Supporting information for
Part mutual information for quantifying direct associations in networks

Juan Zhao[a,1], Yiwei Zhou[a,b,1], Xiujun Zhang[a], Luonan Chen[a,b,c,2]

[a]Key Laboratory of Systems Biology, Innovation Center for Cell Signaling Network, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, University of the Chinese Academy of Sciences, Shanghai 200031, China; [b]School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China; [c]Collaborative Research Center for Innovative Mathematical Modelling, Institute of Industrial Science, University of Tokyo, Tokyo, Japan.

[1]J.Z. and Y.Z. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: lnchen@sibs.ac.cn.

The supporting information includes the follow sections.

Section S1. Properties of part mutual information (PMI)

In this section, we derive the major features of part mutual information (PMI) and also its relations with mutual information (MI) and conditional mutual information (CMI). For random variables X, Y and Z, we assume that X and Y are 2 scalar variables and Z is an n-2 dimensional vector (n>2 is a positive integer), which actually represents a network with n nodes or n variables. All X, Y and Z are in an appropriate outcome space.

Generally, Kullback–Leibler (KL) divergence *D* is defined as

$$D(p\,||\,q) = \sum_x p(x)\log\frac{p(x)}{q(x)} \geq 0$$

between two probability distributions $p(x)$ and $q(x) \in [0,1]$,with $\sum_x p(x) = 1$ and $\sum_x q(x) = 1$, where the equality holds if and only if $p(x) = q(x)$ for all *x* in an outcome space Ω. Here, without the specific explanation, the base of logarithm log is 2. However, for certain cases, we want to analyze the difference between two probability distributions but with $\sum_x p(x) = 1$ and $\sum_x q(x) \leq 1$, i.e., $\sum_x q(x)$ is a partial distribution and may be less than 1. For such cases, we define an extended KL divergence.

Theorem S1 For $p(x)$ and $q(x) \in [0,1]$, with $\sum_x p(x) \geq \sum_x q(x)$ where $x \in \Omega$, the following relation holds

$$\sum_x p(x)\log\frac{p(x)}{q(x)} \geq 0 \quad .$$

The equality holds if and only if $p(x) = q(x)$ for all *x*, where clearly $\sum_x p(x) = \sum_x q(x)$. Note that $\sum_x q(x)\log\frac{q(x)}{p(x)} \geq 0$ generally does not hold for any $q(x)$ and $p(x)$.

Proof: For the natural logarithm of a real number *u*, if *u*>0, we have

$$\ln u \leq u - 1$$

where the equality holds for *u* = 1. Letting Ω denote the set of all *x* for which *p(x)* is non-zero, then we have

$$-\sum_x p(x)\ln\frac{q(x)}{p(x)} \ge -\sum_x p(x)(\frac{q(x)}{p(x)}-1)$$
$$= -\sum_x q(x) + \sum_x p(x) \ge 0$$
where the summation of x is on Ω. Noticing $0\ln 0 = 0$, we have

$$\sum_x p(x)\ln\frac{p(x)}{q(x)} \ge 0.$$

For the equality, clearly, $\frac{q(x)}{p(x)} = 1$ holds for all x in Ω due to the equality $\ln\frac{q(x)}{p(x)} = \frac{q(x)}{p(x)} - 1$.Thus, the equality holds if and only if *p(x)=q(x)* for all $x \in \Omega$, where $\sum_x p(x) = \sum_x q(x)$. The theoretical results hold for any logarithm due to $\log_a x = \frac{\ln x}{\ln a}$ for a positive base a. The theorem also holds provided that *p(x)* and *q(x)* are non-negative numbers, rather than $p(x)$ and $q(x) \in [0,1]$.

■

Clearly, the theorem holds if $\sum_x p(x) = 1$ and $\sum_x q(x) \le 1$ for $p(x)$ and $q(x) \in [0,1]$.

Definition S1 For $p(x)$ and $q(x) \in [0,1]$ with $\sum_x p(x) \ge \sum_x q(x)$ where the random variable *x* is in an outcome space Ω, we define extended KL-divergence from *p(x)* to *q(x)* as

$$D^E(p(x), q(x)) = \sum_x p(x)\log\frac{p(x)}{q(x)} \quad .$$

In this paper, we use the same symbol $D = D^E$ to represent both KL divergence and the extended KL divergence without confusion. From Theorem S1, clearly $D^E(p(x), q(x)) \ge 0$, and the equality holds if and only if *p(x)=q(x)* for all *x*, where $\sum_x q(x) = \sum_x p(x)$. In particular, $D^E(p(x), q(x)) \ge 0$ if $\sum_x p(x) = 1$ and $\sum_x q(x) \le 1$ with $p(x)$ and $q(x) \in [0,1]$.

In this paper, MI(X;Y) is the mutual information (MI) between X and Y. CMI(X;Y|Z) and PMI(X;Y|Z) are the conditional mutual information (CMI) and part mutual information (PMI) between X and Y given Z, respectively. *p(x,y,z)* is the joint probability distribution of X, Y and Z. *p(x)* and *p(x|z)* are the marginal probability distribution of X and conditional probability distribution of X given Z, respectively. Also we assume that X and Y are 2 scalar variables and Z is an n-2 dimensional vector (n>2 is a positive integer), which actually represents a network with n nodes or n variables. All X, Y and Z are in an appropriate outcome space.

We first define MI and CMI. For discrete variables X and Y, MI is calculated as follows

$$\mathrm{MI(X;Y)} = D(p(x,y)\,||\,p(x)p(y)) = \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

where $p(x,y)$ is the joint probability distribution of X and Y, and $p(x)$ and $p(y)$ are the marginal distributions of X and Y, respectively. D is KL divergence from *p(x,y)* to *p(x)p(y)*, where $\sum_{x,y} p(x)p(y) = 1$ and $\sum_{x,y} p(x,y) = 1$. Clearly, MI considers the 'mutual independence' of X and Y, which is defined as

$$p(x)p(y) = p(x,y).$$

When the above mutual independence holds, MI is zero. MI is non-negative, and clearly it equals zero only if the two variables are independent. The above expression is also equivalent to *p(x|y)=p(x) or p(y|x)=p(y).*

On the other hand, CMI for variables X and Y given Z is defined as

$$\mathrm{CMI(X;Y\,|\,Z)} = D(p(x,y,z)\,||\,p(x|z)p(y|z)p(z)) = \sum_{x,y,z} p(x,y,z)\log\frac{p(x,y|z)}{p(x|z)p(y|z)}$$

where $p(x,y|z)$ is the joint conditional probability distribution of *X* and *Y* with the condition *Z*, and *p(x|z)* and *p(y|z)* are conditional marginal probability distributions. Notice that $\sum_{x,y,z} p(x|z)p(y|z)p(z) = 1$ and $\sum_{x,y,z} p(x,y,z) = 1$. Clearly, in contrast to MI or the independence of X and Y, CMI considers the 'conditional independence' of X and Y given Z, which is defined as

$$p(x|z)p(y|z) = p(x,y|z).$$

The above expression is also equivalent to *p(x|y,z)=p(x|z) or p(y|x,z)=p(y|z).*

When the above conditional independence holds, CMI is zero. CMI is also non-negative.

Next, we define a new type of conditional independence of the random variables X and Y given Z, i.e., 'partial independence' of X and Y given Z.

Definition S2. Partial independence of X and Y given Z is defined as follows

$$p^*(x|z)p^*(y|z) = p(x,y|z) \qquad \text{[S1]}$$

where $p^*(x|z)$ and $p^*(y|z)$ are given (1) as

$$p^*(x|z) = \sum_y p(x|z,y)p(y)\,, \quad p^*(y|z) = \sum_x p(y|z,x)p(x)\,.$$

The important property for $p^*(x|z)$ and $p^*(y|z)$ is $p^*(x|z)=p(x|z)$ and $p^*(y|z)=p(y|z)$ if X and Y are independent given Z (i.e., $p(x|z,y)=p(x|z)$ or $p(y|z,x)=p(y|z)$). Also notice that $p^*(x|z)=p(x|z)$ if Y is independent of Z (i.e., $p(y)=p(y|z)$), and $p^*(y|z)=p(y|z)$ if X is independent of Z (i.e., $p(x)=p(x|z)$). These two properties are the key for the partial independence, and can be straightforward derived from the definition of $p^*(x|z)$ and $p^*(y|z)$. Clearly, $p^*(x|z)$ and $p^*(y|z)$ are average values of $p(x|z,y)$ over y and $p(y|z,x)$ over x, respectively. Then, PMI is defined based on this new partial independence Eq.S1.

Definition S3 The part mutual information (PMI) between variables X and Y given Z is defined as

$$\text{PMI(X;Y|Z)} = D(p(x,y,z) \,\|\, p^*(x|z)p^*(y|z)p(z)) \qquad \text{[S2]}$$

where $p(x,y,z)$ is the joint probability distribution of X, Y, and Z, and $D(p(x,y,z)||p^*(x|z)p^*(y|z)p(z))$ represents the extended KL-divergence from $p(x,y,z)$ to $p^*(x|z)p^*(y|z)p(z)$. Notice $\sum_{x,y,z} p(x,y,z) = 1$ but $\sum_{x,y,z} p^*(x|z)p^*(y|z)p(z) \le 1$.

From the definition, PMI can also be rewritten as

$$\begin{aligned}\text{PMI(X;Y|Z)} &= \sum_{x,y,z} p(x,y,z)\log\frac{p(x,y,z)}{p^*(x|z)p^*(y|z)p(z)} \\ &= \sum_{x,y,z} p(x,y,z)\log\frac{p(x,y|z)}{p^*(x|z)p^*(y|z)} \\ &= \sum_{x,y,z} p(x,y,z)\log\frac{p(x,y|z)}{\sum_y p(x|z,y)p(y)\sum_x p(y|z,x)p(x)}\end{aligned} \qquad \text{[S3]}$$

Next, we show that PMI is defined similar to CMI, but has different properties. The major properties are also summarized in Table S15.

Property S1. PMI between X and Y given Z can be decomposed as follows:

$$\begin{aligned}&\mathrm{PMI}(\mathrm{X};\mathrm{Y}\mid\mathrm{Z})=\mathrm{CMI}(\mathrm{X};\mathrm{Y}\mid\mathrm{Z})+D(p(x\mid z)\parallel\sum_{y}p(x\mid z,y)p(y))\\&+D(p(y\mid z)\parallel\sum_{x}p(y\mid z,x)p(\mathrm{x}))\end{aligned} \quad [S4]$$

where D is the extended KL-divergence. In other words,

$$\mathrm{PMI(X;Y|Z)}=\mathrm{CMI(X;Y|Z)}+D(p(x\mid z)\parallel p^{*}(x\mid z))+D(p(y\mid z)\parallel p^{*}(y\mid z)).$$

Proof: Similar to the derivation in (1), based on the definition of PMI, we have

$$\begin{aligned}\mathrm{PMI}(\mathrm{X};\mathrm{Y}\mid\mathrm{Z})&=\sum_{x,y,z}p(x,y,z)\log\frac{p(x,y\mid z)}{\sum_{y}p(x\mid z,y)p(y)\sum_{x}p(y\mid z,x)p(x)}\\&=\sum_{x,y,z}p(x,y,z)\log\frac{p(x,y\mid z)p(x\mid z)p(y\mid z)}{p(x\mid z)p(y\mid z)\sum_{y}p(x\mid z,y)p(y)\sum_{x}p(y\mid z,x)p(x)}\\&=\sum_{x,y,z}p(x,y,z)(\log\frac{p(x,y\mid z)}{p(x\mid z)p(y\mid z)}+\log\frac{p(x\mid z)}{\sum_{y}p(x\mid z,y)p(y)}+\log\frac{p(y\mid z)}{\sum_{x}p(y\mid z,x)p(x)})\\&=\mathrm{CMI}(\mathrm{X};\mathrm{Y}\mid\mathrm{Z})+D(p(x\mid z)\parallel\sum_{y}p(x\mid z,y)p(y))+D(p(y\mid z)\parallel\sum_{x}p(y\mid z,x)p(\mathrm{x}))\\&=\mathrm{CMI}(\mathrm{X};\mathrm{Y}\mid\mathrm{Z})+D(p(x\mid z)\parallel p^{*}(x\mid z))+D(p(y\mid z)\parallel p^{*}(y\mid z))\end{aligned}$$

■

Clearly, conditional mutual information CMI(X;Y|Z) is the first term of PMI as shown in Eq.S4. Intuitively, the second and third terms of Eq,S4 implicitly include the association between X and Y. In other words, PMI(X;Y|Z) is the summation with CMI(X;Y|Z), $D(p(x\mid z)\parallel p^{*}(x\mid z))$ and $D(p(y\mid z)\parallel p^{*}(y\mid z))$ which is actually the association of X and Y. Generally, CMI(X;Y|Z) underestimates of the association between X and Y, but with the additional association $D(p(x\mid z)\parallel\sum_{y}p(x\mid z,y)p(y))=D(p(x\mid z)\parallel p^{*}(x\mid z))$ and $D(p(y\mid z)\parallel\sum_{x}p(y\mid z,x)p(x))=D(p(y\mid z)\parallel p^{*}(y\mid z))$ between X and Y, PMI(X;Y|Z) can quantify the direct association between X and Y in a proper manner.

Property S2. PMI(X;Y|Z) $\geq$ CMI(X;Y|Z) $\geq$ 0 [S5]

Proof: from Property S1, due to non-negative KL divergence, it is obvious that PMI is larger than or equal to CMI, and CMI is also non-negative.

■

Property S2 shows that PMI is generally larger than CMI, and thus has the potential to overcome the underestimation problem of CMI.

Property S3. PMI is symmetric, i.e., PMI(X;Y|Z) = PMI(Y;X|Z) [S6]

Proof: from the definition of PMI, obviously PMI is symmetric.

■

Property S3 shows that PMI is symmetric as the same as MI or CMI.

Property S4. If X and Y are conditionally independent given Z, then PMI(X;Y|Z) =CMI(X;Y|Z) =0.

Proof: from the definition of CMI, it can easily show that CMI(X;Y|Z) =0 if X and Y are conditionally independent given Z, i.e., $p(x,y|z)=p(x|z)p(y|z)$.
For PMI, when X and Y are conditionally independent given Z, we have $p(x|z,y)=p(x|z)$, $p(y|z,x)=p(y|z)$ and $p(x,y|z)=p(x|z)p(y|z)$. Hence

$$\sum_{y} p(x\mid z,y)\,p(y) = \sum_{y} p(x\mid z)\,p(y) = p(x\mid z)\,.$$

In a similar way, we have $\sum_{x} p(y\mid z,x)\,p(x) = p(y\mid z)$ when X and Y are independent given Z. Thus, $p^*(x|z)p^*(y|z)=p(x|z)p(y|z)=p(x,y|z)$, i.e., the partial independence also holds. Hence we have,

$$\begin{aligned}
\mathrm{PMI(X;Y\mid Z)} &= \sum_{x,y,z} p(x,y,z)\log\frac{p(x,y\mid z)}{\sum_{y} p(x\mid z,y)p(y)\sum_{x} p(y\mid z,x)p(x)} \\
&= \sum_{x,y,z} p(x,y,z)\log\frac{p(x\mid z)\,p(y\mid z)}{p(x\mid z)\,\mathrm{p}(y\mid z)} \\
&= \sum_{x,y,z} p(x,y,z)\log 1 \\
&= 0
\end{aligned}$$

■

From Property S4, PMI(X;Y|Z) is minimal, i.e., zero, when X and Y are independent given Z. Thus, PMI can measure the independency of two variables. Property S4 also indicates that both the conditional independence and the partial independence can measure the independence of X and Y given Z.

Property S5. If Z is independent of X and Y, then

PMI(X;Y|Z) = CMI(X;Y|Z) = MI(X;Y).

Proof: for the case when Z is independent of X and Y, we have $p(x,y|z)=p(x,y)$.

Furthermore,

$$p^*(x|z)=\sum_y p(x|z,y)p(y)=\sum_y \frac{p(x,y,z)}{p(z,y)}p(y)=\sum_y \frac{p(x,y)p(z)}{p(z)p(y)}p(y)=\sum_y p(x,y)=p(x)$$

$$p^*(y|z)=\sum_x p(y|z,x)p(x)=\sum_x \frac{p(x,y,z)}{p(z,x)}p(x)=\sum_x \frac{p(x,y)p(z)}{p(z)p(x)}p(x)=\sum_x p(x,y)=p(y)\cdot$$

Thus, $p^*(x|z)=p(x)=p(x|z)$, *and* $p^*(y|z)=p(y)=p(y|z)$. Then, from the above derivation, we have

$$\mathrm{PMI}(\mathrm{X};\mathrm{Y}|\mathrm{Z})=\sum_{x,y,z} p(x,y,z)\log\frac{p(x,y|z)}{p(x|z)p(y|z)}$$
$$=\sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}=\mathrm{MI}(\mathrm{X};\mathrm{Y}|\mathrm{Z})=\mathrm{CMI}(\mathrm{X};\mathrm{Y}|\mathrm{Z})$$

■

Property S5 shows that PMI keeps the same features as CMI(X;Y|Z) and MI(X;Y) when Z is independent of X and Y, which implies that PMI can measure the dependency between X and Y.

If X and Y are strongly associated, we denote the relation as X≈Y, which implies the strongly mutual dependency between X and Y.

Property S6. If Z is strongly associated with X (or Y), i.e., Z ≈ X (or Z ≈ Y), then CMI(X;Y|Z) is equal to zero regardless of the dependence of X and Y given Z, but PMI(X;Y|Z) is generally not zero, and depends on the direct association between X and Y given Z. Specifically, if Z ≈ X or Z ≈ Y, we have $\mathrm{PMI}(\mathrm{X};\mathrm{Y}|\mathrm{Z})=D(p(x|z)\|\sum_y p(x|z,y)p(y))+D(p(y|z)\|\sum_x p(y|z,x)p(x))$, which are greater than or equal to zero.

Proof: Z is strongly associated with X, i.e., $p(x|z)=1$ provided that $p(x,z)\neq 0$, and $p(x|y,z)=1$ provided that $p(x,y,z)\neq 0$. Thus, the summation of *x, y, z* in CMI for those terms with $p(x,y,z)\neq 0$ (note $p(x,y,z)\neq 0$ implies $p(x,z)\neq 0$) is

$$\mathrm{CMI}(\mathrm{X};\mathrm{Y}|\mathrm{Z})=\sum_{x,y,z} p(x,y,z)\log\frac{p(x,y|z)}{p(x|z)p(y|z)}$$
$$=\sum_{x,y,z} p(x,y,z)\log\frac{\frac{p(x,y,z)}{p(z)}}{p(x|z)\frac{p(y,z)}{p(z)}}=\sum_{x,y,z} p(x,y,z)\log\frac{p(x|y,z)}{p(x|z)}$$
$$=\sum_{x,y,z} p(x,y,z)\log\frac{1}{1}=\sum_{x,y,z} p(x,y,z)\log 1=0$$

Noticing that $0\log 0=0$, the remaining terms on *x,y,z* in CMI with $p(x,z)=0$ and $p(x,y,z)=0$ are also equal to zero in such cases. Hence, CMI(X;Y|Z)=0 for the summation of all terms of *x,y,z*.

In the same way, we can also show that CMI(X;Y|Z)=0 when Z is strongly associated with Y, regardless of the direct dependence of X and Y given Z.

On the other hand, when Z is strongly associated with X, we have $p(x|z,y)= 1$ provided that $p(x,y,z) \neq 0$, and $p(x|z)=1$ provided that $p(x,z) \neq 0$. Thus, in a similar way as the above proof of CMI, we can show

$$D(p(x|z) \| p^*(x|z) \geq 0; D(p(y|z) \| p^*(y|z)) \geq 0 .$$

Hence from Property S1 and also due to CMI(X;Y|Z)=0, we have

$$\text{PMI(X;Y|Z)} = D(p(x|z) \| p^*(x|z) + D(p(y|z) \| p^*(y|z)) ,$$

which is generally nonzero, see Property S4). Actually if X and Y are independent given Z, we have $p^*(x|z)=p(x|z)$ and $p^*(y|z)=p(y|z)$ due to $p(x|y,z)=p(x|z)$ and $p(y|x,z)=p(y|z)$. Thus, PMI(X;Y|Z) = 0 if X and Y are independent given Z.

In the same way, we can show that when Z is strongly associated with Y, i.e., $p(y|z,x)= 1$ and $p(y|z)=1$ provided that $p(y,z) \neq 0$ and $p(x,y,z) \neq 0$, PMI becomes

$$\text{PMI(X;Y|Z)} = D(p(x|z) \| p^*(x|z) + D(p(y|z) \| p^*(y|z)) ,$$

which is generally nonzero (it is zero if X and Y are independent given Z, see Property S4).

■

Clearly, when Z is strongly associated with X (or Y), i.e., Z ≈ X (or Z ≈ Y), CMI(X;Y|Z) only measures the information correlated with Z, and the information independent of Z is almost ignored. Thus, knowing Z leaves no uncertainty about X, which results in CMI(X;Y|Z)=0 regardless of the dependence of X and Y given Z. However, the last two terms in PMI(X;Y|Z) can measure that information, which is not correlated with Z and thus makes the accurate evaluation on the direct association of X and Y given Z.

Property S6 indicates that PMI can overcome the underestimation problem of CMI owing to the new partial independence Eq.S1.

Next, we explicitly analyze the conditional independence and partial independence. Based on the analysis of Property S4, we have the following property.

Property S7. If X and Y are conditionally independent given Z, then both the conditional independence and the partial independence hold, i.e., $p(x|z)p(y|z)=p(x,y|z)$ and $p^*(x|z)p^*(y|z)=p(x,y|z)$.

Property S7 can be proven based on the proof of Property S4. Property S7 shows that both conditional independence and partial independence can measure the independence of X and Y given Z. In other words, both CMI and PMI can give the correct results on the independence of X and Y given Z. However, when X (or Y) is strongly associated with Z, in next property, we will show that conditional independence of X and Y given Z always approximately holds wrongly regardless of the dependence of X and Y given Z, but partial independence correctly depends on the association of X and Y given Z for such cases. In other words, PMI can give the correct results on the dependence of X and Y given Z due to its partial independence, but CMI generally cannot give the correct results on the dependence of X and Y given Z due to its conditional independence.

Property S8. Provided that X (or/and Y) and Z are strongly associated, the conditional independence always approximately holds, i.e., $p(x|z)p(y|z)=p(x,y|z)$, even if X and Y are not conditionally independent given Z (i.e., even if X and Y are conditionally dependent given Z), but the partial independence does not necessarily hold for this case, i.e., $p^*(x|z)p^*(y|z) \neq p(x,y|z)$ if X and Y are not independent given Z.

Proof: Assume that X (or Y) and Z are strongly associated. We first analyze the conditional independence.
For the case of $p(x,y,z)\neq 0$, we have $p(x|z) = 1$ and $p(x,y|z) = p(y|z)$. Then,

$$p(x|z)p(y|z)=p(y|z)= p(x,y|z).$$

For the case of $p(x,y,z)=0$, we have $p(x,y|z) = p(x|z)=0$. Then,

$$p(x|z)p(y|z)= p(x,y|z)=0.$$

Thus, provided that X (or Y) and Z are strongly associated, the conditional independence $p(x|z)p(y|z)=p(x,y|z)$ always approximately holds, even if X and Y are not conditionally independent given Z (i.e., even if X and Y are conditionally dependent given Z). Therefore, CMI based on this conditional independence cannot detect the dependence of X and Y given Z for such cases.

On the other hand, when analyzing partial independence, we have $p(y|z,x) = p(y|z)$, $p(x|z,y)=1$ and $p(x,y|z) = p(y|z)$ for the case of $p(x,y,z)\neq 0$. Then, from the summation of $p^*(x|z)$ and $p^*(y|z)$ on y and x, we have

$$p^*(x|z)\leq 1,\ p^*(y|z)\leq p(y|z), \text{ and } p(x,y|z)=p(y|z).$$

Thus,

$$p^*(x|z)p^*(y|z) \leq p(x,y|z).$$

Hence, even if X (or Y) and Z are strongly associated, the partial independence $p^*(x|z)p^*(y|z)=p(x,y|z)$ does not necessarily hold, provided that X and Y are not conditionally independent given Z. Therefore, PMI based on this partial independence can detect the dependence of X and Y given Z for such cases.

We can also prove this property holds if both X and Y are strongly associated with Z.

■

Property S8 indicates that CMI(X;Y|Z) gives the wrong results on the dependency of X and Y given Z for the case when X (or Y) and Z are strongly dependent, but PMI(X;Y|Z) is able to detect their true association. The conceptual illustration of the conditional independence and the partial independence is given in Fig.S1.

Intuitively, $p(x|z)$ is the conditional probability of X given Z, which removes the information of Z, and thus, if X is strongly associated with Z, $p(x|z)$ approaches 1 for the case of $p(x,z) \neq 0$.

On the other hand, recently, causal strength for quantifying causal influence or strength from X to Y was proposed (1), and is defined as

$$C_{X\to Y}(X;Y|Z) = D(p(x,y,z)\,\|\,p_{X\to Y}(x,y,z)) = \sum_{x,y,z} p(x,y,z)\log\frac{p(y|z,x)}{\sum_x p(y|z,x)p(x)}$$

where $p_{X\to Y}(x,y,z) = p(x,z)\sum_x p(y|z,x)p(x)$ is called as the interventional probability distribution with cutting the link $X \to Y$ in a Direct Acyclic Graph (DAG). Notice $\sum_{x,y,z} p_{X\to Y}(x,y,z) \leq 1$. We also inferred the relationship between causal strength and PMI as below.

Property S9. If Z is independent of X, then PMI is equal to causal strength from Y to X, i.e.

$$\text{PMI}(X;Y|Z) = C_{Y\to X}(X;Y|Z). \qquad [S7]$$

If Z is independent of Y, then PMI is equal to causal strength from X to Y, i.e.

$$\text{PMI}(X;Y|Z) = C_{X\to Y}(X;Y|Z). \qquad [S8]$$

Proof: When Z is independent of X, then $p(z,x)=p(z)p(x), p(x|z)=p(x)$. Then we have,

$$\sum_x p(y|z,x)p(x) = \sum_x \frac{p(y,z,x)p(x)}{p(z,x)} = \sum_x \frac{p(y,z,x)p(x)}{p(z)p(x)}$$
$$= \sum_x \frac{p(y,z,x)}{p(z)} = \frac{p(y,z)}{p(z)} = p(y|z)$$

Hence, from definition of PMI, we have

$$\mathrm{PMI}(X;Y\,|\,Z)=\sum_{x,y,z}p(x,y,z)\log\frac{p(x,y\,|\,z)}{\sum_y p(x\,|\,z,y)p(y)\sum_x p(y\,|\,z,x)p(x)}$$

$$=\sum_{x,y,z}p(x,y,z)\log\frac{p(x,y\,|\,z)p(x\,|\,z)p(y\,|\,z)}{p(x\,|\,z)p(y\,|\,z)\sum_y p(x\,|\,z,y)p(y)\sum_x p(y\,|\,z,x)p(x)}$$

$$=\sum_{x,y,z}p(x,y,z)(\log\frac{p(x,y\,|\,z)}{p(x\,|\,z)p(y\,|\,z)}+\log\frac{p(x\,|\,z)}{\sum_y p(x\,|\,z,y)p(y)}+\log\frac{p(y\,|\,z)}{\sum_x p(y\,|\,z,x)p(x)})$$

$$==\sum_{x,y,z}p(x,y,z)(\log\frac{p(x,y\,|\,z)}{p(x\,|\,z)p(y\,|\,z)}+\log\frac{p(x\,|\,z)}{\sum_y p(x\,|\,z,y)p(y)}+\log\frac{p(y\,|\,z)}{p(y\,|\,z)})$$

$$=\sum_{x,y,z}p(x,y,z)\log\frac{p(x,y\,|\,z)}{p(x\,|\,z)p(y\,|\,z)}+\log\frac{p(x\,|\,z)}{\sum_y p(x\,|\,z,y)p(y)}$$

$$=\mathrm{CMI}(X;Y\,|\,Z)+D(p(x\,|\,z)\,\|\sum_y p(x\,|\,z,y)p(y))=C_{Y\to X}(X;Y\,|\,Z)$$

In the same way, when Z is independent of Y, we can prove

PMI(X;Y|Z) = CMI(X;Y|Z)+ $D(p(y\,|\,z)\,\|\sum_x p(y\,|\,z,x)p(x))=C_{X\to Y}(X;Y\,|\,Z)$

■

Eq.S7 and Eq.S8 shows that PMI(X;Y|Z) is equivalent to causal strength(1) when Z is independent of X or Y.

As shown in Eq.S4, PMI has three terms. The first term is CMI which is zero and underestimates the direct dependency if X or Y is strongly associated with Z, i.e., X≈Z or Y≈Z, because knowing Z leaves almost no uncertainty about X or Y. In other words, strong dependency between X and Z (or between Y and Z) makes the dependency between X and Y almost invisible when measuring CMI(X;Y|Z) only (1). PMI is able to overcome all those problems as shown in Properties S1-S9 meanwhile keeping the features of MI and CMI on measuring the direct association of X and Y.

The partial distance correlation (Pdcor) is a power measure for direct associations (5). We also compare the partial distance correlation (Pdcor) defined in (5) for the scaling of the variables on a multi-variable system.

Property S10. CMI(X;Y|Z_1,Z_2)=CMI(X;Y| α Z_1, β Z_2) and PMI(X;Y|Z_1,Z_2)= PMI(X;Y|α Z_1, β Z_2) when $\alpha\neq\beta\neq 0$, but generally Pdcor does not satisfy this property, i.e., Pdcor(X;Y| Z_1,Z_2)$\neq$Pdcor(X;Y|α Z_1, β Z_2), which implies that Pdcor may give the different values depending on the scaling of variables (or units of variables). Here α and β are nonzero (real) numbers.

We give a simple proof that value of PMI will not be changed by different scales of the conditional variables, i.e.,

$$\mathrm{PMI}(X;Y\,|\,\alpha Z_1,\beta Z_2)=\sum_{x,y,z}p(x,y,\alpha z_1,\beta z_2)\log\frac{p(x,y\,|\,\alpha z_1,\beta z_2)}{\sum_y p(x\,|\,\alpha z_1,\beta z_2,y)p(y)\sum_x p(y\,|\,\alpha z_1,\beta z_2,x)p(x)}$$

$$=\sum_{x,y,z}p(x,y,z_1,z_2)\log\frac{p(x,y\,|\,z_1,z_2)}{\sum_y p(x\,|\,z_1,z_2,y)p(y)\sum_x p(y\,|\,z_1,z_2,x)p(x)}$$

$$=\mathrm{PMI}(X;Y\,|\,Z_1,Z_2)$$

where both α and β are nonzero. For CMI, we can have similar proof. It can also easily show that Pdcor may give the different values depending on the scales of the conditioned variables (or units of variables) (see the reference (4,5)).

■

This property implies that Pdcor may give the different values depending on the scaling of variables (or units of variables). In addition, the partial distance correlation (Pdcor) suffers from the false positive problem, i.e., even when two variables X and Y are conditionally independent given variable Z, Pdcor(X;Y|Z) may be non-zero (see the reference (4,5)). In contrast, PMI and CMI can correctly measure such associations. The major properties are summarized in Table S15.

Section S2. Part Mutual Information (PMI) with Gaussian distribution

With the Gaussian assumption of the distribution for (X,Y,Z), PMI has a simple form as indicated in Theorem 1 in the main text.

Theorem 1 Assume that X and Y are two one-dimensional variables, and Z is an n-2 dimensional (n-2 > 0) vector, where n is the dimension of vector (X,Y,Z). Set $x \in$ X, $y \in$ Y, and $z \in$ Z. Letting the vector (X,Y,Z) follow multivariate Gaussian distribution, then PMI between X and Y given Z is simplified as follows

$$\mathrm{PMI}(X;Y\,|\,Z)=\frac{1}{2}(\mathrm{tr}(C^{-1}\Sigma)+\ln C_0-n),$$

where $C^{-1}=\begin{pmatrix} C_{xx} & C_{xy} & C_{xz} \\ C_{xy} & C_{yy} & C_{yz} \\ C_{xz} & C_{yz} & C_{zz} \end{pmatrix}$,

$$C_{xx}=-(\Sigma^{-1})_{xy}((\Sigma^{-1})_{yy}-(\Sigma_3^{-1})_{yy}+\sigma_{yy}^{-1})^{-1}(\Sigma^{-1})_{xy}+(\Sigma^{-1})_{xx},$$

$$C_{xy}=0,$$

$$C_{xz} = -(\Sigma^{-1})_{xy}((\Sigma^{-1})_{yy} - (\Sigma_3^{-1})_{yy} + \sigma_{yy}^{-1})^{-1}((\Sigma^{-1})_{yz} - (\Sigma_3^{-1})_{yz}) + (\Sigma^{-1})_{xz},$$

$$C_{yy} = -(\Sigma^{-1})_{xy}((\Sigma^{-1})_{xx} - (\Sigma_2^{-1})_{xx} + \sigma_{xx}^{-1})^{-1}(\Sigma^{-1})_{xy} + (\Sigma^{-1})_{yy},$$

$$C_{yz} = -(\Sigma^{-1})_{xy}((\Sigma^{-1})_{xx} - (\Sigma_2^{-1})_{xx} + \sigma_{xx}^{-1})^{-1}((\Sigma^{-1})_{xz} - (\Sigma_2^{-1})_{xz}) + (\Sigma^{-1})_{yz},$$

$$C_{zz} = -((\Sigma^{-1})_{yz} - (\Sigma_3^{-1})_{yz})((\Sigma^{-1})_{yy} - (\Sigma_3^{-1})_{yy} + \sigma_{yy}^{-1})^{-1}((\Sigma^{-1})_{yz} - (\Sigma_3^{-1})_{yz}) + ((\Sigma^{-1})_{zz} - (\Sigma_3^{-1})_{zz})$$
$$-((\Sigma^{-1})_{xz} - (\Sigma_2^{-1})_{xz})((\Sigma^{-1})_{xx} - (\Sigma_2^{-1})_{xx} + \sigma_{xx}^{-1})^{-1}((\Sigma^{-1})_{xz} - (\Sigma_2^{-1})_{xz}) + ((\Sigma^{-1})_{zz} - (\Sigma_2^{-1})_{zz}).$$

$\Sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{pmatrix}$ is covariance matrix of variable (X,Y,Z), Σ_1 is covariance matrix of Z, $\Sigma_2 = \begin{pmatrix} \sigma_{xx} & \sigma_{xz} \\ \sigma_{xz} & \sigma_{zz} \end{pmatrix}$ and $\Sigma_3 = \begin{pmatrix} \sigma_{yy} & \sigma_{yz} \\ \sigma_{yz} & \sigma_{zz} \end{pmatrix}$, are the covariance matrix of (X,Z) and (Y,Z). $(\Sigma)_{ij}, i = x, y, z, j = x, y, z$ is the $i \times j$ element of matrix Σ.

$$C_0 = \log\left(\frac{\det\Sigma \det\Sigma_1}{\det\Sigma_2 \det\Sigma_3}\sigma_{xx}\sigma_{yy}((\Sigma^{-1})_{yy} - (\Sigma_3^{-1})_{yy} + \sigma_{yy}^{-1})^{-\frac{1}{2}}((\Sigma^{-1})_{xx} - (\Sigma_2^{-1})_{xx} + \sigma_{xx}^{-1})^{-\frac{1}{2}}\right).$$

The proof of the theorem is given as follows.

Proof for Theorem 1: X follows Gaussian distribution. Hence, its probability density function is

$$f(x) = (2\pi)^{-\frac{1}{2}}\sigma_{xx}^{-\frac{1}{2}}\exp(-\frac{1}{2}(x-\mu_x)\sigma_{xx}^{-1}(x-\mu_x)),$$

where σ_{xx} is the variance of X, and μ_x is the mean of X.

Similarly, the probability density function for Z is

$$f(\mathrm{z}) = (2\pi)^{-\frac{n_z}{2}}\det\Sigma_1^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\mathrm{z}-\mu_z)^{\mathrm{T}}\Sigma_1^{-1}(\mathrm{z}-\mu_z)\right\},$$

where Σ_1 is covariance matrix of Z, and μ_z is the mean of Z.

On the other hand, the joint probability density functions are as follow,

$$f(x,y,z) = (2\pi)^{-\frac{n}{2}}(\det(\Sigma))^{-\frac{1}{2}}\exp(-\frac{1}{2}(u-\mu_u)^{\mathrm{T}}\Sigma^{-1}(u-\mu_u))$$

$$= (2\pi)^{-\frac{n}{2}}(\det(\Sigma))^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(x-\mu_x)(\Sigma^{-1})_{xx}(x-\mu_x)-\frac{1}{2}(y-\mu_y)(\Sigma^{-1})_{yy}(y-\mu_y)\right.$$

$$-\frac{1}{2}(z-\mu_z)^{\mathrm{T}}(\Sigma^{-1})_{zz}(z-\mu_z)-(x-\mu_x)(\Sigma^{-1})_{xy}(y-\mu_y)-(x-\mu_x)(\Sigma^{-1})_{xz}(z-\mu_z)$$

$$\left.-(y-\mu_y)(\Sigma^{-1})_{yz}(z-\mu_z)\right\}$$

$$f(x,z) = (2\pi)^{-\frac{n-1}{2}}(\det(\Sigma_2))^{-\frac{1}{2}}\exp\{-\frac{1}{2}(v-\mu_v)^{\mathrm{T}}\Sigma_2^{-1}(v-\mu_v)\}$$

$$= (2\pi)^{-\frac{n-1}{2}}(\det(\Sigma_2))^{-\frac{1}{2}}\exp\{-\frac{1}{2}(x-\mu_x)(\Sigma_2^{-1})_{xx}(x-\mu_x)-(x-\mu_x)(\Sigma_2^{-1})_{xz}(z-\mu_z)$$

$$-\frac{1}{2}(z-\mu_z)^{\mathrm{T}}(\Sigma_2^{-1})_{zz}(z-\mu_z)\}$$

$$f(y,z) = (2\pi)^{-\frac{n-1}{2}}(\det(\Sigma_3))^{-\frac{1}{2}}\exp\{-\frac{1}{2}(w-\mu_w)^{\mathrm{T}}\Sigma_3^{-1}(w-\mu_w)\}$$

$$= (2\pi)^{-\frac{n-1}{2}}(\det(\Sigma_3))^{-\frac{1}{2}}\exp\{-\frac{1}{2}(y-\mu_y)(\Sigma_3^{-1})_{yy}(y-\mu_y)-(y-\mu_y)(\Sigma_3^{-1})_{yz}(z-\mu_z)$$

$$-\frac{1}{2}(z-\mu_z)^{\mathrm{T}}(\Sigma_3^{-1})_{zz}(z-\mu_z)\}$$

where $u=(x,y,z)^{\mathrm{T}}$, $v=(x,z)^{\mathrm{T}}$, $w=(y,z)^{\mathrm{T}}$. Σ is covariance matrix of μ , i.e. $\Sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{pmatrix}$. Σ_1 is the covariance matrix of Z, $\Sigma_2 = \begin{pmatrix} \sigma_{xx} & \sigma_{xz} \\ \sigma_{xz} & \sigma_{zz} \end{pmatrix}$ and $\Sigma_3 = \begin{pmatrix} \sigma_{yy} & \sigma_{yz} \\ \sigma_{yz} & \sigma_{zz} \end{pmatrix}$ are the covariance matrix of $(x,z)^{\mathrm{T}}$ and $(y,z)^{\mathrm{T}}$.

$(\Sigma)_{ij}$, $i=x,y,z,\ j=x,\mathrm{y},z$ is the $i\times j$ element of the matrix Σ .

Hence

$$p(y\,|\,x,z) = \frac{p(x,y,z)}{p(x,z)}p(x)$$

$$=(2\pi)^{-1}(\sigma_{xx}\frac{\det\Sigma}{\det\Sigma_2})^{-\frac{1}{2}}\exp\{-\frac{1}{2}(x-\mu_x)[(\Sigma^{-1})_{xx}-(\Sigma_2^{-1})_{xx}+\sigma_{xx}{}^{-1}](x-\mu_x)$$

$$-(x-\mu_x)[(\Sigma^{-1})_{xy}(y-\mu_y)+((\Sigma^{-1})_{xz}-(\Sigma_2^{-1})_{xz})(z-\mu_z)]-\frac{1}{2}(y-\mu_y)(\Sigma^{-1})_{yy}(y-\mu_y)$$

$$-(y-\mu_y)(\Sigma^{-1})_{yz}(z-\mu_z)-\frac{1}{2}(z-\mu_z)^{\mathrm{T}}((\Sigma^{-1})_{zz}-(\Sigma_2^{-1})_{zz})(z-\mu_z)$$

If A is a symmetric positive-definite matrix, then

$$\sum_x e^{-\frac{1}{2}\sum_{i,j=1}^{n}A_{ij}x_ix_j+\sum_{i=1}^{n}B_ix_i} = (2\pi)^{\frac{n}{2}}(\det A)^{-\frac{1}{2}}e^{\frac{1}{2}B^TA^{-1}B},$$

and thus, we have

$$\sum_x p(y|z,x)p(x) = (2\pi)^{-1}\left(\sigma_{xx}\frac{\det\Sigma}{\det\Sigma_2}\right)^{-\frac{1}{2}}(2\pi)^{\frac{1}{2}}\left((\Sigma^{-1})_{xx}-(\Sigma_2^{-1})_{xx}+\rho_{xx}^{-1}\right)^{-\frac{1}{2}}\cdot$$

$$\exp\left\{\frac{1}{2}\left[(\Sigma^{-1})_{xy}(y-\mu_y)+\left((\Sigma^{-1})_{xz}-(\Sigma_2^{-1})_{xz}\right)(z-\mu_z)\right]\left[(\Sigma^{-1})_{xx}-(\Sigma_2^{-1})_{xx}+\sigma_{xx}^{-1}\right]^{-1}\right.$$

$$\left.\left[(\Sigma^{-1})_{xy}(y-\mu_y)+\left((\Sigma^{-1})_{xz}-(\Sigma_2^{-1})_{xz}\right)(z-\mu_z)\right]\right\}\cdot$$

$$\exp\left\{-\frac{1}{2}(y-\mu_y)(\Sigma^{-1})_{yy}(y-\mu_y)-(y-\mu_y)(\Sigma^{-1})_{yz}(z-\mu_z)-\frac{1}{2}(z-\mu_z)^T\left[(\Sigma^{-1})_{zz}-(\Sigma_2^{-1})_{zz}\right](z-\mu_z)\right\}$$

$$\sum_y p(x|z,y)p(y) = (2\pi)^{-1}(\sigma_{xx}\frac{\det\Sigma}{\det\Sigma_3})^{-\frac{1}{2}}(2\pi)^{\frac{1}{2}}((\Sigma^{-1})_{yy}-(\Sigma_3^{-1})_{yy}+\sigma_{yy}^{-1})^{-\frac{1}{2}}\cdot$$

$$\exp\{\frac{1}{2}((\Sigma^{-1})_{xy}(x-\mu_x)+((\Sigma^{-1})_{yz}-(\Sigma_3^{-1})_{yz})(z-\mu_z))((\Sigma^{-1})_{yy}-(\Sigma_3^{-1})_{yy}+\sigma_{yy}^{-1})^{-1}$$

$$((\Sigma^{-1})_{xy}(x-\mu_x)+((\Sigma^{-1})_{yz}-(\Sigma_3^{-1})_{yz})(z-\mu_z))\}\cdot$$

$$\exp\{-\frac{1}{2}(x-\mu_x)(\Sigma^{-1})_{xx}(x-\mu_x)-(x-\mu_x)(\Sigma^{-1})_{xz}(z-\mu_z)-\frac{1}{2}(z-\mu_z)^{\mathrm{T}}((\Sigma^{-1})_{zz}-(\Sigma_3^{-1})_{zz})(z-\mu_z)\}$$

Then

$$p(z)\sum_y p(x|y,z)p(y)\sum_x p(y|x,z)p(x)$$

$$=(2\pi)^{-\frac{1}{2}}(\det\Sigma_1)^{-\frac{1}{2}}(\sigma_{yy}\frac{\det\Sigma}{\det\Sigma_3})^{-\frac{1}{2}}(\sigma_{xx}\frac{\det\Sigma}{\det\Sigma_2})^{-\frac{1}{2}}((\Sigma^{-1})_{yy}-(\Sigma_3^{-1})_{yy}+\sigma_{yy}^{-1})^{-\frac{1}{2}}((\Sigma^{-1})_{xx}-(\Sigma_2^{-1})_{xx}+\sigma_{xx}^{-1})^{-\frac{1}{2}}\cdot$$

$$\begin{aligned}
&\exp\{-\frac{1}{2}(x-\mu_x)[-(\Sigma^{-1})_{xy}((\Sigma^{-1})_{yy}-(\Sigma_3^{-1})_{yy}+\sigma_{yy}^{-1})^{-1}(\Sigma^{-1})_{xy}+(\Sigma^{-1})_{xx}](x-\mu_x)\\
&-\frac{1}{2}(y-\mu_y)[-(\Sigma^{-1})_{xy}((\Sigma^{-1})_{xx}-(\Sigma_2^{-1})_{xx}+\sigma_{xx}^{-1})^{-1}(\Sigma^{-1})_{xy}+(\Sigma^{-1})_{yy}](y-\mu_y)\\
&-\frac{1}{2}(z-\mu_z)^{\mathrm{T}}[-((\Sigma^{-1})_{yz}-(\Sigma_3^{-1})_{yz})((\Sigma^{-1})_{yy}-(\Sigma_3^{-1})_{yy}+\sigma_{yy}^{-1})^{-1}((\Sigma^{-1})_{yz}-(\Sigma_3^{-1})_{yz})\\
&+((\Sigma^{-1})_{zz}-(\Sigma_3^{-1})_{zz})-((\Sigma^{-1})_{xz}-(\Sigma_2^{-1})_{xz})((\Sigma^{-1})_{xx}-(\Sigma_2^{-1})_{xx}+\sigma_{xx}^{-1})^{-1}((\Sigma^{-1})_{xz}-(\Sigma_2^{-1})_{xz})\\
&+((\Sigma^{-1})_{zz}-(\Sigma_2^{-1})_{zz})](z-\mu_z)-(x-\mu_x)\cdot 0\cdot(y-\mu_y) \qquad (11)\\
&-(x-\mu_x)[-(\Sigma^{-1})_{xy}((\Sigma^{-1})_{yy}-(\Sigma_3^{-1})_{yy}+\sigma_{yy}^{-1})^{-1}((\Sigma^{-1})_{yz}-(\Sigma_3^{-1})_{yz})+(\Sigma^{-1})_{xz}](z-\mu_z)\\
&-(y-\mu_y)[-(\Sigma^{-1})_{xy}((\Sigma^{-1})_{xx}-(\Sigma_2^{-1})_{xx}+\sigma_{xx}^{-1})^{-1}((\Sigma^{-1})_{xz}-(\Sigma_2^{-1})_{xz})+(\Sigma^{-1})_{yz}](z-\mu_z)
\end{aligned}$$

Let

$$C^{-1}=\begin{pmatrix} C_{xx} & C_{xy} & C_{xz}\\ C_{xy} & C_{yy} & C_{yz}\\ C_{xz} & C_{yz} & C_{zz}\end{pmatrix}$$

where

$$C_{xx}= -(\Sigma^{-1})_{xy}((\Sigma^{-1})_{yy}-(\Sigma_3^{-1})_{yy}+\sigma_{yy}^{-1})^{-1}(\Sigma^{-1})_{xy}+(\Sigma^{-1})_{xx},$$

$$C_{xy}=0,$$

$$C_{xz}=-(\Sigma^{-1})_{xy}((\Sigma^{-1})_{yy}-(\Sigma_3^{-1})_{yy}+\sigma_{yy}^{-1})^{-1}((\Sigma^{-1})_{yz}-(\Sigma_3^{-1})_{yz})+(\Sigma^{-1})_{xz} \quad ,$$

$$C_{yy} = -(\Sigma^{-1})_{xy}((\Sigma^{-1})_{xx} - (\Sigma_2^{-1})_{xx} + \sigma_{xx}^{-1})^{-1}(\Sigma^{-1})_{xy} + (\Sigma^{-1})_{yy} \quad ,$$

$$C_{yz} = -(\Sigma^{-1})_{xy}((\Sigma^{-1})_{xx} - (\Sigma_2^{-1})_{xx} + \sigma_{xx}^{-1})^{-1}((\Sigma^{-1})_{xz} - (\Sigma_2^{-1})_{xz}) + (\Sigma^{-1})_{yz} \quad ,$$

$$C_{zz} = -((\Sigma^{-1})_{yz} - (\Sigma_3^{-1})_{yz})((\Sigma^{-1})_{yy} - (\Sigma_3^{-1})_{yy} + \sigma_{yy}^{-1})^{-1}((\Sigma^{-1})_{yz} - (\Sigma_3^{-1})_{yz}) + ((\Sigma^{-1})_{zz} - (\Sigma_3^{-1})_{zz})$$
$$-((\Sigma^{-1})_{xz} - (\Sigma_2^{-1})_{xz})((\Sigma^{-1})_{xx} - (\Sigma_2^{-1})_{xx} + \sigma_{xx}^{-1})^{-1}((\Sigma^{-1})_{xz} - (\Sigma_2^{-1})_{xz}) + ((\Sigma^{-1})_{zz} - (\Sigma_2^{-1})_{zz})$$

Hence, we have

$$p(z)\sum_y p(x\,|\,y,z)p(y)\sum_x p(y\,|\,x,z)p(x)$$
$$= (\frac{\det C}{\det \Sigma_1})^{\frac{1}{2}}(\sigma_{yy}\frac{\det \Sigma}{\det \Sigma_3})^{-\frac{1}{2}}(\sigma_{xx}\frac{\det \Sigma}{\det \Sigma_2})^{-\frac{1}{2}}((\Sigma^{-1})_{yy} - (\Sigma_3^{-1})_{yy} + \sigma_{yy}^{-1})^{-\frac{1}{2}}((\Sigma^{-1})_{xx} - (\Sigma_2^{-1})_{xx} + \sigma_{xx}^{-1})^{-\frac{1}{2}} \cdot$$
$$(2\pi)^{-\frac{n}{2}}(\det(C))^{-\frac{1}{2}}\exp\{-\frac{1}{2}(u-\mu_u)^{\mathrm{T}}C^{-1}(u-\mu_u)\}$$
$$= C_1 \cdot Q(x,y,z)$$

where

$$C_1 = (\frac{\det \mathrm{C}}{\det \Sigma_1})^{\frac{1}{2}}(\sigma_{yy}\frac{\det \Sigma}{\det \Sigma_3})^{-\frac{1}{2}}(\sigma_{xx}\frac{\det \Sigma}{\det \Sigma_2})^{-\frac{1}{2}}((\Sigma^{-1})_{yy} - (\Sigma_3^{-1})_{yy} + \sigma_{yy}^{-1})^{-\frac{1}{2}}((\Sigma^{-1})_{xx} - (\Sigma_2^{-1})_{xx} + \sigma_{xx}^{-1})^{-\frac{1}{2}},$$
$$Q(x,y,z) = (2\pi)^{-\frac{n}{2}}(\det(C))^{-\frac{1}{2}}\exp\{-\frac{1}{2}(u-\mu_u)^{\mathrm{T}}C^{-1}(u-\mu_u)\}$$

Therefore, PMI equals to

$$\begin{aligned}\mathrm{PMI}(\mathrm{X};\mathrm{Y}\,|\,\mathrm{Z}) &= \sum_{x,y,z} p(x,y,z)\log\frac{p(x,y,z)}{C_1\cdot Q(x,y,z)} \\ &= \sum_{x,y,z} p(x,y,z)\log\frac{p(x,y,z)}{Q(x,y,z)} - \log C_1 \\ &= D(p(x,y,z)\,\|\,Q(x,y,z)) - \log C_1\end{aligned}$$

where $D(p(x,y,z)\,\|\,Q(x,y,z))$ is the KL-divergence between $p(x,y,z)$ and $Q(x,y,z)$. The KL-divergence between two multivariate Gaussian distributions $p(x) = N(x;\mu_p,\Sigma_p)$ and $q(x) = N(x;\mu_q,\Sigma_q)$ can also be written (2) as follows

$$D_{KL}(p\,\|\,q) = \frac{1}{2}\left(tr(\Sigma_q^{-1}\Sigma_p) + (\mu_p-\mu_q)^T\Sigma_q^{-1}(\mu_p-\mu_q) - \ln\left(\frac{\det \Sigma_p}{\det \Sigma_q}\right) - n\right).$$

Therefore, we have

$$\begin{aligned}\mathrm{PMI}(\mathrm{X};\mathrm{Y}\,|\,\mathrm{Z}) &= \frac{1}{2}(\mathrm{tr}(C^{-1}\Sigma) - \log(\frac{\det \Sigma}{\det C}) - n) - \log \mathrm{C}_1 \\ &= \frac{1}{2}(\mathrm{tr}(C^{-1}\Sigma) + \log \mathrm{C}_0 - n)\end{aligned}$$

where

$$C_0 = \log(\frac{\det\Sigma \det\Sigma_1}{\det\Sigma_2 \det\Sigma_3}\sigma_{xx}\sigma_{yy}((\Sigma^{-1})_{yy} - (\Sigma_3^{-1})_{yy} + \sigma_{yy}^{-1})^{-\frac{1}{2}}((\Sigma^{-1})_{xx} - (\Sigma_2^{-1})_{xx} + \sigma_{xx}^{-1})^{-\frac{1}{2}})$$

which proved Theorem 1.

■

Assuming that X and Y are two one-dimensional variables, if (X,Y) follows Gaussian distribution, then the mutual information between X and Y can be calculated by the following formula (3)

$$\mathrm{MI}(X;Y) = -\frac{1}{2}\log\frac{|C(X)|\cdot|C(Y)|}{|C(X,Y)|}, \qquad [S9]$$

where C(X) is variance of variable X, $C(X,Y) = \begin{pmatrix} C(X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(X,Y) & C(X) \end{pmatrix}$ is the covariance matrix of X, and Y. |C(X)| is the determinant of C(X).

Lemma S1 Eq.S9 is equivalent to the following equation

$$\mathrm{MI}(X;Y) = -\frac{1}{2}\log(1-\rho^2) \qquad [S10]$$

where ρ is the Pearson correlation coefficient (PCC) between variables X and Y.

Proof for Lemma S1:

$$|C(X,Y)| = \begin{vmatrix} C(X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(X,Y) & C(Y) \end{vmatrix} = |C(X)|\cdot|C(Y)|\cdot\begin{vmatrix} 1 & \frac{\mathrm{cov}(X,Y)}{|C(X)|} \\ \frac{\mathrm{cov}(X,Y)}{|C(Y)|} & 1 \end{vmatrix}$$

$$= |C(X)|\cdot|C(Y)|(1-(\frac{\mathrm{cov}(X,Y)}{\sqrt{|C(X)|\cdot|C(Y)|}})^2) = |C(X)|\cdot|C(Y)|(1-\rho^2)$$

where cov(X,Y) is the covariance of X and Y. Hence we have

$$\mathrm{MI}(X;Y) = \frac{1}{2}\log\frac{|C(X)|\cdot|C(Y)|}{|C(X,Y)|} = -\frac{1}{2}\log\frac{|C(X,Y)|}{|C(X)|\cdot|C(Y)|}$$

$$= -\frac{1}{2}\log(1-\rho^2)$$

■

Similarly, for the conditional mutual information we have the same result.

Lemma S2 Assume that X and Y are two one-dimensional variables, and Z is an n-2 dimensional (n-2 > 0) vector. Letting (X,Y,Z) follow multi-dimensional Gaussian distribution, then the conditional mutual information between X and Y given Z can be simplified as follow

$$\mathrm{CMI}(X;Y\mid Z) = -\frac{1}{2}\log(1-\rho^2_{XY\cdot Z}), \qquad \text{[S11]}$$

where $\rho^2_{XY\cdot Z}$ is the partial correlation between X and Y given Z.

Proof for Lemma S2 :

When (X,Y,Z) follows multi-dimensional Gaussian distribution, CMI(X;Y|Z) (3) is

$$\mathrm{CMI}(X;Y\mid Z) = \frac{1}{2}\log\frac{|C(X,Z)|\cdot|C(Y,Z)|}{|C(Z)|\cdot|C(X,Y,Z)|} \qquad \text{[S12]}$$

where C(X,Z) is the covariance matrix of (X,Z). |C(X,Z)| is the determinant of C(X,Z).

$C(X,Y,Z) = \begin{pmatrix} C(X) & \mathrm{cov}(X,Y) & \mathrm{cov}(X,Z) \\ \mathrm{cov}(X,Y)^T & C(Y) & \mathrm{cov}(Y,Z) \\ \mathrm{cov}(X,Z)^T & \mathrm{cov}(Y,Z)^T & C(Z) \end{pmatrix}$ is the covariance matrix of (X,Y,Z).

we have

$$|C(Z)|\cdot|C(X;Y\mid Z)| = |C(Z)|\cdot\begin{vmatrix} C(X) & \mathrm{cov}(X,Y) & \mathrm{cov}(X,Z) \\ \mathrm{cov}(X,Y)^T & C(Y) & \mathrm{cov}(Y,Z) \\ \mathrm{cov}(X,Z)^T & \mathrm{cov}(Y,Z)^T & C(Z) \end{vmatrix}$$

$$\begin{aligned}
&= |C(Z)|\cdot(|C(X)|\cdot|C(Y)|\cdot|C(Z)| + 2|\mathrm{cov}(X,Y)|\cdot|\mathrm{cov}(Y,Z)|\cdot|\mathrm{cov}(X,Z)^T| \\
&- |\mathrm{cov}(X,Z)^T|\cdot|C(Y)|\cdot|\mathrm{cov}(X,Z)| - |\mathrm{cov}(Y,Z)|\cdot|\mathrm{cov}(Y,Z)^T|\cdot|C(X)| \\
&- |\mathrm{cov}(X,Y)|\cdot|\mathrm{cov}(X,Y)|\cdot|C(Z)|) \\
&= |C(X)|\cdot|C(Y)|\cdot|C(Z)|^2 + 2|\mathrm{cov}(X,Y)|\cdot|\mathrm{cov}(Y,Z)|\cdot|\mathrm{cov}(X,Z)|\cdot|C(Z)| \\
&- |\mathrm{cov}(X,Z)|^2\cdot|C(Y)|\cdot|C(Z)| - |\mathrm{cov}(Y,Z)|^2\cdot|C(X)|\cdot|C(Z)| - |\mathrm{cov}(X,Y)|^2\cdot|C(Z)|^2 \\
&= |C(X)|\cdot|C(Y)|\cdot|C(Z)|^2 + |\mathrm{cov}(X,Z)|^2\cdot|\mathrm{cov}(Y,Z)|^2 - |\mathrm{cov}(X,Z)|^2\cdot|C(Y)|\cdot|C(Z)| \\
&- |\mathrm{cov}(Y,Z)|^2\cdot|C(X)|\cdot|C(Z)| - (|\mathrm{cov}(X,Y)|\cdot|C(Z)| - |\mathrm{cov}(X,Z)|\cdot|\mathrm{cov}(Y,Z)|)^2
\end{aligned}$$

due to

$$|C(X,Z)|\cdot|C(Y,Z)| = \begin{vmatrix} C(X) & \mathrm{cov}(X,Z) \\ \mathrm{cov}(X,Z)^T & C(Z) \end{vmatrix}\cdot\begin{vmatrix} C(Y) & \mathrm{cov}(Y,Z) \\ \mathrm{cov}(Y,Z)^T & C(Z) \end{vmatrix}$$

$$\begin{aligned}
&= (|C(X)|\cdot|C(Z)| - |\mathrm{cov}(X,Z)|^2)\cdot(|C(Y)|\cdot|C(Z)| - |\mathrm{cov}(Y,Z)|^2) \\
&= |C(X)|\cdot|C(Y)|\cdot|C(Z)|^2 + |\mathrm{cov}(X,Z)|^2\cdot|\mathrm{cov}(Y,Z)|^2 - |\mathrm{cov}(X,Z)|^2\cdot|C(Y)|\cdot|C(Z)| \\
&- |\mathrm{cov}(Y,Z)|^2\cdot|C(X)|\cdot|C(Z)|
\end{aligned}$$

Hence, we have

$$|C(Z)|\cdot|C(X,Y,Z)| = |C(X,Z)|\cdot|C(Y,Z)| - (|\mathrm{cov}(X,Y)|\cdot|C(Z)| - |\mathrm{cov}(X,Z|\cdot|\mathrm{cov}(Y,Z)|)^2.$$

Then, we have

$$\frac{|C(Z)|\cdot|C(X,Y,Z)|}{|C(X,Z)|\cdot|C(Y,Z)|}=\frac{|C(X,Z)|\cdot|C(Y,Z)|-(|\mathrm{cov}(X,Y)|\cdot|C(Z)|-|\mathrm{cov}(X,Z)|\cdot|\mathrm{cov}(Y,Z)|)^2}{|C(X,Z)|\cdot|C(Y,Z)|}$$

$$=1-\frac{(|\mathrm{cov}(X,Y)|\cdot|C(Z)|-|\mathrm{cov}(X,Z)|\cdot|\mathrm{cov}(Y,Z)|)^2}{|C(X,Z)|\cdot|C(Y,Z)|}$$

$$=1-(\frac{|\mathrm{cov}(X,Y)|\cdot|C(Z)|-|\mathrm{cov}(X,Z)|\cdot|\mathrm{cov}(Y,Z)|}{\sqrt{|C(X,Z)|\cdot|C(Y,Z)|}})^2$$

According to the proof of Lemma S1, we have

$|C(X,Z)|=|C(X)|\cdot|C(Z)|\cdot(1-\rho_{XZ}^2)$, and similarly, $|C(Y,Z)|=|C(Y)|\cdot|C(Z)|\cdot(1-\rho_{YZ}^2)$

where ρ_{XZ} and ρ_{YZ} are the Pearson correlation coefficients (PCCs) between X and Z, Y and Z, respectively. Thus we obtain

$$\frac{|C(Z)|\cdot|C(X,Y,Z)|}{|C(X,Z)|\cdot|C(Y,Z)|}=1-(\frac{|\mathrm{cov}(X,Y)|\cdot|C(Z)|-|\mathrm{cov}(X,Z)|\cdot|\mathrm{cov}(Y,Z)|}{\sqrt{|C(X,Z)|\cdot|C(Y,Z)|}})^2$$

$$=1-(\frac{|\mathrm{cov}(X,Y)|\cdot|C(Z)|-|\mathrm{cov}(X,Z|\cdot|\mathrm{cov}(Y,Z)|}{\sqrt{|C(X)|\cdot|C(Z)|\cdot(1-\rho_{XZ}^2)}\cdot\sqrt{|C(Y)|\cdot|C(Z)|\cdot(1-\rho_{YZ}^2)}})^2$$

$$=1-(\frac{\frac{|\mathrm{cov}(X,Y)|}{\sqrt{|C(X)|\cdot|C(Y)|}}-\frac{|\mathrm{cov}(X,Z)|\cdot|\mathrm{cov}(Y,Z)|}{\sqrt{|C(X)|\cdot|C(Z)|}\cdot\sqrt{|C(Y)|\cdot|C(Z)|}}}{\sqrt{(1-\rho_{XZ}^2)}\cdot\sqrt{(1-\rho_{YZ}^2)}})^2$$

$$=1-(\frac{\rho_{XY}-\rho_{XZ}\cdot\rho_{YZ}}{\sqrt{(1-\rho_{XZ}^2)}\cdot\sqrt{(1-\rho_{YZ}^2)}})^2=1-\rho_{XY\cdot Z}^2$$

Finally, CMI(X;Y|Z) is

$$\mathrm{CMI}(X;Y\mid Z)=\frac{1}{2}\log\frac{|C(X,Z)|\cdot|C(Y,Z)|}{|C(Z)|\cdot|C(X,Y,Z)|}$$

$$=-\frac{1}{2}\log\frac{|C(Z)|\cdot|C(X,Y,Z)|}{|C(X,Z)|\cdot|C(Y,Z)|}=-\frac{1}{2}\log(1-\rho_{XY\cdot Z}^2)$$

■

Lemma S1 shows that mutual information (MI) is a transformation of Pearson correlation coefficient (PCC) under the Gaussian assumption. Similarly, Lemma S2 proves the relationship between conditional mutual information (CMI) and the partial correlation under Gaussian assumption.

In Section S1, Property S1 shows the relationship between PMI and CMI, and hence combining with Lemma S2, we can find out the relationship between PMI and partial correlation under Gaussian assumption.

Theorem S2 Assume that X and Y are two one-dimensional variables, and Z is an n-2 dimensional (n-2 > 0) vector. Letting (X,Y,Z) follow multi-dimensional Gaussian distribution, then PMI between X and Y given Z can be simplified as follows

$$\mathrm{PMI(X;Y\,|\,Z)} = -\frac{1}{2}\log(1-\rho^2_{XY\cdot Z}) + D(p(x\,|\,z)\,||\sum_y p(x\,|\,z,y)p(y)) + D(p(y\,|\,z)\,||\sum_x p(y\,|\,z,x)p(\mathrm{x})), \quad [S13]$$

where $\rho_{XY\cdot Z}$ is the partial correlation between X and Y given Z.

Proof: from Property S1, we have

$$\mathrm{PMI(X;Y\,|\,Z)} = \mathrm{CMI(X;Y\,|\,Z)} + D(p(x\,|\,z)\,||\sum_y p(x\,|\,z,y)p(y)) + D(p(y\,|\,z)\,||\sum_x p(y\,|\,z,x)p(\mathrm{x}))$$

and, from Lemma S2, we have $\mathrm{CMI(X;Y\,|\,Z)} = -\frac{1}{2}\log(1-\rho^2_{XY\cdot Z})$. Hence it is obvious that

$$\mathrm{PMI(X;Y\,|\,Z)} = -\frac{1}{2}\log(1-\rho^2_{XY\cdot Z}) + D(p(x\,|\,z)\,||\sum_y p(x\,|\,z,y)p(y)) + D(p(y\,|\,z)\,||\sum_x p(y\,|\,z,x)p(\mathrm{x})).$$

■

From this result, we can actually improve the partial correlation in the same manner as PMI.

Section S3. Algorithm of PMI

Estimating PMI

We estimate probability distribution by partitioning the supports of X, Y and Z into bins of finite size, the number of bins may have some impact of the computing results. As the definition of PMI, we add small ε in all bins to avoid zero division. Because if we do not add ε, we may have the problem that the denominator is zero. Here we set ε to be 0.001, which is very small. We do the same thing when calculating CMI. Note that it is not necessary to add ε for a continuous model, e.g., for the case assuming Gaussian distribution.

PMI based PC-algorithm (PCA-PMI) and web tool

Here, PC-algorithm means path consistency algorithm. According to Theorem 1, PMI can be simplified as Eq.12 under the Gaussian assumption in the main text. This approximation form is intended numerically for calculating PMI on a large-scale system. We use the PMI based path-consistency algorithm illustrated in (3), and the details are below. The algorithm is based upon the assumption of Gaussian distributions on the measured variables.

Algorithm (PCA-PMI)

Input:

Gene expression matrix $\boldsymbol{A=\{a_1,a_2...a_n\}}$, where $\boldsymbol{a_i}$ is the expression vector of Gene $\boldsymbol{i}$.

Parameter for dependence threshold λ.

Output:

Inferred gene network $\boldsymbol{G}$,

Order of inferred network L.

Step-1. Initialization. Generate the complete connected network $\boldsymbol{G_0}$ for all genes (i.e. the clique graph of all genes). Set $L := -1$.

Step-2. $L := L+1$; For a nonzero edge $\boldsymbol{G}_0(i,j) \neq 0$, select adjacent genes connected with both genes i and j. Compute the number T of the adjacent genes (not including genes i and j) .

Step-3. Set $\boldsymbol{G} := \boldsymbol{G}_0$. If $T < L$, stop. If $T \geq L$, select out L genes from these T genes and let them as $\boldsymbol{K} = [k_1, \cdots, k_L]$. The number of all selections for $\boldsymbol{K}$ is C_T^L. Compute the L-order $\mathrm{PMI}(i, j \mid \boldsymbol{K})$ for all C_T^L selections, and choose the maximal one denoting as $\mathrm{PMI}_{\max}(i, j \mid \boldsymbol{K})$. If $\mathrm{PMI}_{\max}(i, j \mid \boldsymbol{K}) < \lambda$, set $\boldsymbol{G}(i,j) = 0$. Here, we use covariance matrices of the measured variables, i.e., Σ, Σ_1, Σ_2, and Σ_3 shown in Theorem 1.

Step-4. If $\boldsymbol{G} = \boldsymbol{G}_0$, stop; If $\boldsymbol{G} \mathrel{!}= \boldsymbol{G}_0$, set $\boldsymbol{G}_0 := \boldsymbol{G}$ and return to Step-2.

All the source codes and web tool of our algorithm of PCA-PMI can be accessed at http://www.sysbio.ac.cn/cb/chenlab/software/PCA-PMI.

PMI based PC-algorithm in Kernel version (kPCA-PMI) and web tool

The covariance matrix in PCA-PMI algorithm can be replaced by Kernel matrix. Here we apply the distance covariance (dCov) (4) as our Kernel function. Then we can have PMI based PC-algorithm in Kernel version (kPCA-PMI).

Algorithm (kPCA-PMI)

Input:

Gene expression matrix $\boldsymbol{A=\{a_1,a_2...a_n\}}$, where $\boldsymbol{a_i}$ is the expression vector of Gene $\boldsymbol{i}$.

Parameter for dependence threshold λ.

Output:

Inferred gene network ***G***,

Order of inferred network *L*.

Step-1. Initialization. Generate the complete connected network $\boldsymbol{G_0}$ for all genes (i.e. the clique graph of all genes). Set $L := -1$.

Step-2. $L := L+1$; For a nonzero edge $\boldsymbol{G}_0(i, j) \neq 0$, select adjacent genes connected with both genes *i* and *j*. Compute the number *T* of the adjacent genes (not including genes *i* and *j*) .

Step-3. Set $\boldsymbol{G} := \boldsymbol{G}_0$. If $T < L$, stop. If $T \geq L$, select out *L* genes from these *T* genes and let them as $\boldsymbol{K} = [k_1, \cdots, k_L]$. The number of all selections for ***K*** is C_T^L. Compute the *L*-order $\mathrm{PMI}(i, j \mid \boldsymbol{K})$ for all C_T^L selections, and choose the maximal one denoting as $\mathrm{PMI}_{\max}(i, j \mid \boldsymbol{K})$. If $\mathrm{PMI}_{\max}(i, j \mid \boldsymbol{K}) < \lambda$, set $\boldsymbol{G}(i, j) = 0$. Here, we use distance covariance matrices (4) of the measured variables, i.e., Σ, Σ_1, Σ_2, and Σ_3 are defined as follows:

$$\Sigma = \begin{pmatrix} Kernel_{xx} & Kernel_{xy} & Kernel_{xz} \\ Kernel_{xy} & Kernel_{yy} & Kernel_{yz} \\ Kernel_{xz} & Kernel_{yz} & Kernel_{zz} \end{pmatrix},$$

$$\Sigma_1 = Kernel_{zz}, \quad \Sigma_2 = \begin{pmatrix} Kernel_{xx} & Kernel_{xz} \\ Kernel_{xz} & Kernel_{zz} \end{pmatrix}, \text{ and } \Sigma_3 = \begin{pmatrix} Kernel_{yy} & Kernel_{yz} \\ Kernel_{yz} & Kernel_{zz} \end{pmatrix},$$

where $Kernel_{xy} = dCov(x, y)$ and the details about dCov are given in (4).

Step-4. If $\boldsymbol{G} = \boldsymbol{G}_0$, stop; If $\boldsymbol{G} \mathrel{!=} \boldsymbol{G}_0$, set $\boldsymbol{G}_0 := \boldsymbol{G}$ and return to Step-2.

All the source codes and web tool of our algorithm of PCA-PMI can be accessed at http://www.sysbio.ac.cn/cb/chenlab/software/PCA-PMI.

Generally, it is not a trivial task to choose the best threshold λ for each dataset. In this work, we suggest the following process to choose an appropriate threshold λ. Specifically, we measure the convergence of the inferred network on one observed dataset, starting from

$\lambda = 0$ to $\lambda = \lambda_{\max}$ with a small (positive) step size $\Delta\lambda$. When the difference D (positive value) of the two inferred networks by two consecutive λ s is sufficiently small, e.g., $\frac{D}{n(n-1)/2} \le \varepsilon$, we use that latest λ as the appropriate threshold value to infer the network from the dataset. Here, D is defined by the generalized Euclidean distance between the two adjacent matrices of the two respective networks inferred by the algorithm. Note that the number of total elements in an adjacent matrix for a network with *n* nodes is n^2 and the number of elements in the lower triangular matrix of the adjacent matrix is *n(n-1)/2*. Instead of the Euclidean distance, other distance, e.g., correlation-based distance, can be used to choose an appropriate threshold.

In addition, it may take longer CPU time if the size of the network is over thousands of nodes, depending on the network structure. Thus, for large-scale problems (e.g., whole genome network), it is suggested to adopt GPU for efficient parallel computation.

Section S4. Equitability of PMI using simulated data.

(6) suggests that statistics used to measure associations should satisfy the property of equitability, which means that a good statistics should give significant and similar scores to different types of associations with equal noise, against the random relation. (7) proved that MI is more equitable than estimates of MIC. Hence, we also examined the equitability of PMI in the same way as the previous work (7). We generated three datasets with 1000 and 5000 data points. Ten relationships with the form Y=f(X,Z)+aη, were simulated, in contrast to the random case or independent case between X and Y. η is the normally distributed noise, and the details can be found in Table S1. PMI was calculated by the simulated datasets. We plotted the curve of PMI against $1-R^2\{f(x,z);y\}$ which is the same as in (7). Here R^2 is PCC between Y and f(x). Fig.S3A and Fig.S3B show that values of PMI decrease when noises increase, but the signals are significant comparing to the random relation. For the relationship of independency (random) between X and Y, clearly values of PMI are always near zero. For the equally noisy relationships, values of PMI are also close to each other. With the increase of the noise, the curves of different relationships are very consistent, particularly for the relationships of linear, quadratic and cubic.

References

1. Janzing D, Balduzzi D, Grosse-Wentrup M, & Schölkopf B (2013) Quantifying causal influences. *The Annals of Statistics* 41(5):2324-2358.
2. Roberts SJ & Penny WD (2002) Variational Bayes for generalized autoregressive models. *Ieee T Signal Proces* 50(9):2245-2257.
3. Zhang X, *et al.* (2012) Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics* 28(1):98-104.
4. Szekely GJ, Rizzo ML, & Bakirov NK (2007) Measuring and testing dependence by correlation of distances. Ann Stat 35(6):2769-2794.

5. Szekely GJ, Rizzo ML, & Bakirov NK (2007) Measuring and testing dependence by correlation of distances. Ann Stat 35(6):2769-2794.
6. Reshef DN, et al. (2011) Detecting novel associations in large data sets. Science 334(6062):1518-1524.
7. Kinney JB & Atwal GS (2014) Equitability, mutual information, and the maximal information coefficient. Proceedings of the National Academy of Sciences of the United States of America 111(9):3354-3359.

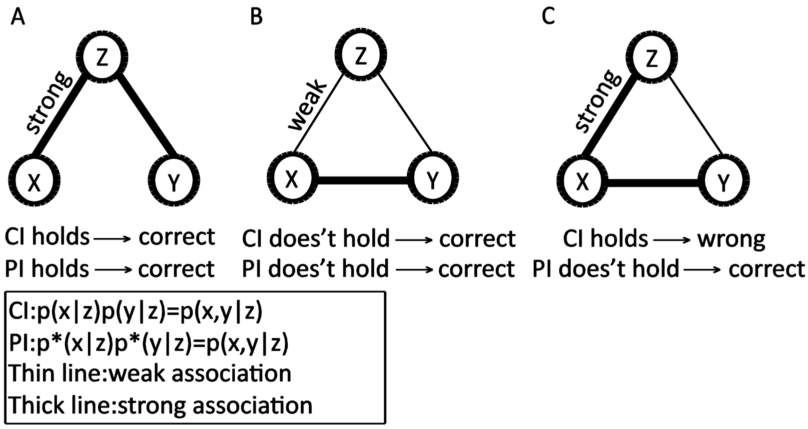

Fig.S1. The conceptual illustration of conditional independence and partial independence of X and Y given Z. (A) When there is no direct association between X and Y, both conditional independence (CI) and partial independence (PI) hold (correct). (B) If Z is weakly associated with X and Y, both conditional independence and partial independence does not hold (correct) for the case of a direct association between X and Y. (C) If Z is strongly associated with X and/or Y, conditional independence always approximately holds (wrong) but partial independence does not hold (correct) for the case of a direct association between X and Y given Z.

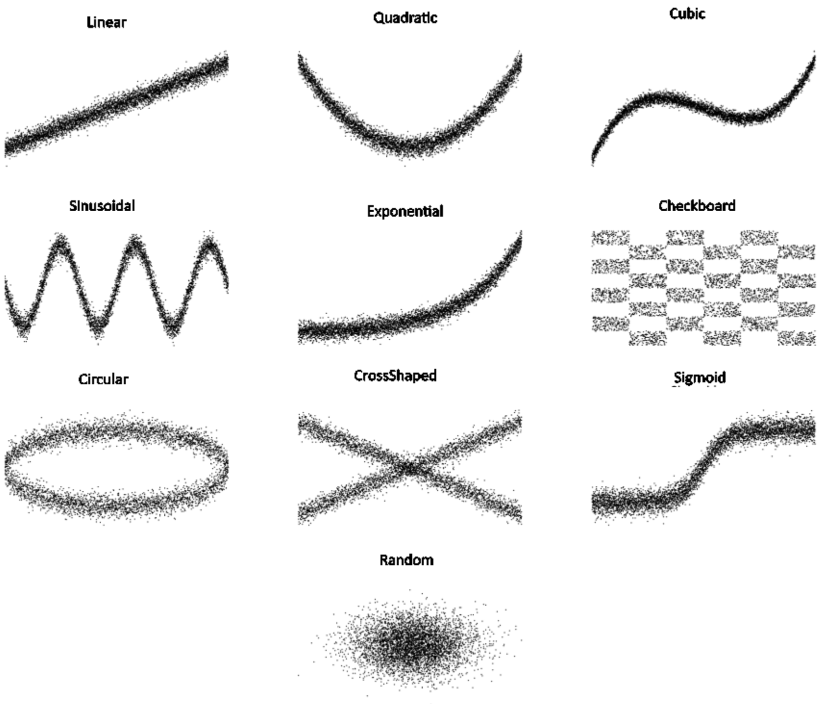

Fig.S2. The relations for the simulation in Table 1, and Fig.2, Fig.S3.

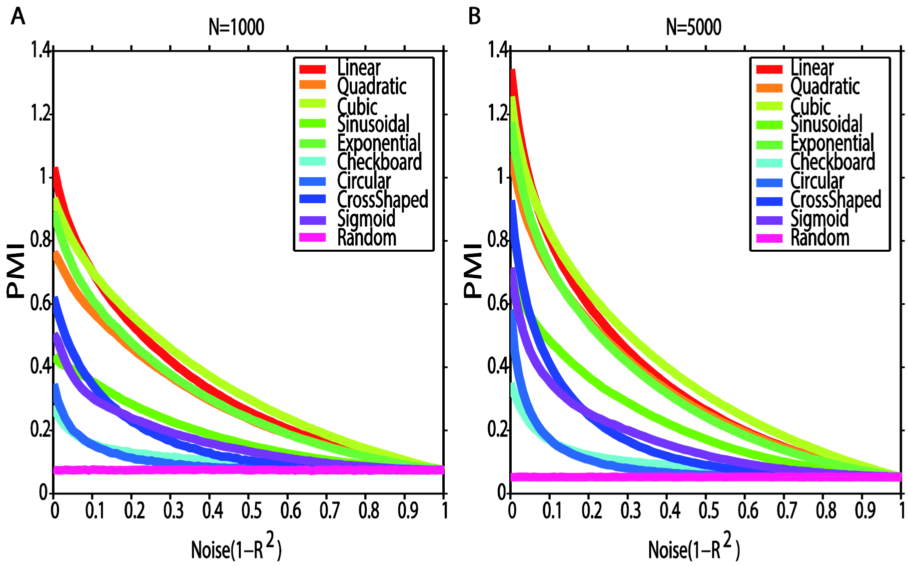

Fig.S3. Tests for R^2-equitability of PMI. We simulated ten relationships with the form $Y = f(X, Z) + a\eta$, and the details of the function f can be found in Table S3, where η is uniformly distributed noise from -1 to 1. The noise was quantified as 1-R^2, and values of PMI were plotted against the noise. (A) and (B) are the curves of PMI against the noise with 1000 data points and 5000 data points, respectively.

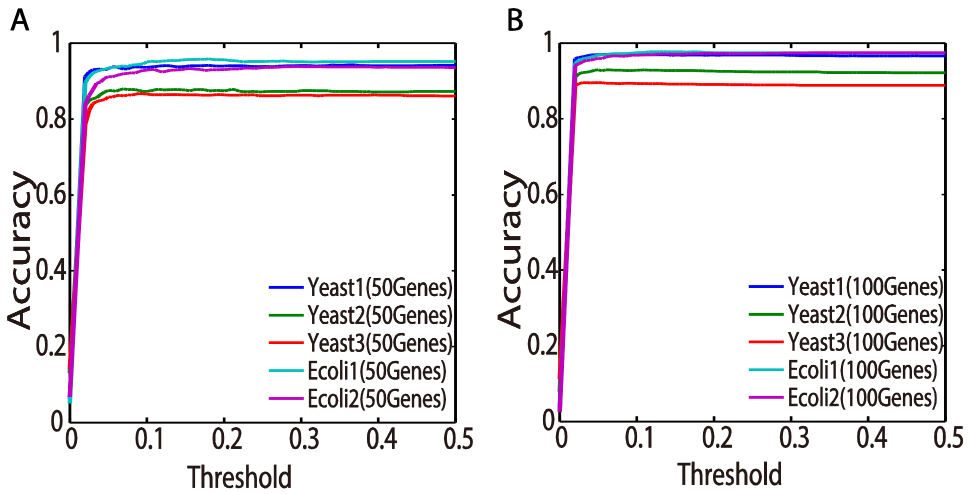

Fig.S4. The accuracy with different thresholds of PMI using all gene expression data with 50 and 100 genes. The threshold is ranged from zero to one.

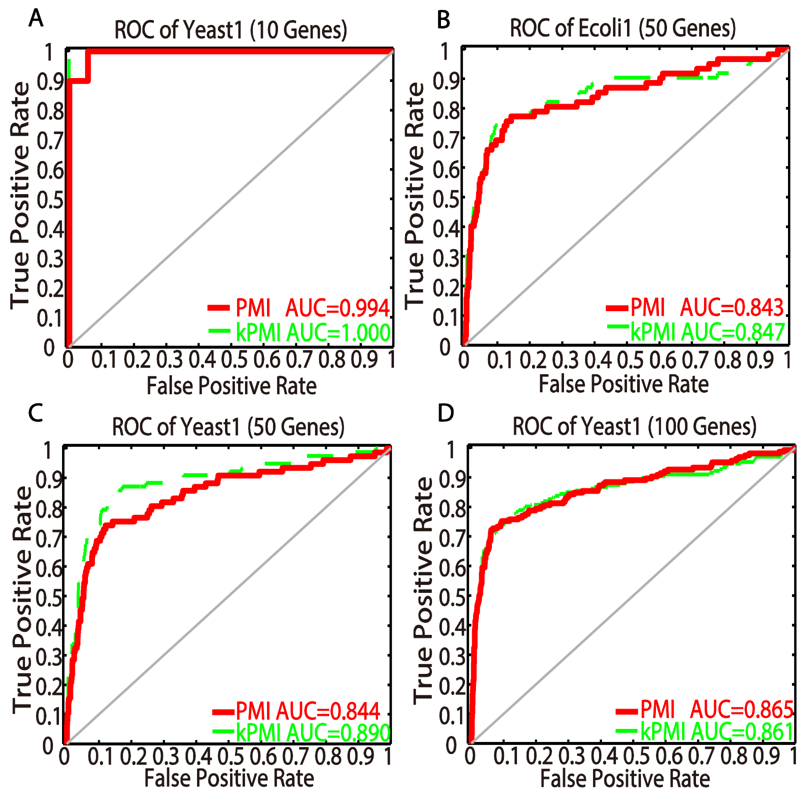

Fig.S5. Comparison the ROC curve for Yeast1 with 10, 50 and 100 genes and Ecoli1 with 50 genes. The red line is calculated by PCA-PMI, while the green line is calculated by kPCA-PMI.

Table S1. Relationship types for Tables 1-2 and Figs.2-3 in the main text

Relation type	Formula
Linear	$Y=X+Z+1+a\eta$
Quadratic	$Y=X^2+Z+1+a\eta$
Cubic	$Y=X(X-2)(X+2)+Z+a\eta$
Sinusoidal	$Y=\sin(\pi X)+Z+3+a\eta$
Exponential	$Y=2^X-Z+2.5+a\eta$
Checkerboard	$Y=\begin{cases}1+2\alpha+\beta+0.5Z, X\in[2k-1,2k]\\ 2\alpha+\beta+0.5Z, X\in[2k,2k+1]\end{cases}+0.5Z+a\eta$
Circular	$Y=(-1)^{\theta}\times\sqrt{\max(X)^2-X^2}+Z+4+a\eta$
Cross-Shaped	$Y=(-1)^{\theta}\times X+0.5Z+a\eta$
Sigmoid	$Y=\frac{e^{\pi X}}{1+e^{\pi X}}-Z+a\eta$
Random	$Y=10\eta+\frac{Z}{20}+10+a\eta$

θ follows a Binomial Distribution B(n,p) with n=1 and p=0.5. α is a random integer uniformly sampled from $\{0,\pm1,\pm2,...,\pm\max(\lceil X \rceil)\}$,where $\lceil X \rceil$ is the integer no larger than X. β follows a uniform distribution U(0,1). We simulated both uniform distribution and normal distribution examples. For uniform distribution, we set Z϶[-1,1], η is uniformly distributed in [-1,1]. For the case X is independent of Z, we set X uniformly distributed in [-1,1], while for the case X is strongly associated with Z, we set X=0.01η+Z. For normal distribution, we set that Z follows normal distribution with mean 0 and standard deviation 1, and η is normally distributed noise with mean 0 and standard deviation 1. a is noise amplitude. For the case X is independent of Z, we set that X follows standard normal distribution, while for the case X is strongly associated with Z, we set X=0.01η+Z.

Table S2. P-values for Table 1 when X and Z are independent

Relation	PMI(X;Y|Z)	CMI(X;Y|Z)	PC(X;Y|Z)	PS(X;Y|Z)
Linear	0.0	0.0	0.0	0.0
Quadratic	0.0	0.0	0.35	0.28
Cubic	0.0	0.0	0.0	0.0
Sinusoidal	0.0	0.0	0.0	0.0
Exponential	0.0	0.0	0.0	0.0
Checkerboard	0.0	0.0	0.0	0.0
Circular	0.0	0.0	0.39	0.39
Cross-Shaped	0.0	0.0	0.05	0.05
Sigmoid	0.0	0.0	0.0	0.0
Random	0.33	0.33	0.10	0.07

Here PS(X;Y|Z) is partial Spearman correlations of X and Y given Z. All the values are the matched p-values in Table 1A in the main text. We randomize the order of Y for T times (here T=1000) and get 1000 null datasets. P-values of PMI and CMI are defined as the proportion of the statistics yielded by null datasets greater than the one yielded by true dataset in the whole null datasets. P-values of Partial Pearson Correlation and Partial Spearman Correlation are directly gained from Fisher-z Transformation.

Table S3. P-values for Table 1 when X and Z are strongly associated

Relation	PMI(X;Y|Z)	CMI(X;Y|Z)	PC(X;Y|Z)	PS(X;Y|Z)
Linear	0.0	0.0	0.0	0.0
Quadratic	0.0	0.41	0.11	0.22
Cubic	0.0	0.76	0.0	0.0
Sinusoidal	0.0	0.01	0.10	0.84
Exponential	0.0	0.97	0.08	0.0
Checkerboard	0.0	0.06	0.67	0.75
Circular	0.0	0.99	0.81	0.87
Cross-Shaped	0.0	0.0	0.80	0.74
Sigmoid	0.0	1.0	0.21	0.0
Random	0.64	0.20	0.21	0.24

Here PS(X;Y|Z) is partial Spearman correlations of X and Y given Z. All the values are the matched p-values in Table 1B in the main text. We randomize the order of Y for T times (here T=1000) and get 1000 null datasets. P-values of PMI and CMI are defined as the proportion of the statistics yielded by null datasets greater than the one yielded by true dataset in the whole null datasets. P-values of Partial Pearson Correlation and Partial Spearman Correlation are directly gained from Fisher-z Transformation.

Table S4. Comparing PMI with CMI and PC when Z is independent of X

Relation	PMI(X;Y|Z)	CMI(X;Y|Z)	PC(X;Y|Z)	PS(X;Y|Z)
Linear	**0.78***	**0.63***	**1***	**0.95***
Quadratic	**0.49***	**0.39***	0.06	0.04
Cubic	**0.38***	**0.35***	**0.39***	**0.75***
Sinusoidal	**0.33***	**0.25***	0.04	**0.09***
Exponential	**0.56***	**0.41***	**0.88***	**0.89***
Checkerboard	**0.10***	**0.09***	0.03	**0.03***
Circular	**0.22***	**0.09***	0.02	0.02
Cross-Shaped	**0.42***	**0.40***	0.04	0.04
Sigmoid	**0.36***	**0.24***	**0.93***	**0.92***
Random	0.06	0.06	0.03	0.03

Here PS(X;Y|Z) is partial spearman correlations of X and Y given Z.

* implies statistically significant in terms of P-value (see Table S6).

Table S5. Comparing PMI with CMI and PC when X is strongly associated with Z

Relation	PMI(X;Y|Z)	CMI(X;Y|Z)	PC(X;Y|Z)	PS(X;Y|Z)
Linear	**1.72***	**0.02***	**1***	**0.89***
Quadratic	**0.68***	0.01	0.02	0.02
Cubic	0.55	0	**0.03***	**0.02***
Sinusoidal	**0.95***	**0.01***	0.03	0.01
Exponential	**0.57***	0	0.03	0.06
Checkerboard	**0.18***	0.01	0.03	0.02
Circular	**0.72***	0.01	0.03	0.03
Cross-Shaped	**0.62***	0.01	0.03	0.02
Sigmoid	**0.84***	0.01	0.03	**0.64***

Random	0.21	0.01	0.03	0.03

Here PS(X;Y|Z) is partial spearman correlations of X and Y given Z.

* implies statistically significant in terms of P-value (see Table S7).

Table S6. P-values for Table S4

Relation	PMI(X;Y|Z)	CMI(X;Y|Z)	PC(X;Y|Z)	PS(X;Y|Z)
Linear	0.0	0.0	0.0	0.0
Quadratic	0.0	0.0	0.90	0.90
Cubic	0.0	0.0	0.0	0.0
Sinusoidal	0.0	0.0	0.36	0.0
Exponential	0.0	0.0	0.0	0.0
Checkerboard	0.0	0.01	0.06	0.01
Circular	0.0	0.0	0.70	0.65
Cross-Shaped	0.0	0.0	0.45	0.87
Sigmoid	0.0	0.0	0.0	0.0
Random	0.38	0.42	0.21	0.33

Here PS(X;Y|Z) is partial spearman correlations of X and Y given Z. All the values are the matched p-values in Table S4. We randomize the order of Y for T times (here T=1000) and get 1000 null datasets. P-values of PMI and CMI are defined as the proportion of the statistics yielded by null datasets greater than the one yielded by true dataset in the whole null datasets. P-values of Partial Pearson Correlation and Partial Spearman Correlation are directly gained from Fisher-z Transformation.

Table S7. P-values for Table S5

Relation	PMI(X;Y|Z)	CMI(X;Y|Z)	PC(X;Y|Z)	PS(X;Y|Z)
Linear	0.0	0.0	0.0	0.0
Quadratic	0.0	0.11	0.58	0.39
Cubic	0.42	0.12	0.0	0.0
Sinusoidal	0.0	0.04	0.37	0.85
Exponential	0.0	0.54	0.58	0.05
Checkerboard	0.0	0.50	0.18	0.12
Circular	0.0	0.36	0.69	0.48
Cross-Shaped	0.0	0.37	0.51	0.95
Sigmoid	0.0	0.52	0.43	0.0
Random	0.17	0.71	0.42	0.53

Here PS(X;Y|Z) is partial spearman correlations of X and Y given Z. All the values are the matched p-values in Table S5. We randomize the order of Y for T times (here T=1000) and get 1000 null datasets. P-values of PMI and CMI are defined as the proportion of the statistics yielded by null datasets greater than the one yielded by true dataset in the whole null datasets. P-values of Partial Pearson Correlation and Partial Spearman Correlation are directly gained from Fisher-z Transformation.

Table S8. Various relation types between X and Y given Z for Tables S9-S12

Relation type	Formula
Linear + Sinusoidal	$Y=X+\sin(\pi X)+Z+1$
Quadratic + Sinusoidal	$Y=X^2+\sin(2\pi X)+Z+1$
Cubic + Sinusoidal	$Y=X^3+\sin(\pi X)+Z+\pi$
Sinusoidal + Exponential	$Y=\sin(\pi X)+2^X+Z+3$
Sinusoidal(High frequency)	$Y=\sin(10\pi X)+Z+1$
Exponential + Quadratic	$Y=10^X+X^2+Z+2$
Exponential + Cubic	$Y=e^X+X^3+0.5Z$
Quadratic + Exponential + Sinusoidal	$Y=X^2+2^X+\sin(\pi X)+Z+4$

$Z\epsilon[-1,1]$, for the case that X is independent of Z, we set X uniformly distributed in [-1,1], while for the case that X is strongly associated with Z, we set $X=0.01\eta+Z$, where η is normally distributed noise with mean 0 and standard deviation 1.

Table S9. Comparing PMI with CMI and PC when Z is independent of X for relation types in Table S8

Relation	PMI(X;Y|Z)	CMI(X;Y|Z)	PC(X;Y|Z)	PS(X;Y|Z)
Linear + Sinusoidal	**1.25***	**0.98***	**0.93***	**0.93***
Quadratic + Sinusoidal	**0.80***	**0.60***	**0.36***	**0.36***
Cubic + Sinusoidal	**1.31***	**0.99***	**0.95***	**0.95***
Sinusoidal + Exponential	**1.24***	**0.94***	**0.91***	**0.91***
Sinusoidal(High frequency)	**0.31***	0.07	**0.08***	**0.08***
Exponential + Quadratic	**0.88***	**0.85***	**0.82***	0.87
Exponential + Cubic	**1.18***	**1.12***	**0.97***	**0.99***

Quadratic + Exponential + Sinusoidal	**1.22***	**0.99***	**0.86***	**0.85***

Here PS(X;Y|Z) is partial spearman correlations of X and Y given Z.

* implies statistically significant in terms of P-value (see Table S11).

Table S10. Comparing PMI with CMI and PC when X is strongly associated with Z for relation types in Table S8

Relation	PMI(X;Y|Z)	CMI(X;Y|Z)	PC(X;Y|Z)	PS(X;Y|Z)
Linear + Sinusoidal	**2.24***	0.02	0.03	0.02
Quadratic + Sinusoidal	**1.74***	**0.02***	0.03	0.01
Cubic + Sinusoidal	**2.50***	0.02	0.03	**0.13***
Sinusoidal + Exponential	**2.32***	0.02	0.03	0.02
Sinusoidal(High frequency)	**1.53***	**0.02***	0.02	0.01
Exponential + Quadratic	**1.98***	0.01	0.03	0.03
Exponential + Cubic	**2.39***	0.01	0.05	**0.96***
Quadratic + Exponential + Sinusoidal	**2.26***	0.01	0.03	0

Here PS(X;Y|Z) is partial spearman correlations of X and Y given Z.

* implies statistically significant in terms of P-value (see Table S12).

Table S11. P-value for Table S9

Relation	PMI(X;Y|Z)	CMI(X;Y|Z)	PC(X;Y|Z)	PS(X;Y|Z)
Linear + Sinusoidal	0	0	0	0
Quadratic + Sinusoidal	0	0	0	0
Cubic + Sinusoidal	0	0	0	0
Sinusoidal + Exponential	0	0	0	0
Sinusoidal(High frequency)	0	0.96	0.01	0.01
Exponential + Quadratic	0	0	0	9.11
Exponential + Cubic	0	0	0	0
Quadratic + Exponential + Sinusoidal	0	0	0	0

Here PS(X;Y|Z) is partial spearman correlations of X and Y given Z. All the values are the matched p-values in Table S9. We randomize the order of Y for T times (here T=1000) and obtain 1000 null datasets. P-values of PMI and CMI are defined as the proportion of the statistics yielded by null datasets greater than the one yielded by true dataset in the whole null datasets. P-values of Partial Pearson Correlation and Partial Spearman Correlation are directly gained from Fisher-z Transformation.

Table S12. P-value for Table S10

Relation	PMI(X;Y|Z)	CMI(X;Y|Z)	PC(X;Y|Z)	PS(X;Y|Z)
Linear + Sinusoidal	0	0.53	0.67	0.47
Quadratic + Sinusoidal	0	0.02	0.64	0.94
Cubic + Sinusoidal	0	0.22	0.84	0
Sinusoidal + Exponential	0	0.17	0.64	0.66
Sinusoidal(High frequency)	0	0	0.25	0.84
Exponential + Quadratic	0	0.79	0.81	0.29
Exponential + Cubic	0	1	0.30	0
Quadratic + Exponential + Sinusoidal	0	0.78	0.77	0.95

Here PS(X;Y|Z) is partial spearman correlations of X and Y given Z. All the values are the matched p-values in Table S10. We randomize the order of Y for T times (here T=1000) and obtain 1000 null datasets. P-values of PMI and CMI are defined as the proportion of the statistics yielded by null datasets greater than the one yielded by true dataset in the whole null datasets. P-values of Partial Pearson Correlation and Partial Spearman Correlation are directly gained from Fisher-z Transformation.

Table S13. Comparing PMI with CMI and PC when X is moderately correlated with Z (both Z and η follow uniform distribution)

Relation	PMI(X;Y|Z)	CMI(X;Y|Z)	PC(X;Y|Z)	PS(X;Y|Z)
Linear	**2.01***	**0.04***	**1***	**0.87***
Quadratic	**1.42***	0.02	0.03	**0.16***
Cubic	**1.68***	**0.04***	**0.75***	**0.95***
Sinusoidal	**1.55***	0.02	**0.14***	**0.88***
Exponential	**1.51***	0.03	**0.22***	**0.75***
Checkerboard	**0.37***	**0.03***	0.03	0.02
Circular	**0.89***	**0.02***	0.03	0.03
Cross-Shaped	**0.77***	**0.03***	0.03	**0.28***

Sigmoid	**1.63***	**0.04***	**0.37***	**0.76***
Random	0.23	0.02	0.03	0.03

Here PS(X;Y|Z) is partial spearman correlations of X and Y given Z.

* implies statistically significant in terms of P-value.

Table S14. Comparing PMI with CMI and PC when X is moderately correlated with Z (both Z and η follow normal distribution)

Relation	PMI(X;Y|Z)	CMI(X;Y|Z)	PC(X;Y|Z)	PS(X;Y|Z)
Linear	**1.54***	**0.03***	**1***	**0.88***
Quadratic	**0.90***	**0.02***	**0.02***	**0.06***
Cubic	**1.04***	**0.02***	0.11	**0.21***
Sinusoidal	**1.08***	**0.02***	0.03	**0.04***
Exponential	**1.00***	**0.02***	**0.08***	**0.77***
Checkerboard	**0.28***	0.01	0.02	0.02
Circular	**0.64***	**0.02***	0.03	0.03
Cross-Shaped	**0.53***	0.01	0.03	**0.26***
Sigmoid	**0.93***	0.01	**0.09***	**0.79***
Random	0.18	0.01	0.03	0.03

Here PS(X;Y|Z) is partial spearman correlations of X and Y given Z.

* implies statistically significant in terms of P-value.

Table S15. Major properties of Part Mutual Information (PMI) and partial independence

Property	Description
Property S1	PMI(X;Y|Z)=CMI(X;Y|Z)+D($p(x\|z)\|\|p^*(x\|z)$)+ D($p(y\|z)\|\|p^*(y\|z)$)
Property S2	PMI(X;Y|Z) ≥ CMI(X;Y|Z) ≥ 0
Property S3	PMI(X;Y|Z) = PMI(Y;X|Z)
Property S4	If X⊥Y| Z, then PMI(X;Y|Z) = CMI(X;Y|Z) = 0
Property S5	If Z⊥X and Z⊥Y, then PMI(X;Y|Z)=CMI(X;Y|Z)=MI(X;Y)
Property S6	If Z≈X or/and Z≈Y, then CMI(X;Y|Z) = 0, but PMI(X;Y|Z) = D($p(y\|z)\|\|p^*(y\|z)$) + D($p(x\|z)\|\|p^*(x\|z)$) is generally non-zero (PMI(X;Y|Z) is zero when X⊥Y| Z)
Property S7	If X⊥Y|Z, then both $p(x,y\|z)=p(x\|z)p(y\|z)$ and $p(x,y\|z)=p^*(x\|z)p^*(y\|z)$ hold.

Property S8	If X≈Z, then $p(x,y\|z)=p(x\|z)p(y\|z)$ always approximately holds even if X and Y are dependent given Z, but not for $p(x,y\|z)=p^{*}(x\|z)p^{*}(y\|z)$.

X⊥Y| Z implies $p(x|y,z)=p(x|z)$ or $p(y|x,z)=p(y|z)$; X⊥Y implies that $p(x|y)=p(x)$ or $p(y|x)=p(y)$; X≈Y implies that X and Y are strongly dependent; $p^{*}(x|z)=\sum_{y} p(x|z,y)p(y)$, $p^{*}(y|z)=\sum_{x} p(y|z,x)p(x)$.