Supplementary File for "A New Knowledge Mining and Root Cause Analysis Methodology for Multivariate Time Series"

THIS part provides the supplementary material to support the research in the paper.

I. PRELIMINARIES

Notations adopted in this paper are summarized in Table S-I.

II. MODELING WITH SBTPN

The SBTPN defined in Section IV of the paper has two key characteristics: the integration of Petri net (PN) transitions into the probability distribution of Bayesian networks (BNs), and the introduction of synergy effects into BNs.

The necessity of the first key characteristic of SBTPN, namely the integration of PN transitions into the probability distribution of BNs, is discussed first. Given that BNs can perform interpretable probabilistic reasoning under uncertainty, they are employed as the fundamental architecture for the proposed RCA method. However, BNs do not explicitly emphasize "AND/OR" relations, which increases the complexity of constructing BNs. Consider the following two chemical reactions: 1) Reaction a: $A \rightarrow D$, where A and D are the reactant and product of a, respectively; and 2) Reaction b: $B+C \rightarrow D$, where B and C are the reactants of b, and D is its product. The causal relations among A, B, C, and D can be expressed by the following causal rules: R_1: IF "Reactant A exists" THEN "Product D appears" and R_2: IF "Reactant B exists" AND "Reactant C exists" THEN "Product D appears". Fig. S1(a) illustrates the idea of modeling the above reactions using a BN. Note that, throughout this paper, $\hat{x}$ or $\check{x}$ denotes the state of a variable, place or transition corresponding to x. For a variable or place, its state is "true" or "false". As to a transition, its state is "enabled" or "disabled". From Fig. S1(a), the BN fails to directly depict the "AND" relation between variables B and C. This leads to high complexity in constructing the conditional probability table (CPT) M_1.

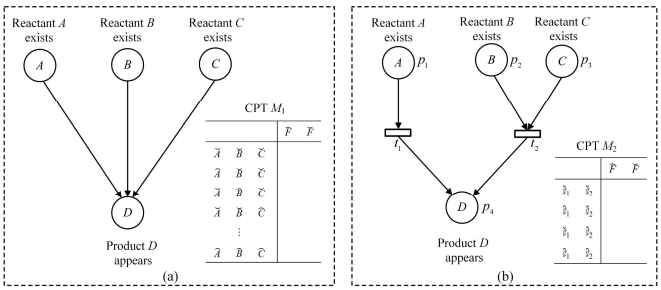

Fig. S1. Difference between a BN and the modeling paradigm to be proposed. (a) BN model instance and (b) Modeling paradigm to be proposed.

As shown in Fig. S1(b), the complexity of constructing a CPT can be reduced by integrating the causal variables within the same "AND" structure into a single one. In Fig. S1(b), a PN place (represented by a circle) is used to depict a causal or effect variable. A PN transition (represented by a rectangle) is used to model a causal rule. The causal variable is expressed as the input place of a transition. Causal variables that point to the same transition have an "AND" relation. The effect variable is expressed as the output place of a transition. A solid arc represents the flow relation from a causal variable to a transition (or from a transition to an effect variable). As shown in Fig. S1(b), transitions t_1 and t_2 are used to model rule R_1: IF "Reactant A exists" THEN "Product D appears" and rule R_2: IF "Reactant B exists" AND "Reactant C exists" THEN "Product D appears", respectively. The probability that proposition "Product D appears" holds depends on the probability that either R_1 or R_2 is activated. Therefore, CPT M_2 can be determined by identifying whether R_1 or R_2 is activated. Clearly, the complexity of constructing M_2 is lower than that of constructing M_1.

From the perspective of knowledge representation, the "AND" relation can be represented by integrating all cause variables within the same "AND" structure into an additional BN variable. However, from the perspective of knowledge mining, causal discovery methods aim to discover the causal relations among observed variables. Because the additional variable does not exist in the original data, it is impossible to directly discover the causal relation between it and other variables in the BN from the data. Modeling "AND" relations with PN transitions does not suffer from this limitation, because they are only used to constrain the causal relations among observed variables. By associating all cause variables in an "AND" structure with a PN transition, and incorporating the transition as an additional node into BN, the complexity of constructing the CPT for BN can be reduced. This approach also enables BN to depict "AND/OR" relations and synergy effects during RCA.

The importance of the second key characteristic of SBTPN, namely the introduction of synergy effects into BNs, is discussed next. Causal models, including BNs and PNs, fail to depict synergy effects. They treat a synergy factor as either a causal or non-causal variable of the effect one, which can affect the reliability of causality-driven RCA. Consider the chemical reaction a: $A \overset{B}{\rightarrow} C$, where A is the reactant, and B and C are the inhibitor (a substance that slows down a chemical reaction or prevents it from occurring) and product, respectively. Essentially, A is the cause of the production of C, but B is not its cause. However, B can affect the production of C. Specifically, reaction a is harder to carry out under the action of inhibitor B, i.e., B has an inhibitory effect on a. In a rule-based system, such phenomena can be expressed as: A non-causal variable can affect the probability that a reasoning

TABLE S-I
NOTATIONS

Symbol	Description
X	Set of variables.
P	Set of places.
T	Set of transitions.
${}^{\circ}t$	Set of places that have synergy effects on transition t.
p°	Set of transitions affected by the synergy effects of place p.
$\beta(p)$	Variable associated with place p.
$\Pr(\cdot)$	Probability function.
${}^{\bullet}v / v^{\bullet}$	Pre-set/post-set of a node $v \in P \cup T$.
$\hat{\boldsymbol{p}} / \check{\boldsymbol{p}}$	State of place p, indicating that $\beta(p)$ is present/absent.
$\hat{\boldsymbol{t}} / \check{\boldsymbol{t}}$	State of transition t, indicating that t is enabled/disabled.
a_{ij}	Attention score that is used to determine whether x_i is the cause of x_j.
w_{ij}	Kernel weight that is used to calculate the time delay of the impact of x_i on x_j.
e_{ij}	Time delay of the impact of x_i on x_j.
ζ	Total number of time steps of a time series.
μ_{ij}^{τ}	Impact of x_i on x_j at time step τ.
$\Phi(x_i, x_j, x_k)$	Strength of the "AND" relation between x_i and x_j pointing to x_k.
$\mathbb{N}_h$	$\{1, 2, ..., h\}$.
$\Psi({}^{\bullet}t_k; t_k^{\bullet}; p_x) / \Upsilon({}^{\bullet}t_k; t_k^{\bullet}; p_x)$	Strength of the catalytic/inhibitory effect of p_x on t_k.
η_1 / η_2	Threshold for determining a catalytic/ inhibitory effect.
$\mathbb{C}(t_k) / \mathbb{I}(t_k)$	Set of places that have catalytic/inhibitory effects on t_k.

rule holds. In this paper, if a non-causal variable can reduce (increase) the probability that a reasoning rule holds, the variable is said to have an inhibitory (catalytic) effect on the rule.

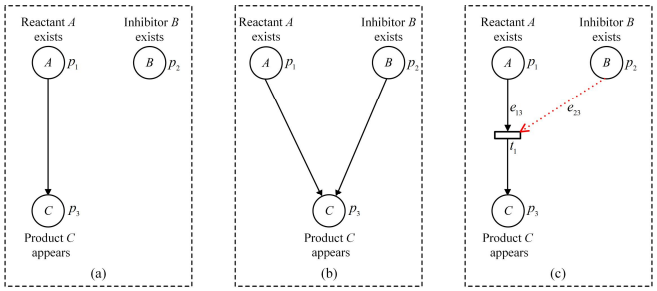

Fig. S2. Knowledge modeling for the chemical reaction. (a) and (b) BN modelling paradigms, and (c) proposed modelling paradigm.

Figs. S2(a) and (b) show the results of modeling the aforementioned reaction based on BNs. Next, the reason for the existence of *C* is discussed based on the causal structures shown in Figs. S2(a) and (b). Consider the case where reactant *A*, inhibitor *B*, and product *C* are present. From Fig. S2(a), reactant *A* is the reason for the presence of product *C* in this case. It can be inferred from Fig. S2(b) that in this case, both reactant *A* and inhibitor *B* are the reasons for the existence of product *C*. In fact, inhibitor *B* can block the causal relation between reactant *A* and product *C*. Therefore, in this case, neither reactant *A* nor inhibitor *B* is the reason for the presence of product *C*. From Figs. S2(a) and (b), traditional BNs fail to depict synergy effects, thus affecting the credibility of RCA. To characterize the impact of synergy effects on RCA, this paper uses PNs to model the synergy effects constrained by time. Specifically, a synergy relation is defined from a place to a transition to depict the synergy effects of non-causal variables on causal rules. The weight assigned to the synergy relation indicates the time delay in the generation of the synergy effect. A red dotted (green dashed) arc represents an inhibitory (catalytic) effect. Fig. S2(c) shows the proposed modelling paradigm, which can accurately model the synergy effects constrained by time.

III. SBTPN STRUCTURE MINING

Fig. S3 shows the proposed methodology. Its basic idea is as follows: 1) learning simple temporal causal relations from MTS; 2) mining the "AND/OR" relations constrained by time based on the discovered causal relations; 3) discovering the synergy effects constrained by time based on the identified "AND/OR" relations; and 4) implementing RCA based on the "AND/OR" relations and synergy effects.

A. Learning Simple Temporal Causal Relations

This study leverages the temporal causal discovery framework (TCDF) proposed in [1] to learn simple temporal causal relations. It is a 1D dilated convolutional neural network and its structure is called Attention-based Dilated Depthwise Separable Temporal Convolutional Network (AD-DSTCN). As shown in Fig. S3, each time series x_j $j \in \{1, 2, ..., m\}$ is associated with an AD-DSTCN j that is composed of m dilated convolutional neural networks (DCNNs) with 1D kernels. The input of DCNN $i \in \{1, 2, ..., m\}$

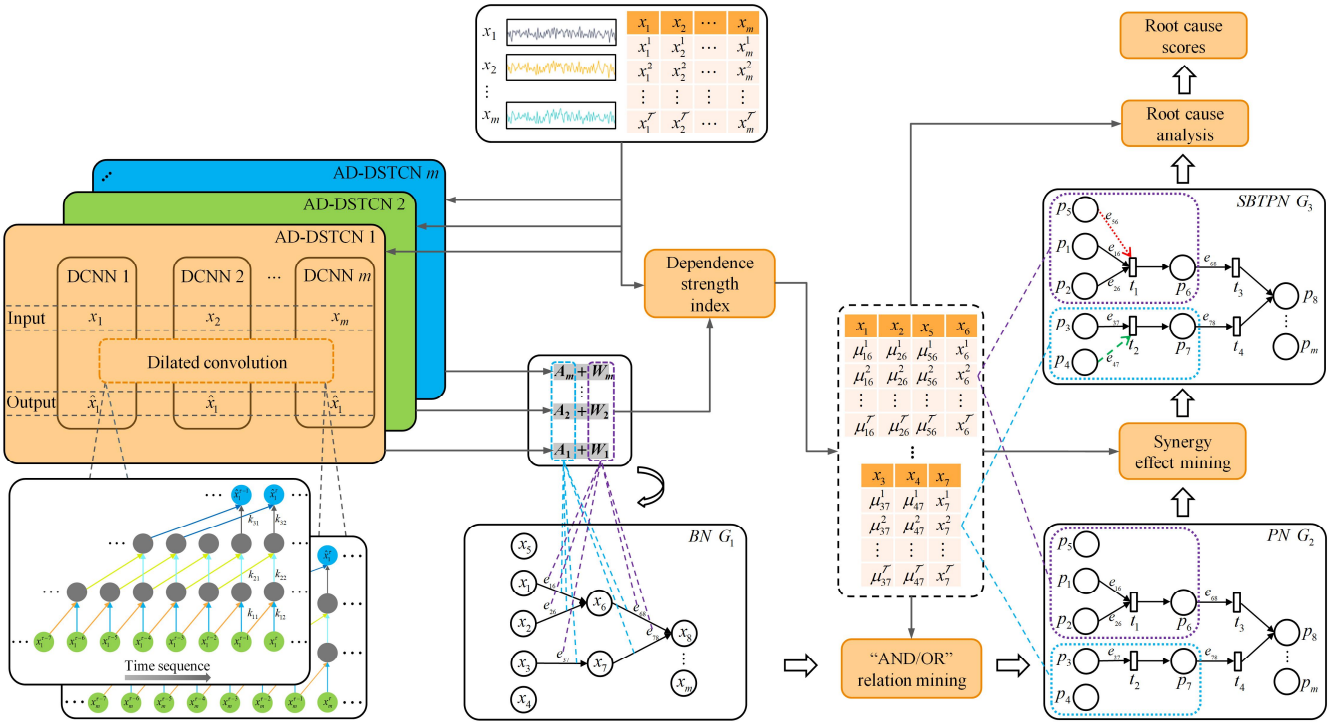

Fig. S3. Schematic diagram of the proposed methodology.

in AD-DSTCN j is a time series x_i and its outputs are the predictive value $\hat{x}_j$, the attention score a_{ij} that is used to determine whether x_i is the cause of x_j, and the kernel weight w_{ij} that is used to calculate the time delay of the impact of x_i on x_j. Vectors $A_j = [a_{1j}, a_{2j}, ..., a_{mj}]$ and $W_j = [w_{1j}, w_{2j}, ..., w_{mj}]$ are used to determine the causes of x_j and the time delays of causal effects between x_j and its causes, respectively. As seen in Fig. S3, TCDF outputs a simple temporal causal graph G_1.

The idea for discovering simple temporal causal relations is discussed in Algorithm 1.

Algorithm 1 Discovering simple temporal causal relation

Input: A dataset ∂ composed of MTS $X = \{x_1, x_2, ..., x_m\}$.

Output: A simple temporal causal graph G_1.

1: Build a TCDF consisting of m AD-DSTCNs for X;
2: **for** $j = 1, 2, ..., m$ **do**
3: Construct AD-DSTCN j whose input is ∂ and goal is to predict x_j;
4: Generate an attention score vector A_j and a kernel weight vector W_j using AD-DSTCN j;
5: Using A_j to identify the direct causes of x_j;
6: Using W_j to determine the time delays of x_j's direct causes affecting x_j;
7: **end for**
8: **return** a simple temporal causal graph G_1

B. *Mining "AND/OR" Relations Constrained by Time*

This study leverages the causal graph G_1 in Fig. S3 learned by TCDF to mine the "AND/OR" relations constrained by time.

Let $Parents(G_1, x_k)$ denote the set of causes of time series x_k in G_1. Given two time series x_i, $x_j \in Parents(G_1, x_k)$, the function $\Phi(x_i, x_j, x_k)$ in (4) in the paper is used to determine whether there is an "AND" relation between x_i and x_j pointing to x_k. This work supposes that this "AND" relation exists only if $\Phi(x_i, x_j, x_k) > 0.5$ and $\Phi(x_j, x_i, x_k) > 0.5$. For $\forall\ Q_i \subset Parents(G_1, x_k)$ such that $\forall x_i, x_j \in Q_i$: $\Phi(x_i, x_j, x_k) > 0.5 \wedge \Phi(x_j, x_i, x_k) > 0.5$, Q_i is added to a set denoted by $AND(G_1, x_k)$, which depicts all the "AND" relations among the causes of x_k. For $\forall x_l \in Parents(G_1, x_k) - AND(G_1, x_k)$, it is added to a set denoted by $T_{associ}(G_1, x_k)$ along with Q_i, which depicts all the "AND/OR" relations among the causes of x_k. $T_{associ}(G_1, x_k)$ can be mapped to a place set $T_{associ}(G_1, p_k) = \{p_l, Q_i^*\}$ such that $\beta(p_k) = x_k$, $\beta(p_l) = x_l$, and $\forall p_j \in Q_i^*\ \exists x_j \in Q_i$: $\beta(p_j) = x_j$. Each Q_i^* (or p_l) in $T_{associ}(G_1, p_k)$ is associated with a transition t_j satisfying $\dot{t}_j = Q_i^*$ (or $\dot{t}_j = p_l$) and $t_j^{\cdot} = p_k$. For $\forall p_i \in \dot{t}_j$, the directed arcs from p_i to t_j and from t_j to p_k are defined. The weight e_{ik} on the arc from p_i to t_j represents the time delay of variable $\beta(p_i)$ affecting variable $\beta(p_k)$, which can be calculated by Definition 2 in the paper. Then, PN G_2 in Fig.

S3, which considers the "AND/OR" relations constrained by time, can be obtained.

Example 1: For the variables x_1, x_2, x_3, x_6, x_7, and x_8 in the causal graph G_1 in Fig. S3, we have $Parents(G_1, x_1) = \varnothing$, $Parents(G_1, x_2) = \varnothing$, $Parents(G_1, x_3) = \varnothing$, $Parents(G_1, x_6) = \{\{x_1\}, \{x_2\}\}$, $Parents(G_1, x_7) = \{\{x_3\}\}$, and $Parents(G_1, x_8) = \{\{x_6\}, \{x_7\}\}$. Suppose $\Phi(x_1, x_2, x_6) > 0.5$, $\Phi(x_2, x_1, x_6) > 0.5$, $\Phi(x_6, x_7, x_8) < 0.5$, and $\Phi(x_7, x_6, x_8) < 0.5$, we have $AND(G_1, x_1) = \varnothing$, $AND(G_1, x_2) = \varnothing$, $AND(G_1, x_3) = \varnothing$, $AND(G_1, x_6) = \{\{x_1, x_2\}\}$, $AND(G_1, x_7) = \varnothing$, and $AND(G_1, x_8) = \varnothing$. Then, we can get $T_{associ}(G_1, x_1) = \varnothing$, $T_{associ}(G_1, x_2) = \varnothing$, $T_{associ}(G_1, x_3) = \varnothing$, $T_{associ}(G_1, x_6) = \{\{x_1, x_2\}\}$, $T_{associ}(G_1, x_7) = \{\{x_3\}\}$, and $T_{associ}(G_1, x_8) = \{\{x_6\}, \{x_7\}\}$. These "AND/OR" relations can be depicted by PN G_2 in Fig. S3, where $\dot{t}_1 = \{p_1, p_2\}$, $t_1^{\cdot} = \{p_6\}$, $\dot{t}_2 = \{p_3\}$, $t_2^{\cdot} = \{p_7\}$, $\dot{t}_3 = \{p_6\}$, $t_3^{\cdot} = \{p_8\}$, $\dot{t}_4 = \{p_7\}$, and $t_4^{\cdot} = \{p_8\}$. The time delays corresponding to these "AND/OR" relations, which can be derived by using Definition 2, are expressed as $E(p_1, p_6) = e_{16}$, $E(p_2, p_6) = e_{26}$, $E(p_3, p_7) = e_{37}$, $E(p_6, p_8) = e_{68}$, and $E(p_7, p_8) = e_{78}$.

The idea for mining the "AND/OR" relations constrained by time is realized via Algorithm 2.

Algorithm 2 Mining "AND" relations constrained by time

Input: MTS X with dataset ∂; causal graph G_1 learned by Algorithm 1.

Output: A PN-based temporal causal graph G_2.

1: **for** $\forall x_k \in X$ **do**
2: **for** $\forall\ Q_i \subset Parents(G_1, x_k)$ **do**
3: **if** $\forall x_i, x_j \in Q_i:\ \Phi(x_i, x_j, x_k) > 0.5\ \wedge\ \Phi(x_j, x_i, x_k) > 0.5$ **then**
4: Add Q_i to $AND(G_1, x_k)$;
5: **end if**
6: **end for**
7: **for** $\forall Q_i, Q_j \subset AND(G_1, x_k) \wedge Q_i \subset Q_j$ **do**
8: Remove Q_i from $AND(G_1, x_k)$;
9: **end for**
10: Add $AND(G_1, x_k)$ to $T_{associ}(G_1, x_k)$;
11: **for** $\forall x_l \in Parents(G_1, x_k) - AND(G_1, x_k)$ **do**
12: Add x_l to $T_{associ}(G_1, x_k)$;
13: **end for**
14: Map $T_{associ}(G_1, x_k)$ to $T_{associ}(G_1, p_k)$;
15: **for** $\forall p_i \in T_{associ}(G_1, p_k)$ **do**
16: Let $E(p_i, p_k) = e_{ik}$, where e_{ik} is calculated by Definition 2;
17: **end for**
18: **end for**
19: **return** a PN-based temporal causal graph G_2

C. Discovering Synergy Effects Constrained by Time

Based on PN G_2 in Fig. S3, this section proposes a method for mining the synergy effects constrained by time.

Neglecting the direction of the directed edges in G_2 yields an undirected graph. For transition t_k in the undirected graph, let $RelatedPlaces(t_k)$ be the set of all the places located in the same connected component as t_k, and $UnrelatedPlaces(t_k) = P - RelatedPlaces(t_k)$ be the set of the places that may have synergy effects on t_k. Given place $p_x \in UnrelatedPlaces(t_k)$, the function $\Psi(\dot{t}_k; t_k^{\cdot}; p_x)$ ($\Upsilon(\dot{t}_k; t_k^{\cdot}; p_x)$) in (5) in the paper is used to determine whether p_x has a catalytic (inhibitory) effect on t_k. Let η_1 and η_2 denote the thresholds for determining a catalytic effect and an inhibitory one, respectively, which satisfy $\eta_1, \eta_2 \in [0.5, 1)$. If $\Psi(\dot{t}_k; t_k^{\cdot}; p_x) > \eta_1$, p_x has a catalytic effect on t_k and a green dashed arc from p_x to t_k is defined. If $\Upsilon(\dot{t}_k; t_k^{\cdot}; p_x) > \eta_2$, p_x has an inhibitory effect on t_k and a red dotted arc from p_x to t_k is defined. We set $\eta_1 = 0.51$ and $\eta_2 = 0.54$. Given $p_j = t_k^{\cdot}$, the weight e_{xj} on the dashed/dotted arc from p_x to t_k indicates the time delay of variable $\beta(p_x)$ affecting variable $\beta(p_j)$, which can be calculated by Definition 2 in the paper. Then, SBTPN G_3 in Fig. S3, which considers the "AND/OR" relations and synergy effects constrained by time, can be derived.

Example 2: For the transitions t_1 and t_2 in PN G_2 in Fig. S3, we have $RelatedPlaces(t_1) = RelatedPlaces(t_2) = \{p_1, p_2, p_3, p_6, p_7, p_8\}$ and $UnrelatedPlaces(t_1) = UnrelatedPlaces(t_2) = \{p_4, p_5\}$. If $\Psi(p_3; p_7; p_4) > \eta_1$ and $\Upsilon(p_1, p_2; p_6; p_5) > \eta_2$, p_4 has a catalytic effect on t_2 and p_5 has an inhibitory effect on t_1. As shown in SBTPN G_3 in Fig. S3, a green dashed arc from p_4 to t_2 and a red dotted arc from p_5 to t_1 are defined to represent the above catalytic and inhibitory effects, respectively. Their corresponding time delays, which can be calculated by Definition 2, are expressed as $E(p_4, p_7) = e_{47}$ and $E(p_5, p_6) = e_{56}$, respectively.

The procedure for mining the synergy effects constrained by time is summarized in Algorithm 3.

Algorithm 3 Mining synergy effects constrained by time

Input: MTS X with dataset ∂; PN G_2 learned by Algorithm 2; thresholds η_1 and η_2 used to determine synergy effects.

Output: An SBTPN G_3.

1: **for** $\forall t_k \in G_2$ **do**
2: **for** $\forall p_x \in UnrelatedPlaces(t_k)$ **do**
3: **if** $\Psi(\dot{t}_k; t_k^{\cdot}; p_x) > \eta_1$ **then**
4: Define a green dashed arc from p_x to t_k;
5: **end if**
6: **if** $\Upsilon(\dot{t}_k; t_k^{\cdot}; p_x) > \eta_2$ **then**
7: Define a red dotted arc from p_x to t_k;
8: **end if**

9: Denote by p_j the output place of t_k;
10: Let $E(p_x, p_j) = e_{xj}$, where e_{xj} is calculated by Definition 2;
11: **end for**
12: **end for**
13: **return** an SBTPN G_3

Noise can interfere with the discovery of causal relations, especially "AND" relations and synergy effects. Specifically, for the "AND" and "OR" structures pointing to the same effect variable, identifying the "AND" structures becomes more difficult as the number of the "OR" structures increases. This is because the "OR" structure can weaken the influence of the cause variables within the "AND" structure on the effect variable, making it difficult to discover the causal relations among these variables and thus identify the entire "AND" structure. Moreover, noise is a prerequisite for the existence of synergy effects. If a causal effect is deterministic (i.e., always occurs), synergy effect will not exist. The amount of synergy information carried by synergy variables increases as the noise ratio increases in a certain range, thus making the synergy effects easier to identify.

IV. SBTPN-BASED RCA

This section proposes an RCA method considering "AND/OR" relations and synergy effects based on SBTPN.

For a place p_j in an SBTPN, let $CauseSet(p_j)$ denote the places corresponding to the direct or indirect causes of variable $\beta(p_j)$, and $\hat{\boldsymbol{p}}_j$ represent that $\beta(p_j)$ is in an abnormal state. Consider SBTPN G_3 in Fig. S3. We have $CauseSet(p_8) = \{p_1, p_2, p_3, p_6, p_7\}$. Any variable $\beta(p_i)$ satisfying $p_i \in CauseSet(p_j)$ may be the root cause of the abnormality of $\beta(p_j)$. $\Pr(\hat{\boldsymbol{p}}_i|\hat{\boldsymbol{p}}_j)$ depicts the possibility that $\hat{\boldsymbol{p}}_i$ is the root cause of $\hat{\boldsymbol{p}}_j$. For a transition $t_k \in p_i^{\bullet}$ located in the causal path from $\beta(p_i)$ to $\beta(p_j)$, if it is affected by synergy effects, the value of $\Pr(\hat{\boldsymbol{p}}_i|\hat{\boldsymbol{p}}_j)$ is changed. Let $\Pr(\hat{\boldsymbol{p}}_i|\hat{\boldsymbol{p}}_j)'$ denote the probability that $\hat{\boldsymbol{p}}_i$ is the root cause of $\hat{\boldsymbol{p}}_j$ considering all synergy effects imposed on t_k, which can be characterized by (8) in the paper.

Let $\delta = \max\{\Pr(\hat{\boldsymbol{p}}_i|\hat{\boldsymbol{p}}_j)' \mid p_i \in CauseSet(p_j)\}$ such that $\delta \geq 0.5$. It is a benchmark used to identify the potential root causes of $\beta(p_j)$ from $CauseSet(p_j)$. For $\forall p_i \in CauseSet(p_j)$, if $\Pr(\hat{\boldsymbol{p}}_i|\hat{\boldsymbol{p}}_j)' \in [0.5, \delta]$, p_i is added to a set denoted by $Candidate(p_j)$. It records the potential root causes of $\beta(p_j)$. Let $TranSet(p_a, p_b)$ denote the set of transitions located in the causal path from variable $\beta(p_a)$ to variable $\beta(p_b)$. For $\forall p_q \in Candidate(p_j)$, if $\exists t_u \in p_q^{\bullet} \cap TranSet(p_q, p_j)$ such that $p_s \in {}^{\bullet}t_u \wedge p_s \notin Candidate(p_j)$, p_q is removed from $Candidate(p_j)$. Let $I(p)$ denote the set of input places of p. Consider SBTPN G_3 in Fig. S3. We have $I(p_8) = \{p_6, p_7\}$. For $\forall p_1, p_2, ..., p_u \in Candidate(p_j)$ such that $p_1 \in I(p_j)$ and $p_k \in I(p_{k-1})$ $(k = 2,3,...,u)$, if $p_g \notin Candidate(p_j)$ for $\forall p_g \in I(p_u)$, p_u is added to a set denoted by $RootCause(p_j)$. The variables corresponding to the places in $RootCause(p_j)$ are the root causes of $\beta(p_j)$.

Example 3: Consider SBTPN G_3 in Fig. S3. Suppose $\delta = \Pr(\hat{\boldsymbol{p}}_6|\hat{\boldsymbol{p}}_8)' \geq 0.5$. If $\Pr(\hat{\boldsymbol{p}}_2|\hat{\boldsymbol{p}}_8)'$, $\Pr(\hat{\boldsymbol{p}}_3|\hat{\boldsymbol{p}}_8)'$, $\Pr(\hat{\boldsymbol{p}}_6|\hat{\boldsymbol{p}}_8)'$, $\Pr(\hat{\boldsymbol{p}}_7|\hat{\boldsymbol{p}}_8)' \in [0.5, \delta]$, we have $Candidate(p_8) = \{p_3, p_6, p_7\}$. Since $p_6 \in I(p_8)$, $p_6 \in Candidate(p_8)$, $p_1, p_2 \in I(p_6)$, and $p_1, p_2 \notin Candidate(p_8)$, p_6 is added to $RootCause(p_8)$. Since $p_7 \in I(p_8)$, $p_7 \in Candidate(p_8)$, $p_3 \in I(p_7)$, $p_3 \in Candidate(p_8)$, and $I(p_3) = \varnothing$, p_3 is added to $RootCause(p_8)$. Thus, the root causes of $\beta(p_8)$ are $\beta(p_3)$ and $\beta(p_6)$.

The process for discovering the root causes of abnormal events is realized via Algorithm 4.

Algorithm 4 RCA for abnormal events

Input: MTS X with dataset ∂; SBTPN G_3 learned by Algorithm 3; place set P_A corresponding to abnormal events.

Output: Root causes of abnormal events.

1: **for** $\forall p_j \in P_A$ **do**
2: **for** $\forall p_i \in CauseSet(p_j)$ **do**
3: Calculate $\Pr(\hat{\boldsymbol{p}}_i|\hat{\boldsymbol{p}}_j)'$ using (6)-(8);
4: **end for**
5: Let $\delta = \max\{\Pr(\hat{\boldsymbol{p}}_i|\hat{\boldsymbol{p}}_j)' | p_i \in CauseSet(p_j)\}$;
6: **if** $\delta < 0.5$ **then**
7: **continue**;
8: **end if**
9: **for** $\forall p_i \in CauseSet(p_j)$ **do**
10: **if** $\Pr(\hat{\boldsymbol{p}}_i|\hat{\boldsymbol{p}}_j)' \in [0.5, \delta]$ **then**
11: Add p_i to $Candidate(p_j)$;
12: **end if**
13: **end for**
14: **for** $\forall p_q \in Candidate(p_j)$ **do**
15: **if** $\exists t_u \in p_q^{\bullet} \cap TranSet(p_q, p_j)$ such that $p_s \in {}^{\bullet}t_u \wedge p_s \notin Candidate(p_j)$ **then**
16: Remove p_q from $Candidate(p_j)$;
17: **end if**
18: **end for**
19: **for** $\forall p_1, p_2, ..., p_u \in Candidate(p_j)$ such that $p_1 \in I(p_j) \wedge p_k \in I(p_{k-1})$ $(k = 2,3,...,u)$ **do**
20: **if** $p_g \notin Candidate(p_j)$ for $\forall p_g \in I(p_u)$ **then**
21: Add p_u to $RootCause(p_j)$;
22: **end if**
23: **end for**
24: **end for**
25: **return** the root causes of P_A

EXPERIMENTAL DETAILS

A. Parameter Sensitivity Analysis

This section utilizes the solar panel manufacturing case presented in Section VII of the paper to discuss the sensitivity of SBTPN to parameter settings. SBTPN is sensitive to threshold values η_1 for determining catalytic effects and η_2 for determining inhibitory effects, where $\eta_1, \eta_2 \in [0.5,1)$. From Fig. S4, SBTPN loses its ability to identify synergy effects when the values of η_1 or η_2 are high. This is because a stronger catalytic (inhibitory) effect results in a stronger correlation between the corresponding catalytic (inhibitory) factor and the effect variable, thereby increasing the likelihood of misinterpreting the catalytic (inhibitory) effect as a causal one. Moreover, a low value of η_2 negatively impacts the performance of SBTPN in identifying synergy effects. This is because noise can interfere with the identification of synergy effects, and a low synergy identification threshold is unable to eliminate this interference. From the above analysis, the thresholds for identifying synergy effects should be determined cautiously.

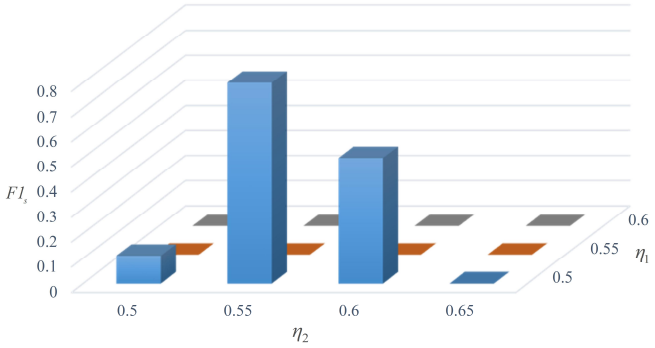

Fig. S4. Performance of SBTPN in identifying synergy effects under different combinations of η_1 and η_2.

B. Principles of RCA

This section defines the principles of RCA based on the compared methods discussed in Section VII-A of the paper.

For an anomalous variable x_j^τ in Fig. S5, if its direct/indirect cause $x_i^{\tau-k}$ is also in an abnormal state such that k is consistent with the time delay of x_i affecting x_j given in Fig. S5, $x_i^{\tau-k}$ is added to a set denoted by $PotenRoot(x_j^\tau)$. $PotenRoot(x_j^\tau)$ denotes the potential root causes of x_j^τ. Consider the case that the products x_{20}^τ, $x_{17}^{\tau-1}$, $x_{14}^{\tau-3}$, and $x_{10}^{\tau-6}$ in Fig. S5 are all unqualified. By the temporal causal relations in Fig. S5, we have $PotenRoot(x_{20}^\tau) = \{x_{17}^{\tau-1}, x_{14}^{\tau-3}\}$. Let $AntTran(x_j)$ denote the set of the transitions that directly/indirectly point to variable x_j, $DireCause(t)$ denote the set of t's input places-associated variables, and $Cause(t)$ denote the set of the variables in $DireCause(t)$ and their direct/indirect causes. Consider the variable x_{14} and the transition t_{11} in Fig. S5. We have $AntTran(x_{14}) = \{t_1, t_2, t_5, t_6\}$, $DireCause(t_{11}) = \{x_{14}\}$, and $Cause(t_{11}) = \{x_1, x_2, x_3, x_4, x_{10}, x_{11}, x_{14}\}$. The following four rules are adopted to determine the true root causes of x_j^τ.

Rule 1: For a transition $t \in AntTran(x_j)$ such that t is affected by inhibitory effects, the abnormality of $x_i \in Cause(t)$ is not the cause of the abnormality of x_j. Therefore, x_i is removed from $PotenRoot(x_j^\tau)$.

Rule 2: For a transition $t \in AntTran(x_j)$, if $\exists x_a, x_b \in DireCause(t)$ such that at least one of them is in a normal state, the abnormality of $x_i \in Cause(t)$ is not the cause of the abnormality of x_j. Thus x_i is removed from $PotenRoot(x_j^\tau)$.

Rule 3: For a transition $t \in AntTran(x_j)$, if $\exists x_i \in Cause(t)$ such that it is in a normal state, the abnormality of the direct/indirect cause x_h of x_i is not the cause of the abnormality of x_j. Therefore, x_h is removed from $PotenRoot(x_j^\tau)$.

Rule 4: By the above three rules, $PotentRoot(x_j^\tau)$ is updated. For the updated $PotentRoot(x_j^\tau)$, if $\exists x_i \in PotenRoot(x_j^\tau)$ such that x_k is not the cause of x_i for $\forall x_k \in PotenRoot(x_j^\tau) \setminus x_i$, x_i is the root cause of x_j^τ.

The temporal causal structure in Fig. S5 is used to illustrate the above rules. Consider the following cases.

Case 1: x_{20}^τ, $x_{17}^{\tau-1}$, $x_{14}^{\tau-3}$, and $x_{10}^{\tau-5}$ are all unqualified, and $x_{29}^{\tau-2}$ is true (i.e., $x_{14}^{\tau-3}$ is quality checked at time $\tau-2$). We have $PotenRoot(x_{20}^\tau) = \{x_{10}^{\tau-5}, x_{14}^{\tau-3}, x_{17}^{\tau-1}\}$. By Rule 1, we can get that the abnormality of $x_{10}^{\tau-5}$ or $x_{14}^{\tau-3}$ is not the cause of the abnormality of x_{20}^τ. According to Rule 4, the cause of x_{20}^τ being unqualified is $x_{17}^{\tau-1}$ being unqualified.

Case 2: x_{14}^τ, $x_{10}^{\tau-2}$, and $x_1^{\tau-3}$ are all unqualified, but $x_2^{\tau-3}$ is qualified. We have $PotenRoot(x_{14}^\tau) = \{x_1^{\tau-3}, x_{10}^{\tau-2}\}$. According to Rule 2, the abnormality of $x_1^{\tau-3}$ is not the cause of the abnormality of x_{14}^τ. By Rule 4, the cause of x_{14}^τ being unqualified is $x_{10}^{\tau-2}$ being unqualified.

Case 3: x_{20}^τ, $x_{17}^{\tau-1}$, and $x_{10}^{\tau-5}$ are all unqualified, but $x_{14}^{\tau-3}$ is qualified. We have $PotenRoot(x_{20}^\tau) = \{x_{10}^{\tau-5}, x_{17}^{\tau-1}\}$. By Rule 3, the abnormality of $x_{10}^{\tau-5}$ is not the cause of the abnormality of x_{20}^τ. According to Rule 4, the cause of x_{20}^τ being unqualified is $x_{17}^{\tau-1}$ being unqualified.

The Syn500 dataset discussed in the paper is used to evaluate the performance of the compared methods in RCA. This study focuses on the abnormalities of end products x_{20}, x_{21}, and x_{22} in Fig. S5. According to the causal relations in Fig. S6(a) and Rules 1-4, the true root causes of abnormalities

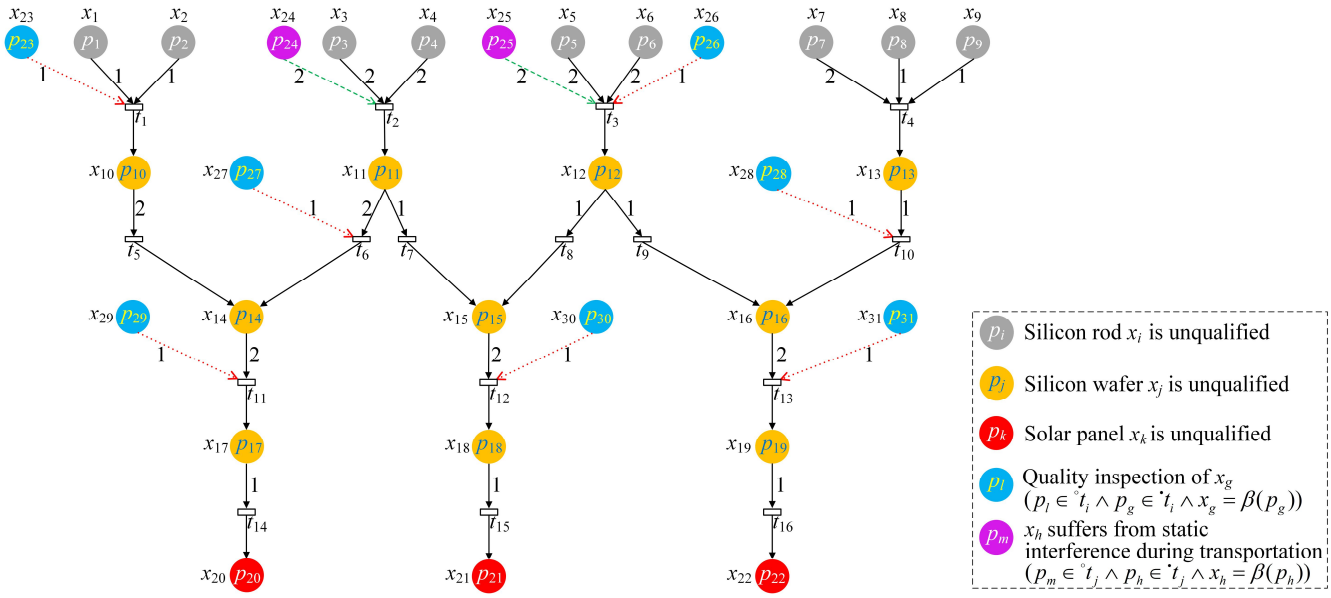

Fig. S5. Quality anomalies of solar panel manufacturing process.

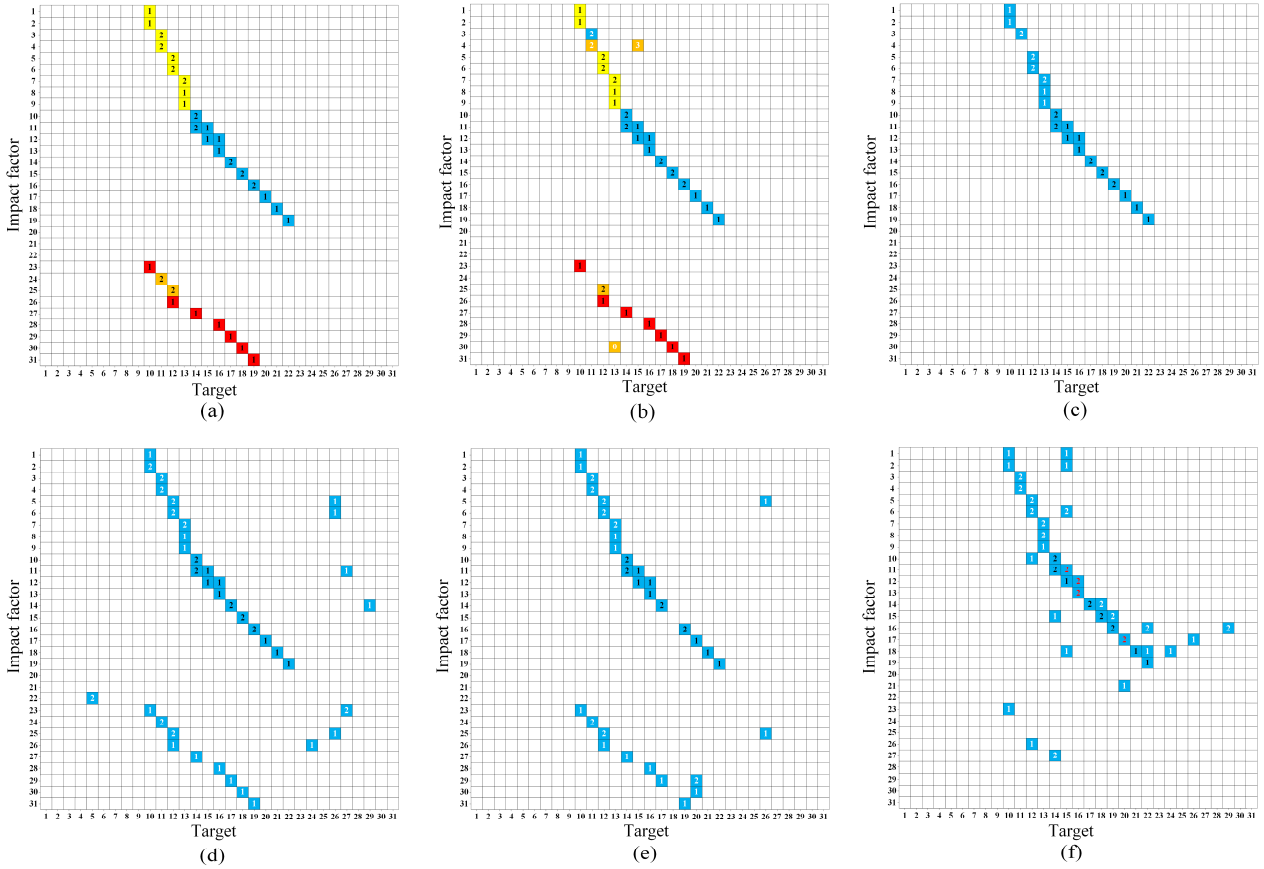

Fig. S6. Causality matrices generated by all compared methods. (a) Ground truth, (b) SBTPN, (c) TCDF, (d) MTE, (e) PCMCI, and (f) MLP-GC.

of x_{20}, x_{21} and x_{22} on Syn500 can be identified. Since all the competing methods fail to consider "AND" relations as well as synergy effects, the root causes of these three variables predicted by them on Syn500 can be obtained by using the causal relations in Figs. S6(c)-(f) as well as Rules 3 and 4. The root causes of these variables predicted by the proposed SBTPN on Syn500 can be derived by using the causal relations in Fig. S6(b) and Algorithm 4.

REFERENCE

[1] M. Nauta, D. Bucur, and C. Seifert, "Causal discovery with attention-based convolutional neural networks," *Mach. Learn. Knowl. Extr.*, vol.1, no.1, pp.312–340, 2019.