.2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} 0.6 \\ &= 0.61 \\ \Pr\{X_{4,2}\} &= a_{4,2;2}a_{2;1,1}b_{1,1} + a_{4,2;2}a_{2;5}a_{5;3}a_{3;1,2}b_{1,2} = 0.39. \end{aligned}$$

According to (51), we have

$$\Pr\{X_{5,1}\} = a_{5,1;3}a_{3;4}a_{4;2}a_{2;1,1}b_{1,1} + a_{5,1;3}a_{3;1,2}b_{1,2} = 0.42$$
$$\Pr\{X_{5,2}\} = a_{5,2;3}a_{3;4}a_{4;2}a_{2;1,1}b_{1,1} + a_{5,2;3}a_{3;1,2}b_{1,2} = 0.58.$$

From the above calculations, we know that the prior probabilities of the mild and severe damages of Alphatown are 61% and 39%, respectively, while the prior probabilities of the mild and severe damages of Betaville are 42% and 58%, respectively.

According to (56), we have

$$\begin{aligned} &\Pr\{X_{4,1}|X_{5,1}\} \\ &= \frac{a_{5,1;3}a_{3;4,1}a_{4,1;2}a_{2;1,1}b_{1,1} + a_{4,1;2}a_{2;5,1}a_{5,1;3}a_{3;1,2}b_{1,2}}{a_{5,1;3}a_{3;4}a_{4;2}a_{2;1,1}b_{1,1} + a_{5,1;3}a_{3;1,2}b_{1,2}} \\ &= 0.2867 \\ &\quad \Pr\{X_{4,2}|X_{5,1}\} = 0.7143 \\ &\quad \Pr\{X_{4,1}|X_{5,2}\} = 0.8448 \\ &\quad \Pr\{X_{4,2}|X_{5,2}\} = 0.1552. \end{aligned}$$

That is, conditional on the evidence that the damage of Betaville is mild, the probability of the severe damage to Alphatown increases from 39% to 71.43%; conditional on the evidence that the damage to Betaville is severe, the probability of the severe damage of Alphatown is reduced from 39% to 15.52%.

2) M-DUCG With DCGs: Based on Fig. 6, according to (4) and (5), we have

$$\begin{aligned} X_{4k} &= F_{4k;2}X_2 \\ &= F_{4k;2}(Z_{2;1}F_{2;1}B_1 \\ &\qquad + Z_{2;5}F_{2;5}(F_{5;3}(Z_{3;1}F_{3;1}B_1 + Z_{3;4}F_{3;4}X_4))) \\ X_{5j} &= F_{5j;3}X_3 \\ &= F_{5j;3}(Z_{3;1}F_{3;1}B_1 \\ &\qquad + Z_{3;4}F_{3;4}(F_{4;2}(Z_{2;1}F_{2;1}B_1 + Z_{2;5}F_{2;5}X_5))). \end{aligned}$$

It is noted that X_4 and X_5 on the right side of the above equations are repeated variables and should be treated as null according to Assumption 4. Thus, we have

$$X_{4k} = F_{4k;2}\big(Z_{2;1}F_{2;1}B_1 + Z_{2;5}F_{2;5}F_{5;3}Z_{3;1}F_{3;1}^{\{4\}}B_1\big)$$
$$X_{5j} = F_{5j;3}\big(Z_{3;1}F_{3;1}B_1 + Z_{3;4}F_{3;4}F_{4;2}Z_{2;1}F_{2;1}^{\{5\}}B_1\big).$$

It is noted that the conditional events $Z_{n;i}$ still exist. As described in [15, Sec. 3.2], $(Z_{n;i}+\bar{Z}_{n;i})$ should be multiplied with above expressions. As given in Section II

$$\begin{aligned}(\overline{Z}_{2;5} + Z_{2;5}) &= (\overline{Z}_{3;1} + Z_{3;1})\\ &= (Z_{2;1} + \overline{Z}_{2;1})\\ &= (Z_{3;4} + \overline{Z}_{3;4})\\ &= (B_{1,1} + B_{1,2}).\end{aligned}$$

In addition, it is easy to know $(B_{1,1} + B_{1,2})^y = (B_{1,1} + B_{1,2})$ for $y \geq 1$. Thus, noting that all $r_{n;i} = 1$, we have

$$\begin{aligned}X_{4k} &= (Z_{2;1} + \bar{Z}_{2;1})(\bar{Z}_{2;5} + Z_{2;5})(\bar{Z}_{3;1} + Z_{3;1})F_{4k;2}\\ &\quad\cdot\big(Z_{2;1}F_{2;1}B_1 + Z_{2;5}F_{2;5}F_{5;3}Z_{3;1}F_{3;1}^{\{4\}}B_1\big)\\ &= (B_{1,1} + B_{1,2})F_{4k;2}\big(B_{1,1}F_{2;1}B_1\\ &\qquad + B_{1,2}F_{2;5}F_{5;3}B_{1,2}F_{3;1}^{\{4\}}B_1\big)\\ &= F_{4k;2}\big(B_{1,1}\big(B_{1,1}F_{2;1}B_1 + B_{1,2}F_{2;5}F_{5;3}F_{3;1}^{\{4\}}B_1\big)\\ &\quad + B_{1,2}\big(B_{1,1}F_{2;1}B_1 + B_{1,2}F_{2;5}F_{5;3}F_{3;1}^{\{4\}}B_1\big)\big)\\ &= F_{4k;2}\big(\big(F_{2;1,1}^{\{5\}}B_{1,1}\big) + \big(F_{2;5}^{\{1\}}F_{5;3}F_{3;1,2}^{\{4\}}B_{1,2}\big)\big)\\ &= A_{4k;2}(A_{2;1,1}B_{1,1} + A_{2;5}A_{5;3}A_{3;1,2}B_{1,2}) \qquad (57)\\ X_{5j} &= A_{5j;3}(A_{3;4}A_{4;2}A_{2;1,1}B_{1,1} + A_{3;1,2}B_{1,2}). \qquad (58)\end{aligned}$$

It is seen that the two equations are exactly the same as (50) and (51), respectively, which result in (52) and (56). Obviously, the final numerical results are the same, which means that the M-DUCG algorithm presented in this paper to deal with DCGs gets the exactly same inference results as using ordinary method to deal with sum of DAGs.

The benefit of applying M-DUCG algorithm is that we can deal with DUCG with DCGs uniformly without recognizing whether this DUCG can be decomposed as a set of weighted DAGs or not. Actually, the M-DUCG methodology can be uniformly applied to all cases, including DAG, DCG, conditional causalities and incomplete DUCG cases.

C. Mixture of S-DUCG and M-DUCG

The mixture of S-DUCG and M-DUCG is DCUG, and is shown in Fig. 3. In other words, DUCG may include both S-DUCG and M-DUCG simultaneously. Sometimes, S-DUCG or M-DUCG only is also called DUCG in the case without confusion. Actually, the method of DUCG dealing with DCGs is just the application of the methods of S-DUCG and M-DUCG illustrated in Sections VI-A1 and VI-A2 in this section, respectively. To shorten the length, the details are omitted.

VII. Conclusion

In this paper, the DUCG inference algorithms presented in [15]–[19] are extended from DAGs to DCGs. The method to deal with DCGs is presented. The mathematical definition of DUCG with or without DCGs is discussed in detail.

It is pointed out that a DUCG with DCGs may be the combination of a set of exclusive and exhaustive DAGs. The DUCG with this type of DCGs is obviously the representation of JPD over a set of random variables.

In general, a complete DUCG with or without DCGs can represent a JPD over a set of random variables $\{V_1, V_2, \ldots, V_n\}$, because any $\Pr\{V_{1j_1}V_{2j_2}\ldots V_{nj_n}\}$ can be exactly calculated with the method presented in this paper, while satisfying $\sum_{j_1,j_2,\ldots,j_n}\Pr\{V_{1j_1}V_{2j_2}\ldots V_{nj_n}\} = 1$.

In the case of incomplete DUCG with or without DCGs, the DUCG represents a part of JPD, because only some of $\Pr\{V_{1j_1}V_{2j_2}\ldots V_{nj_n}\}$ can be exactly calculated.

Assumption 4 presented in this paper is the key to break DCGs. It is important to note that the application of Assumption 4 in M-DUCG is different from that in S-DUCG, because in S-DUCG, $r_{n;i}$ associated with the eliminated parent variable V_i should be eliminated from r_n of child variable X_n. Moreover, when the discard of causality chain is just at the input of logic gates, the reconstruction of LGS_i and $F_{n;i}$ must be made, in which the state simple and product simple should be achieved.

Examples are provided to illustrate the significance and importance of DCGs, the mathematical meaning of DCGs, and the method to deal with DCGs.

Appendix I
Details of (34)

$$\begin{aligned}&\Pr\{B_{1,2}X_{2,2}X_{3,2}B_{4,2}\}\\ &= \Pr\left\{\begin{array}{l}((r_{2;1}/r_2)(r_{3;4}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))\\ \quad A_{2,2;1,2}B_{1,2}A_{3j;4,2}B_{4,2}\\ +((r_{2;1}/r_2)(r_{3;2}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))\\ \quad A_{3,2;2,2}A_{2,2;1,2}B_{1,2}B_{4,2}\\ +((r_{2;3}/r_2)(r_{3;4}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))\\ \quad A_{2,2;3,2}A_{3,2;4,2}B_{4,2}B_{1,2}\end{array}\right\}\\ &= ((r_{2;1}/r_2)(r_{3;4}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))\\ &\quad\times a_{2,2;1,2}b_{1,2}a_{3,2;4,2}b_{4,2}\\ &\quad+((r_{2;1}/r_2)(r_{3;2}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))\\ &\quad\times a_{3,2;2,2}a_{2,2;1,2}b_{1,2}b_{4,2}\\ &\quad+((r_{2;3}/r_2)(r_{3;4}/r_3) + (1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))\\ &\quad\times a_{2,2;3,2}a_{3,2;4,2}b_{4,2}b_{1,2}\\ &= ((0.5/1.5)(1/2) + (1/3)(1/1.5)(1/2))\\ &\quad\times 0.8\times 0.1\times 0.7\times 0.2\\ &\quad+((0.5/1.5)(1/2) + (1/3)(1/1.5)(1/2))\\ &\quad 0.8\times 0.8\times 0.1\times 0.2\\ &\quad+((1/1.5)(1/2) + (1/3)(1/1.5)(1/2))\\ &\quad\times 0.9\times 0.7\times 0.2\times 0.1 = 0.01227.\end{aligned}$$

Appendix II
Details of (42)

$$\begin{aligned}
&\Pr\{B_1X_2X_3\}\\
&= \Pr\left\{B_1\begin{pmatrix}P_{2;1}B_1P_{3;4}B_4+(\overline{P}_{3;4}+P_{3;4}\overline{B}_4)P_{3;2}P_{2;1}B_1\\ +(\overline{P}_{2;1}+P_{2;1}\overline{B}_1)P_{2;3}P_{3;4}B_4\end{pmatrix}\right\}\\
&= \Pr\left\{\begin{array}{l}P_{2;1}B_1P_{3;4}B_4+(\overline{P}_{3;4}+P_{3;4}\overline{B}_4)P_{3;2}P_{2;1}B_1\\ +\overline{P}_{2;1}B_1P_{2;3}P_{3;4}B_4\end{array}\right\}\\
&= p_{2;1}b_1p_{3;4}b_4+(\overline{p}_{3;4}+p_{3;4}\overline{b}_4)p_{3;2}p_{2;1}b_1\\
&\quad +\overline{p}_{2;1}b_1p_{2;3}p_{3;4}b_4\\
&= 0.8\times 0.1\times 0.7\times 0.2+(0.3+0.7\times 0.8)0.8\times 0.8\times 0.1\\
&\quad +0.2\times 0.1\times 0.9\times 0.7\times 0.2 = 0.06876.
\end{aligned}$$

Appendix III
Details of (46)

$$\begin{aligned}
&\Pr\{B_{1,2}X_{2,2}X_{3,2}\}\\
&= \Pr\left\{\begin{array}{l}((r_{2;1}/r_2)(r_{3;4}/r_3)+(1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))\\ \quad A_{2,2;1,2}B_{1,2}A_{3,2;4}B_4\\ +((r_{2;1}/r_2)(r_{3;2}/r_3)+(1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))\\ \quad A_{3,2;2,2}A_{2,2;1,2}B_{1,2}\\ +((r_{2;3}/r_2)(r_{3;4}/r_3)+(1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))\\ \quad A_{2,2;3,2}A_{3,2;4}B_4B_{1,2}\end{array}\right\}\\
&= ((r_{2;1}/r_2)(r_{3;4}/r_3)+(1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))\\
&\quad \times a_{2,2;1,2}b_{1,2}a_{3,2;4}b_4\\
&\quad +((r_{2;1}/r_2)(r_{3;2}/r_3)+(1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))\\
&\quad \times a_{3,2;2,2}a_{2,2;1,2}b_{1,2}\\
&\quad +((r_{2;3}/r_2)(r_{3;4}/r_3)+(1/3)(r_{2;3}/r_2)(r_{3;2}/r_3))\\
&\quad \times a_{2,2;3,2}a_{3,2;4}b_4b_{1,2}\\
&= ((0.5/1.5)(1/2)+(1/3)(1/1.5)(1/2))\\
&\quad \times 0.8\times 0.1\begin{pmatrix}0 & 0 & 0.7\end{pmatrix}\begin{pmatrix}0.7\\0.1\\0.2\end{pmatrix}\\
&\quad +((0.5/1.5)(1/2)+(1/3)(1/1.5)(1/2))0.8\times 0.8\times 0.1\\
&\quad +((1/1.5)(1/2)+(1/3)(1/1.5)(1/2))\\
&\quad \times 0.9\begin{pmatrix}0 & 0 & 0.7\end{pmatrix}\begin{pmatrix}0.7\\0.1\\0.2\end{pmatrix}0.1 = 0.02649.
\end{aligned}$$

Acknowledgment

The authors would like to thank the reviewers and the editor for their comments and suggestions.

References

[1] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, 1st ed. Cambridge, MA, USA: MIT Press, Nov. 2009.

[2] P. Larrañaga and S. Moral, "Probabilistic graphical models in artificial intelligence," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 1511–1528, Mar. 2011.

[3] L. Hongmei, H. Wenning, G. Wenyan, and C. Gang, "Survey of probabilistic graphical models," in *Proc. 10th Web Inf. Syst. Appl. Conf.*, Nanjing, China, Nov. 2013, pp. 275–280.

[4] P. Larrañaga, H. Karshenas, C. Bielza, and R. Santana, "A review on probabilistic graphical models in evolutionary computation," *J. Heuristics*, vol. 18, no. 5, pp. 795–819, Oct. 2012.

[5] A. Kimmig, L. Mihalkova, and L. Getoor, "Lifted graphical models: A survey," *Mach. Learn.*, May 2014, doi: 10.1007/s10994-014-5443-2.

[6] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA, USA: Morgan Kaufmann, 1988.

[7] F. V. Jensen, *An Introduction to Bayesian Networks*. New York, NY, USA: Springer-Verlag, 1996.

[8] D. Heckerman and J. S. Breese, "Causal independence for probability assessment and inference using Bayesian networks," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 26, no. 6, pp. 826–831, Nov. 1996.

[9] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems*. New York, NY, USA: Springer-Verlag, 1999.

[10] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2003.

[11] J. Pearl and S. Russell, "Bayesian networks," in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed., 2nd ed. Cambridge, MA, USA: MIT Press, 2003.

[12] I. Rish *et al.*, "Adaptive diagnosis in distributed systems," *IEEE Trans. Neural Netw.*, vol. 16, no. 5, pp. 1088–1109, Sep. 2005.

[13] P. A. D. Castro and F. J. Von Zuben, "Learning ensembles of neural networks by means of a Bayesian artificial immune system," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 304–316, Feb. 2011.

[14] M. M. Adankon, M. Cheriet, and A. Biem, "Semisupervised learning using Bayesian interpretation: Application to LS-SVM," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 513–524, Apr. 2011.

[15] Q. Zhang, "Dynamic uncertain causality graph for knowledge representation and reasoning: Discrete DAG cases," *J. Comput. Sci. Technol.*, vol. 27, no. 1, pp. 1–23, Jan. 2012.

[16] Q. Zhang, C. Dong, Y. Cui, and Z. Yang, "Dynamic uncertain causality graph for knowledge representation and probabilistic reasoning: Statistics base, matrix, and application," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 4, pp. 645–663, Apr. 2014.

[17] Q. Zhang, "Dynamic uncertain causality graph for knowledge representation and reasoning: Continuous variable, uncertain evidence and failure forecast," *IEEE Trans. Syst., Man, Cybern., Syst.*, doi: 10.1109/TSMC.2015.2392711, 2015.

[18] Q. Zhang and S. Geng, "Dynamic uncertain causality graph applied to dynamic fault diagnoses of large and complex systems," *IEEE Trans. Rel.*, doi: 10.1109/TR.2015.2416332, 2015.

[19] C. Dong, Y. Wang, Q. Zhang, and N. Wang, "The methodology of dynamic uncertain causality graph for intelligent diagnosis of vertigo," *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 162–174, Jan. 2014.

[20] C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller, "Context-specific independence in Bayesian networks," in *Proc. 12th Int. Conf. UAI*, Aug. 1996, pp. 115–123.

[21] Q. Zhang, "Probabilistic reasoning based on dynamic causality trees/diagrams," *Rel. Eng. Syst. Safety*, vol. 46, no. 3, pp. 209–220, 1994.

[22] K. P. Murphy, "Dynamic Bayesian networks: Representation, inference and learning," Ph.D. dissertation, Comput. Sci. Division, UC Berkeley, Berkeley, CA, USA, Jul. 2002.

[23] R. V. Borges, A. d'Avila Garcez, and L. C. Lamb, "Learning and representing temporal knowledge in recurrent networks," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2409–2421, Dec. 2011.

[24] D. T. Mirikitani and N. Nikolaev, "Recursive Bayesian recurrent neural networks for time-series modeling," *IEEE Trans. Neural Netw.*, vol. 21, no. 2, pp. 262–274, Feb. 2010.

[25] P. Spirtes, "Directed cyclic graphical representations of feedback models," in *Proc. 11th Conf. UAI*, pp. 491–498, Aug. 1995.

[26] J. Pearl and R. Dechter, "Identifying independencies in causal graphs with feedback," in *Proc. 12th Int. Conf. UAI*, pp. 420–426, Aug. 1996.

[27] T. Richardson, "A discovery algorithm for directed cyclic graphs," in *Proc. 12th Int. Conf. UAI*, pp. 454–461, Aug. 1996.

[28] B. J. Frey and D. J. C. MacKay, "A revolution: belief propagation in graphs with cycles," in *Advances in Neural Information Processing Systems 10*. MIT Press: Cambridge, MA, 1998.

[29] B. Milch, B. Marthi, D. Sontag, S. Russell, D. L. Ong, and A. Kolobov, "Approximate inference for infinite contingent Bayesian networks," in *Proc. 10th Int. Workshop Artif. Intell. Statist.*, Barbados, Jan. 2005.

[30] Y. Hu, X. Tu, X. Cui, and Q. Cheng, "A inferential method of uncertain knowledge based on Bayes-network," (in Chinese) *Comput. Integr. Manuf. Syst.*, vol. 7, no. 12, pp. 62–68, 2001.

[31] D. Heckerman, D. M. Chickering, C. Meek, R. Rounthwaite, and C. Kadie, "Dependency networks for inference, collaborative filtering, and data visualization," *J. Mach. Learn. Res.*, vol. 1, pp. 49–75, Oct. 2000.

[32] J. Deng, B. Choi, J. Xu, H. Hu, and S. S. Bhowmick, "Incremental maintenance of the minimum bisimulation of cyclic graphs," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 11, pp. 2536–2550, Nov. 2013.

[33] C. Achayuthakan, C. J. Dent, J. W. Bialek, and W. Ongsakul, "Electricity tracing in systems with and without circulating flows: Physical insights and mathematical proofs," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 1078–1087, May 2010.

[34] I. A. Krepkiy, "The sandpile groups of chain-cyclic graphs," *J. Math. Sci.*, vol. 200, no. 6, pp. 698–709, Aug. 2014.

[35] J. Williams and R. Lau, "Approximate evaluation of marginal association probabilities with belief propagation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 4, pp. 2942–2959, Oct. 2014.

[36] K. P. Murphy, "The Bayes net toolbox for MATLAB," *Comput. Sci. Statist.*, vol. 33, no. 2, pp. 331–351, Oct. 2001.

Qin Zhang was born in Chongqing, China, in 1956. He received the B.S., M.S., and Ph.D. degrees from Tsinghua University, Beijing, China, in 1982, 1984, and 1989, respectively, all in nuclear engineering.

He was a Visiting Scholar with the University of Tennessee, Knoxville, TN, USA, and the University of California at Los Angeles, CA, USA, from 1987 to 1989, where he was involved in research on system reliability engineering and fault diagnoses. He was a Post-Doctoral Fellow with the School of Economics and Management, Tsinghua University, from 1990 to 1991, where he was an Associate Professor with the Institute of Nuclear Energy Technology (INET), from 1991 to 1993. He was the Department Manager of China Technology Innovation Corporation, and the General Manager of United Corporation for Technology Innovation, Xiamen, China, from 1993 to 1997. He was the Head of Chongqing Science and Technology Commission from 1997 to 2003, and the Deputy Commissioner of the State Intellectual Property Office of China from 2003 to 2009. Since 2009, he has been an Executive Secretary of the China Association for Science and Technology. He is currently a Professor with INET and the School of Computer Science and Engineering, Beihang University, Beijing.