Shannon information is defined for characterizing the uncertainty information of classical probabilistic distributions. As an uncertainty measure it is generally believed to be positive. This holds for any information quantity from two random variables because of the polymatroidal axioms. However, it is unknown why there is negative information for more than two random variables on finite dimensional spaces. It is first shown that the negative tripartite Shannon mutual information implies specific Bayesian network representations of its joint distribution. Then it is shown that the negative Shannon information is obtained from general tripartite Bayesian networks with quantum realizations. This provides a device-independent witness of negative Shannon information. Finally the result is extended for general networks. The present result shows new insights in the network compatibility from non-Shannon information inequalities.

1. Introduction

For a given discrete Markov process a fundamental problem is how to characterize the information produced in such a statistical process. Suppose the statistical outcomes of a set of possible events \( \{ x_1, \ldots, x_n \} \) with respective occurrence of frequencies as \( p_1, \ldots, p_n \in [0, 1] \). Is there a measure of how uncertain is the outcome except for its distribution? For any such a quantity denoted as \( H(p_1, p_2, \ldots, p_n) \), it is reasonable to satisfy the following axioms: 1) \( H \) is a continuous function in each variable \( p_i \) of the probability; 2) \( H \) is an increasing function of sample number for the uniform distribution; 3) \( H \) is weighted summation of its single values if any one choice is changed into two. These axioms imply the unique entropy given by Shannon\(^{[1]} \) as

\[
H(p_1, p_2, \ldots, p_n) = - \sum_{i=1}^{n} p_i \log p_i.
\]

This features the average uncertainty of a given statistical process.

The Shannon entropy shows a remarkable application for mutual information of two discrete random variables as

\[
I(X; Y) = H(X) + H(Y) - H(X, Y)
\]

where the mutual information means the uncertainty between two variables \( X \) and \( Y \). It is a measure of the information to which knowledge of one variable reduces uncertainty about the other. These entropy functions satisfy the polymatroidal axioms\(^{[2]} \) of \( H(X), I(X; Y) \geq 0 \) for any finite dimensional variables \( X \) and \( Y \). Shannon actually shows general information inequalities which are the “physical laws” for characterizing the fundamental limits in classical communications and compression.\(^{[3]} \) However, this intrigues a surprising feature of negative information for three or more discrete variables beyond the polymatroidal axioms. Each one in fact presents a so-called non-Shannon inequality.\(^{[3]} \) One primitive example is from the mutual information contained in three discrete variables \( X, Y, \) and \( Z \) on finite dimensional spaces as

\[
I(X; Y; Z) = I(X; Y) - I(X; Y|Z)
= H(X) + H(Y) + H(Z) + H(X, Y, Z)
- H(X, Y) - H(Y, Z) - H(Z) - H(X, Y, Z)
\]

where \( I(X; Y|Z) = H(Y, Z) - H(Z) - H(X, Y, Z) \) denotes the mutual information conditional on the outcome of variable \( Z \). One example is shown in Figure 1. The proper local measurements on the network (a) may generate a joint probability distribution \( P_{xyz} \) with the notation of \( P_{xyz} = \frac{1}{4}[000] + \frac{1}{4}[011] + \frac{1}{4}[101] + \frac{1}{4}[110] \), where \([xyz]\) denotes joint event \( (X = x, Y = y, Z = z) \), and the probability of joint outcome \([xyz]\) = \([000],[011],[101],[110]) \) or \([110] \) is \( \frac{1}{4} \). This probability distribution yields to a negative information of \( I(X; Y; Z) = -1 \). Especially, the new correlations are built for two independent parties Alice and Bob assisted by the other’s local operations, that is, the mutual information conditional on the outcome of \( Z \) is given by \( I(X; Y|Z) = 1 \) while the mutual information \( I(X; Y) \) is zero. This may imply a simple explanation of the Shannon negative information from Bayesian networks.\(^{[4]} \) This intrigues a natural problem for characterizing general negative Shannon information.

In Hilbert space formulation of Bayesian network in Figure 1a, each bipartite edge is replaced by an entanglement,\(^{[5,6]} \) as shown in Figure 1b. Under the proper postulates of quantum state representation, quantum measurement, Born rule and tensor decom-
of composes systems, the quantum probability is given by 
\[ p(x, y, z) = \text{Tr}(M_x \otimes M_y \otimes M_z) \rho_1 \otimes \rho_2 \], where \( \{ M_x \}, \{ M_y \}, \text{ and } \{ M_z \} \) denotes respective quantum measurements of Alice, Bob, and Matchmaker, and \( \text{Tr}(\cdot) \) denotes the trace operation of matrix. This quantum probability shows not only the similar features of classical statistics of \( I(X; Y; Z) \leq 0 \), but also the quantumness of entanglement,\(^{3,7}\) that is, two independent parties Alice and Bob can build quantum entanglement assisted by the other’s local operations and classical communication.\(^{8,9}\) This provides a simple physical model for verifying Bayesian network and the new quantumness of entanglement assisted by other party from its statistical distribution.

Our motivation in this work is to investigate a general problem of the Bayesian network compatibility of the negative Shannon mutual information. For a given tripartite joint probability distribution with negative mutual information on finite sample spaces, we first classify all the compatible Bayesian networks. We show there are intrinsic network configurations for these distributions, that is, chain network consisting of two edges or triangle network consisting of three edges. This implies a device-independent verification of negative Shannon mutual information using quantum networks. The feature of negative Shannon mutual information is generic for any tripartite quantum entangled network or general multipartite networks.

2. Result

2.1. Negative Shannon Mutual Information in Bayesian Network Model

We first introduce some notations of Bayesian networks.\(^{14}\) A graph \( G \) consists of a vertex (or node) set \( V \), and an edge (or link) set \( E \). The vertices in a given graph are corresponding to measurable variables, and the edges denote certain relationships that hold in pairs of variables. A bi-directed edge denotes the existence of unobserved common causes. These edges will be marked as curved arcs with two arrowheads, as shown in Figure 2. If all edges are directed, we then have a directed graph.

Directed graph may include directed cycles. One example is given by \( X \rightarrow Y \rightarrow Z \rightarrow X \) with \( X, Y, Z \in V \), which represents

Figure 2. Schematic Markov compatibility of negative Shannon mutual information. a) Classical chain network consisting of two latent variables \( \lambda \) and \( \gamma \). b) Quantum chain network consisting of two bipartite states \( \rho_1 \) and \( \rho_2 \) on Hilbert space \( H \). c) Classical triangle network consisting of three latent variables \( \lambda, \gamma, \) and \( \eta \). d) Classical tripartite network consisting of one latent variable \( \lambda \).
mutual causation or feedback processes. The self-loops (e.g., $X \rightarrow X$) are not allowed in what follows. A graph that contains no directed cycle is called acyclic. A graph that is both directed and acyclic is called a directed acyclic graph (DAG). A family in a graph is a set of nodes containing a node and all its parents, where the parents of one node mean all nodes which are connected to it.

Denote a probability distribution as $p_x = \sum p_{x|y} \mathbb{P}[x]$ with random variable $X$ on a finite sample space $\mathcal{X}$, where $[x]$ denotes the event of $X = x$ and $p_x$ denotes the probability of the outcome $x$. The notation $\sum$ does not mean the summation but a notation of union of all possible events of a given probability distribution. Similar notations will be used for multivariate joint distributions $p_{x_1, \ldots, x_n}$ on finite sample space $\mathcal{X}_1 \times \ldots \times \mathcal{X}_n$. Consider the task of specifying an arbitrary joint distribution $p_{x_1, \ldots, x_n} = \sum_{x_1 \ldots x_n} p_{x_1 \ldots x_n} [x_1 \ldots x_n]$ for $n$ random variables, $X_1, \ldots, X_n$ on finite sample spaces $\mathcal{X}_1, \ldots, \mathcal{X}_n$, respectively, and $[x_1 \ldots x_n]$ denotes the joint event of $X_i = x_i$, $i = 1, \ldots, n$. The basic Bayes rule allows us to decompose $p_{x_1 \ldots x_n}$ into

$$p_{x_1 \ldots x_n} = \prod_j p_{x_j|x_1 \ldots x_{j-1}}$$

where $p_{x_j|x_1 \ldots x_{j-1}}$ denotes the probability of outcome $x_j$ conditional on the outcomes of predecessors $x_1, \ldots, x_{j-1}$. Suppose that each $x_j$ is dependent on a small subset $pa(x_j)$ of its predecessors. We have the following definition.

**Definition 1** (Markovian Parents[4]). Let $V = \{X_1, \ldots, X_n\}$ be an ordered set of measurable variables, and let $P_{x_1, \ldots, x_n}$ be the joint probability distribution on these variables. A set of variables $PA_j$ is said to be Markovian parents of $X_j$ if $PA_j$ is a minimal set of predecessors of $X_j$ that renders $X_j$ independent of all its other predecessors, that is,

$$p_{x_j|x_1 \ldots x_{j-1}} = p_{x_j|x_1 \ldots x_{j-1}}$$

and such that no proper subset of $PA_j$ satisfies Equation (5).

Definition 1 implies for each $X_j$ there is a set $PA_j$ of preceding variables for determining its probability. This can be represented by DAG, where $PA_j$ denotes all the parent nodes toward the node $X_j$. Definition 1 provides a simple recursive algorithm for constructing such a DAG for a given $p_{x_1, \ldots, x_n} = \sum_{x_1 \ldots x_n} p_{x_1 \ldots x_n}[x_1 \ldots x_n]$ as follows:

**Algorithm 1.**

i) Starting with the pair $(X_i, X_j)$, we draw an arrow from $X_i$ to $X_j$ if and only if the two variables are dependent.

ii) For $X_i$, we draw an arrow from either $X_i$ or $X_j$ to $X_i$ if $X_j$ is dependent of $X_i$ or $X_j$.

iii) For $j \geq 3$, one can select any minimal set $PA_j$ of $X_j$‘s possible predecessors. And then, draw an arrow from each member in $PA_j$ to $X_j$.

This follows an iterative algorithm to get a DAG of Bayesian network compatible with the given distribution $P_{x_1, \ldots, x_n}$. It has been shown that $PA_j$ is unique for a given distribution $P_{x_1, \ldots, x_n}$.[4]

From the Reichenbach’s common cause principle,[4,11] it allows a Markovian decomposition as

$$p_{x_1 \ldots x_n} = \prod_j p_{x_j|x_1 \ldots x_{j-1}}$$

(6)

The Markovian dependence can be represented by the directed acyclic graph (DAG) of Bayesian networks.[4]

The DAG shows probabilistic and statistical importance for data mining and efficient inferences. A basic problem in statistics theory is to explore the related DAG for a given statistical distribution. The most traditional way to explore possible DAGs from observations is based on the Markov decomposition in Equation (6) and the faithfulness assumption.[4,12]

**Definition 2** (Markov compatibility[4]). If a probability function $p_{x_1 \ldots x_n}$ admits the factorization of Equation (6) relative to DAG $G$, then $G$ and $P_{x_1 \ldots x_n}$ are compatible.

In classical realization of Bayesian networks, each edge is represented by one measurable variable on proper measurable space, and each outcome depends on all the related variables.[4]

The joint distribution is a multifarious function of all outcomes. This allows us to decompose $p_{x_1 \ldots x_n}$ with one measurable variable $\lambda$[7] as:

$$p_{x_1 \ldots x_n} = \int \lambda p(x_1, \ldots, x_n | \lambda) \mu(\lambda) d\lambda$$

(7)

where $(\Omega, \mu(\lambda))$ denotes the measurable space of unobservable latent variable $\lambda$, and $\mu(\lambda)$ denotes the probability of $\lambda$, and $p(x_j|\lambda)$ denotes the characteristic function of outcome $x_j$ conditional on the variable $\lambda$. This kind of Bayesian networks with latent variables shows non-trivial constraints on its correlations,[13,14]

For a joint distribution of two random variables $X$ and $Y$ on finite sample spaces $\mathcal{X} \times \mathcal{Y}$, Equation (1) is used for featuring the common uncertainty of both variables.[4] The non-negative of $H$ and $I$ from the polymatroidal axioms[2] is useful for solving the Markov compatibility with single latent variable in Equation (7). Instead, the entropy function in Equation (3) shows the tripartite mutual uncertainty only for Markov chains,[11] that is,

$$I(X; Y; Z) \geq 0$$

(8)

if $(X, Y, Z)$ (under any order) consists of a Markov chain[4]. This inspires a generalized unordered Markov condition of $I(X; Y; Z) \geq 0$. The joint probability distribution can be generated from one single latent variable as Equation (7). Remarkably, there are joint distributions implying negative Shannon mutual information, that is, $I(X; Y; Z) < 0$. One example is shown in Figure 2a. The coarse-grained single-variable model in Equation (7) does not imply any intrinsic feature of this case. Instead, we prove new Markov compatibilities for these distributions using Bayesian networks.[4]

Case 1. From the definition in Equation (3) the first scenarios satisfies the following constrains:

$$I(X; Y) = 0, I(X; Y|Z) > 0.$$

(9)

Different from the Markov conditional independence of $p(x, y|z) = p(x|z)p(y|z)$,[4] the present condition in Equation (9) implies a conditional dependence, that is, two independent random variables $X$ and $Y$ on $\mathcal{X} \times \mathcal{Y}$ can build new correlations

$$I(X; Y; Z) = 0.$$
conditional on the outcome of variable $Z$. It can be mathematically formulated as $p(x, y) = p(x)p(y|x)$ and $p(x, y|z) \neq p(x|z)p(y|z)$. The so-called anti-Markov condition provides a primitive explanation of nonnegative Shannon mutual information with $I(X; Y; Z) < 0$.

Especially, consider the example shown in Figure 2a. Combined with classical Birkhoff transformation\cite{15} (e.g., a doubly stochastic matrix $a_{ij}$ which satisfies each column or each row consists of a probability distribution, that is, $\sum a_{ij} = \sum a_{iy} = 1$ and $a_{ij} \geq 0$), the joint distribution $P_{xyz}$ acting on the operators $P_{xyz}$ satisfying the condition in Equation (9) it is compatible with quantum networks consisting of two generalized Einstein–Podolsky–Rosen (EPR) states:\cite{5} $|\phi_i\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and $|\phi_2\rangle_{BC} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, as shown in Figure 2b. A simple local measurement strategy implies a quantum joint distribution

$$P_{xyz} = \frac{1}{4}[|00\rangle + \frac{1}{4}|01\rangle + \frac{1}{4}|10\rangle + \frac{1}{4}|11\rangle]$$

where Alice and Bob perform respective projection measurements $\{M_x, x \in \{0, 1\}\}$ and $\{M_y, y \in \{0, 1\}\}$ while the other performs local measurement $\{M_z = |z\rangle\langle z|, z \in \{00, ..., 11\}\}$. This provides a possible experimental verification of any negative Shannon information satisfying Equation (9). As a directive result, any joint distribution in Equation (14) generated from local measurements on the quantum chain network in Figure 2b with any bipartite states $\rho_A$ and $\rho_B$ is compatible with the classical distribution in Equation (11), that is,

$$S_x = S_q$$

where $S_x$ consists of all classical probability distributions in Equation (11) with respect to any two measurable variables $\lambda$ and $\gamma$, or equivalently represented by $S_x = \{P^{00}_{xy}, I(X; Y) = 0, I(X; Y|Z) > 0\}$; $S_q$ consists of all the quantum probability distributions in Equation (14) derived from any two quantum states $\rho_A$ and $\rho_B$, that is, $S_q = \{P^{00}_{xy}, I(X; Y) = 0, I(X; Y|Z) > 0\}$. The equality in Equation (15) means that there is no quantum nonlocality beyond classical networks for quantum chain network if each party has only one set of local measurements. This means that for a tripartite chain network its classical realization with two independent variables can simulate all quantum correlations from any realization of two independent entangled states with one measurement setting per party. This is different from previous results with more than one measurement settings on chain networks.\cite{9,10} It is also different from the triangle network consisting of three variables with one measurement setting.\cite{15} This may inspire another interesting problem for what the network ingredients may inspire the quantum nonlocality for a general network.

**Case 2.** For dependent random variables $X$ and $Y$ on respective finite sample spaces $\mathcal{X}$ and $\mathcal{Y}$, it suggests the second scenarios for $I(X; Y; Z) \leq 0$ as

$$I(X; Y) > 0, I(X; Y|Z) > 0$$

One example is shown as

$$P_{xyz} = \frac{1}{8}[|00\rangle + \frac{1}{8}|01\rangle + \frac{1}{8}|10\rangle + \frac{1}{8}|11\rangle + \frac{1}{8}|20\rangle + \frac{1}{8}|21\rangle + \frac{1}{8}|30\rangle + \frac{1}{8}|31\rangle]$$

For special case of two particles $A$ and $B$, a state $\rho$ on Hilbert space $H_A \otimes H_B$ is entangled if it cannot be decomposed into

$$\rho = \sum_i p_i \rho_A^{(i)} \otimes \rho_B^{(i)}$$

where $\{p_i\}$ is a probability distribution, $\rho_A^{(i)}$ and $\rho_B^{(i)}$ are states of respective particle $A$ and $B$. Similar definitions may be extended for multiple particles.\cite{16}

In Hilbert space formulation, a finite-dimensional pure state is represented by a normalized vector $|\phi\rangle$ in Hilbert space $H_a$.\cite{6} An ensemble of pure states $\{\phi_i\}$ with a mixing probability $p_i$ is represented by a density matrix $\rho = \sum_p p_i |\phi_i\rangle \langle \phi_i|$ on Hilbert space $H_A$. Here, $\rho$ is positive semidefinite matrix with unit trace. The multipartite quantum system is defined on the tensor of local states, that is, the tensor of Hilbert space $\mathcal{S}_{1, \cdots, n} \otimes H_A$. Any measurement acting on $H_A$ consists of an ensemble $\{M_{ij}\}$ of projection operators or generalized positive semidefinite operators satisfying $\sum M_{ij} = 1_A$ with the identity operator $1_A$. After all the local measurements on a given state $\rho$ on Hilbert space $\mathcal{S}_{1, \cdots, n} \otimes H_A$, from Born rule, the quantum joint probability is given by

$$p_{x_1 \cdots x_n} = \text{Tr}(M_{x_1} \otimes \cdots \otimes M_{x_n})\rho$$

for the joint outcome $x_1, \ldots, x_n$.\cite{12}

**Result 1.** Any joint distribution satisfying the condition in Equation (9) is compatible with a chain Bayesian network.
which has \( I(X; Y) = 1 \) and \( I(X; Y|Z) = 2 \). Here, the conditional mutual information \( I(X; Y|Z) \) is defined in Equation (3). By using the binary representation of \( i, g : i \mapsto i_{1}i_{2} \), we may construct a compatible network model\(^1\) by two steps. One is to split \( P_{xyz} \) into two joint distributions of \( P_{x_{1}y_{1}z} = \frac{1}{2}[00] + \frac{1}{2}[11] \) and \( P_{x_{2}y_{2}z} \) in Equation (10) by using local classical transformation \( g \), that is,

\[
P_{g^{-1}(x,y,z)} = P_{x_{1}y_{1}z} \times P_{x_{2}y_{2}z}
\]

under the inverse mapping \( g^{-1} \) of \( g \) for \( x \) and \( y \), where the joint distribution \( P_{x_{1}y_{1}z} \) satisfies \( I(X_{1};X_{2}) = I(X;Y) \), and the joint distribution \( P_{x_{2}y_{2}z} \) satisfies \( I(X_{1};Y_{2}) = 0 \) and \( I(X_{2};Y_{1}|Z) = 2 \). Note that the joint distribution \( P_{x_{1}y_{1}z} \) can be generated by one latent variable \( \lambda_{i} \) from Equation (7) while \( P_{x_{2}y_{2}z} \) is compatible with the chain network using Result 1. This fact implies a new compatible network consisting of three latent variables, as shown in Figure 2c.

In general, for a given distribution \( P_{xyz} \) satisfying the condition in Equation (16), suppose there is classical transformation \( g : x \mapsto x_{1}x_{2} \) such that

\[
P_{g^{-1}(x,y,z)} = P_{x_{1}y_{1}z} \times P_{x_{2}y_{2}z}
\]

where the joint distribution \( P_{x_{2}y_{2}z} \) satisfies the condition in Equation (9). Under this assumption, we can obtain the following result.

**Result 2.** For any joint distribution on finite sample spaces satisfying the condition in Equation (16), there is a compatible triangle network if the decomposition in Equation (19) holds.

A recent result shows triangle quantum network shows nonlocal correlations beyond all classical realizations under local measurement assumptions\(^1\) that is, by performing one set of local measurements on a triangle quantum network consisting of three entangled states there are quantum tripartite joint probability \( P_{xyz} \), which cannot be generated from any classical triangle network consisting of three measureable variables and local measurements. Interestingly, all of their quantum distributions satisfy the condition in Equation (16). This implies the inequivalence of quantum and classical realizations of triangle network in Figure 2c. Thus the decomposition in Equation (19) provides a sufficient condition to verify triangle network, and may be evaluated by using numeric methods.\(^1\)

The other is a tripartite network consisting of one latent variable as shown in Figure 2d. One example is the distribution \( P_{xyz} = \frac{1}{2}[001] + \frac{1}{2}[010] + \frac{1}{2}[100] \), which cannot be generated from all networks in Figure 2a–c.\(^3\)

This yields to a further problem to distinguish different configurations of triangle networks. Here, we present an informational method as:\(^3\)

\[
I(X; Y; Z) \leq \min \{ H(X|Y, Z), H(Y|X, Z), H(Z|X, Y) \}
\]

**Example 1.** Consider a Greenberger–Horne–Zeilinger (GHZ)-type distribution:

\[
P_{ghz} = a[000] + (1 - a)[111]
\]

where \( a \in [0, 1] \). This distribution can be generated by local projection measurements under the computation basis \( \{0, 1\} \) on a GHZ state\(^2\): \( \sqrt{a}(000) + \sqrt{1 - a}(111) \). It is easy to prove that \( I(X; Y, Z) \geq 0 \) for \( a \geq 0 \). Moreover, we show that this distribution is generated from a triangle network in Figure 2d by violating the inequality in Equation (21) for \( a > 0 \).

**Example 2.** Consider a mixture of GHZ-type distribution and W-type distribution as

\[
P_{xyz} = pP_{ghz} + (1 - p)P_{w}
\]

where \( P_{w} = a[001] + b[010] + (1 - a - b)[100] \) and \( p \in [0, 1] \). Here, the W-type distribution \( P_{w} \) can be generated by local projection measurements under the computation basis \( \{0, 1\} \) on a W state\(^3\): \( \frac{1}{\sqrt{2}} [(001) + (010) + (100)] \). It follows from \( I(X; Y, Z) \leq 0 \) for \( p \leq 0.814 \) assisted by numeric evaluations. Moreover, the distribution is generated from a triangle network in Figure 2d by violating the inequality in Equations (20) or (21) for \( p \geq 0.836 \).

**Example 3.** Consider a generalized W-type distribution\(^2\):

\[
P_{w} = a[001] + b[010] + (1 - a - b)[100]
\]

where \( a, b \geq 0 \) and \( 0 \leq a + b \leq 1 \). We show that this distribution implies a negative Shannon mutual information of \( I(X; Y, Z) < 0 \) for any \( a, b \).\(^1\) Unfortunately, this distribution cannot be verified by using the inequalities in Equations (20) and (21). Instead, we verify any one of the following distributions:\(^3\)

\[
P_{xyz} = a[001] + b[010] + c[100] + d[011]
\]

and its permutations using the second-order inflation method\(^2\) as shown in Figure 3, where \( a, b, c, d \geq 0 \) and \( a + b + c + d = 1 \). Here, the triangle network consisting of \( X_{1}, Y_{1}, Z_{1} \) and \( \lambda_{1}, \gamma_{1}, \eta_{1} \) is the
first-order inflation of the initial network consisting of X, Y, Z and variables \( \lambda, \gamma, \eta \). Moreover, the triangle network consisting of \( X_1, Y_1, Z_2 \) and \( \lambda_2, \gamma_2, \eta_2 \) is the second-order inflation of the initial network.

### 2.2. Generic Negative Shannon Mutual Information

Results 1 and 2 show specific network decompositions can be arisen from of negative Shannon mutual information. Our consideration here is for exploring its converse problem, that is, what information can be learned from a given tripartite network configuration. In general, for a given tripartite Bayesian network \( \mathcal{N} \) (consisting of one, two, or three independent measurable variables), define the minimal tripartite mutual information as

\[
I_{\text{min}}(X; Y; Z) = \min_{P_{\text{sys}}} \{I(X; Y; Z)\}
\]  

where the minimum is over all the possible distributions \( P_{\text{sys}} \) generated from \( \mathcal{N} \). Informally, any one in a tripartite Bayesian network can locally generate additional information for others than its absence. This shows generic negative Shannon mutual information for tripartite networks.\(^{[14]}\)

**Result 3.** Any distribution from tripartite Bayesian networks satisfies

\[
I_{\text{min}}(X; Y; Z) \leq 0
\]  

Result 3 intrigues an interesting indicator of any tripartite network by defining the tripartite information increasing as

\[
\Delta := \max\{I(X; Y | Z) - I(X; Y)\}
\]

\[
= -I_{\text{min}}(X; Y; Z)
\]  

This can be used to characterize how much information can be built by local operations and classical communication (LOCC) of one party. It is of a fundamental rule of management science. One example is the organizational theory\(^{[23]}\) where the indicator \( \Delta \) can be used to characterize the increasing information by group behaviors associated with the network relationship. This may be extended and applied for general organizational theory beyond the scope of this paper.

### 2.3. Negative Shannon Mutual Information from General Networks

Results 1–3 shows the negative Shannon mutual information of three variables imply the compatible Bayesian networks. A natural problem is to explore general networks. Our method here is from the so-called multipartite independent networks\(^{[24]}\) that is, there are some nodes which have not shared any entanglement. Especially, consider an \( n + m \)-partite quantum network \( \mathcal{N}_n \) consisting of generalized EPR states\(^{[5]}\) \( |\phi_i\rangle = \cos \theta_i |00\rangle + \sin \theta_i |11\rangle \) with \( \theta_i \in (0, \frac{\pi}{2}) \), \( i = 1, \ldots, N \). Our goal here is to consider the multipartite independent network.\(^{[26]}\) Especially, assume that there are \( n \) number of nodes \( A_1, \ldots, A_n \) in \( \mathcal{N}_n \) with \( n \geq 2 \) such that each pair of them has not shared any entanglement, as shown in Figure 4.

Denote \( X_i \in \mathcal{X}_i \) and \( Y_j \in \mathcal{Y}_j \) as the respective outcomes of \( A_i \) and \( B_j \) under local projection measurements, \( i = 1, \ldots, n; j = 1, \ldots, m \). Our main result here is to prove that any general \( n \)-independent quantum network \( \mathcal{N}_n \) (with \( n \geq 3 \)) shows different features beyond \( \mathcal{N}_2 \), that is, the chain network in Figure 2b. Specially, we show that both negative and positive mutual information can be generated from \( \mathcal{N}_n \) with \( n \geq 3 \). Here, by using Equation (3) iteratively and the equality \( H(X_1, \ldots, X_n | Y_{n+1}) = H(X_1, \ldots, X_{n+1}) - H(X_{n+1}) \), the mutual information of multivariate is defined as

\[
I(X_1; \ldots; X_n; Y) = I(X_1; \ldots; X_n) - I(X_1; \ldots; X_n | Y)
\]  

where \( I(X_1; \ldots; X_n) \) denotes the mutual information of variables \( X_1, \ldots, X_n \) and can be defined by

\[
I(X_1; \ldots; X_n) = \sum_{\text{odd } s \leq n} \sum_{1 \leq j_1 < \cdots < j_s \leq n} H(X_{j_1}, \ldots, X_{j_s}) - \sum_{\text{even } s \leq n} \sum_{1 \leq j_1 < \cdots < j_s \leq n} H(X_{j_1}, \ldots, X_{j_s}) - I(X_1; \ldots; X_n | Y)
\]

denotes the mutual information of \( X_1, \ldots, X_n \) conditional on the outcomes of \( Y = Y_1 \ldots Y_n \) and can be defined as

\[
I(X_1; \ldots; X_n) = \sum_{\text{odd } s \leq n} \sum_{1 \leq j_1 < \cdots < j_s \leq n} H(X_{j_1}, \ldots, X_{j_s}) - \sum_{\text{even } s \leq n} \sum_{1 \leq j_1 < \cdots < j_s \leq n} H(X_{j_1}, \ldots, X_{j_s})
\]
We first show some local operations may generate positive Shannon information for others as: \[ I(X_1; \ldots; X_n) = 0 \]
\[ I(X_1; \ldots; X_n | Y) > 0 \] (33)
where the joint probability distribution \( P_{X_1 \ldots X_n} \) is obtained from local measurements on the network \( \mathcal{N}_n \). This implies negative Shannon mutual information as
\[ I(X_1; \ldots; X_n; Y) < 0 \] (34)
Moreover, there are some local operations may generate negative Shannon mutual information for others as:
\[ I(X_1; \ldots; X_n; Y) = 0 \]
\[ I(X_1; \ldots; X_n; Y) < 0 \] (35) (36)
This implies positive Shannon mutual information as
\[ I(X_1; \ldots; X_n; Y) > 0 \] (37)
which is different from Equation (34). Thus the general \( k \)-independent quantum network with \( k \geq 3 \) can generate both negative and positive Shannon mutual information beyond Result 1 for tripartite chain network even if both have similar Bell nonlocality. This intriguing new kinds of non-Shannon-type information for any independent set.

**Example 4.** One example is the joint distribution in Equation (10). Another example is given by
\[ P_{xy} = \sum_{ij} p_i q_j [x_i y_j] \] (38)
with three finite-dimensional random variables \( X, Y, Z \), where \( \{ p_i \} \) and \( \{ q_j \} \) are probability distributions. Consider classical transformation \( g : z_j \mapsto x_j y_j \), it follows
\[ P_{xyz} = \sum_{ijk} p_i q_j g_k [x_i y_j z_k] = \sum_{ijk} p_i [x_i] \sum_{jk} g_k [y_j z_k] \] (39)
which raises two independent joint distributions of \( \sum_{i} p_i [x_i] \) and \( \sum_{j} q_j [y_j] \). It is easy to check that the joint distribution \( P_{xyz} \) satisfying the constriction in Equation (9), that is, a negative information of \( I(X; Y; Z) < 0 \). From Result 1, there is a tripartite chain network compatible with the distribution in Equation (38). From Equation (33) the present example can be extended for general star network which is compatible with the joint distribution
\[ P_{x_1 \ldots x_n y} = \sum_{i_1 \ldots i_n} p_{i_1} \ldots p_{i_n} [x_{i_1} \ldots x_{i_n} y_{i_1} \ldots i_n] \], (40)
with \( n + 1 \) finite-dimensional variables \( X_1, \ldots, X_n, Y \), where \( \{ p_i \} \) are probability distributions, \( j = 1, \ldots, n \).

**Example 5.** For the triangle network one example is given by Equation (17). Another is given by
\[ P_{xyz} = \sum_{ijk} p_{i,j} q_{i,k} r_{i} [x_{i} y_{j} z_{k}] \] (41)
with three finite-dimensional variables \( X, Y, Z \), where \( \{ p_{i,j} \}, \{ q_{i,k} \}, \) and \( \{ r_{i} \} \) are probability distributions. Consider local mappings \( g_i : x_i \mapsto (i, j), g_k : y_k \mapsto (j, k) \), and \( g_{i,k} : z_{i,k} \mapsto (k, i) \). It follows a new joint distribution
\[ P_{xy}^{(ik)} = \sum_{ijk} p_{i,j} g_{i,k} [x_{i} y_{j} z_{k}] = \sum_{ijk} \sum_{ii} p_{i,j} \sum_{jj} q_{i,k} \sum_{kk} r_{i} \] (42)
which raises three independent joint distributions of \( \sum_{i,j} p_{i,j} \sum_{j} q_{i,k} \sum_{k} r_{i} \), and \( \sum_{i} r_{i} \). This means the joint distribution in Equation (42) satisfy the conditions in Equations (16) and (19). From Result 2, there is a tripartite triangle network consisting of three variables compatible with the distribution in Equation (41).}

### 3. Conclusions

For tripartite Shannon mutual information Results 1 and 2 imply a compatible chain network for special decomposition of its joint distribution in two cases. This means the negative Shannon mutual information hide different network configurations compatible with specific decompositions of joint distributions. This is different from recent results for featuring higher-order statistical correlations which take use of Euler diagram corresponding to Shannon information. The main reason is that negative Shannon mutual information can be featured by using non-Shannon inequalities beyond Euler diagram. Here, the set bounded by Shannon-type information inequalities is denoted as polymatroidal region. A general problem is to determine whether all the polymatroids are entropic. Especially, the negative Shannon information of three random variables implies the existence of non-entropic polymatroids on the boundary. This is further extended for four or more random variables, which even allow unconstrained non-Shannon-type information inequalities. A natural problem is then to explore the Markov compatibility of general non-Shannon-type information. Another problem is to explore different formations of Shannon-type information.

In summary, we provided an operational characterization of negative Shannon mutual information. The main idea is inspired by Bayesian networks. We have investigated the intrinsic network compatibility of all tripartite joint distributions. Similar results are proved for its quantum realizations. This provided a general method for experimentally verifying negative Shannon information in a device-independent manner. These results should be interesting in the information theory, deep learning, quantum nonlocality, and quantum networks.

### Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

### Acknowledgements

The author is grateful to Jingyun Fan, Yali Mao, and Zhengda Li for discussions. This work was supported by the National Natural Science Foundation of China (Nos. 62172341, 61772437), and Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology (Grant No. SIQSE202105).
Conflict of Interest

The author declares no conflict of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Keywords

Bayesian network, Markov compatibility, network inflation, non-Shannon information, quantum network, Shannon information

Received: July 1, 2022
Revised: September 16, 2022
Published online: October 17, 2022

[1] C. E. Shannon, Bell Syst. Tech. J. 1948, 27, 379.
[2] S. Fujishige, Info. Contr. 1978, 39, 55.
[3] Z. Zhang, R. W. Yeung, IEEE Trans. Inf. Theory 1997, 43, 1982.
[4] J. Pearl, Causality, Models, Reasoning, and Inference, 2nd ed., Cambridge University Press, Cambridge 2009.
[5] A. Einstein, B. Podolsky, N. Rosen, Phys. Rev. 1935, 47, 777.
[6] M. A. Nielsen, I. L. Chuang, Quantum Computation and Quantum Information, 10th ed., Cambridge University Press, Cambridge 2010.
[7] J. S. Bell, Phys. 1964, 1, 195.
[8] M. Zukowski, A. Zeilinger, M. A. Horne, A. K. Ekert, Phys. Rev. Lett. 1993, 71, 4287.
[9] D. Cavalcanti, M. L. Almeida, V. Scarani, A. Acin, Nat. Commun. 2011, 2, 184.
[10] C. Branciard, N. Gisin, S. Pironio, Phys. Rev. Lett. 2010, 104, 170401.
[11] W. Feller, An Introduction to Probability Theory and Its Applications, 3rd ed., Vol. 1, J. Wiley & Sons, New York 1968.
[12] P. Spirtes, N. Glymour, R. Scheines, Causation, Prediction, and Search, 2nd ed., MIT Press, Cambridge, MA 2001.
[13] J. Tian, J. Pearl, in Uncertainty in Artificial Intelligence, Proceedings of the Eighteenth Conference (Eds: A. Darwiche, N. Friedman), Morgan Kaufmann, San Francisco, CA 2002, pp. 519–527.
[14] C. Kang, J. Tian, in Proceedings of the Seventeenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-01), MIT Press, Cambridge, MA 2006, pp. 233–240.
[15] Supplementary materials includes all proofs of the main results and refs. [3, 5, 6, 14, 15, 16, 17, 18, 19, 20, 21].
[16] D. Bohm, Phys. Rev. 1952, 85, 166.
[17] G. Birkhoff, Univ. Nac. Tucuman Rev. Ser. A 1946, 5, 147.
[18] T. Fritz, New J. Phys. 2012, 14, 103001.
[19] M.-O. Renou, Y. Wang, S. Borreii, S. Beigi, N. Gisin, N. Brunner, Phys. Rev. Lett. 2019, 123, 070403.
[20] M.-X. Luo, arXiv:2107.05846, 2021.
[21] F. Lad, G. Sanfilippo, G. Agró, Stat. Sci. 2015, 30, 40.
[22] E. Wolfe, R. W. Spekkens, T. Fritz, J. Causal Inference 2019, 7, 20170020.
[23] M.-X. Luo, Phys. Rev. Lett. 2018, 120, 140402.
[24] R. Horodecki, P. Horodecki, M. Horodecki, K. Horodecki, Rev. Mod. Phys. 2009, 81, 865.
[25] M.-O. Renou, E. Bäumer, S. Borreii, N. Brunner, N. Gisin, S. Beigi, Phys. Rev. Lett. 2019, 123, 140401.
[26] P. Spirtes, R. Scheines, C. Meek, T. Richardson, C. Glymour, H. Hoijtink, A. Boomsmma, TETRAD 3: Tools for Causal Modeling, Program, Psychology Press, London, England 1996.
[27] W. Dür, G. Vidal, J. I. Cirac, Phys. Rev. A 2000, 62, 062314.
[28] M. Navascués, E. Wolfe, J. Causal Inference 2020, 8, 70.
[29] D. M. Greenberger, M. A. Horne, A. Zeilinger, in Bell's Theorem, Quantum Theory and Conceptions of the Universe (Ed: M. Kafatos), Kluwer, Dordrecht 1989, pp. 69–72.
[30] J. L. Kmetz, The Information Processing Theory of Organization Managing Technology Accession in Complex Systems, 1st ed., Ashgate Pub Ltd, Farnham, UK 1998.
[31] R. Yeung, IEEE Trans. Inf. Theory 1997, 43, 1924.
[32] F. Matus, IEEE Trans. Inf. Theory 2006, 44, 236.
[33] N. Ay, Discrete Appl. Math. 2009, 157, 2439.
[34] V. S. Yepez, R. P. Sagar, H. G. Laguna, Few-Body Syst. 2017, 58, 158.