

Spatio-temporal graphical counterfactuals: an overview

Mingyu KANG¹, Duxin CHEN^{2*}, Ziyuan PU³, Jianxi GAO^{4,5} & Wenwu YU^{2,6,7*}

¹*School of Cyber Science and Engineering, Southeast University, Nanjing 210096, China*

²*School of Mathematics, Southeast University, Nanjing 210096, China*

³*School of Transportation, Southeast University, Nanjing 210096, China*

⁴*Department of Computer Science, Rensselaer Polytechnic Institute, Troy NY 12180, USA*

⁵*Center for Network Science and Technology, Troy NY 12180, USA*

⁶*Frontiers Science Center for Mobile Information Communication and Security, Nanjing 210096, China*

⁷*Purple Mountain Laboratories, Nanjing 211102, China*

Received 5 July 2024/Revised 1 August 2025/Accepted 8 September 2025/Published online 6 February 2026

Abstract Counterfactual thinking is a crucial yet challenging topic for artificial intelligence to learn knowledge from data and ultimately improve performance for new scenarios. Many research studies, including the potential outcome model (POM) and the structural causal model (SCM), have been proposed to address this. However, their modeling, theoretical foundations, and application approaches often differ. Moreover, there is a lack of graphical approaches for inferring spatio-temporal counterfactuals, that account for spatial and temporal interactions among multiple units. Thus, in this work, we aim to present a survey that compares and discusses different counterfactual models, theories and approaches. Additionally, we propose a unified graphical causal framework to infer spatio-temporal counterfactuals.

Keywords counterfactual inference, spatio-temporal graphical model, potential outcome model, structural causal model, causality

Citation Kang M Y, Chen D X, Pu Z Y, et al. Spatio-temporal graphical counterfactuals: an overview. *Sci China Inf Sci*, 2026, 69(4): 141201, <https://doi.org/10.1007/s11432-024-4752-6>

1 Introduction

How can we enable an intelligent machine to think about and answer a spatio-temporal counterfactual question? For example, sometimes we want to know: if a different investment strategy had been implemented, would we have obtained higher returns? Or, in computer networks, what changes would occur in network load if a node's configuration had never been changed? Or would someone still purchase a product even if they had never seen the advertisement? These questions generally follow a spatio-temporal counterfactual pattern: "Given the observed . . . , if . . . had been done, how would . . . have evolved?". Thus, according to the pattern, time must be considered first, because the questions are posed now, but the counterfactual actions are supposed to be performed in the past. Second, counterfactual outcomes are unobservable. The outcomes of some actions are observed when they are performed on a real-world system, thus, these outcomes are factual. But the counterfactual outcomes are not observed, and they are imagined by supposing that different actions had been taken. Third, the real-world system can be observed as a multivariate time series, and there are widespread multivariate interactions within it [1, 2].

To enable intelligent machines to mimic counterfactual thinking, the potential outcome model (POM) [3] and structural causal model (SCM) [4–7] have been proposed as two frameworks with different foundations. POM uses stable unit treatment value assumption (SUTVA), consistency assumption, positivity assumption and ignorability assumption as foundations [8–11]. If these assumptions are hold, one can use matching or imputation approaches to infer counterfactuals without confounders. In contrast, SCM builds a causal ladder, i.e., abduction, action and prediction, to infer counterfactuals through graphical language d-separation [4, 5, 7]. With the graphical language d-separation, the causal Markov assumption, faithfulness assumption and causal sufficiency assumption are proposed to guarantee the correctness of counterfactual inference on SCM [12–16].

However, the distinct foundations between POM and SCM lead to differences in counterfactual inference. These differences include the following. (i) POM assumes positivity, whereas SCM does not. Thus, the counterfactuals in POM are factual and can be found in observations certainly. In contrast, counterfactuals in SCM can be out

* Corresponding author (email: chendx@seu.edu.cn, wyyu@seu.edu.cn)

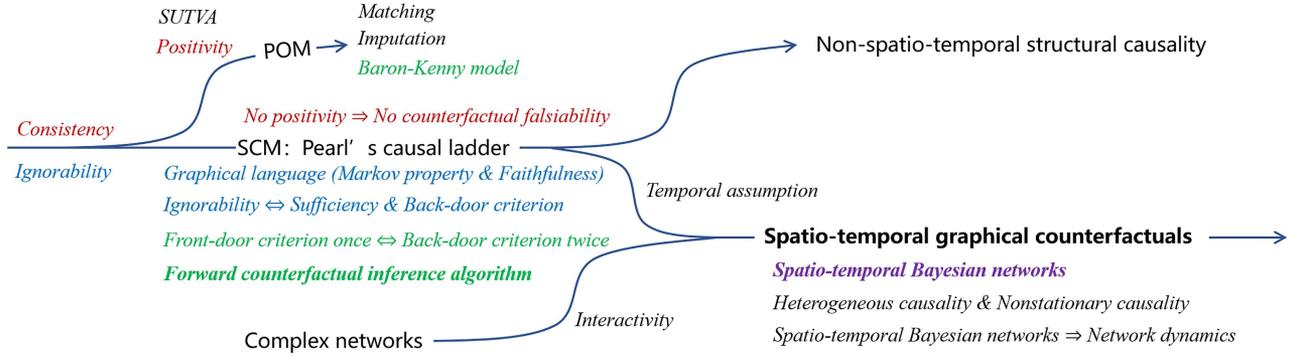


Figure 1 (Color online) Towards spatio-temporal graphical counterfactuals.

of observations and are therefore unfalsifiable, that is, their correctness cannot be empirically validated. (ii) POM explicitly assumes consistency, but this assumption is rejected for SCM [7]. Actually, SCM also implicitly relies on the consistency assumption to ensure that the counterfactual outcome corresponds to one of the possibly unobserved potential outcomes. (iii) SCM uses a type of graphical language to deconfound with certain rules, but POM mainly relies on human-guided deconfounding, which makes it difficult to satisfy the ignorability. In contrast, SCM can deconfound via the back-door criterion [4, 5, 17, 18] under the causal sufficiency assumption, and it is equivalent to ignorability. (iv) POM faces limitations to analyze indirect causation (mediation), thus POM typically relies on the Baron-Kenny model [19, 20] for assistance. By contrast, SCM facilitates this analysis through the front-door criterion [4, 5, 17, 18], which decomposes a causal pathway into two segments: from the source variable to the mediator, and from the mediator to the target variable. Back-door adjustment is then applied twice, assuming no additional mediators are presented.

Moreover, both POM and SCM currently lack the ability to model the spatio-temporal interactions among system units. For example, the SUTVA requires no interaction between system units in POM. Similarly, SCM, when based on the traditional Bayesian network, does not handle multivariate time series and often overlooks nonstationarity. This limitation motivates us to identify a common point to unify the two frameworks and strengthen inference for spatio-temporal counterfactuals. Our recent work [21, 22] has proposed a spatio-temporal Bayesian networks (STBNs) model based on a temporal assumption that the cause precedes or parallels the effect in time. It is found that the famous complex networks [23–25], a type of network dynamics used to model spatio-temporal interactions, are also a special case of STBNs.

Thus, we further conduct this work to build a spatio-temporal graphical counterfactual framework based on STBNs. This framework requires two components: (i) STBNs used as the foundation of SCM, (ii) a forward counterfactual inference algorithm designed on the decomposability of the front-door criterion. We then discuss how to get the two points, following the storyline as shown in Figure 1. Thus, the main contributions of this work are as follows.

- (1) A survey is conducted to discuss the frameworks from POM to SCM, including their theoretical foundations and application approaches. This part is organized from Sections 2 to 4.
- (2) A forward counterfactual inference algorithm is designed here to recursively and autonomously infer counterfactuals through causal graphical languages. This part is presented as Algorithm 1 in Subsection 4.3.
- (3) An overview of the spatio-temporal graphical counterfactual framework is proposed to discuss the spatio-temporal Bayesian networks, their nonstationarity, and the relationship with complex networks. This part is organized in Section 5.

2 Potential outcome model

The counterfactuals are defined as the unobserved potential outcomes in POM, as shown in Table 1. Let Y be the potential outcomes of the target variable, and T be the treatment variable. $T = 1$ represents assignment to the treatment group, and $T = 0$ represents assignment to the control group in a randomized controlled trial involving N units. X represents covariates that have relationship with Y and T , and are controlled for randomized trials. POM models the general process from causal estimands (T , Y and X) to “Science” [26]. In this case, for different treatments T , only one potential outcome, $Y(1)$ or $Y(0)$, can be observed, and another one is unobserved, namely the counterfactual.

Table 1 Data for POM example.

Units i	Covariates X	Potential outcomes		Treatment T
		$Y(1)$	$Y(0)$	
1	x_1	$y_1(1)$?	1
2	x_2	?	$y_2(0)$	0
3	x_3	?	$y_3(0)$	0
4	x_4	$y_4(1)$?	1
5	x_5	$y_5(1)$?	1
\vdots	\vdots	\vdots	\vdots	\vdots
N	x_N	?	$y_N(0)$	0

2.1 Inferring counterfactuals through POM

To infer the counterfactuals using POM, one common approach is exact sample matching. For a unit i , one can search the samples with exactly matching covariates X but treatment T is different, and then estimate the counterfactual outcome $Y_i(1 - T_i)$. But it is difficult to conduct when the dimension of X is high, as it requires a large number of samples to support the matching. Thus, approximate matching is commonly better, e.g., caliper matching [27–30] and propensity score matching [11, 31–33]. They search for a group of similar samples to the target unit with one or more measurements, then calculate counterfactuals.

Another common approach is data-driven imputation, which views counterfactuals as missing values, and interpolates them by fitting the observational data. Linear or nonlinear regressions [34–37] are typically used to interpolate missing values based on patterns learned from the observational data. In contrast, tensor decomposition approaches [38–40] recover the missing values by decomposing and reconstructing a sparse data tensor. Deep generative models [41–47] have also been used to capture data uncertainty and generate missing values by randomly sampling. Note that, the imputed missing values produced by generative models are not unique. Instead, they constitute a distribution of plausible values, which distinguishes them from regression-based and tensor decomposition approaches.

2.2 Foundational assumptions of POM

Before using the above approaches within the POM framework, four assumptions must be accepted. The first is SUTVA, which is defined as follows.

Assumption 1 (SUTVA [10]). In Table 1, (1) there is no interference between units, that is, neither $Y_i(1)$ nor $Y_i(0)$ is affected by what action any other unit received; (2) there are no hidden versions of treatments, that is, no matter how unit i received treatment $T = 1$, the outcome that would be observed would be $Y_i(1)$, and similarly for treatment $T = 0$.

SUTVA guarantees the identical independence between the experimental units, and for unit i , the outcome Y_i is only up to its treatment T_i , not the others'. However, this assumption is not always satisfied, e.g., in social networks, where interactions occur at varying times.

The second assumption is consistency, defined as follows.

Assumption 2 (Consistency [8, 9]). In Table 1, if the i -th unit is selected for a treatment T_i , the observed value of Y_i , neither $Y_i(1)$ nor $Y_i(0)$, is the same for all assignments of treatments to the other experimental units.

Another form to describe the consistency assumption through expectation $\mathbb{E}[\cdot]$ is

$$\begin{aligned} \mathbb{E}[Y(1)|T = 1] &= \sum_{i=1}^N y_i \times t_i = \mathbb{E}[Y|T = 1], \\ \mathbb{E}[Y(0)|T = 0] &= \sum_{i=1}^N y_i \times (1 - t_i) = \mathbb{E}[Y|T = 0], \end{aligned} \tag{1}$$

where y_1, \dots, y_N are the samples for N units, and $t_1, \dots, t_N \in \{0, 1\}$ are the values of treatment variable T , as shown in Table 1. This is the same as the case of taking a conditional expectation with respect to X . Thus, the consistency assumption guarantees that the missing counterfactual values in Subsection 2.1 can be interpolated by the other observational values.

The third assumption is positivity, defined as follows.

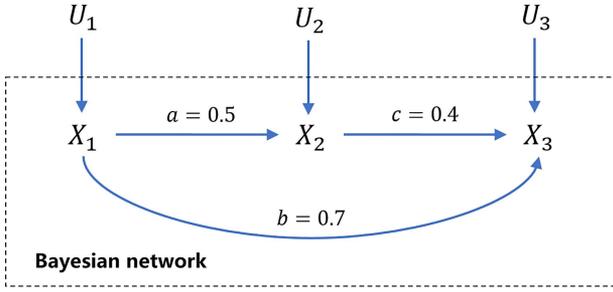


Figure 2 (Color online) Example of SCM [5].

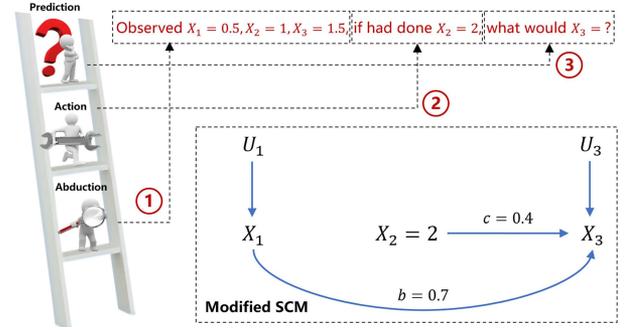


Figure 3 (Color online) Diagram of Pearl's causal ladder [5].

Assumption 3 (Positivity [11]). In Table 1, $0 < P(T = 1|X = x) < 1$.

This means that, for $X = x$, some treatments are always assigned to $T = 0$ or $T = 1$ at random. Otherwise, the counterfactuals for $T = 1$ could not be obtained if all units with $X = x$ are assigned with $T = 0$, and this is the same as the counterfactuals for $T = 0$.

The fourth assumption is ignorability, defined as follows.

Assumption 4 (Ignorability [8,9]). In Table 1, $T \perp\!\!\!\perp Y(1), Y(0)|X$ for each unit.

Thus, the ignorability assumption implies that all the confounders have been observed in X , that is, no hidden confounder would interfere with the observations of $Y(1), Y(0)$. Thus, the ignorability assumption is also called the unconfoundedness assumption. This assumption is usually made to avoid the preference for experimental manipulation, which would guarantee randomization in the experiments [48].

3 Structural causal model

In contrast to POM, SCM provides a graphical language for intuitively discussing causality, thereby avoiding complex human-guided technical operations. An SCM is built on a set of variables, including endogenous variables $\mathbf{V} = \{X_1, \dots, X_n\}$ and exogenous variables $\mathbf{U} = \{U_1, \dots, U_n\}$. Note that endogenous variables are observable from data, but exogenous variables are not. Moreover, a set of functions $\mathbf{F} = \{f_1, \dots, f_n\}$ can also be used to describe the functional relationship between these variables.

For example, as shown in Figure 2, the SCM can be functionally described as

$$\begin{aligned} f_1 : X_1 &= U_1, \\ f_2 : X_2 &= aX_1 + U_2, \\ f_3 : X_3 &= bX_1 + cX_2 + U_3, \end{aligned} \tag{2}$$

where $a = 0.5, b = 0.7, c = 0.4$. U_1, U_2, U_3 are all additive noise, and they are mutually independent.

In an SCM, the functional relationship must conform to the directed acyclic graphical constraint. The graph is originally based on the concept of Bayesian network [12–16]. And later, it is also called a causal graphical model to highlight its causal structure [4,5,17,18,49]. And recently, it is also called a causal network in the field of causal discovery [50–54]. But actually, they are the same in the framework of SCM, and they are all directed acyclic graphs (DAGs) that represent Markovian knowledge. Thus, to avoid misunderstanding, we here use the name of Bayesian network uniformly, due to the fact that causal graphical model and causal network are not always a Bayesian network in Figure 2. For example, direct cyclic graphs [55–58], Markov networks [59,60], and full time graph [6], they are also causal graph or causal network.

3.1 Pearl's causal ladder

To infer the counterfactuals through SCM, there are three steps.

- (1) **Abduction.** Infer the values of exogenous variables \mathbf{U} from the observational data.
- (2) **Action.** Perform an intervention on SCM, e.g., $\text{do}(X_2 = 2)$ on SCM in Figure 2, and then, X_2 would be assigned with value 2, and the arrows to X_2 would be modified, as shown in Figure 3.
- (3) **Prediction.** Use the modified SCM to recalculate the counterfactuals of the target variable.

As shown in Figure 3, consider the following query “observed $X_1 = 0.5, X_2 = 1, X_3 = 1.5$, if had done $X_2 = 2$, what would $X_3 = ?$ ”. Obviously, this is a counterfactual question. Thus, following Pearl’s causal ladder, as shown in Figure 3, we first infer the values of exogenous variables $\mathbf{U} = \{U_1, U_2, U_3\}$ from the observational data $X_1 = 0.5, X_2 = 1$ and $X_3 = 1.5$, as follows:

$$\begin{aligned} U_1 &= 0.5, \\ U_2 &= 1 - 0.5 \times 0.5 = 0.75, \\ U_3 &= 1.5 - 0.7 \times 0.5 - 0.4 \times 1 = 0.75. \end{aligned} \quad (3)$$

Then, $\text{do}(X_2 = 2)$ to obtain a modified SCM. And finally, recalculate the counterfactual of X_3 , that is,

$$X_3(\text{do}(X_2 = 2)) = 0.5 \times 0.7 + 2 \times 0.4 + 0.75 = 1.9, \quad (4)$$

where $X_3(\text{do}(X_2 = 2))$ is different compared to the observational X_3 .

The intervention, also named as do-operator [4], can be generalized into the form of probability, if the modularity assumption is introduced as follows.

Assumption 5 (Modularity [4]). If a set of variables $\{X_{j_1}, \dots, X_{j_p}\} \subset \mathbf{V}$ is intervened, then for each variable $X \in \mathbf{V}$, it is obtained that

(1) if $X \notin \{X_{j_1}, \dots, X_{j_p}\}$, then $P(X|\mathbf{Pa}(X))$ remains unchanged, where $\mathbf{Pa}(X)$ denotes the set of predecessors of X in a Bayesian network, also called the causal parents of X ;

(2) if $X \in \{X_{j_1}, \dots, X_{j_p}\}$, then $P(X = x|\mathbf{Pa}(X)) = 1$ if x is the value set by intervention to X , otherwise, $P(X = x|\mathbf{Pa}(X)) = 0$.

This means that if a set of interventions is applied to X_{j_1}, \dots, X_{j_p} , their values would be fixed. The connections between them and their causal parents would be broken, and their causal parents would not affect them anymore in the modified SCM, just like in Figure 3. Moreover, this also means that the probability distributions of the other variables that are not intervened upon would not change.

3.2 Foundational assumptions of Bayesian network

As presented above, to clearly define causality, each SCM is associated with a Bayesian network satisfying the directed acyclic graphical constraint. Thus, the causality in the Bayesian network provides the foundation for counterfactuals. Thus, we discuss the foundational assumptions of Bayesian networks in the following sections.

First of all, a concept of d-separation can be defined on a Bayesian network \mathcal{G} as follows.

Definition 1 (d-separation [12–16]). Let \mathbf{X} , \mathbf{Y} and \mathbf{Z} be three disjoint subsets of endogenous variables $\mathbf{V} = \{X_1, \dots, X_n\}$, and let \mathbf{p} be any path from a node in \mathbf{X} to a node in \mathbf{Y} regardless of direction. \mathbf{Z} is said to block \mathbf{p} if and only if there is a node $v \in \mathbf{p}$ satisfying one of the following items:

(1) v has v -structure (two nodes $a, b \in \mathbf{p}$ pointing to v , namely $a \rightarrow v \leftarrow b$), and neither v nor its any descendants are in \mathbf{Z} ;

(2) v in \mathbf{Z} and v does not have v -structure.

Then, \mathbf{Z} d-separate \mathbf{X} and \mathbf{Y} , denoted as $\mathbf{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{Y} | \mathbf{Z}$, if \mathbf{Z} blocks any \mathbf{p} .

Then, two assumptions, causal Markov (or Markov property) and faithfulness, are introduced as follows.

Assumption 6 (Causal Markov [12–16]). Probability distribution P is Markovian to a Bayesian network \mathcal{G} if

$$\mathbf{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{Y} | \mathbf{Z} \Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}, \quad (5)$$

where \mathbf{X} , \mathbf{Y} and \mathbf{Z} are three disjoint subsets.

Assumption 7 (Faithfulness [12–16]). Probability distribution P is faithful to a Bayesian network \mathcal{G} if

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z} \Rightarrow \mathbf{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{Y} | \mathbf{Z} \quad (6)$$

for all disjoint subsets \mathbf{X} , \mathbf{Y} and \mathbf{Z} .

It is intuitive to understand the two assumptions: the Bayesian network \mathcal{G} corresponds one-to-one to the independence structure of the probability distribution P in the observational data. Thus, if they are satisfied, the joint distribution can be factorized according to the Bayesian network, as follows:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \mathbf{Pa}(X_i)), \quad (7)$$

where $\text{Pa}(X_i)$ is the predecessors of X_i in \mathcal{G} , that is also called as causal parents.

Moreover, a Bayesian network is also assumed to be causally sufficient.

Assumption 8 (Causal sufficiency [12–16]). Variables \mathbf{V} are said to satisfy causal sufficiency if there is no hidden variable that is a common cause of two or more variables in \mathbf{V} .

This means that if a Bayesian network (or an SCM) is causally sufficient, there is no hidden path connecting two endogenous variables in \mathbf{V} , and all variable information is collected sufficiently to support the network. Actually, the ignorability assumption in POM (see Assumption 4) also implies the causal sufficiency, and this assumption is weaker than ignorability. But note that causal sufficiency is hard to satisfy in practice, because many real-world systems (e.g., the economic and climate systems) are complex; thus, we usually cannot collect all the information needed to describe them. Thus, some research studies focus on this issue. For example, the fast causal inference algorithm and its variants [61–66] are proposed to discover causality in the presence of hidden variables. And some other related studies [67–69] investigate causal discovery with soft intervention, which intervenes in the SCM without altering the network structure. Related studies [70–73] have also proposed approaches for identifying optimal and efficient adjustment sets, in which all variables are observable, minimal and valid, and or in which removing any variables would invalidate the adjustment.

4 Differences between POM and SCM

In this section, the fundamental differences between POM and SCM are discussed, focusing on three aspects. (i) Counterfactual falsifiability: POM assumes positivity; therefore, counterfactuals in POM are, in principle, supported by observations and can be empirically falsified. In contrast, counterfactuals in SCM may lie outside the support of the observed data and are thus unfalsifiable. (ii) Ignorability and the back-door criterion: POM requires the ignorability assumption to ensure unconfoundedness and the validity of counterfactual inference. In practice, satisfying ignorability typically requires identifying appropriate covariates. This is a complex, human-guided process. By contrast, SCM achieves unconfoundedness through causal sufficiency and back-door adjustment, which are equivalent to the ignorability assumption. SCM facilitates counterfactual inference through graphical language. (iii) Direct causation and indirect causation: POM struggles to process the indirect causation, and it typically relies on the Baron-Kenny model. In contrast, SCM has the advantage of adjusting for the front-door path (indirect causation) by equivalently applying the back-door criterion twice. We further leverage this advantage to design a forward counterfactual inference algorithm that autonomously infers spatio-temporal counterfactuals. The following are the details.

4.1 Counterfactual falsifiability

As shown in Figure 3, consider the following query, “Observed $X_1 = 0.5, X_2 = 1, X_3 = 1.5$, if had done $X_2 = 2$, what would $X_3 = ?$ ”. To answer this query, the positivity assumption requires the existence of at least two observational samples that share the same values $X_1 = 0.5$ but differ in X_2 . Specially, one with $X_2 = 1$, and another one with $X_2 = 2$ (see Assumption 3). And then, under the consistency assumption (see Assumption 2), we can infer counterfactuals by exactly matching the two samples. In contrast to SCM, we can infer the counterfactuals through Pearl’s causal ladder, even if we only have a single sample, as shown in Subsection 3.1. Thus, how to falsify the counterfactuals?

This is a challenging problem, as counterfactuals are, by definition, unobserved (see “?” in Table 1), and we need at least two samples for comparison. In practice, we cannot return to the past to apply a different treatment, nor can we perfectly replicate the same experiment at a different time. Thus, the inferred counterfactuals in SCM cannot be falsified without the assumptions of positivity and consistency, and this view is rejected ambiguously in Pearl’s book [7].

Another explanation is that, without positivity, but with consistency only in POM, counterfactuals can be extrapolated by regressive approaches, and they are falsifiable. For example, as shown in Figure 4, if there are T and Y only, and the observational data $(T, Y) = \{(0, 0.5), (1, 1), (2, 2.5), (3, 2)\}$ is collected, the linear regressive model can be fitted as $Y = 0.5T + 0.5$. We then query “if had done $T = 4$, what would $Y = ?$ ”. The counterfactual answer is $Y = 2.5$, and the regressive function can falsify the answer because it does not go beyond the observation. Thus, the consistency assumption actually claims that counterfactuals can be falsified by observations. This is also true for SCM, which uses regressive approaches to infer counterfactuals.

Moreover, there may be ambiguity regarding the preset Bayesian network (see Figure 2). To build an SCM, a Bayesian network must first be known. However, how to discover a probabilistic network from a single sample?

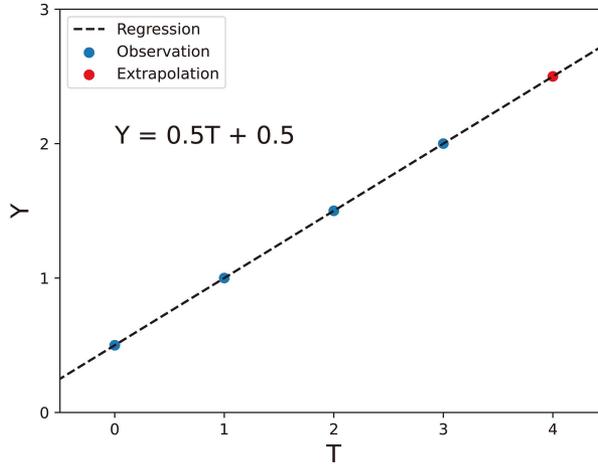


Figure 4 (Color online) Diagram of data extrapolation.

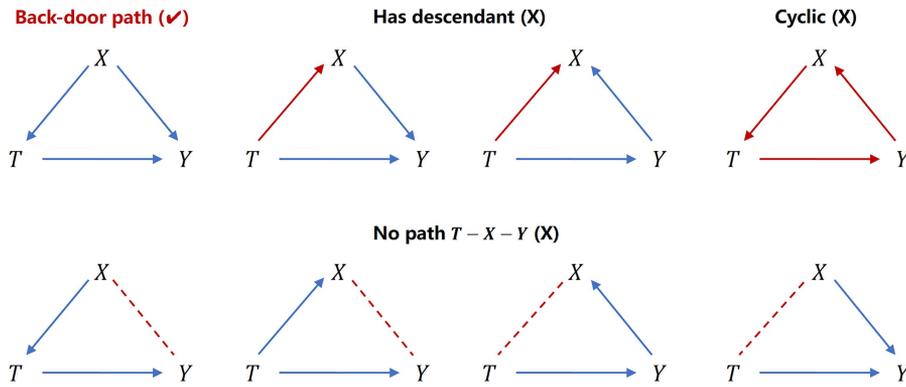


Figure 5 (Color online) Diagram of a back-door path.

This is usually impossible. A reasonable explanation is that the network structure can be inferred from the system’s mechanisms (e.g., physical mechanisms, communication connections) or from graphical knowledge learned from other domains, rather than from probability distribution. However, if it is accepted, the definition of causality could be broader and more ambiguous. This is why probabilistic and causal Bayesian networks are often indistinguishable, as stated in Pearl’s book [7].

Thus, SCM provides a way to think about counterfactuals, emphasizing that the thinking process is counterfactual, not the thinking result, because the result cannot always be falsified without the positivity and consistency assumptions. In contrast, POM emphasizes that the thinking result is counterfactual. This is a fundamental difference in the problem of counterfactual falsifiability.

4.2 Ignorability and back-door criterion

In the view of statistics, the ignorability assumption (see Assumption 4) is equivalent in calculation to the causal sufficiency assumption (see Assumption 8) with a back-door criterion. The formal definition of the back-door criterion can be found in these papers and books [4, 5, 17, 18]. Here we present an alternative formulation with respect to Table 1, defined as follows.

Definition 2 (Back-door criterion [4, 5, 17, 18]). In Table 1, the covariate X is said to satisfy the back-door criterion, if a Bayesian network is built on all the variables relative to X, Y, T , and there is an ordered pair $T \rightarrow Y$ that satisfies

- (1) no variable in X is a descendant of T ;
- (2) X blocks every path between T and Y that contains an arrow into T .

Thus, for X, Y, T , only one structure can satisfy the back-door criterion, as shown in Figure 5. And in the

back-door path, it is obtained that

$$\begin{aligned}
 & P(Y|\text{do}(T = 1)) \\
 &= \sum_x P(Y|\text{do}(T = 1), X = x)P(X = x|\text{do}(T = 1)) \\
 &= \sum_x P(Y|T = 1, X = x)P(X = x),
 \end{aligned} \tag{8}$$

which is the same in case of $\text{do}(T = 0)$ [4, 5, 17, 18]. Then, if causal sufficiency holds (i.e., no hidden variables confound X, Y and T), the expectation of the potential outcomes in Table 1 can be calculated as

$$\begin{aligned}
 & \mathbb{E}[Y|\text{do}(T = 1)] \\
 &= \sum_y y \times P(Y = y|\text{do}(T = 1)) \\
 &= \sum_y \sum_x y \times P(Y = y|T = 1, X = x)P(X = x) \\
 &= \sum_x P(X = x) \sum_y y \times P(Y = y|T = 1, X = x) \\
 &= \mathbb{E}[\mathbb{E}[Y|T = 1, X]].
 \end{aligned} \tag{9}$$

This is the same in POM, as follows:

$$\begin{aligned}
 & \mathbb{E}[Y(1)] \\
 &= \mathbb{E}[\mathbb{E}[Y(1)|X]] \quad /* \text{ Law of full expectation} */ \\
 &= \mathbb{E}[\mathbb{E}[Y(1)|T = 1, X]] \quad /* \text{ Ignorability} */ \\
 &= \mathbb{E}[\mathbb{E}[Y|T = 1, X]], \quad /* \text{ Consistency} */
 \end{aligned} \tag{10}$$

which is the same in case of $\text{do}(T = 0)$. Thus, with comparing (9) and (10), $\mathbb{E}[Y|\text{do}(T = 1)] = \mathbb{E}[Y(1)]$, and $\mathbb{E}[Y|\text{do}(T = 0)] = \mathbb{E}[Y(0)]$, in the two frameworks, but they start from different assumptions. Thus, the ignorability assumption can be decomposed as the causal sufficiency assumption and the back-door criterion, and the causal sufficiency is relatively trivial to satisfy. The ignorability is more likely to be a technical assumption that needs many controlled trials to deconfound, while the back-door path is clear to search in a Bayesian network. A potential, more difficult technical issue is accurately discovering the network structure.

4.3 Direct and indirect causation

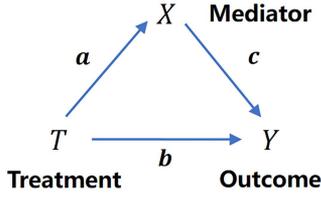
Suppose that there is a causal ordered pathway from one treatment variable to another outcome variable. Then, if there is at least one variable between the two endpoints, the pathway is called indirect causation; otherwise, it is called directed causation. Indirect causation, also called the mediating effect, has been discussed in depth in Rubin's papers [74–76]. Here, the Baron-Kenny model [19, 20] is introduced to present a mediating analysis for a potential outcome using a graph. As shown in Figure 6, $T \rightarrow Y$ is direct causation, and $T \rightarrow X \rightarrow Y$ is indirect causation.

To analyse indirect causation, the causal-steps approach [19] is proposed by building three linear regression models, as shown in Figure 6. Suppose that the regressive models are fitted sufficiently, the indirect causation would be detected if satisfy

- (1) coefficients a, b and c are significant;
- (2) $|b| < |b'|$.

Another approaches are to test the significance of $H_0 : b' - b = 0$ [77–79]. Or, to test the significance of $H_0 : ab = 0$ [80, 81]. If the correct regressive models are built, counterfactuals can be calculated, just as in SCM in Figure 2. Moreover, they are all regression-based approaches that are limited for use in large multivariate datasets, because one needs to test whether a covariate is a confounder or a mediator, and this process would be seriously time-consuming.

However, the test process can be faster with a Bayesian network in an SCM. Here, we propose an algorithm for forward counterfactual inference, as shown in Algorithm 1. To clearly demonstrate the principle of Algorithm 1, we first introduce the front-door criterion as follows.



$$\begin{aligned}
 X &= aT + U_X \\
 Y &= bT + cX + U_Y \\
 Y &= b'T + U_{total}
 \end{aligned}$$

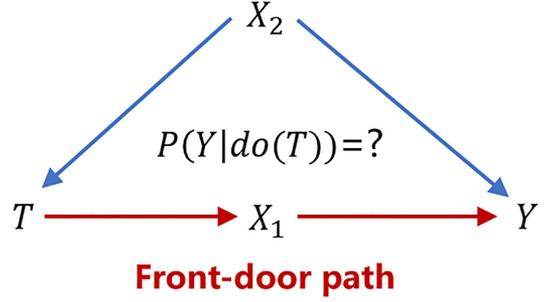


Figure 6 (Color online) Diagram of mediating analysis [19]. Here, a, b, c are coefficients and U_X, U_Y are biases. b' and U_{total} are the coefficient and bias, respectively, for the total effect.

Figure 7 (Color online) Diagram of front-door adjustment [5]. Here, T is a treatment variable and Y is a target variable. X_1 and X_2 are both covariates.

Algorithm 1 Forward counterfactual inference.

Input: A Bayesian network \mathcal{G} built on $\mathbf{V} = \{X_1, \dots, X_n\}$ with one target X_i and multiple sources X_{j_1}, \dots, X_{j_p} .

Output: Distribution $P(X_i | \text{do}(X_{j_1}), \dots, \text{do}(X_{j_p}))$.

- 1: Search one or multiple ordered causal pathways in \mathcal{G} from the sources X_{j_1}, \dots, X_{j_p} to the target X_i ;
- 2: **if** X_{j_1}, \dots, X_{j_p} are all the direct causation for X_i **then**
- 3: Search back-door paths (common causes) starting from X_{j_1}, \dots, X_{j_p} and X_i reversely in \mathcal{G} ;
- 4: **if** no back-door path within them **then**
- 5: **return** $P(X_i | X_{j_1}, \dots, X_{j_p})$;
- 6: **else**
- 7: For every back-door path, select a group of direct predecessors, $\mathbf{W} \subset \mathbf{V}$, and then, according to the back-door criterion (see Definition 2);
- 8: **return** $\sum_w P(X_i | X_{j_1}, \dots, X_{j_p}, W)P(W)$;
- 9: **end if**
- 10: **end if**
- 11: Otherwise, for q pathways starting from X_{j_p} , select q first mediators $X_{j_{p_1}}, \dots, X_{j_{p_q}}$. Then, a group of source-mediator pairs, e.g., $(X_{j_p}, X_{j_{p_q}})$, can be obtained;
- 12: Factorize $P(X_i | \text{do}(X_{j_1}), \dots, \text{do}(X_{j_p}))$ as

$$\begin{aligned}
 &P(X_i | \text{do}(X_{j_1}), \dots, \text{do}(X_{j_p})) \\
 &= \sum_{x_{j_{1:}j_{p_q}}} P(X_i | \text{do}(X_{j_1}), \dots, \text{do}(X_{j_{2_1}}), \dots, \text{do}(X_{j_{p_q}})) \\
 &\quad \times \prod_p \prod_q P(X_{j_{p_q}} | \text{do}(X), X \in \mathbf{Pa}(X_{j_{p_q}}) \subseteq \{X_{j_1:j_p}\}),
 \end{aligned} \tag{11}$$

where $\mathbf{Pa}(X_{j_{p_q}})$ is the causal parents of $X_{j_{p_q}}$;

- 13: Solve each distribution $P(\cdot | \text{do}(\cdot), \dots, \text{do}(\cdot))$ in (11) by Algorithm 1 recursively;
 - 14: **return** $P(X_i | \text{do}(X_{j_1}), \dots, \text{do}(X_{j_p}))$.
-

Definition 3 (Front-door criterion [4, 5, 17, 18]). The covariate X is said to satisfy the front-door criterion, if a Bayesian network is built on all the variables relative to X, Y, T , and there is an ordered pair $T \rightarrow Y$ that satisfies

- (1) X intercepts all directed paths from T to Y ;
- (2) there is no back-door path from T to X ;
- (3) all back-door paths from X to Y are blocked by T .

Then, a classic example is provided to show the front-door adjustment using Theorem 3.4.1 in [5], as shown in Figure 7 here. If given a treatment variable T and a target variable Y , with two covariates X_1 and X_2 . Then one can adjust the distribution $P(Y | \text{do}(T))$ with the front-door criterion as follows:

$$\begin{aligned}
 P(Y | \text{do}(T)) &= \sum_{x_1} P(X_1 | \text{do}(T)) P(Y | \text{do}(X_1)) \\
 &= \sum_{x_1} P(X_1 | T) \sum_{t'} P(Y | X_1, T') P(T').
 \end{aligned} \tag{12}$$

Thus, we can conclude from this equation:

- (1) If there is one or multiple ordered causal pathways from T to Y and there is at least one mediator on the way, then the causal pathway can be decomposed into two causal pathways through that mediator, e.g., $T \rightarrow X_1$ and $X_1 \rightarrow Y$ in (12);

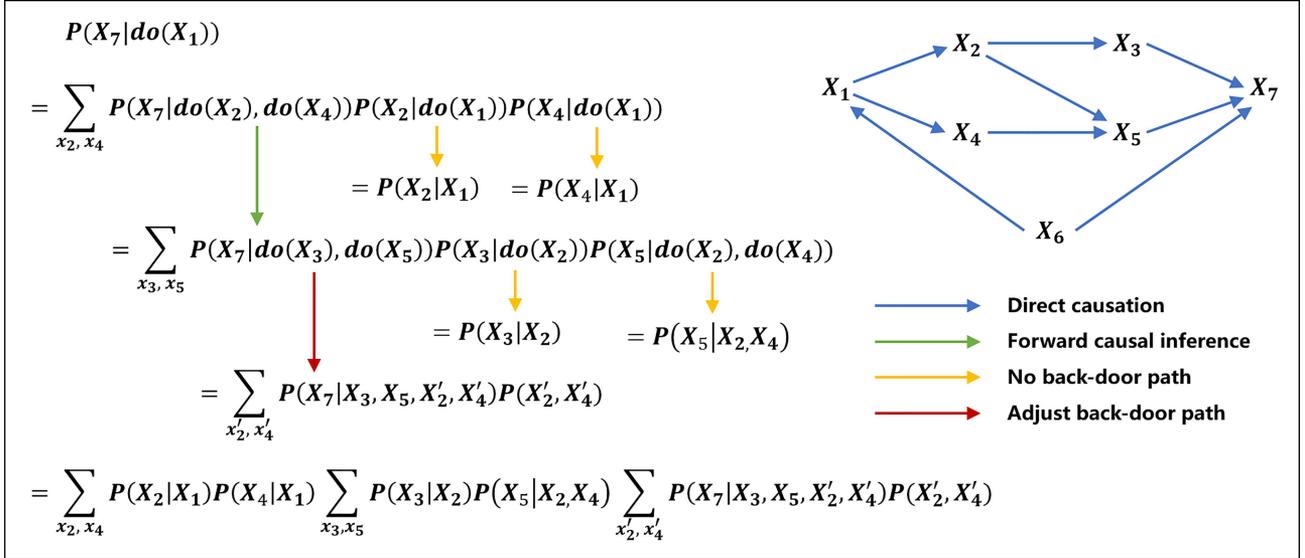


Figure 8 (Color online) Example of the forward counterfactual inference algorithm. Here, X_i is the same as X'_i to distinguish between the two recursions.

(2) If there is no mediator on all causal pathways, then adjustment using the back-door criterion. Of course, if there is no back-door path, no need to adjust.

Thus, the front-door principle is non-atomic and decomposable. With this more foundational principle in mind, we propose the forward counterfactual algorithm to autonomously apply it, as presented in Algorithm 1. The algorithm can be applied to an arbitrary set of variables $\{X_1, \dots, X_n\}$ that satisfy the causal sufficiency condition. We also provide an example to illustrate the inference process, as shown in Figure 8.

In Algorithm 1, the approaches to realize the codes in lines 1 and 3 can be various, e.g., some variants of the depth-first-search algorithm and breadth-first-search algorithm [82]. Thus, their time complexities are both $\mathcal{O}(|\mathbf{V}| + |\mathbf{E}|)$, if a Bayesian network $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ is given, where \mathbf{V} is the node set and \mathbf{E} is the edge set. Meanwhile, if the computational cost of factorization and adjustment are overlooked, the time complexity of the code in line 11 in Algorithm 1 is $\mathcal{O}(|\mathbf{V}|)$, because we only need to traverse all mediators in each pathway forward, and this process can be performed in parallel. Thus, the total time complexity of Algorithm 1 is $\mathcal{O}(|\mathbf{V}|^2 + |\mathbf{V}||\mathbf{E}|)$. Thus, if the network structure of the Bayesian network is large and can be discovered accurately, the efficiency of SCM to infer counterfactuals would be better with polynomial time complexity.

5 Spatio-temporal graphical counterfactuals

Then, for the examples of counterfactual questions at the beginning of this article, that is, if a different investment strategy had been implemented, would we obtain higher returns? And, in computer networks, what changes would occur in network load if a node's configuration had never been changed? Would an individual still purchase the product had they not been exposed to the advertisement? In these real-world scenarios, interactive behaviors with lagged causal effects exist widely. POM cannot answer these questions, because it does not allow the interactive behaviors between experimental units (see SUTVA in Assumption 1), even with some temporal quasi-experimental approaches (e.g., regression discontinuity design [83,84] and differences in differences [85,86]). This applies to SCM as well, because the foundational Bayesian network is a DAG that also does not allow mutual connections between network nodes. Thus, to answer these counterfactual queries for interactive units and lagged causal effects, we propose a concept of spatio-temporal graphical counterfactuals.

5.1 Spatio-temporal Bayesian networks

Before the SCM, various causal graphical models have been proposed to model the spatio-temporal causality. As shown in Figure 9, a temporal Bayesian network [59,87] is proposed to model the momentary causal dependency in a first-order Markov process. Further, dynamical Bayesian networks [59] are proposed to represent a combination of multiple temporal Bayesian networks. Full time graph [6] is proposed to summarize temporal Bayesian dependencies with complete timestamps. Full time graph allows the instantaneous causal effects between two units (or variables)

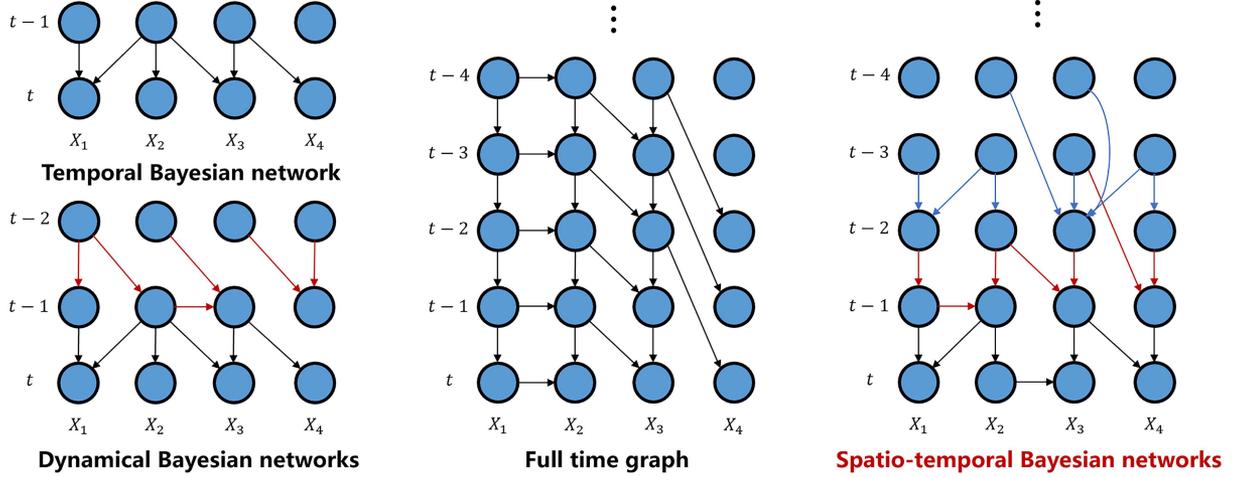


Figure 9 (Color online) Diagrams of various causal graphs.

Table 2 Comparisons of various causal graphs.

	Temporal Bayesian network	Dynamical Bayesian networks	Full time graph	Spatio-temporal Bayesian networks
High-order causation	×	×	✓	✓
Instantaneous causation	×	✓	✓	✓
Nonstationary causation	×	✓	×	✓

in a high-order Markov process, but it only supports strictly stationary causality, that is, causal dependencies that do not change over time. Moreover, the full time graph can be uniquely decomposed into multiple high-order temporal Bayesian networks, named UCN, if the instantaneous causal effects are not allowed [21]. The comparisons of these causal graphs are shown in Table 2.

Thus, based on the characteristics of these causal graphs, STBNs are defined as a group of DAGs indexed by different timestamps, as shown in Figure 9. In STBNs, nodes are not equivalent to variables. They represent the state of the variables at some time steps. Under the assumptions of causal Markov (see Assumption 6) and faithfulness (see Assumption 7), the joint distribution of all nodes can be factorized as

$$P(X_{1,t}, \dots, X_{n,t}, X_{1,t-1}, \dots, X_{n,t-T+1}) = \prod_{\tau=0}^{T-1} \prod_{i=1}^n P(X_{i,t-\tau} | \mathbf{Pa}(X_{i,t-\tau})), \quad (13)$$

where X_1, \dots, X_n are the variables, and the timestamps are in the range of $0 \sim T - 1$. Theoretically, STBNs allow an infinite number of time steps, but in practice, they are typically finite.

To guarantee the global and unique DAG in STBNs, a temporal assumption is introduced as follows.

Assumption 9 (Temporal assumption). In STBNs, the cause node precedes or occurs concurrently with the effect node.

This assumption does not allow the directed cause-effect pairs from now to the past, e.g., $X_{i,t-\tau_1} \rightarrow X_{j,t-\tau_2}, \tau_1 < \tau_2$. And this guarantees a functionally identifiable network structure without Markov equivalence classes [6, 21, 88]. Moreover, the causal sufficiency assumption (see Assumption 8) is also needed to guarantee the unconfoundedness.

Then, if STBNs are given, the spatial-temporal counterfactuals can be inferred through Algorithm 1, because the global and local STBNs are both DAGs. Compared with naive Bayesian networks, STBNs incorporate spatio-temporal concepts, and the inference process naturally proceeds forward in time. This enables momentary interventions at every time step, and all interventions work sequentially rather than simultaneously, as in Algorithm 1. If this property is considered thoroughly, the inference algorithm can be faster.

5.2 Nonstationarity of STBNs

Nonstationarity is a significant topic in research on time series. However, the reasons for it are likely to be different from the view of causality. Three of the most popular reasons are as follows.

(1) **Trending.** The time series has ascending or descending trending for long terms.

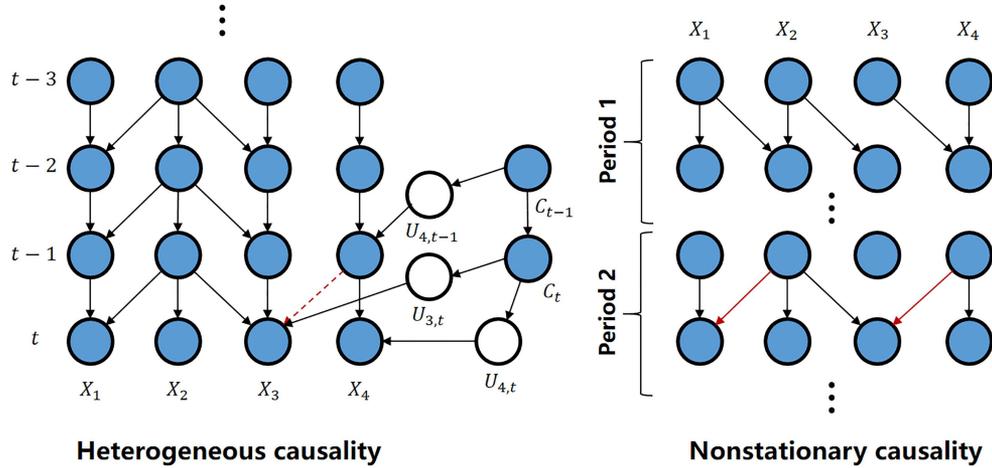


Figure 10 (Color online) Diagrams of nonstationary causality and heterogeneous causality. Here, $C_{t-\tau} = t - \tau, \tau = 0, \dots, T - 1$ are usually assumed. Thus, they are observable, instead of a hidden variable.

(2) **Time-varying variance.** The variances of the additive noise are time-varying, a phenomenon often referred to as heterogeneous causality in [89,90].

(3) **Nonstationary causality.** The structures of STBNs change over time, and at least once, as shown in Figure 10.

For trending time series, an effective approach is to calculate the differences of the time series during preprocessing. This approach obtains the differential time series, that is, the observations of the system dynamics, as in (14). This implies that the dynamics is stable over the period, and that the structures of STBNs do not change, that is, the differential time series are stationary. Thus, if the dynamics equations, e.g., (14), are not equal to zero, the causal links in STBNs are identifiable through the Jacobian matrix of dynamics [91–94].

In the case of time-varying variance, causal identifiability relies on the mutual independence of exogenous additive noises [88]. Specifically, if no causal connection exists between two additive noise nodes, the causation is identifiable. Thus, an effective approach is to relate the time-varying noises with a common timestamp, treating the timestamps at each time step as observable confounders $C_{t-\tau}, \tau = 0, \dots, T - 1$ [89,90,95,96]. As shown in Figure 10, there are three causal pathways, that is, $C_{t-1} \rightarrow U_{4,t-1} \rightarrow X_{4,t-1}$, $C_{t-1} \rightarrow C_t \rightarrow U_{3,t} \rightarrow X_{3,t}$, and $C_{t-1} \rightarrow C_t \rightarrow U_{4,t} \rightarrow X_{4,t}$. Thus, if C_t and C_{t-1} are hidden, the spurious connection between $X_{4,t-1}$ and $X_{3,t}$ would be detected, like the dotted line in Figure 10. However, C_t and C_{t-1} are usually known as the timestamp t and $t - 1$, thus, C_t and C_{t-1} block the fork paths between $X_{4,t-1}, X_{3,t}$ and $X_{4,t}$. Then, the noise variables $U_{4,t-1}, U_{3,t}$ and $U_{4,t}$ are mutually independent. Thus, causal identifiability is preserved. Note that, however, the premise to do so is that any hidden confounders associated with time-varying variance can be expressed as smooth or non-smooth functions of timestamps [90].

Nonstationary causality presents a significant challenge, particularly in scenarios involving frequently changing structures. To identify the nonstationary causality, one approach is to use sliding windows to identify causation within different time periods, as shown in Figure 10. For each sliding window, it is assumed that the causal structure is invariant, and then, a group of samples would be input to identification algorithms, e.g., [21,22,97–100], to identify an invariant structure during this period. However, if network changes are frequent, the samples used for identification over a short period would be few. In contrast, we usually need sufficient samples to assess the significance of causal connections. Thus, another reasonable approach is to first identify the change points, and then, determine the causality for the corresponding periods [101,102]. However, this still does not work in the extreme case of frequently changing, because the samples between two change points may be few. Moreover, there are other approaches that view nonstationary causality as a probabilistic normalized flow, and then use variational Bayesian inference to estimate it globally [103–105].

5.3 STBNs and complex networks

Here, we suppose STBNs model has been built, and it is associated with an interactive time-series process of multivariate (or multiple units). Then, we introduce a complex network, another network model, to examine the

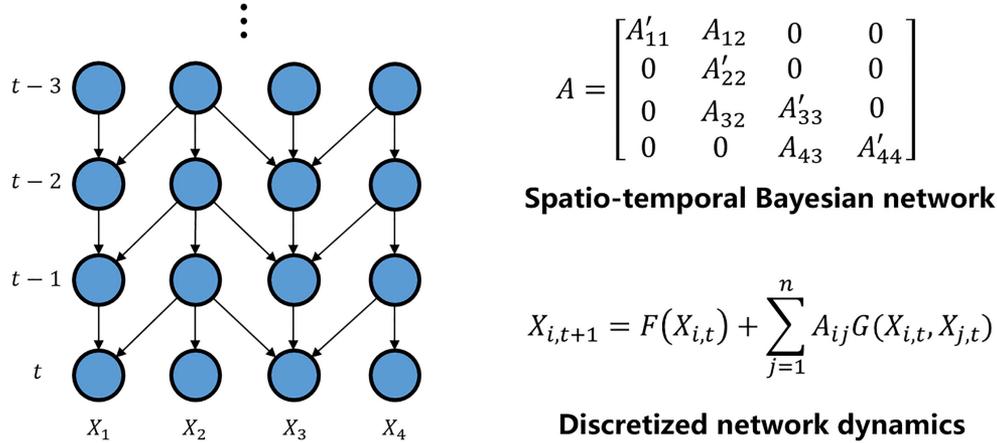


Figure 11 (Color online) Diagram of discretized network dynamics. Here, A is the adjacency matrix of the first-order STBNs, and $A'_{ii}, i = 1, \dots, n$ represents the joined causal effect of F and G .

causal interactions in a complex system. The network dynamics for variables X_1, \dots, X_n can be defined as

$$\frac{dX_{i,t}}{dt} = F(X_{i,t}) + \sum_{j=1}^n A_{ij} G(X_{i,t}, X_{j,t}), \quad (14)$$

where $F(\cdot)$ is the self-governed function of X_i , and $G(\cdot, \cdot)$ is the interactive function for node pairs X_i and X_j . Moreover, $A_{ij} \geq 0$ is the connection between X_i and X_j . Thus, if Eq. (14) is discretized, we can find it to be the same as first-order STBNs on stationary time series, as shown in Figure 11.

First-order STBNs exhibit advantageous properties when the underlying network dynamics process certain structural features. For example, determining how to affect a node in STBNs by “intervening” another node, the practical inference steps may not be long, due to the small-world property [106, 107] or the scale-free property [108] of complex networks. Moreover, counterfactual outcomes would not change because network systems have tolerance from external “interventions” [109, 110]. We would also know the synchronization [111–114], controllability [115–122], resilience [23, 123] of STBNs. We would also be able to sparsely identify the network structure even if the network size is large [24, 25], because the independence is equivalent to predictability in STBNs [21]. Also, the network size of STBNs can be reduced, because many complex networks can be simplified and still provide an insightful description of the causality of interest [124, 125]. This reduction supports the design of a counterfactual inference algorithm on smaller networks, thereby improving inference efficiency.

6 Conclusion

This work focuses on the spatio-temporal graphical counterfactuals and presents an overview of their theoretical foundations and application approaches. To discuss the theoretical basis, a survey is conducted, and the definition of counterfactuals is based on the concept of potential outcomes in the POM framework. We then introduce the SCM framework, which infers counterfactuals via graphical languages equivalent to those in POM. Further, to infer counterfactuals on intelligent machine autonomously, a forward counterfactual inference algorithm is designed in this work, and it can recursively solve the counterfactual probability distribution, $P(X_i | \text{do}(X_{j_1}), \dots, \text{do}(X_{j_p}))$, with polynomial time complexity if multi-nodes interventions are conducted on X_{j_1}, \dots, X_{j_p} . With this algorithm, spatio-temporal graphical counterfactuals can be inferred, given two elements: the network structure of STBNs and the algorithm identifying it. This work discusses structure identification algorithms under various nonstationary conditions and examines the feasibility of improving algorithm efficiency from the perspective of complex networks.

Acknowledgements This work was supported in part by National Key R&D Program of China (Grant No. 2022ZD0120004), Youth Scientist Project of the Ministry of Science and Technology of China (Grant No. 2025YFF0524100), National Natural Science Foundation of China (Grant Nos. 62233004, 62273090, 62073076, T2541017, 52525204, 52572354), Zhishan Youth Scholar Program of the Southeast University, Jiangsu Provincial Scientific Research Center of Applied Mathematics (Grant No. BK20233002), Basic Research Program of Jiangsu (Grant No. BK20253020), Open Research Project of the State Key Laboratory of Industrial Control Technology, China (Grant No. ICT2025B54), US National Science Foundation (Grant No. 2047488), and Rensselaer-IBM AI Research Collaboration.

References

- 1 Yu W W, Chen D X, Liu H Z, et al. Systems science in the new era: intelligent systems and big data. Sci China Inf Sci, 2024, 67: 136201

- 2 Ding D R, Han Q L, Ge X H, et al. Privacy-preserving filtering, control and optimization for industrial cyber-physical systems. *Sci China Inf Sci*, 2025, 68: 141201
- 3 Imbens G W, Rubin D B. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge: Cambridge University Press, 2015
- 4 Pearl J. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press, 2009
- 5 Pearl J, Glymour M, Jewell N P. *Causal Inference in Statistics: A Primer*. Hoboken: John Wiley & Sons, 2016
- 6 Peters J, Janzing D, Schölkopf B. *Elements of Causal Inference: Foundations and Learning Algorithms*. Cambridge: MIT Press, 2017
- 7 Pearl J, Mackenzie D. *The Book of Why: the New Science of Cause and Effect*. New York: Basic Books, 2018
- 8 Rubin D B. Inference and missing data. *Biometrika*, 1976, 63: 581–592
- 9 Rubin D B. Bayesian inference for causal effects: the role of randomization. *Ann Statist*, 1978, 6: 34–58
- 10 Rubin D B. Randomization analysis of experimental data: the Fisher randomization test comment. *J Am Statist Assoc*, 1980, 75: 591–593
- 11 Rosenbaum P R, Rubin D B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983, 70: 41–55
- 12 Pearl J. Bayesian networks: a model of self-activated memory for evidential reasoning. In: *Proceedings of the 7th Conference of the Cognitive Science Society*, University of California, 1985
- 13 Pearl J. Fusion, propagation, and structuring in belief networks. *Artif Intell*, 1986, 29: 241–288
- 14 Pearl J. Evidential reasoning using stochastic simulation of causal models. *Artif Intell*, 1987, 32: 245–257
- 15 Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann, 1988
- 16 Geiger D, Verma T, Pearl J. Identifying independence in Bayesian networks. *Networks*, 1990, 20: 507–534
- 17 Pearl J. Causal diagrams for empirical research. *Biometrika*, 1995, 82: 669–688
- 18 Pearl J. Causal inference in statistics: an overview. *Statist Surv*, 2009, 3: 96–146
- 19 Baron R M, Kenny D A. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Personality Soc Psychol*, 1986, 51: 1173–1182
- 20 Hayes A F. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. 3rd ed. New York: Guilford Publications, 2022
- 21 Kang M, Chen D, Meng N, et al. Identifying unique spatial-temporal Bayesian network without Markov equivalence. *IEEE Trans Artif Intell*, 2024, doi: 10.1109/TAI.2024.3483188
- 22 He Y, Kang M, Chen D, et al. NHCE: a neural high-order causal entropy algorithm for disentangling coupling dynamics. *IEEE Trans Netw Sci Eng*, 2024, 11: 5930–5942
- 23 Gao J, Barzel B, Barabási A L. Universal resilience patterns in complex networks. *Nature*, 2016, 530: 307–312
- 24 Gao T T, Yan G. Autonomous inference of complex network dynamics from incomplete and noisy data. *Nat Comput Sci*, 2022, 2: 160–168
- 25 Prasse B, van Mieghem P. Predicting network dynamics without requiring the knowledge of the interaction graph. *Proc Natl Acad Sci USA*, 2022, 119: e2205517119
- 26 Rubin D B. Causal inference using potential outcomes. *J Am Statist Assoc*, 2005, 100: 322–331
- 27 Althausen R P, Rubin D. The computerized construction of a matched sample. *Am J Sociology*, 1970, 76: 325–346
- 28 William G C, Rubin D B. Controlling bias in observational studies: a review. *Indian J Statist Ser A*, 1973, 35: 417–446
- 29 Rubin D B. Multivariate matching methods that are equal percent bias reducing, I: some examples. *Biometrics*, 1976, 32: 109–120
- 30 Stuart E A, Rubin D B. Matching with multiple control groups with adjustment for group differences. *J Educational Behav Stat*, 2008, 33: 279–306
- 31 Rubin D B, Thomas N. Matching using estimated propensity scores: relating theory to practice. *Biometrics*, 1996, 52: 249–264
- 32 Rubin D B. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med*, 1997, 127: 757–763
- 33 Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *J Economic Surv*, 2008, 22: 31–72
- 34 Pearson K. LIII. On lines and planes of closest fit to systems of points in space. *London Edinburgh Dublin Philos Mag J Sci*, 1901, 2: 559–572
- 35 Cox D R. The regression analysis of binary sequences. *J R Statist Soc Ser B-Statist Methodol*, 1958, 20: 215–232
- 36 Sims C A. Macroeconomics and reality. *Econometrica*, 1980, 48: 1–48
- 37 Blanchard O J, Quah D. The dynamic effects of aggregate demand and supply disturbances. *Am Economic Rev*, 1988, 79: 655–673
- 38 Kolda T G, Bader B W. Tensor decompositions and applications. *SIAM Rev*, 2009, 51: 455–500
- 39 Acar E, Dunlavy D M, Kolda T G, et al. Scalable tensor factorizations for incomplete data. *Chemometrics Intell Lab Syst*, 2011, 106: 41–56
- 40 Aggarwal C C. *Recommender Systems*. Berlin: Springer, 2016
- 41 Kingma D P, Welling M. Auto-encoding variational Bayes. 2013. ArXiv:1312.6114
- 42 Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: *Proceedings of Advances in Neural Information Processing Systems*, 2014
- 43 Rezende D, Mohamed S. Variational inference with normalizing flows. In: *Proceedings of International Conference on Machine Learning*, 2015. 1530–1538
- 44 Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. In: *Proceedings of Advances in Neural Information Processing Systems*, 2020. 6840–6851
- 45 Song J, Meng C, Ermon S. Denoising diffusion implicit models. In: *Proceedings of International Conference on Learning Representations*, 2021
- 46 Song Y, Sohl-Dickstein J, Kingma D P, et al. Score-based generative modeling through stochastic differential equations. In: *Proceedings of International Conference on Learning Representations*, 2021
- 47 Kang M, Zhu R, Chen D, et al. A cross-modal generative adversarial network for scenarios generation of renewable energy. *IEEE Trans Power Syst*, 2024, 39: 2630–2640
- 48 Rubin D B. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educational Psychol*, 1974, 66: 688–701
- 49 Pearl J. [Bayesian analysis in expert systems]: comment: graphical models, causality and intervention. *Statist Sci*, 1993, 8: 266–269
- 50 Glymour C, Zhang K, Spirtes P. Review of causal discovery methods based on graphical models. *Front Genet*, 2019, 10: 524
- 51 Scholkopf B, Locatello F, Bauer S, et al. Toward causal representation learning. *Proc IEEE*, 2021, 109: 612–634

- 52 Vowels M J, Camgoz N C, Bowden R. D'ya Like DAGs? A survey on structure learning and causal discovery. *ACM Comput Surv*, 2022, 55: 1–36
- 53 Assaad C K, Devijver E, Gaussier E. Survey and evaluation of causal discovery methods for time series. *J Artif Intell Res*, 2022, 73: 767–819
- 54 Hasan U, Hossain E, Gani M O. A survey on causal discovery methods for i.i.d. and time series data. 2023. ArXiv:2303.15027
- 55 Lacerda G, Spirtes P, Ramsey J, et al. Discovering cyclic causal models by independent components analysis. In: *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2008. 366–374
- 56 Spirtes P. Directed cyclic graphical representations of feedback models. In: *Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 1995. 491–498
- 57 Forré P, Mooij J M. Markov properties for graphical models with cycles and latent variables. 2017. ArXiv:1710.08775
- 58 Bongers S, Forré P, Peters J, et al. Foundations of structural causal models with cycles and latent variables. *Ann Statist*, 2021, 49: 2885–2915
- 59 Koller D, Friedman N. *Probabilistic Graphical Models: Principles and Techniques*. Cambridge: MIT Press, 2009
- 60 Barnard A, Page D. Causal structure learning via temporal markov networks. In: *Proceedings of the 9th International Conference on Probabilistic Graphical Models*, 2018. 13–24
- 61 Spirtes P, Meek C, Richardson T. Causal inference in the presence of latent variables and selection bias. In: *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence (UAI)*, 1995. 499–506
- 62 Spirtes P. An anytime algorithm for causal inference. In: *Proceedings of the 8th International Workshop on Artificial Intelligence and Statistics*, 2001. 278–285
- 63 Colombo D, Maathuis M H, Kalisch M, et al. Learning high-dimensional directed acyclic graphs with latent and selection variables. *Ann Statist*, 2012, 40: 294–321
- 64 Colombo D, Maathuis M H. Order-independent constraint-based causal structure learning. *J Mach Learn Res*, 2014, 15: 3921–3962
- 65 Ogarrio J M, Spirtes P, Ramsey J. A hybrid causal search algorithm for latent variable models. In: *Proceedings of the 8th International Conference on Probabilistic Graphical Models*, 2016. 368–379
- 66 Raghu V K, Ramsey J D, Morris A, et al. Comparison of strategies for scalable causal discovery of latent variable models from mixed data. *Int J Data Sci Anal*, 2018, 6: 33–45
- 67 Kocaoglu M, Jaber A, Shanmugam K, et al. Characterization and learning of causal graphs with latent variables from soft interventions. In: *Proceedings of Advances in Neural Information Processing Systems*, 2019
- 68 Jaber A, Kocaoglu M, Shanmugam K, et al. Causal discovery from soft interventions with unknown targets: characterization and learning. In: *Proceedings of Advances in Neural Information Processing Systems*, 2020. 9551–9561
- 69 Li A, Jaber A, Bareinboim E. Causal discovery from observational and interventional data across multiple environments. In: *Proceedings of Advances in Neural Information Processing Systems*, 2023. 16942–16956
- 70 Rotnitzky A, Smucler E. Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *J Mach Learn Res*, 2020, 21: 1–86
- 71 Runge J. Necessary and sufficient graphical conditions for optimal adjustment sets in causal graphical models with hidden variables. In: *Proceedings of Advances in Neural Information Processing Systems*, 2021. 15762–15773
- 72 Triantafillou S, Cooper G. Learning adjustment sets from observational and limited experimental data. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021. 9940–9948
- 73 Smucler E, Sapienza F, Rotnitzky A. Efficient adjustment sets in causal graphical models with hidden variables. *Biometrika*, 2022, 109: 49–65
- 74 Imbens G W, Rubin D B. Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann Statist*, 1997, 25: 305–327
- 75 Mealli F, Rubin D B. Assumptions allowing the estimation of direct causal effects. *J Econom*, 2003, 112: 79–87
- 76 Rubin D B. Direct and indirect causal effects via potential outcomes. *Scand J Stat*, 2004, 31: 161–170
- 77 MacKinnon D P, Dwyer J H. Estimating mediated effects in prevention studies. *Eval Rev*, 1993, 17: 144–158
- 78 MacKinnon D P, Lockwood C M, Williams J. Confidence limits for the indirect effect: distribution of the product and resampling methods. *Multivariate Behav Res*, 2004, 39: 99–128
- 79 MacKinnon D P, Fairchild A J, Fritz M S. Mediation analysis. *Annu Rev Psychol*, 2007, 58: 593–614
- 80 Sobel M E. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociol Methodol*, 1982, 13: 290–312
- 81 Sobel M E. Some new results on indirect effects and their standard errors in covariance structure models. *Sociol Methodol*, 1986, 16: 159–186
- 82 Cormen T H, Leiserson C E, Rivest R L, et al. *Introduction to Algorithms*. 4th ed. Cambridge: MIT Press, 2022
- 83 Hahn J, Todd P, Klaauw W. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 2001, 69: 201–209
- 84 McCrary J. Manipulation of the running variable in the regression discontinuity design: a density test. *J Econom*, 2008, 142: 698–714
- 85 Bertrand M, Duffo E, Mullainathan S. How much should we trust differences-in-differences estimates? *Q J Economics*, 2004, 119: 249–275
- 86 Lechner M. The estimation of causal effects by difference-in-difference methods. *Found Trends Econom*, 2011, 4: 165–224
- 87 Sun J, Taylor D, Boltt E M. Causal network inference by optimal causation entropy. *SIAM J Appl Dyn Syst*, 2015, 14: 73–106
- 88 Hoyer P, Janzing D, Mooij J M, et al. Nonlinear causal discovery with additive noise models. In: *Proceedings of Advances in Neural Information Processing Systems*, 2008
- 89 Zhang K, Huang B, Zhang J, et al. Causal discovery from nonstationary/heterogeneous data: skeleton estimation and orientation determination. In: *Proceedings of International Joint Conference on Artificial Intelligence*, 2017. 1347–1353
- 90 Huang B, Zhang K, Zhang J, et al. Causal discovery from heterogeneous/nonstationary data. *J Mach Learn Res*, 2020, 21: 1–53
- 91 Barzel B, Barabási A L. Universality in network dynamics. *Nat Phys*, 2013, 9: 673–681
- 92 Brouillard P, Lachapelle S, Lacoste A, et al. Differentiable causal discovery from interventional data. In: *Proceedings of Advances in Neural Information Processing Systems*, 2020. 21865–21877
- 93 He Y, Cui P, Shen Z, et al. DARING: differentiable causal discovery with residual independence. In: *Proceedings of the 27th ACM*

- SIGKDD Conference on Knowledge Discovery & Data Mining, 2021. 596–605
- 94 Zhang A, Liu F F, Ma W C, et al. Boosting causal discovery via adaptive sample reweighting. In: Proceedings of International Conference on Learning Representations, 2023
- 95 Huang B, Zhang K, Schölkopf B. Identification of time-dependent causal model: a Gaussian process treatment. In: Proceedings of International Joint Conference on Artificial Intelligence, 2015. 3561–3568
- 96 Huang B, Zhang K, Gong M, et al. Causal discovery and forecasting in nonstationary environments with state-space models. In: Proceedings of International Conference on Machine Learning, 2019. 2901–2910
- 97 He Y, Geng Z. Causal network learning from multiple interventions of unknown manipulated targets. 2016. ArXiv:1610.08611
- 98 Peters J, Bühlmann P, Meinshausen N. Causal inference by using invariant prediction: identification and confidence intervals. *J Royal Statist Soc Ser B*, 2016, 78: 947–1012
- 99 Pfister N, Bühlmann P, Peters J. Invariant causal prediction for sequential data. *J Am Statist Assoc*, 2018, 114: 1264–1276
- 100 Pamfil R, Sriwattanaworachai N, Desai S, et al. DYNOTEARS: structure learning from time-series data. In: Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics, 2020. 1595–1605
- 101 Adams R P, MacKay D J C. Bayesian online changepoint detection. 2007. ArXiv:0710.3742
- 102 Aminikhanghahi S, Cook D J. A survey of methods for time series change point detection. *Knowl Inf Syst*, 2017, 51: 339–367
- 103 Rezende D, Mohamed S. Variational inference with normalizing flows. In: Proceedings of International Conference on Machine Learning, 2015. 1530–1538
- 104 Papamakarios G, Nalisnick E, Rezende D J, et al. Normalizing flows for probabilistic modeling and inference. *J Mach Learn Res*, 2021, 22: 1–64
- 105 Javaloy A, Sanchez-Martin P, Valera I. Causal normalizing flows: from theory to practice. In: Proceedings of Advances in Neural Information Processing Systems, 2023. 58833–58864
- 106 Watts D J, Strogatz S H. Collective dynamics of ‘small-world’ networks. *Nature*, 1998, 393: 440–442
- 107 Newman M E J, Watts D J. Renormalization group analysis of the small-world network model. *Phys Lett A*, 1999, 263: 341–346
- 108 Barabási A L, Albert R. Emergence of scaling in random networks. *Science*, 1999, 286: 509–512
- 109 Albert R, Jeong H, Barabási A L. Error and attack tolerance of complex networks. *Nature*, 2000, 406: 378–382
- 110 Gao J, Buldyrev S V, Havlin S, et al. Robustness of a network of networks. *Phys Rev Lett*, 2011, 107: 195701
- 111 Boccaletti S, Kurths J, Osipov G, et al. The synchronization of chaotic systems. *Phys Rep*, 2002, 366: 1–101
- 112 Arenas A, Díaz-Guilera A, Kurths J, et al. Synchronization in complex networks. *Phys Rep*, 2008, 469: 93–153
- 113 Boccaletti S, Almindral J A, Guan S, et al. Explosive transitions in complex networks’ structure and dynamics: percolation and synchronization. *Phys Rep*, 2016, 660: 1–94
- 114 Ghosh D, Frasca M, Rizzo A, et al. The synchronized dynamics of time-varying networks. *Phys Rep*, 2022, 949: 1–63
- 115 Wang X F, Chen G. Pinning control of scale-free dynamical networks. *Physica A-Statist Mech its Appl*, 2002, 310: 521–531
- 116 Lu J H, Chen G R. A time-varying complex dynamical network model and its controlled synchronization criteria. *IEEE Trans Automat Contr*, 2005, 50: 841–846
- 117 Yu W, Chen G, Lü J. On pinning synchronization of complex dynamical networks. *Automatica*, 2009, 45: 429–435
- 118 Yu W, DeLellis P, Chen G, et al. Distributed adaptive control of synchronization in complex networks. *IEEE Trans Automat Contr*, 2012, 57: 2153–2158
- 119 Yu W, Chen G, Lü J, et al. Synchronization via pinning control on general complex networks. *SIAM J Control Optim*, 2013, 51: 1395–1416
- 120 Yu W, Ren W, Zheng W X, et al. Distributed control gains design for consensus in multi-agent systems with second-order nonlinear dynamics. *Automatica*, 2013, 49: 2107–2115
- 121 Liu Y Y, Slotine J J, Barabási A L. Controllability of complex networks. *Nature*, 2011, 473: 167–173
- 122 Li A, Cornelius S P, Liu Y Y, et al. The fundamental advantages of temporal networks. *Science*, 2017, 358: 1042–1046
- 123 Liu X, Li D, Ma M, et al. Network resilience. *Phys Rep*, 2022, 971: 1–108
- 124 Gao J. Intrinsic simplicity of complex systems. *Nat Phys*, 2024, 20: 184–185
- 125 Thibeault V, Allard A, Desrosiers P. The low-rank hypothesis of complex systems. *Nat Phys*, 2024, 20: 294–302