Bell nonlocality in networks

Armin Tavakoli,[1,2,*] Alejandro Pozas-Kerstjens,[3,4,†] Ming-Xing Luo,[5,‡] and Marc-Olivier Renou[6,§]

[1]*Institute for Quantum Optics and Quantum Information - IQOQI Vienna,*
Austrian Academy of Sciences, Boltzmanngasse 3, 1090 Vienna, Austria
[2]*Institute for Atomic and Subatomic Physics, Vienna University of Technology, 1020 Vienna, Austria*
[3]*Departamento de Análisis Matemático, Universidad Complutense de Madrid, 28040 Madrid, Spain*
[4]*Instituto de Ciencias Matemáticas (CSIC-UAM-UC3M-UCM), Madrid, Spain*
[5]*School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China*
[6]*ICFO - Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*

Bell's theorem proves that quantum theory is inconsistent with local physical models. It has propelled research in the foundations of quantum theory and quantum information science. As a fundamental feature of quantum theory, it impacts predictions far beyond the traditional scenario of the Einstein-Podolsky-Rosen paradox. In the last decade, the investigation of nonlocality has moved beyond Bell's theorem to consider more sophisticated experiments that involve several independent sources which distribute shares of physical systems among many parties in a network. Network scenarios, and the nonlocal correlations that they give rise to, lead to phenomena that have no counterpart in traditional Bell experiments, thus presenting a formidable conceptual and practical challenge. This review discusses the main concepts, methods, results and future challenges in the emerging topic of Bell nonlocality in networks.

CONTENTS

* armin.tavakoli@oeaw.ac.at
† physics@alexpozas.com
‡ mxluo@swjtu.edu.cn
§ marc-olivier.renou@icfo.eu
The underlined authors share the first authorship

I. INTRODUCTION

Bell's theorem [1], derived in 1964, constitutes a milestone in the understanding of quantum theory. The theorem states that no physical theory based only on local variables can account for all the predictions of quantum theory. Motivated both by foundational interest and the development of quantum information science, the last thirty years have witnessed intensive research on the topic of Bell nonlocality (see Refs. [2, 3] for reviews, and also Ref. [4]). For a long time, research focus remained on simple experiments in the spirit of the Einstein-Podolsky-Rosen paradox [5], where two parties perform measurements on their respective shares of a two-particle state. However, recent years have seen a major development in the field: concepts in Bell nonlocality are applied to many-party scenarios featuring several independent sources that each distribute a physical state. The independence of the sources reflects a network structure over which parties are connected. This endeavour, referred to as Bell nonlocality in networks, concerns a substantially broader manifestation of nonlocality as compared to that arising in the intensively researched Einstein-Podolsky-Rosen scenario.

The motivation behind the research program on Bell nonlocality in networks is two-fold. Firstly, from the foundational point of view, it is interesting to understand quantum theory and its relationship to other physical models in more sophisticated and qualitatively new scenarios that naturally continue the spirit of Bell's theorem. Secondly, from the point of view of quantum information science, it is becoming increasingly relevant to apply quantum phenomena in networks towards information processing tasks. This is due the technological developments towards scalable quantum networks (see e.g. Refs. [6–8] for an overview of recent progress on the development of a quantum internet). Moreover, Bell nonlocality in networks also brings novel forms of entanglement to the forefront of the physical process. When many *a priori* independent shares of entangled states are distributed in a quantum network, a party that holds multiple shares originating from different sources can perform entangled measurements (i.e. projections of distinct quantum systems in an entangled basis) to *a posteriori* distribute entanglement between distant, initially independent, systems in the network. This is fundamentally based on the process of entanglement swapping [9], which has no counterpart in traditional tests of Bell inequalities. However, as it has turned out, the analysis of Bell nonlocality in networks cannot fall back on the long-established ideas and tools for standard Bell inequality experiments. Instead, the topic has turned out a formidable challenge which demands new lines of thought.

This article presents a general review of the present knowledge of Bell nonlocality in networks. It is also intended to serve as a self-contained introduction to the topic, as a highlight of its most notable progress, and as pointer to its main open problems. In the remainder of this introduction, we first briefly present standard Bell scenarios, then introduce network Bell scenarios and finally give an overview of the main methods used to analyse correlations in networks. In Section II, we discuss in depth the two smallest, and seminal, examples of networks, which are known as the bilocal scenario and the triangle scenario. In Section III, we present general methodology for analysing classical and quantum correlations in networks. In Section IV, we discuss how the network configuration imposes ultimate limits of physical predictions. In Section V, we review experiments demonstrating network nonlocality. In Section VI, we outline a number of additional topics in network nonlocality and closely related subjects. Finally, in Section VII we conclude the review and present a list of interesting open problems and research directions.

A. Standard Bell scenarios

A standard Bell scenario features several parties who perform measurements on the respective shares of a system originally emitted from a single physical source. For simplicity, let us focus on the bipartite scenario, although the definitions and conclusions naturally generalise to more parties. The measurement events are assumed to be spacelike separated and therefore independent. Our parties, call them Alice and Bob, freely and independently choose a measurement setting each, denoted x and y respectively. The outcome of their measurements are respectively labelled a and b (see Figure 2a). If Alice and Bob repeat the experiment a large number of times, the relative frequencies they obtain for the events (a, b) when their settings are (x, y) accurately correspond to a conditional probability distribution $p(a, b|x, y)$, which we also simply refer to as P. This probability distribution is often referred to as the correlations shared by Alice and Bob. A priori, the only assumption made on the correlations stems from relativity, namely that the output of either party reveals no information about the input of the other party. This so-called no-signaling principle is formalised as

$$\forall b, x, y: \quad \sum_a p(a, b|x, y) = p(b|x, y) \stackrel{\text{NS}}{=} p(b|y),$$
$$\forall a, x, y: \quad \sum_b p(a, b|x, y) = p(a|x, y) \stackrel{\text{NS}}{=} p(a|x). \tag{1}$$

A local-variable theory is an attempt at giving a physical explanation for the correlations P based on the principle that the outcome of Alice (resp. Bob) only depends on her (resp. his) input and some stochastic property that is carried with the shares of the system due to their common past. This stochastic property, historically called a local hidden variable[1], is denoted λ and subject to some unknown probability density $\mu(\lambda)$. Therefore, we say that the correlations admit a local model (i.e. they satisfy Bell's local causality assumption) if they can be written as

$$p(a, b|x, y) = \int d\lambda \mu(\lambda) p(a|x, \lambda) p(b|y, \lambda), \tag{2}$$

---

[1] Since λ does not actually need to be hidden, we will simply refer to it as a local variable.
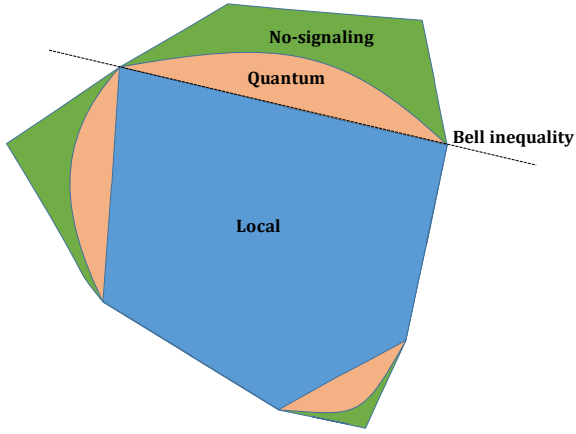
FIG. 1. Illustration of the local polytope, the quantum set of correlations (convex) and the no-signaling polytope. Optimal Bell inequalities are hyperplanes corresponding to a facet of the local polytope. More generally, any hyperplane delimiting the local polytope constitutes a valid Bell inequality.

where, importantly, Alice's and Bob's choices of settings are assumed to be independent from the local variable [i.e. $\mu(\lambda|x,y) = \mu(\lambda)$]. In this sense, any correlation consistent with the no-signaling principle that does not admit a local model is said to be nonlocal.

When the number of inputs and outputs is finite, the set of correlations with a local model admits a simple characterisation in terms of a polytope [10], i.e. a high-dimensional convex polyhedron. We will therefore refer to this set as the local polytope. The polytope structure follows from the fact that since any randomness in the response functions of Alice and Bob can be absorbed in $\mu(\lambda)$ one may, without loss of generality, assume that the response functions are deterministic, i.e. that $p(a|x,\lambda) \in \{0,1\}$ and $p(b|y,\lambda) \in \{0,1\}$. Therefore, each choice of deterministic response functions constitutes a vertex of a polytope in the space of correlations. Since there are only finitely many deterministic responses, one can without loss of generality take λ to belong to a finite set and replace the integral in Eq. (2) by a summation. Due to the freedom of varying $\mu(\lambda)$, the achievable correlations in a local model are the convex combinations of these vertices, thus giving rise to the local polytope (see the blue polytope in Figure 1). This allows one to decide whether a given $p(a,b|x,y)$ admits a local model either by determining the existence of a suitable $\mu(\lambda)$ (a problem which can be cast as a linear program [11, 12]) or, alternatively, by determining all the linear inequalities that characterise the facets of the local polytope and checking that they are all satisfied. The latter inequalities (or more generally, any hyperplane which separates the local polytope) are known as Bell inequalities (see Figure 1 for an illustration). Proving that at least one Bell inequality is violated is sufficient for asserting that correlations are nonlocal.

The simplest and most popular Bell inequality is the so-called Clauser-Horne-Shimony-Holt (CHSH) Bell inequality [13]. The CHSH inequality concerns the simplest nontrivial Bell scenario, namely that in which Alice and Bob each choose between just two measurements, $x \in \{0,1\}$

and $y \in \{0,1\}$, and produce binary outcomes $a \in \{0,1\}$ and $b \in \{0,1\}$. In this scenario, the nontrivial faces of the local polytope are, up to re-labelings, given by a single inequality:

$$\mathcal{S}_{\text{CHSH}} := \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leq 2, \quad (3)$$

where $\langle A_x B_y \rangle = \sum_{a,b} (-1)^{a+b} p(a,b|x,y)$ is the joint expectation value of the observations of Alice and Bob when their settings are x and y, respectively. Using the shorthand notation $\hat{A}_\pm := A_0 \pm A_1$, the CHSH expression becomes $\mathcal{S}_{\text{CHSH}} = \langle \hat{A}_+ B_0 + \hat{A}_- B_1 \rangle$. For local models, the expectation value is taken over λ. Then, using the triangle inequality, Eq. (3) follows from $|B_y| = 1$ and $|\hat{A}_+| + |\hat{A}_-| \leq 2$.

This simple argument highlights that Bell inequalities themselves are rather unremarkable. What is, however, remarkable, is the fact that they can be violated in quantum theory. The correlations predicted by quantum theory are given by the Born rule

$$p(a,b|x,y) = \text{Tr}\left(A_{a|x} \otimes B_{b|y} \rho\right), \quad (4)$$

for some measurements $\{A_{a|x}\}$ and $\{B_{b|y}\}$[2] and a quantum state ρ. The general statement that quantum theory eludes local models is known as Bell's theorem [1]. In order to find a quantum violation of the CHSH inequality, let Alice and Bob share the maximally entangled two-qubit state $|\phi^+\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right)$ and let them perform locally anticommuting measurements, corresponding to the observables $A_0 = \sigma_1$, $A_1 = \sigma_3$, $B_0 = \frac{\sigma_1 + \sigma_3}{\sqrt{2}}$ and $B_1 = \frac{\sigma_1 - \sigma_3}{\sqrt{2}}$, where by $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ we denote the Pauli matrices. The expectation values, computed from the Born rule, are $\langle A_x B_y \rangle = \frac{(-1)^{xy}}{\sqrt{2}}$. This gives a violation of the CHSH inequality as $\mathcal{S}_{\text{CHSH}} = 2\sqrt{2} > 2$, which proves the nonlocal nature of correlations in quantum theory. In fact, quantum theory does not allow for violations larger than $\mathcal{S}_{\text{CHSH}} = 2\sqrt{2}$ [14].

The existence of quantum models, which enable stronger correlations than local-variable ones, raises the question of the ultimate limits on correlations in Bell scenarios. Are there stronger-than-quantum correlations? The no-signaling principle is considered as the minimal requirement that correlations must respect in a Bell scenarios. Therefore, by only demanding that correlations are no-signaling, one obtains the most general notion of nonlocal correlations (see Figure 1). Their relevance is showcased in the celebrated Popescu-Rohrlich distribution, which reads $p(a,b|x,y) = \frac{1}{4}\left[1 + (-1)^{a+b+xy}\right]$ and is consistent with the no-signaling principle described in Eq. (1). It gives the algebrically maximal violation of the CHSH inequality, $\mathcal{S}_{\text{CHSH}} = 4$ [15].

The first experimental violation of the CHSH inequality was reported already in 1972 [16]. In 1982, it was followed by a more faithful demonstration [17]. In recent years, quantum violations of CHSH-type inequalities were reported in rigorous experiments that closed the various experimental loopholes that otherwise could have been exploited in order to maintain a local-variable model [18–21].

---

[2] These are in general positive operator-valued measures, i.e. the measurement operators are non-negative and sum to identity.
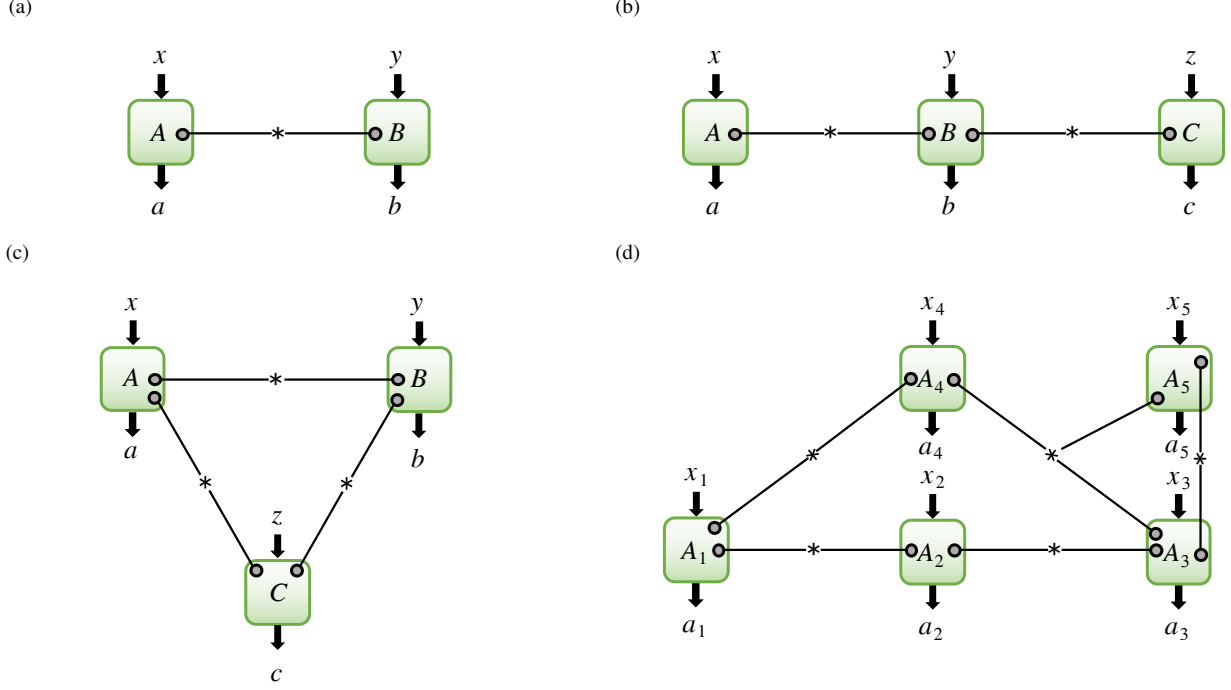
FIG. 2. Illustrations of network Bell scenarios. (a) The standard Bell scenario, with two parties that share a common source. This is the simplest scenario for studying nonlocality and it is the focus of previous reviews [2, 3]. (b) The bilocal scenario, where two extremal parties each share a source with a central party. This is the scenario that underlies quantum repeaters. (c) The triangle scenario, with three parties are connected to each other through bipartite sources. (d) A more complex network with cycles, multipartite sources, and conditionally independent parties.

B. Networks: Beyond the standard Bell scenario

Standard Bell scenarios have inspired the development of network Bell scenarios. Network Bell scenarios come with a crucial conceptual difference: they feature more than one source. Each source in a network, labeled $i = 1, \ldots, m$, is assumed to be independent from the rest. Although most of the works surveyed in this review consider networks with only bipartite sources, it is equally permissible to consider sources that emit more than two subsystems. Notice that this source independence assumption is irrelevant in standard Bell scenarios since these only involve a single source. This has direct implications in the interpretation of what the local variables represent. In standard Bell scenarios one may either view the local variable λ as a physical influence carried with particles due to a shared past (this is Bell's original interpretation [1]) or as a free and unlimited resource of shared randomness made accessible to the parties. However, in network Bell scenarios, due to the presence of independent sources, one can no longer consider the two interpretations equivalent: global shared randomness is no longer a free resource. In this sense, network Bell scenarios are more restricted than their traditional counterparts.

While the independence of the sources in general constitutes an additional assumption, it is arguably a natural one when considering distant sources that are built and operated separately. Each source distributes shares of a physical system to a given subset of the parties present in the network.

Therefore, some parties can hold several, initially independent, shares originating from different sources. The manner in which sources connect the parties can be described as a set of relations $i \rightarrow j$ indicating that source i emits a share to party j. The parties, labeled $j = 1, \ldots, n$, receive random and independent inputs x_j based on which they apply a measurement to their system share(s) and obtain an outcome a_j. In the limit of many rounds, this gives rise to correlations in the network that are written $p(a_1, \ldots, a_n | x_1, \ldots, x_n)$[3]. In Figure 2 we illustrate a few examples of network Bell scenarios.

In local-variable models for the correlations in the network, the independence of the m sources is manifested by associating a stochastic local variable, λ_i, to each of them. The outcome of party j is modelled as a stochastic response conditioned on their input x_j and the local variables that they have received. Naturally generalising Eq. (2), the local correlations achievable in a network take the form

$$p(\bar{a}|\bar{x}) = \int d\lambda_1 \mu_1(\lambda_1) \ldots \int d\lambda_m \mu_m(\lambda_m)$$
$$\times\, p(a_1|x_1, \bar{\lambda}_1) \ldots p(a_n|x_n, \bar{\lambda}_n), \tag{5}$$

where $\bar{a} = (a_1, \ldots, a_n)$, $\bar{x} = (x_1, \ldots, x_n)$, $\bar{\lambda}_j$ denotes the collection of local variables associated to the sources that send

---

[3] For networks with three parties (as in Figures 2b and 2c) we typically use the more handy notations x, y, z to denote inputs and a, b, c to denote outputs. The probability distribution is then written $p(a, b, c|x, y, z)$.
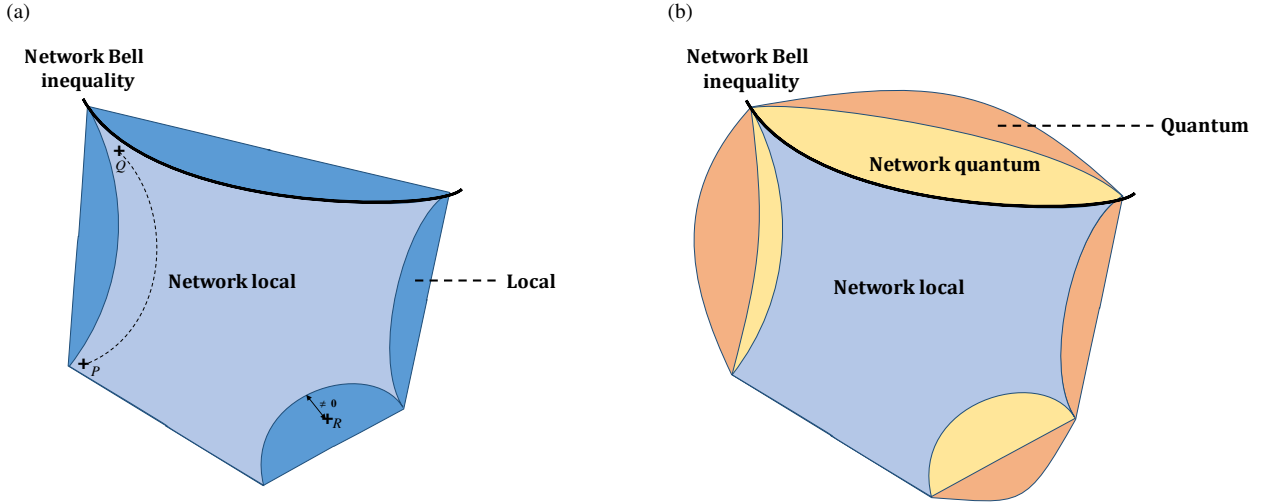
FIG. 3. Illustration of sets of network correlations. (a) The set of network local correlations is contained in the set of local correlations, so there exist Bell-local correlations that violate network Bell inequalities. Thus, when restricting to correlations generated by local systems, a violation of a network Bell inequality can be interpreted as a certification that the correlations were not generated according to the network structure. (b) The set of network local correlations is also contained in the set of network quantum correlations, which is itself contained in the set of quantum correlations without independent sources. Therefore, network Bell inequalities, once the underlying network configuration is fixed, have a meaning analogous to that of standard Bell inequalities (i.e. identifying nonlocal correlations). The sets of network correlations are non-convex (not even star-convex), but any two points P, Q can be connected in a natural curved way [22]. For local variables, one asymptotically recovers the set boundaries [23], which can be described by polynomial expressions. The sets are also closed, so every point R outside it is at some distance to the set, hence upholds some degree of tolerance to noise.

systems to party j and $\mu_i(\lambda_i)$ denotes the probability density of λ_i. These models are called network local models. The independence of the sources is the key feature of the above definition, which distinguishes it from standard multipartite Bell scenarios [24]. In analogy with standard Bell scenarios, the network local models can without loss of generality be restricted to a finite alphabet, i.e. each λ_i can be taken to have a finite cardinality and the corresponding $\mu_i(\lambda_i)$ can be restricted to a probability distribution [25]. In Ref. [22] it was shown how to determine an upper bound on the alphabet size of the local variables for any given network based on the number of inputs and outputs and the network structure.

The independence of the sources causes the set of correlations attainable in a network to acquire a more complex structure than the local polytope (see Figure 3a). A few qualitative properties of the set of local correlations in a network are

(i) The set is contained in the local polytope.

(ii) The set is not convex.

(iii) The set is closed, connected and characterised by a finite number of polynomial inequalities.

The first property is self-evident, because the local polytope corresponds to relaxing the independence of the sources in Eq. (5). However, property (ii) highlights a crucial difference between the network local set and the local polytope. The non-convexity of the set of local correlations in networks [26] has important consequences. It makes the characterisation of correlations in networks a much more difficult task

than in standard Bell scenarios, for both analytical and numerical considerations. A particular aspect of this is that optimal Bell inequalities are no longer separating hyperplanes. The reason is that the deterministic strategies that constitute the vertices of the local set are also the vertices of the network local set, and the facets that connect these vertices are the linear Bell inequalities of the local polytope (see Figure 3). Consequently, in network scenarios, one must consider nonlinear Bell inequalities. However, property (iii) stipulates that these nonlinear Bell inequalities can be restricted to polynomial expressions [27]. Furthermore, since the network local set is closed, every point outside the network local set upholds some (perhaps very small) degree of tolerance to noise.

Characterising network local correlations is a central problem in the field of network nonlocality, in the same way that characterising Bell-local correlations is the central problem in the field of Bell nonlocality [3]. A common practice in the latter is analysing whether correlations satisfy all Bell inequalities defined by the facets of the local polytope: the violation of any of those inequalities certifies that the correlations are nonlocal. In contrast, in the case of nonlocality in networks the violation of a Bell-like inequality can signal two different phenomena. On one hand, if there exists a guarantee of the underlying network structure, the violation can be interpreted in a way analogous to that of standard Bell inequalities (i.e. that the parties have employed non-classical resources to generate their correlations). This interpretation can be observed in Figure 3b. On the other, and as is shown in Figure 3a, there exists Bell-local correlations that do not satisfy the polynomial Bell inequalities that characterise the set of local correlations com-

patible with a given network. This means that network Bell inequalities can, in some cases, be interpreted also as witnesses of the underlying network structure.

In quantum models of network correlations, the independence of the sources leads one to associate an independent quantum state, ρ_i, to each of the sources. Each party performs a measurement $\{A^{(j)}_{a_j|x_j}\}$ on their respective shares. The correlations are then given by the Born rule

$$p(\bar{a}|\bar{x}) = \mathrm{Tr}\left[\left(A^{(1)}_{a_1|x_1} \otimes \ldots \otimes A^{(n)}_{a_n|x_n}\right)(\rho_1 \otimes \ldots \otimes \rho_m)\right], \quad (6)$$

where the respective Hilbert spaces are suitably ordered as given by the network configuration. Note that each state ρ_i can be entangled and that the tensor product structure $\rho_1 \otimes \ldots \otimes \rho_m$ indicates that there is neither entanglement nor classical correlation between the various states, as implied by the independence of the sources. Alike the network local set, the set of network quantum correlations is non-convex and closed (see Figure 3b). However, while it naturally contains the network local set, it is not the case that it also subsumes the local polytope (or vice versa): there exist correlations that, while being possible to generate in a network with quantum sources, cannot be generated by the same number of parties if they all share a global local variable, and vice versa[4].

Moreover, in analogy with no-signaling correlations generalising quantum correlations in standard Bell scenarios, one may consider what are the most general correlations that can arise in a given network. However, it turns out that the correlations implied by a theory-independent treatment of networks featuring independent sources are far less straightforward to characterise and require the introduction of concepts outside the scope of this review (see, for instance, Refs. [28, 29]). Despite this, in Section IV we discuss available methods for deriving constraints that follow exclusively from the structure of the network, irrespective of the nature of the sources.

C. Brief overview on network methods

While methods for analysing local, quantum and no-signaling correlations in standard Bell nonlocality greatly benefit from the convexity of these correlation sets [3], methods for analysing their counterparts in networks in general require significantly different approaches. To these purposes a variety of methods have been developed and investigated. They all come with different advantages and drawbacks, most notably concerning generality (ideally applying to any correlation in any network), strength (ideally constituting optimal correlation criteria) and practicality (ideally being computable with commonly accessible resources). While these methods will later be reviewed individually, we here provide a brief

	Disguised Bell nonlocality	Algebraic inequalities	Inflation	Explicit decompositions
Type of networks	Any	With condit. independence	Any	With condit. independence
Example	Triangle	Star	Triangle	Bilocal
Section	II A	II B 1, III A	III B	II B 2, II B 3

TABLE I. Selection of methods for characterising correlations that admit local variable models in networks discussed in this review, along with the features that identify the newtorks they are applicable to, and examples of such families.

overview of some selected methods accompanied with pointers to the specific sections where they are further discussed.

Firstly, Table I contains a selection of methods dedicated to studying network local correlations. Many approaches exist, which range from the explicit parametrisation of network local models [30, 31] to the use of hierarchies of relaxations of the sets of compatible correlations [32], or the analytical construction of network Bell inequalities [26, 30, 33]. The methods contained in the table, and additional ones, are collected in Sections II and III.

Secondly, Table II collects techniques that allow the characterisation not only of network local models, but also of quantum models in networks of the form of Eq. (6) and theory-independent network correlations. On one hand, network quantum correlations are particularly interesting since they encompass the network nonlocal correlations that are expected to be attainable in experiments. As in the case of network local correlations, there exist characterisations of very different nature, with examples being the use of functional analysis or infinite hierarchies of relaxations. These methods are contained, mainly, in Section III. On the other hand, theory-independent correlations are the most general correlations that can be generated in the network with space-like separation, and thus their study can be understood as the analysis of the limits that the network structure imposes in the correlations that can be produced in it. The techniques that characterise these correlations are reviewed mostly in Section IV.

II. ELEMENTARY NETWORK SCENARIOS

In this section we focus on the two quintessential examples of network Bell scenarios that arise in the simplest multipartite setting, namely that involving only three parties. There are only three qualitatively different network configurations based on three parties. (i) All parties are connected to a single source. This is the standard tripartite Bell scenario, which is not the focus of this review (see instead, e.g., Ref. [3]). (ii) One party is separately connected to each of the two other parties. This three-on-a-line scenario is illustrated in Figure 2b and it is commonly referred to as the bilocal scenario[5]. (iii)

---

[4] Examples of both cases will be discussed throughout the review. Eqs. (27)-(28) represent illustrations of the former, while the GHZ distribution of Eq. (25) is an example of the latter.

[5] The term "bilocal scenario" does not mean that the scenario only can be analysed in terms of network local models. Instead, it is common in the lit-

	Scalar extension	Classical inflation	Quantum inflation	Entropy cones	Covariance matrices	Finner inequalities
Local	✓	✓	✓	✓	✗	✗
Quantum	✓	✗	✓	✓	✓	✗
Theory-independent	✗	✓	✗	✓	✳	✓
Section	III A 3	III B 1, IV A	III B 2	III D, IV A	IV B	IV C

TABLE II. Methods for characterising network correlations of different natures and the types of correlations they can constrain. Given that the sets of local, quantum and theory-independent correlations compatible with a same network follow a strict inclusion order, a method designed for constraining correlations of one type can naturally constrain correlations of all the corresponding sub-types (i.e., higher up in the table). The ✓ signs in the table denote instead the existence of variants of the method that distinguish the selected correlations from more general ones (i.e., lower in the table). Analogously, the ✗ sign denotes that the method is not capable of singling out the selected correlations from more general ones. The ✳ sign denotes the fact that the ability of the method to constrain network correlations has only been proven in certain networks, and its applicability to the general case remains unknown.

All three parties are pairwise connected to each other. This network is illustrated in Figure 2c and is known as the triangle scenario. The bilocal scenario and the triangle scenario are the focus of this section. However, before discussing them in detail, we use these two scenarios in order to discuss the relationship between standard Bell nonlocality and network nonlocality.

A. Bell nonlocality disguised as network nonlocality

There are many, qualitatively different, manifestations of network nonlocality (see, for instance, the discussions in Refs. [27, 30, 34, 35]). This is reflected in one of the central questions of the field, namely: which forms of network nonlocality ought to be viewed as genuine to the network structure on which they arise? While all forms of network nonlocality elude the model in Eq. (5), some manifestations are more contrived in the sense that they trace back to standard Bell nonlocality. Here, following Ref. [27], we discuss a few typical examples of how standard Bell nonlocality may disguise itself as network nonlocality.

One manifestation of disguised Bell nonlocality is that only a part of the network demonstrates Bell nonlocality. As a simple example, consider the so-called bilocal scenario (recall, Figure 2b), in which parties Alice, Bob and Charlie obtain the correlations $p(a, b, c|x, y, z)$. Imagine that Alice and Bob decide to test the CHSH inequality without any regard to Charlie. Therefore, only the marginal distribution $p(a, b|x, y)$ will be relevant. In a network local model (5), in which the corre-

_____

erature to use the term bilocal scenario to refer to any realisation of the network in Figure 2b, while bilocal correlations maintains its standard meaning as correlations of the form given by Eq. (8).

lations take the form in Eq. (8), this distribution becomes

$$p(a, b|x, y) = \sum_c \left[\int d\lambda_1 d\lambda_2 \mu_1(\lambda_1)\mu_2(\lambda_2) \right.$$
$$\left. \times p(a|x, \lambda_1)p(b|y, \lambda_1, \lambda_2)p(c|z, \lambda_2) \right]$$
$$= \int d\lambda_1 \mu_1(\lambda_1)p(a|x, \lambda_1) \int d\lambda_2 \mu_2(\lambda_2)p(b|y, \lambda_1, \lambda_2)$$
$$= \int d\lambda_1 \mu(\lambda_1)p(a|x, \lambda_1)p(b|y, \lambda_1), \quad (7)$$

where in the second line we have used that $\sum_c p(c|z, \lambda_2) = 1$ since it is a probability distribution, and in the third line we have used that $\int d\lambda_2 \mu(\lambda_2)p(b|y, \lambda_1, \lambda_2) = p(b|y, \lambda_1)$. The result is precisely the standard local model in Eq. (2). Therefore, the bipartite marginal $p(a, b|x, y)$ admits a local model if the distribution $p(a, b, c|x, y, z)$ admits a network local model. Importantly, if $p(a, b|x, y)$ is nonlocal in the standard sense, then the distribution $p(a, b, c|x, y, z)$ has no network local model [30]. Thus, even extreme cases like the distribution obtained when Alice and Bob violate a Bell inequality and Charlie outputs a random number does not admit a network local model. This is an instance of standard Bell nonlocality disguised as network nonlocality. Notably, this is not restricted to our example of the bilocal scenario but readily extends to any network.

Disguised Bell nonlocality is not always as explicit as in the above example. Sometimes, it may be concealed and distributed among all the parties in the network. A prime example is the possibility to establish a correspondence between the correlations arising in a standard Bell scenario and correlations arising in a four-party network in the shape of a line [27, 30]. In the Bell scenario, Alice and Bob each use a random number generator to decide their inputs x and y (see Figure 4a). Then, they perform their measurements and obtain the correlations $p(a, b|x, y)$. These correlations can always be mapped to a network scenario in which the inputs are replaced by new parties, connected respectively to Alice and Bob via new bipartite sources. This yields a four-party line network in which parties have no inputs and x and y now act as the outputs of the new parties (see Figure 4b1). It is now possible to generate network correlations $p(x, a, b, y)$ that are network local (resp. network quantum) if and only if $p(a, b|x, y)$ is lo-
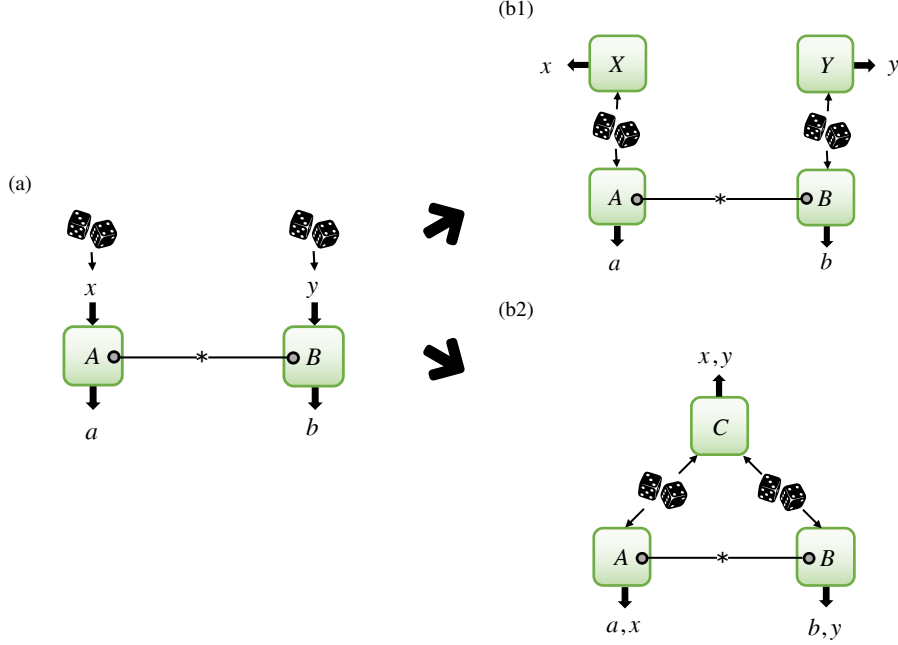
FIG. 4. Bell nonlocality disguised as network nonlocality. (a) The standard bipartite Bell scenario, in which the inputs are obtained from two independent random number generators. (b1) Four-party line network without inputs, isomorphic to the standard Bell scenario by reinterpreting the outputs of the random number generators as outputs of new parties. Note that this reinterpretation does not require any physical modification of a setup. (b2) Triangle network strategy to disguise the standard Bell scenario.

cal (resp. quantum). This follows from re-interpreting Alice's and Bob's private random number generators as sources of shared randomness from which the new parties in the network deterministically produce their output. Thanks to the shared randomness, Alice and Bob can then recover x and y respectively and proceed to perform their standard Bell inequality test (see Ref. [27] for a proof). More generally, such an argument applies to any standard Bell experiment or network Bell experiment with inputs in which the input of one party is replaced by a new party and a new source that connects only these two parties. Interestingly, this also highlights an important difference between network nonlocality and standard Bell nonlocality: the latter always requires the parties to have inputs whereas the above examples show that, with the source independence assumption, the former is possible even when no party has an input.

In a similar spirit, another type of map is possible between the Bell scenario and a network Bell scenario. This time, the inputs of both Alice and Bob are interpreted as the outputs of a single new party who is independently connected to Alice and Bob via two new sources. This leads to a triangle-shaped network (see Figure 4b2). Again, the argument relies on interpreting Alice's and Bob's random number generators as classical sources that distribute, say, one bit of shared randomness. This allows the new party, call it Charlie, to learn both the settings of Alice and Bob, which form his output, $\tilde{c} = (x, y)$. Similarly, since Alice and Bob have now received x and y from the respective new sources, they perform e.g. a test of the CHSH inequality and output $\tilde{a} = (a, x)$ and $\tilde{b} = (b, y)$. If the correlations $p(a, b|x, y)$ violate the CHSH inequality,

then the correlations $p(\tilde{a}, \tilde{b}, \tilde{c})$ do not admit a local model in the triangle network [27]. In the case of maximal quantum violation of the CHSH inequality, the network nonlocal distribution $p(\tilde{a}, \tilde{b}, \tilde{c})$ is known as the Fritz distribution, which is given explicitly in Eq. (A1). However, all the nonlocality stems from the entanglement distributed between Alice and Bob and therefore is, in essence, that of a CHSH inequality test. In fact, such disguised Bell nonlocality can elude network local models even when the three sources are subject to classical correlations [36]. Arguments of the type seen here extend to more general network scenarios as well.

Notice that these different examples of disguised network nonlocality can be combined and therefore apply to a vast variety of network scenarios. This further emphasises the importance of distinguishing more genuine forms of network nonlocality. Finally, it is interesting to point out that, within quantum theory, entangled measurements may not be necessary in order go beyond disguised network nonlocality. This plausibility is due to the fact that quantum theory features a phenomenon known as "nonlocality without entanglement", whereby there exist measurements whose eigenstates are product states but nevertheless cannot be reliably distinguished by two separated parties, even when classical communication is allowed [37].

B. The bilocal scenario

The bilocal scenario is the simplest network scenario. It is of particular relevance since it corresponds to the sce-

nario underlying entanglement swapping experiments (see e.g. Refs. [9, 38, 39]). It features three parties who receive random and independent inputs x, y and z, respectively, and produce outcomes a, b and c. The experiment gives rise to correlations contained in a conditional probability distribution of the form $p(a, b, c|x, y, z)$. The network features two independent sources: one emits a system shared between Alice and Bob, and the other emits a system shared between Bob and Charlie (see Figure 2b). Therefore, Alice and Charlie *a priori* share no correlations. Due to the lack of a shared source between Alice and Charlie, it is said that they are conditionally independent. Conditional independence refers to independence with respect to the inputs of a party and the systems received from the sources[6]. This means that regardless of the physical model considered in the bilocal scenario, the marginal distribution between Alice and Charlie must factor into $p(a, c|x, z) = p(a|x)p(c|z)$. Conditional independence is the key property exploited in many of the methods that analyse correlations in networks featuring it among different parties.

While nonlocality in the bilocal scenario has been studied earlier in the context of entanglement swapping, the concept of network local correlations in the bilocal scenario was first proposed in Ref. [26]. The probability distribution $p(a, b, c|x, y, z)$ admits a network local model (or, for simplicity, a bilocal model) if it can be written in the form

$$p(a, b, c|x, y, z) = \int d\lambda_1 \mu_1(\lambda_1) \int d\lambda_2 \mu_2(\lambda_2)$$
$$\times p(a|x, \lambda_1)p(b|y, \lambda_1, \lambda_2)p(c|z, \lambda_2). \quad (8)$$

This is Eq. (5) in the context of the bilocal scenario. In order to show that $p(a, b, c|x, y, z)$ admits a bilocal model, one must find an explicit decomposition of the form (8). In contrast, to show that $p(a, b, c|x, y, z)$ does not admit a bilocal model, one must show that no decomposition of the form (8) exists. This is often a nontrivial matter. Nevertheless, there are to date several approaches to detect non-bilocal correlations. On the one hand, one may opt for methods tailored for more general networks, such as the inflation technique (detailed in Section III B) or scalar extension (detailed in Section III A 3). On the other hand, one can attempt to construct explicit bilocal Bell inequalities, i.e. inequalities defined for probability distributions in the bilocal scenario that are satisfied by all bilocal correlations. Here, we focus on bilocal Bell inequalities and save the discussion of more general methodology, of which the bilocal scenario is a special case, for later.

Before proceeding, it is relevant to note that standard entanglement swapping protocols can enable a restricted form of nonlocality in the bilocal scenario. In such a protocol, Bob performs a standard entanglement swapping measurement [40] and then Alice and Charlie test the CHSH inequality

———————

[6] In the bilocal scenario, Alice and Charlie are conditionally independent because, even when their respective inputs and the systems sent by the sources are known, the joint distribution of their outcomes factorises. Note that, in contrast, Alice's and Charlie's outcomes are not independent if one conditions on Bob's outcome.

conditioned on just one of Bob's outcomes [41]. Motivated by benchmarking the success of entanglement swapping, rather than by nonlocality in networks, many such experiments have been performed before Bell nonlocality in networks emerged as a topic of its own (detailed in Section V). The focus of this section will, however, be on more general approaches to the bilocal scenario, which do not target direct tests of standard Bell inequalities between Alice and Charlie via entanglement swapping protocols.

1. Nonlocality and the Bell state measurement

The first example of a bilocal Bell inequality, that was conceptually different from testing a standard Bell inequality after entanglement swapping, was reported in Ref. [26] and later shown to constitute a special instance of a more general bilocal Bell inequality derived in Ref. [30]. Here we present the latter bilocal Bell inequality. Consider the bilocal scenario with Alice and Charlie each having two possible inputs, $x, z \in \{0, 1\}$ and each producing one of two possible outputs $a, c \in \{0, 1\}$. Bob has a fixed input and produces one of four possible outputs, which for simplicity are represented as two bits $b = (b_0, b_1) \in \{0, 1\}^2$. The correlations can therefore be written as $p(a, b, c|x, z)$ where the label y has been omitted. Next, define linear combinations of the probabilities as

$$I_+ := \frac{1}{4}\langle \hat{A}_+ B^0 \hat{C}_+ \rangle = \frac{1}{4} \sum_{a,b,c,x,z} (-1)^{a+b_0+c} p(a, b, c|x, z),$$

$$I_- := \frac{1}{4}\langle \hat{A}_- B^1 \hat{C}_- \rangle = \frac{1}{4} \sum_{a,b,c,x,z} (-1)^{a+b_1+c+x+z} p(a, b, c|x, z),$$
$$\quad (9)$$

with shorthand notations $\hat{A}_\pm := A_0 \pm A_1$ and $\hat{C}_\pm := C_0 \pm C_1$, where A_x and C_z are the observables of Alice and Charlie. In any physical model respecting the independence of the sources, one can use that $|\langle B^0 \rangle|, |\langle B^1 \rangle| \leq 1$ to obtain $|I_\pm| \leq \frac{1}{4}\langle |\hat{A}_\pm| \rangle \langle |\hat{C}_\pm| \rangle$, where the expectation values are over λ_1 and λ_2 respectively. Then, the Cauchy-Schwarz inequality gives

$$\sqrt{|I_+|} + \sqrt{|I_-|} \leq \frac{1}{2}\sqrt{\langle |\hat{A}_+| + |\hat{A}_-| \rangle}\sqrt{\langle |\hat{C}_+| + |\hat{C}_-| \rangle}. \quad (10)$$

The expressions under the square roots are similar to those encountered previously in our discussion of the CHSH inequality. In a network local model, it holds that $\langle |\hat{A}_+| + |\hat{A}_-| \rangle \leq 2$ and similarly for the second square root. This leads to the bilocal Bell inequality of BRGP [30], which reads

$$\mathcal{S}_{\text{BRGP}} := \sqrt{|I_+|} + \sqrt{|I_-|} \leq 1. \quad (11)$$

Observing $\mathcal{S}_{\text{BRGP}} > 1$ in an experiment implies that no bilocal model is possible[7]. Importantly, however, a violation of

———————

[7] As in the case of inequalities in Bell scenarios, in reality Eq. (11) represents a whole family of inequalities, generated by relabeling inputs and outputs

the BRGP inequality is not to be considered a necessary condition for non-bilocal correlations; it is highly plausible that there exists correlations that do not admit a bilocal model that nevertheless do not violate the BRGP inequality.

Moreover, Ref. [30] presents two more bilocal Bell inequalities, each for different values of Bob's available inputs and outputs. While these are mathematically similar to the above BRGP inequality, they are tailored for quantum protocols based on separable measurements and partial entanglement swapping measurements, respectively.

The most interesting quantum protocols that violate the BRGP inequality are based on entanglement swapping, i.e. a procedure in which Bob measures his two independent systems in an entangled basis in such a way that the state shared between Alice and Charlie becomes entangled. Let each source emit a copy of the maximally entangled state, i.e. Alice and Bob share $|\phi^+\rangle_{AB_1}$ and Bob and Charlie share $|\phi^+\rangle_{B_2C}$. Bob performs a complete Bell state measurement of his two particles. This paradigmatic measurement (famous from e.g. quantum teleportation [42] and dense coding [43]) is a projection in a basis $\{|\psi_{b_0b_1}\rangle\}$ of maximally entangled two-qubit states, namely

$$
\begin{aligned}
|\psi_{00}\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}, && |\psi_{01}\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}, \\
|\psi_{10}\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}, && |\psi_{11}\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}.
\end{aligned}
\tag{12}
$$

These states are also interchangably denoted $\{|\phi^+\rangle, |\psi^+\rangle, |\phi^-\rangle, |\psi^-\rangle\}$ respectively. Depending on his output, Bob renders Alice's and Charlie's initially independent qubits in one of these maximally entangled states, $|\psi_{b_0b_1}\rangle_{AC}$; but they cannot know which of the four states they hold due to the randomness in Bob's outcome. If Alice and Charlie each measure the anticommuting observables $A_0 = C_0 = \frac{\sigma_1 + \sigma_3}{\sqrt{2}}$ and $A_1 = C_1 = \frac{\sigma_1 - \sigma_3}{\sqrt{2}}$, one obtains the violation $\mathcal{S}_{\mathrm{BRGP}} = \sqrt{2} > 1$. Hence, no bilocal model can account for the predictions of quantum theory. In particular, it is not known whether a larger violation of the BRGP inequality is possible in quantum theory. Interestingly, however, it is also possible to violate the BRGP inequality (but to the lesser magnitude of $\mathcal{S}_{\mathrm{BRGP}} = 2^{1/4} > 1$) using only separable measurements for Bob [44]. Thus, perhaps surprisingly, a violation does not imply entanglement swapping in the quantum network. The reason for this traces back to disguised Bell nonlocality; separable measurements can be used to achieve standard Bell nonlocality between Alice-Bob and Bob-Charlie respectively, which to some extent enables violations of the BRGP inequality.

The Bell state measurement is a natural component of a quantum protocol for violating the BRGP inequality. However, as shown in Refs. [45, 46], any such protocol can fundamentally be understood in terms of the CHSH inequality. The key observation is that while the Bell state measurement itself indeed enables entanglement swapping, it

---

in the parties' measurements. A violation of any of them is a sufficient condition for incompatibility with a bilocal model.

acts in a separable manner on the level of the quantities $I_\pm$. Specifically $\sum_{b_0,b_1}(-1)^{b_0}|\psi_{b_0b_1}\rangle\langle\psi_{b_0b_1}| = \sigma_1 \otimes \sigma_1$ and $\sum_{b_0,b_1}(-1)^{b_1}|\psi_{b_0b_1}\rangle\langle\psi_{b_0b_1}| = \sigma_3 \otimes \sigma_3$. If the sources emit arbitrary two-qubit states ρ_{AB_1} and ρ_{B_2C} respectively, one finds that

$$
\mathcal{S}_{\mathrm{BRGP}} \leq \frac{1}{2}\sqrt{\mathcal{S}_{\mathrm{CHSH}}(A_0, A_1)_{\rho_{AB_1}}}\sqrt{\mathcal{S}_{\mathrm{CHSH}}(C_0, C_1)_{\rho_{B_2C}}},
\tag{13}
$$

where the quantities under the square roots are the CHSH parameters (3) obtained when Alice or Charlie tests the CHSH inequality with Bob and the latter performs the measurements σ_1 and σ_3. Refs. [45, 46] used this reduction to the CHSH scenario in order to prove that all pure entangled states can violate the BRGP inequality. Moreover, these works gave a general criterion (similar to the Horodecki criterion for the CHSH inequality [47]) for the magnitude of the quantum violation obtainable from a given pair of two-qubit states. Such an analysis, restricted to a particular class of two-qubit states, was derived earlier in Ref. [48].

An important aspect of Bell-type experiments is the quantification of nonlocal correlations, which can be achieved in several ways. In standard Bell scenarios two operational approaches are particularly notable: (i) determining the amount of isotropic noise that the entangled state can be exposed to before the correlations admit a local model (see e.g. Refs. [11, 49]), and (ii) determining the amount of superluminal communication with which a local model must be supplemented in order to simulate quantum correlations (see e.g. Refs. [50–55]). Exploration of the second approach in quantum networks is an interesting problem which has hitherto only been explored for particular protocols based on entanglement swapping in the bilocal scenario [56]. The former approach, however, straightforwardly extends to network scenarios. For the violation of the BRGP inequality, consider that each source emits a mixture of the maximally entangled state and isotropic noise,

$$
\rho_v^{\mathrm{iso}} := v|\phi^+\rangle\langle\phi^+| + \frac{1-v}{4}\mathbb{1},
\tag{14}
$$

where $v \in [0, 1]$ is the visibility of the state in each source. The critical visibility per source required to violate the inequality is $v_{\mathrm{crit}} = \frac{1}{\sqrt{2}}$. This follows immediately from the connection with the CHSH inequality, for which the critical visibility also is $v_{\mathrm{crit}} = \frac{1}{\sqrt{2}}$. Importantly, however, the BRGP inequality offers an advantage over a conventional entanglement swapping protocol (i.e. when Alice and Charlie test the CHSH inequality post-selected on one of Bob's outputs): in such a protocol, the critical visibility per source required to violate the CHSH inequality between Alice and Charlie is $v_{\mathrm{crit}} = 2^{-\frac{1}{4}}$.

2. Nonlocality and the Elegant Joint Measurement

Whereas quantum nonlocality in the bilocal scenario is expected to be qualitatively different from standard Bell nonlocality, we have seen that quantum protocols for the BRGP inequality essentially boil down to separate tests of the CHSH

inequality. This raises the question of whether non-bilocal correlations that bear no immediate resemblance to standard Bell nonlocality can be generated using entanglement swapping measurements different from the Bell state measurement.

One such entanglement swapping measurement was proposed in Ref. [57]. This, so-called Elegant Joint Measurement, is a projection of two qubits in a basis of iso-entangled states subject to the symmetry that if either qubit is lost, the four reduced states form a regular tetrahedron inside the Bloch sphere. The Elegant Joint Measurement, $\{|e_b\rangle\}_{b=1}^4$, is a projection onto the following basis of partially entangled states:

$$
\begin{aligned}
|e_1\rangle &:= \frac{1}{\sqrt{8}}(-1-i, 0, -2i, -1+i)^\dagger, \\
|e_2\rangle &:= \frac{1}{\sqrt{8}}(1-i, 2i, 0, 1+i)^\dagger, \\
|e_3\rangle &:= \frac{1}{\sqrt{8}}(-1+i, 2i, 0, -1-i)^\dagger, \\
|e_4\rangle &:= \frac{1}{\sqrt{8}}(1+i, 0, -2i, 1-i)^\dagger.
\end{aligned} \tag{15}
$$

This measurement was first introduced in Ref. [58] in the context of encoding quantum information in antiparallel spins. It also admits another operational interpretation as the optimal projective two-qubit measurement for discriminating the reduced states of either qubit [59]. This, together with its connection to symmetric informationally complete measurements [60], has enabled its generalisation to higher-dimensional systems [59]. Another generalisation extends it to a one-parameter family of two-qubit measurements in which the radius of the reduced-state tetrahedron can be continuously tuned. This allows one to construct measurements intermediate between the Elegant Joint Measurement and the Bell state measurement [31]. A relevant open problem consists in generalising the Elegant Joint Measurement to systems of many qubits.

The Elegant Joint Measurement enables quantum correlations in the bilocal scenario that are qualitatively different from those obtained through the Bell state measurement, and can be detected through the violation of bilocal Bell inequalities that do not, in any apparent manner, reduce to a known standard Bell inequality. Such bilocal Bell inequalities were presented in Ref. [31]. Consider a scenario in which Alice and Charlie have three inputs each and binary outcomes, while Bob has a single input and four outcomes. Bilocal correlations in this scenario are restricted by the following bilocal Bell inequality:

$$
\mathcal{S}_{\mathrm{TGB}} := \frac{1}{3}\left(\sum_{y=z}\langle B^y C_z\rangle - \sum_{x=y}\langle A_x B^y\rangle\right) \\
- \sum_{x\neq y\neq z\neq x}\langle A_x B^y C_z\rangle \leq 3, \quad (16)
$$

which is valid when all the (one-, two- and three-party) expectation values that do not appear in the inequality are zero[8].

This inequality has been proven under additional symmetry assumptions on the bilocal model and also been evidenced numerically for general bilocal models [31]. Nevertheless, a fully general analytical proof is still lacking. Note that while the inequality is linear, it is nevertheless a nontrivial constraint on bilocal correlations due to the restrictions imposed on the auxiliary expectation values. Notably Ref. [31] also proposed another, substantially different, bilocal Bell inequality for the same scenario. This inequality is nonlinear, does not involve auxiliary constraints and also serves as a quantum Bell inequality, i.e. it can also be used to limit network quantum correlations in the experiment.

The bilocal Bell inequality (16) is tailored to a quantum protocol based on Bob implementing the Elegant Joint Measurement. If the sources each distribute a singlet state and Alice and Charlie each measure the three Pauli observables, one obtains the violation $\mathcal{S}_{\mathrm{TGB}} = 4$. In contrast, no violation of the inequality is known based on the Bell state measurement. The strength of the known violation can be quantified in terms of its noise robustness when the sources instead emit two-qubit Werner states[9] [61]. The critical visibility for a violation is $v_{\mathrm{crit}} = \frac{\sqrt{37}-1}{6} \approx 0.847$, but the best known bilocal model for the full distribution has a critical visibility of $v_{\mathrm{crit}} \approx 0.791$. This indicates that tighter bilocal Bell inequalities could still be possible. Furthermore, the optimality of the quantum violation remains an open problem. Finally, an interesting observation is that a violation is obtained at a visibility per source which is lower than the visibility required for any of the swapped entangled states between Alice and Charlie to violate the CHSH inequality.

3. Explicit bilocal decompositions

The discussion of bilocal Bell inequalities is centered about detecting correlations that do not admit a bilocal model. Conversely, how can one detect correlations that do admit a bilocal model? One answer is to find all bilocal Bell inequalities, thus characterising the boundary of the bilocal set of correlations, and checking that they are all satisfied. This is, however, not a practical approach since no method is known for systematically finding these inequalities. A more straightforward answer is to construct an explicit bilocal decomposition (8). To this end, one benefits substantially from the previously discussed fact that the general bilocal model (8) can without loss of generality be taken as a deterministic bilocal model, i.e.

$$
p(a, b, c | x, y, z) = \sum_{\lambda_1, \lambda_2} p_1(\lambda_1) p_2(\lambda_2) \\
\times D^{\mathrm{A}}_{\lambda_1}(a|x) D^{\mathrm{B}}_{\lambda_1\lambda_2}(b|y) D^{\mathrm{C}}_{\lambda_2}(c|z), \quad (17)
$$

----

where p_1 and p_2 are arbitrary probability distributions and (D^A, D^B, D^C) are deterministic response functions, indexed for each of the parties by their relevant local variable. Denote the number of inputs and outputs for Alice by s_A and o_A: there are only $s_A^{o_A}$ deterministic response functions for Alice, and similarly for Bob and Charlie. Therefore, one can limit the cardinality of the relevant local variable accordingly. Consequently, the matter of simulating a given correlation is reduced to finding a suitable pair of probability distributions p_1 and p_2. Nonetheless, this can still not be straightforwardly evaluated due to the non-convexity of the problem.

It is possible to re-formulate the deterministic bilocal model (17) such that it becomes more handy to search for an explicit bilocal decomposition. This possibility is explored in Ref. [30], where bilocal models are analysed on the level of correlators; as opposed to probabilities. Although this approach readily extends to some larger networks and also to other input/output scenarios, we exemplify it here for a simple case and note that its generalisation is conceptually straightforward, albeit sometimes notationally cumbersome.

Let Alice, Bob and Charlie have binary outputs and associate the bit strings $\alpha = \alpha_1 \ldots \alpha_{s_A}$, $\beta = \beta_1 \ldots \beta_{s_B}$ and $\gamma = \gamma_1 \ldots \gamma_{s_C}$ which indicate the deterministic responses of each party for each measurement. We write $q_{\alpha\beta\gamma}$ for the probability of adopting this deterministic strategy. From these probabilities, define correlators through the discrete Fourier transform

$$e_{ijk} = \sum_{\alpha,\beta,\gamma} (-1)^{\alpha \cdot i + \beta \cdot j + \gamma \cdot k} q_{\alpha\beta\gamma}, \qquad (18)$$

where i, j and k also are bit strings. The correlators are now one-to-one with the probabilities. Following common practice, one may define $A_x = (-1)^{\alpha_x}$, $B_y = (-1)^{\beta_y}$ and $C_z = (-1)^{\gamma_z}$ and therefore write the correlators as expectation values

$$e_{ijk} = \langle A_1^{i_1} \ldots A_{s_A}^{i_{s_A}} B_1^{j_1} \ldots B_{s_B}^{j_{s_B}} C_1^{k_1} \ldots C_{s_C}^{k_{s_C}} \rangle. \qquad (19)$$

The correlators can now be divided into two sets: those that are fixed by the given $p(a, b, c|x, y, z)$ and those that are not. This handy distinction can be made because some correlators are physically relevant (e.g. $\langle A_1 B_1 C_2 \rangle$) whereas others are not (e.g. $\langle A_1 B_1 C_2 C_1 \rangle$). The formers correspond to the cases when i, j and k each contain at most one element "1" and the latters correspond to all other cases, which may now be viewed as degrees of freedom in the bilocal model. Restricting to these degrees of freedom, the conditions for a bilocal simulation can be expressed straightforwardly. Firstly, since the probabilities must remain positive, one reverses the Fourier transform in (18) and imposes that for $\forall \alpha, \beta, \gamma$

$$q_{\alpha\beta\gamma} = \frac{1}{2^{s_A + s_B + s_C}} \sum_{i,j,k} (-1)^{\alpha \cdot i + \beta \cdot j + \gamma \cdot k} e_{ijk} \geq 0. \qquad (20)$$

Secondly, to respect the independence of the sources, one imposes that Alice and Charlie are conditionally independent. This amounts to the nonlinear constraint

$$\forall i, k : \quad e_{i\bar{0}k} = e_{i\bar{0}\bar{0}} e_{\bar{0}\bar{0}k}, \qquad (21)$$

where $\bar{0} = 0 \ldots 0$. This method has been employed in Refs. [30, 31] to efficiently construct bilocal simulations of noisy quantum correlations. The main reason that the method is more efficient than brute-force search over p_1 and p_2 in Eq. (17) is that the degrees of freedom in the bilocal model are separated from the constraints implied by $p(a, b, c|x, y, z)$ and that the independence of the sources now appers as a quadratic constraint (21) on only some of the correlators. Its practical usefulness in larger networks is yet to be explored.

C. The triangle network

Going beyond the bilocal scenario, one may add a third, independent, source connecting Alice and Charlie, resulting in the triangle network captured in Figure 2c. This is the simplest nontrivial scenario that does not feature conditional independence between any pair of parties. Consequently, the main ideas leveraged in Section II B to analyse the bilocal scenario do not apply to the triangle scenario, and finding nonlocal correlations in this network (beyond Fritz's example, discussed in Section II A) is a difficult task. Following directly from the generic definition in Eq. (5), local models in the triangle network admit the form

$$p(a, b, c|x, y, z) = \int d\lambda_1 \mu_1(\lambda_1) \int d\lambda_2 \mu_2(\lambda_2) \int d\lambda_3 \mu_3(\lambda_3)$$
$$\times p(a|x, \lambda_1, \lambda_2) p(b|y, \lambda_2, \lambda_3) p(c|z, \lambda_1, \lambda_3). \qquad (22)$$

While we already have seen in Section II A that quantum theory can elude such models, we here focus on qualitatively different approaches to analyzing nonlocality the triangle network.

1. Two interesting distributions

An important example of quantum nonlocality in the triangle scenario was reported in Ref. [35]. There, the authors propose a quantum probability distribution incompatible with any triangle local model (22), which in no apparent way reduces to disguised Bell nonlocality, as was the case for the Fritz distribution [see Section II A and Eq. (A1)]. Such nonlocality is therefore, arguably, inherently different. However, the arguments suggesting that the distribution cannot be reduced to a form of disguised Bell nonlocality are only qualitative and intuitive. This stresses the need for formalising a rigorous concept of genuine network nonlocality that goes beyond failure to admit the network local model of Eq. (5).

The distribution is obtained from the following quantum experiment, where the parties always perform the same measurement (so one can omit the inputs x, y and z) with four possible outcomes, labeled by $a, b, c \in \{\bar{0}, \bar{1}_0, \bar{1}_1, \bar{2}\}$. Each source distributes the state $\frac{|10\rangle + |01\rangle}{\sqrt{2}}$, and every party measures in the

same basis, given by the vectors

$$|\bar{0}\rangle := |00\rangle,$$
$$|\bar{1}_0\rangle := c|01\rangle + s|10\rangle, \qquad |\bar{1}_1\rangle := s|01\rangle - c|10\rangle, \quad (23)$$
$$|\bar{2}\rangle := |11\rangle,$$

with c, s being real parameters that satisfy $c^2 + s^2 = 1$. The resulting probability distribution, P_{RGB4}, is explicitly given in Eq. (A3) in Appendix A. Notice that the measurement can be implemented in two steps. In a first step, the parties measure the operators associated to $\bar{0}$, $\bar{1}$, $\bar{2}$ where $\bar{1} := |01\rangle\langle01| + |10\rangle\langle10|$. Second, whenever $\bar{1}$ is obtained, the parties measure $\bar{1}_0$, $\bar{1}_1$, which are the only entangled measurement operators. This reformulation allows to interpret this experiment in terms of token-counting strategies, following Ref. [62]. The states $|0\rangle$, $|1\rangle$ can be interpreted as the absence or the presence of a token. Thus, the sources can be thought of as sending one token each, to the left ($|01\rangle$) or to the right ($|10\rangle$) parties they are connected to, in superposition. A party first measures the total number of tokens received, obtaining $\bar{0}$, $\bar{1}$ or $\bar{2}$ tokens, and then performs an additional entangled measurement. As each source distributes one token, the sum of token numbers obtained by all the parties is always three. As we further explain in Section III C, the rigidity property of classical token-counting strategies in the triangle, which demonstrates that there is essentially a unique way to classically simulate the distribution of the token counts, allows to prove that the resulting probability distribution is nonlocal for $0.785 \lesssim c^2 < 1$.

The key reason that allows the distribution P_{RGB4} to be proven nonlocal is that it is tailored to uphold convenient properties. While this is not always desirable, especially when considering experimental implementations, there exist proposals for experiments that demonstrate single-photon nonlocality based on the nonlocality of P_{RGB4} [63]. However, more often than not, one wants to assert the nonlocality of a probability distribution of interest, perhaps arising from more intuitive measurements in the triangle network. The price of adopting this perspective, i.e. to investigate the nonlocality of quantum correlations arising from more natural measurements (including fully entangled ones) is that rigorous proofs of nonlocality are difficult. Interestingly, however, the naively most natural candidate for the task, namely the Bell state measurement (12), performed by all three parties on shared maximally entangled states, leads to correlations that admit a triangle local model[10] [57]. Therefore, the quest for quantum triangle nonlocality motivates the use of other natural entanglement swapping measurements.

Refs. [57, 64] investigated the case of all three parties performing the Elegant Joint Measurement (15) on maximally entangled states emitted by all three sources. This leads to a

very symmetric distribution [64], defined by

$$p_3 = \frac{25}{256}, \qquad p_2 = \frac{1}{256}, \qquad p_1 = \frac{5}{256}, \quad (24)$$

where the subindex denotes the number of parties that obtain the same outcome. Based on its symmetry, the fact that its realisation requires complete entanglement swapping measurements and the inability of current approaches at simulating it in a network local model [64], this distribution is conjectured to be network nonlocal in a manner that cannot be reduced to disguised Bell nonlocality. Current techniques for identifying network nonlocal correlations [32] have, however, failed to prove the conjecture so far. Also, sophisticated numerics for constructing network local models (further discussed in Section VI D) have thus far failed to disprove the conjecture [65]. Therefore, the question remains open.

2. Inflation in the triangle network

Inflation, introduced by Wolfe et al. [32][11], is a powerful concept that enables the analysis of correlations in any network and, as will be discussed in Section VI G, also in more general causal structures. The triangle network is the simplest network for which the methods developed for characterising bilocal correlations, which mostly relied on the conditional independence of Alice and Charlie, fail. This makes the triangle network the natural scenario to apply inflation methods. In fact, many notable results in the triangle network, described below, have been obtained via inflation. Before summarising these results, we briefly illustrate the general principles behind inflation, which are applied in different ways for particular types of correlations throughout the review.

Inflation through example: Inflation is an attempt at proof through contradiction. If a distribution can be generated with some sources and measurement devices in a particular network scenario, then one can consider which kinds of distributions one would be able to generate if given access to multiple copies of such sources and measurement devices. The networks that are generated by arranging those copies are called inflations of the original network. The question of whether a probability distribution can be generated in a given network is translated into finding distributions compatible with the inflation networks (this is, satisfying all factorisation constraints implied by the inflated network and all symmetries consequence from elements in the inflated network being copies of each other) that, once marginalised over sources and measurement devices that reproduce (collections of) parts of the original network, coincide with the (product of) corresponding marginals of the original distribution. If this is achieved, the inflation is said to admit a compatible distribution. The original distribution cannot be generated in the original network if a compatible distribution in any of its inflations is proved not to exist.

---

[10] The distribution can be realised when the sources distribute a four-valued classical variable each, two of the parties output the value of the variable received that is shared with the remaining party, and this remaining party generates its output according to a deterministic strategy defined in advance.

---

[11] Note that inflation-like arguments appear earlier in the literature [28].
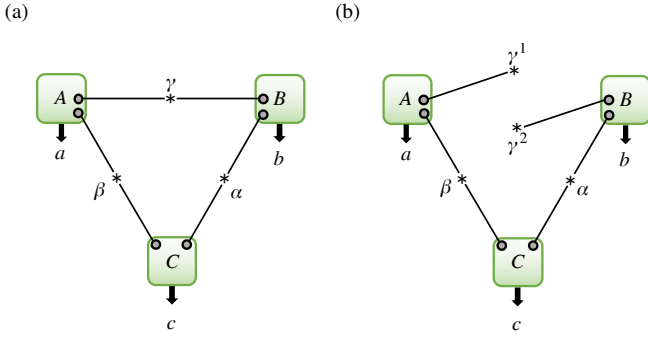
(a)   (b)

FIG. 5. Illustration of inflation in the triangle scenario. (a) is the triangle network, formed by three parties and three bipartite sources. By measuring the systems that are sent from their respective sources, the parties produce outcomes a, b and c according to the probability distribution P_{obs}^{ABC}. (b) A possible inflation considers two copies of the source γ. One sends a system only to Alice, and the other one sends a system only to Bob. By measuring their respective systems the parties produce outcomes according to the distribution P_{inf}^{ABC}. In both cases, Alice and Charlie receive the same systems from the same type of sources, so $P_{obs}^{AC} = P_{inf}^{AC}$. This is not the case for Alice and Bob, since $P_{inf}^{AB} = P_{inf}^A P_{inf}^B$ (which also equates to $P_{obs}^A P_{obs}^B$), while $P_{obs}^{AB} \neq P_{obs}^A P_{obs}^B$ in general.

Let us illustrate this reasoning in a simple example. Following Ref. [32], we show that a shared random bit (also called the GHZ distribution), given by

$$P_{GHZ} := \frac{1}{2}([000] + [111]),\,^{12} \qquad (25)$$

cannot be generated in the triangle scenario. Assume by contradiction that the distribution is realisable in the triangle network. This means that there exist sources α, β, γ and measurement devices A, B, C, arranged as illustrated in Figure 5a, that generate P_{GHZ}. Consider next the inflation in Figure 5b, where two copies of the source γ are available: one sends a system to Alice, and the other sends a system to Bob. The sub-network composed of Alice, Charlie, and the sources that connect to them, is the same in the original network and in the inflation since γ^1 is an exact copy of γ, and thus a distribution compatible with the inflation must satisfy $P_{inf}^{AC} = P_{GHZ}^{AC} = \frac{1}{2}([00] + [11])$. The same reasoning applies to the sub-network composed of Bob, Charlie and their sources, so $P_{inf}^{BC} = P_{GHZ}^{BC} = \frac{1}{2}([00] + [11])$. Moreover, given that, in the inflation, Alice and Bob are not connected by a common source, distributions that can be generated in the inflated scenario satisfy $P_{inf}^{AB} = P_{inf}^A P_{inf}^B$ (which, furthermore, can be identified with $P_{GHZ}^A P_{GHZ}^B$ because the individual sub-networks composed of each party and their respective sources are the same in the inflation and in the original scenario). These three conditions are incompatible: if there are no occurrences of $a \neq c$ nor of $b \neq c$ in the inflated network because

of the first two conditions, it must be the case that $a = b$ every time the three parties measure their shares. However, this is impossible if P_{inf}^{AB} is of a product form as per the third condition. One has reached a contradiction, which proves that the premise "P_{GHZ} can be generated in the triangle scenario" is false.

Note that the proof above has made no assumption about the nature of the sources α, β and γ. Therefore, the conclusion reached is independent of whether they distribute local variables, quantum states or states of more general probabilistic theories. In fact, the concept of inflation is applied in different ways depending on whether one wants to characterise local, quantum, or theory-independent correlations in networks. All these constructions are detailed later in the review, in Sections III B 1, III B 2 and IV A, respectively.

Results in the triangle from inflation: The original idea of inflation appears in Ref. [32], followed by a series of applications leading to novel results. Not only Ref. [32] recovers the known result that the distribution (25) does not admit a local model in the triangle [66], but also provides the first proof that the so-called W distribution, or P_W, defined as[13]

$$P_W := \frac{1}{3}([100] + [010] + [001]), \qquad (26)$$

does not admit such a model. Moreover, Ref. [32] explains how inflation can be used to transform consistency inequalities satisfied by all probability distributions into compatibility inequalities in networks, presenting the complete characterisation of the set of binary-outcome probability distributions that admit a particular inflation of the triangle scenario. This is, Ref. [32] provides a first complete description of a nontrivial relaxation[14] of the set of binary-outcome correlations that admit a triangle local model.

The characterisation above was later found in Ref. [34] not to admit quantum violations. This lead to the conjecture that, for probability distributions with no inputs and binary outcomes, the sets of distributions generated in the triangle network by classical variables and by quantum systems coincide. Furthermore, Ref. [34] provides, using larger inflations, the first inequalities for the triangle network (with no inputs and four outcomes per party) that admit quantum violations. While the original argument of Ref. [27] only established network nonlocality in the triangle for a small set of individual distributions, the inequalities in Ref. [34] allow one to witness network nonlocality in a region around these paradigmatic distributions, thus providing noise-tolerant witnesses of nonlocality.

Ref. [67] presents a modification of the original inflation argument that allows the analysis not only of classical network models, but also quantum network models of the type

---

12 We use the shorthand $[abc]$ for denoting the product of Kronecker deltas $\delta_{A,a}\delta_{B,b}\delta_{C,c}$, where A, B and C are the random variables that represent the outcomes of Alice's, Bob's, and Charlie's measurements, respectively.

13 This distribution corresponds to the coordinated action of all parties outputting 0 except one, which outputs 1.

14 Let us stress the fact that the complete characterisation is of a relaxation of the set of triangle local correlations. There may exist binary-outcome correlations that do not admit a model of the form of Eq. (22) but satisfy all the inequalities of the characterisation provided in Ref. [32].

of Eq. (6). This "quantum inflation" argument is employed to prove that the W distribution can neither be generated by measuring quantum states in the triangle scenario, and to find the amount of noise required in order for it to admit a quantum model. When considering distributions with inputs, the authors prove that the Mermin-GHZ distribution (see [67, Eq. (12)] for its definition) does not admit a quantum model in the triangle. For the two cases above, and exploiting duality

theory of semidefinite programming, they also find witnesses of incompatibility with triangle quantum models in the form of polynomial expressions. Finally, Ref. [67] finds bounds on well-known multipartite Bell inequalities in the classical and quantum triangle networks, illustrating the existence of gaps between different sets of correlations. A reproduction of the results for two famous multipartite Bell inequalities is given below:

Mermin's Inequality [68]

$$\langle A_1 B_0 C_0 \rangle + \langle A_0 B_1 C_0 \rangle + \langle A_0 B_0 C_1 \rangle - \langle A_1 B_1 C_1 \rangle \leq \begin{cases} 2 & \mathcal{L}^{\triangle}, \mathcal{L}^{\varhexagon} \\ 3.085^* & \mathcal{Q}^{\triangle} \\ 4 & \mathcal{Q}^{\varhexagon}, \mathcal{NS}^{\triangle}, \mathcal{NS}^{\varhexagon} \end{cases} \tag{27}$$

Svetlichny's Inequality [24]

$$\begin{aligned} \langle A_1 B_0 C_0 \rangle + \langle A_0 B_1 C_0 \rangle + \langle A_0 B_0 C_1 \rangle - \langle A_1 B_1 C_1 \rangle \\ - \langle A_0 B_1 C_1 \rangle - \langle A_1 B_0 C_1 \rangle - \langle A_1 B_1 C_0 \rangle + \langle A_0 B_0 C_0 \rangle \end{aligned} \leq \begin{cases} 4 & \mathcal{L}^{\triangle}, \mathcal{L}^{\varhexagon} \\ 4.405^* & \mathcal{Q}^{\triangle} \\ 4\sqrt{2} & \mathcal{Q}^{\varhexagon} \\ 8 & \mathcal{NS}^{\triangle}, \mathcal{NS}^{\varhexagon} \end{cases} \tag{28}$$

where $\triangle$ denotes the triangle scenario, $\varhexagon$ refers to the standard tripartite Bell scenario, and $\mathcal{L}$, $\mathcal{Q}$ and $\mathcal{NS}$ represent the sets of local, quantum and theory-independent correlations that can be generated in each of the scenarios, respectively.

The bounds with the asterisk are not necessarily tight, and thus it remains open whether they can be lowered by means of more computational power or using other techniques different from inflation. In particular, a very important problem that remains to be solved is having a clear understanding of the relations between correlations that do not admit network local models and correlations that are nonlocal according to other, standard definitions of multipartite nonlocality [24].

It is also possible to employ inflation arguments to establish theory-independent limits on the correlations that are achievable in networks, providing a notion of no-signaling in networks. In a nutshell, no-signaling in networks amounts to the inability of one party to learn the local operations of another (we defer a more detailed discussion on this concept to Section IV A 2). By inflating the network, necessary conditions on this inability can be established. Ref. [69] applies such ideas to the triangle scenario without inputs and binary outcomes, obtaining simple analytical constraints on no-signaling correlations in the network.

III. METHODS FOR GENERAL NETWORKS

The fauna of networks rapidly grows more diverse as one ventures beyond the three-party scenarios considered in the previous section. Here we review methods of considerable generality that apply to the characterisation of correlations in broad, or even fully general, classes of networks. Throughout

the discussion, we exemplify the various methods in concrete network scenarios.
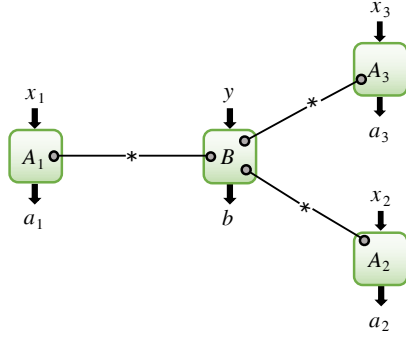
A. Conditional independence

Any network in which at least two parties do not receive shares from the same source upholds a degree of conditional independence. This property is independent from the physical model: by marginalising over some parties, other parties become independent. Evidently, most networks are of this type; notable examples are parties that are pairwise connected in a chain configuration or parties that are pairwise connected in a star configuration (see Figure 6). In this section, we discuss methods for constructing Bell inequalities for networks that feature conditional independence.

1. Variable elimination methods

We have already seen how to exploit conditional independence in order to construct a Bell inequality for the bilocal scenario. Specifically, the construction of the BRGP inequality (11) hinges on the ability of exploiting the conditional independence of Alice and Charlie by eliminating Bob through a suitable ansatz for a Bell inequality. Methods relying on this approach apply to a breadth of different networks and input/output scenarios.

A generalisation of the BRGP inequality, using methods that closely parallel the original work, was presented in Ref. [70]. In this work, explicit Bell inequalities were derived for networks featuring an arbitrary number of parties under
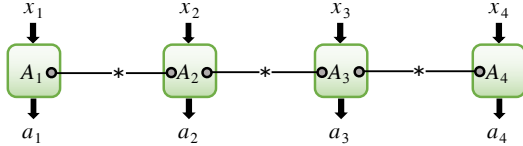
(a)

(b)

FIG. 6. Families of networks with conditional independences. (a) Star network with $m = 3$ sources. Upon tracing out the central party, the remaining become independent of each other. (b) Chain network with $m = 3$ sources. Tracing any non-extremal party separates the remaining in two sets that are independent of each other.

the restriction that they all have binary inputs and outputs. The method applies to any network that features conditional independence and is centered on identifying the largest set of mutually unconnected parties, i.e. a set of k parties such that no two of them receive a share from the same source. The bilocal scenario is a special case corresponding to $k = 2$. Another example is the network illustrated Figure 2d in which e.g. parties $\{A_2, A_5\}$ form a set of $k = 2$ conditionally independent parties. Once k has been determined, one can construct correlation quantities and a Bell inequality ansatz analogous to that encountered in the BRGP inequality (i.e. $I_{\pm}$ in Eq. (9) and the left-hand side of Eq. (11), respectively). In further analogy, one can then eliminate all but the k conditionally independent parties and exploit their independence to establish a network Bell inequality. The maximal quantum violation of these inequalities is based on maximally entangled states and Greenberger-Horne-Zeilinger (GHZ) states [71], and is found to also parallel that encountered for the BRGP inequality. Later, more generalisations were found by using the stabiliser formalism [72], which allows to find nonlinear Bell inequalities in the bilocal scenario that are tailored for non-maximally entangled states.

Another avenue for constructing network Bell inequalities by exploiting conditional independence is through the network extension method developed in Ref. [73]. The premise of this method is that one has an n-party network, in which all parties produce binary outputs, for which one already has a Bell inequality. Then, one can extend the network by adding one source that connects (say) party number n to a new party, $n + 1$, who has binary inputs and outputs. The added party is therefore conditionally independent of the first $n - 1$ parties.

This fact can be leveraged to construct a Bell inequality for the extended network by means of solving a quantifier elimination problem. While quantifier elimination is in general challenging, several concrete examples of solutions in particular networks are presented in Ref. [73]. The practical advantage of the network extension method is that it can be applied iteratively. In particular, one may start from a standard Bell scenario, for which the construction of a Bell inequality is straightforward, and iteratively expand it to a variety of different network configurations. Notably, this method is limited to expanding networks through the addition of bipartite sources. This limitation was, however, overcome in a subsequent work [74] which showed how the extension method could in principle be used to derive Bell inequalities for any network that does not feature a cycle. Notably, however, the network Bell inequalities obtained by this method are, typically, tailored for violation using quantum protocols based on separable measurements.

A third approach to network Bell inequalities is based on variable elimination, and is notably not restricted to binary outputs. This approach is based on interpreting network Bell inequalities as instances of a marginal problem [75]. It is known that also standard Bell inequalities can be viewed in this framework [76, 77]. To see how this interpretation arises, consider that Alice and Bob each perform measurements x and y with a_x and b_y denoting the outcomes. Is it possible to explain the probability distribution observed in the Bell experiment as the marginal distribution of a larger, unphysical, joint probability distribution $\vec{p} = p(a_1, \ldots, a_{|x|}, b_1, \ldots, b_{|y|})$ corresponding to a scenario in which Alice and Bob are able to perform all their respective measurements in each round? Notice that $\vec{p}$ contains several unobservable events because Alice and Bob each only perform one measurement per round in a Bell experiment. The answer to this question is equivalent to the existence of a local model and can be decided by eliminating the unobservable events from $\vec{p}$ using e.g. Fourier-Motzkin elimination [78], which can be evaluated as a linear program. The resulting constraints on the observable events in $\vec{p}$ (compatible with normalisation, $\sum_j p_j = 1$, and positivity, $\vec{p} \geq 0$) are precisely the Bell inequalities. Now, considering network Bell scenarios, one encounters a similar situation but with the key difference that it now features also nonlinear constraints on the larger probability distribution $\vec{p}$. These nonlinear constraints stem from conditional independence, i.e. that the marginal distribution of two unconnected parties factorises. As shown in Ref. [75], one can consider the unobservable events in $\vec{p}$ that feature nonlinearities and find the smallest number of free variables required to linearise the problem. Then it can, once again, be solved by means of Fourier-Motzkin elimination. In this manner, one can systematically construct Bell inequalities for networks with conditional independence.

2. Stars, chains, and beyond

An important example of a multiparty network is a star-shaped configuration. Such a star network is composed of

a central party, that is separately connected, via m bipartite sources, to m branch parties (see Figure 6a). Correlations in the network arise through the central party jointly measuring the m independent shares received from the sources and the branch parties locally measuring the single shares received from the corresponding sources. Nonlocality in star networks was first investigated in Ref. [33] in a scenario where the central party has a fixed input and produces an output consisting of m bits while each branch party has binary inputs and outputs. Naturally extending the ideas behind the BRGP inequality, this construction of network Bell inequalities is based on defining linear combinations I_j (specifically $j = 1, \ldots, 2^{m-1}$ different ones) of the observed probabilities in a manner analogous to that encountered in Eq. (9) for the bilocal scenario. This leads to the following Bell inequality for star networks

$$\mathcal{S}_{\text{TSCA}} \coloneqq \sum_j |I_j|^{1/m} \leq 2^{m-2}. \qquad (29)$$

In particular, for $m = 2$, this reduces to the BRGP inequality of Eq. (11). By considering complete, m-partite entanglement swapping, quantum theory can violate these inequalities regardless of the number of branches in the network. The violation emerges by each source distributing a maximally entangled two-qubit state and the central party jointly measuring the m independent qubits in a basis of GHZ states, i.e. a basis composed of states equivalent to $\frac{|0\rangle^{\otimes m} + |1\rangle^{\otimes m}}{\sqrt{2}}$ up to local unitary transformations. Conditioned on the output, this renders the m, initially independent, branch parties in a globally entangled state. Suitable local measurements then lead to the quantum violation $\mathcal{S}_{\text{TSCA}} = 2^{m-2}\sqrt{2}$ [33]. However, it is not known whether these are the maximal violations. Notably, if the sources instead emit isotropic states of the form of Eq. (14), a violation is possible when the visibility per source is at least $v = \frac{1}{\sqrt{2}}$, independently of m. Moreover, a modification of the inequality (29) has also been developed [33] which admits an identical noise tolerance per source but is tailored to be violated by a quantum protocol based on separable measurements; this modification may be thus viewed as wirings of many independent standard Bell inequality tests.

The star inequalities (29) have themselves been generalised in several ways. Ref. [79] extended the analysis to star networks that feature multipartite sources and showed that networks featuring such sources exhibit nonlocality with an exponentially improving tolerance to noise, in a manner analogous to that known for standard multipartite Bell inequalities [68]. Another generalisation is one in which a standard Bell inequality for binary-outcome scenarios is mapped to a Bell inequality for a star network. This map has the property that the local bound of the original Bell inequality also appears as the local bound of the associated network Bell inequality. Moreover, every quantum strategy in the original Bell scenario can be adapted into a quantum strategy in the network scenario which returns the same degree of violation [80]. This method allows one to first investigate a standard Bell inequality and then map the results to the more challenging network scenario. Such constructions may rightfully be viewed as examples of disguised Bell nonlocality, as in Section II A.

Star networks are also interesting from the point of view of activation of nonlocality. It is known that some entangled states that do not violate any standard Bell inequality (see e.g. Ref. [81] for a discussion of such states) can have their nonlocality activated when distributed in a star network [82, 83]. This situation is further discussed in Section VI C, but notably the analysis is framed in the context of standard multipartite Bell inequalities, i.e. it does not invoke the independence of the sources. It is therefore expected that star networks, that actively exploit the independence of the sources, should be able to reveal the nonlocality of even noisier states. In spite of this general belief, there presently exists no nontrivial example of quantum correlations in a network with independent sources that reveals the nonlocality of a Bell-local state.

Another natural type of network is shaped like a chain. These networks notably represent quantum repeaters [84, 85], on which most long distance quantum communication protocols are based. A chain network is composed of m sources and $n = m + 1$ parties in such a way that the source number i connects to parties number i and $i + 1$. An example is illustrated in Figure 6b. Beyond the bilocal scenario, nonlocality in chain networks was first investigated in Ref. [30] for the special case of $n = 4$ parties and further discussed in Ref. [27]. This was then extended to arbitrarily long chains in Ref. [86], where, in particular, the authors consider scenarios in which only the left-most and right-most party have binary inputs and outputs while all middle parties have a fixed input and four outputs. A natural realisation of this scenario is through complete entanglement swapping in a sequence for a two-qubit state, i.e. the entanglement initially shared between parties 1 and 2 is swapped sequentially throughout the chain so that it ends up being shared between party 1 and party n (see also Ref. [87]). The nonlocality of this scenario can naturally be detected via an end-to-end CHSH violation. However, for sources of visibility v and perfect swappings, the end-to-end visibility of the swapped state decreases exponentially as v^m, which makes this CHSH violation (or any protocol based on it, such as device independent quantum key distribution) hard for large m: the critical single source visibility $v = 2^{-\frac{1}{m+1}}$ scales exponentially. This analysis, based on standard Bell nonlocality, does not include the source independence assumption. Including it, Ref. [86] introduces linear network Bell inequalities which obtain the same visibility threshold. In contrast to this scaling, the network extension method (see Section III A 1) was applied to the chain network with binary inputs and outputs, obtaining a network inequality violated with a constant critical visibility per source of $\frac{1}{\sqrt{2}}$ [73]. The quantum protocol, however, does not exploit entanglement swapping, so the same features can be obtained via parallel violations of the CHSH inequality for each of the sources. The importance of this protocol raises the question of the existence of a network inequality exploiting the source independence assumption which (i) a violation would single out entanglement swapping (this is, that it cannot be violated with separable measurements) and (ii) would obtain a better asymptotic scaling in the visibility of the sources required.

Other, related, types of networks have also been consid-

ered in the literature. Ref. [88] developed Bell inequalities for a particular network structure that features m sources that each emit an m-partite state. In a subsequent work [89] it was shown that these networks can support quantum nonlocality when all the observers who hold more than one independent share of a system have no input. Ref. [75] derived an explicit Bell inequality and its quantum violation for a four-party network featuring two sources that each emits three-partite states. Finally, Ref. [70] presented Bell inequalities and quantum violations for polygon-shaped networks and a number of other network configurations.

3. Semidefinite relaxations

When speaking of numerical characterisations, semidefinite programming [90, 91] is a powerful tool for studying quantum correlations. Many different types of quantum correlations, arising in different scenarios with different assumptions, can be bounded by means of semidefinite programming relaxations (see e.g. Refs. [92–101]). In particular, the so-called Navascués-Pironio-Acín (NPA) relaxation hierarchy allows to compute increasingly strong necessary conditions for a given probability distribution to admit a quantum model in a standard Bell experiment [92, 93]. In essence, the NPA hierarchy relaxes the problem

$$
\begin{aligned}
\min \quad & f(p) \\
\text{s.t.} \quad & p \in \mathcal{Q},
\end{aligned}
\tag{30}
$$

where $f(p)$ is a linear functional of the elements of a quantum probability distribution $p = p(\bar{a}|\bar{x}) \in \mathcal{Q}$ [e.g. a Bell operator like those in Eq. (3) or Eq. (27)], by the hierarchy of problems

$$
\begin{aligned}
\min \quad & f(p) \\
\text{s.t.} \quad & \Gamma^{(k)}(p) \succeq 0,
\end{aligned}
\tag{31}
$$

where $k = 1, 2, \ldots, \infty$, and $\Gamma^{(k)}(p)$ are called moment matrices, built from expectation values of the operators characterising the parties' measurements. A rigorous definition of these moment matrices and examples of their use in the context of Bell nonlocality can be found in Refs. [92, 93, 102]. The sets of correlations $\mathcal{Q}^{(k)} = \{p : \Gamma^{(k)}(p) \succeq 0\}$ are relaxations of the set of quantum correlations, $\mathcal{Q}$, that satisfy $\mathcal{Q}^{(1)} \supseteq \mathcal{Q}^{(2)} \supseteq \cdots \supseteq \mathcal{Q}^{(\infty)} \supset \mathcal{Q}$ [93]. Thus, in practice, instead of solving the problem (30) which is difficult and, in general, undecidable [103, 104], one provides increasingly tight bounds on the actual solution by solving Eq. (31) for increasing k. Each of these problems, in contrast with Eq. (30), can be cast as a semidefinite program, which can be solved efficiently. In fact, these problems are common in a wide variety of disciplines [105, 106], and multiple software packages exist that generate and solve them [107–114].

While the theoretical arguments of the NPA construction are also satisfied by network quantum correlations, in general the additional constraints on the moment matrices that are imposed by the network structure break the connection between Eq. (31) and semidefinite programming. A simple example is

the already studied bilocal scenario (recall Figure 2b and Section II B). The fact that there is no source connecting Alice and Charlie implies that in the bilocal network any expectation value that does not involve operators of Bob factorises into the separate expectation values of Alice's and Charlie's operators. This property translates into nonlinear constraints between the elements of the $\Gamma^{(k)}(p)$, which are not admitted in semidefinite programming. Therefore, while Eq. (31) can be adapted to correlations in quantum networks by adding the necessary constraints, an efficient method for solving the problem does not exist in general.

It is noteworthy that semidefinite programs can be used to address quantum nonlocality in any network, without being based on the NPA hierarchy [67] (more on this later). Here, however, we focus on an adaptation of the NPA hierarchy that applies to networks which feature conditional independence, in order to bound the corresponding set of quantum correlations [115]. This adaptation has the form of Eq. (31), but it uses extended moment matrices that generalise those appearing in the NPA hierarchy. Typically, the rows and columns of moment matrices are indexed by a product of the operators associated to each party, and each cell in the matrix is associated to the expectation value of the product of the operator indexing the column and the (adjoint of the) operator indexing the row. In Ref. [115], not only products of operators are used to index the rows and columns of moment matrices, but also products of operators and scalar quantities associated to expectation values. For this reason, the method is called scalar extension. The fact that the rows and columns are indexed by nonlinear quantities results in cells inside the moment matrix also representing nonlinear quantities, so polynomial constraints in the probability distribution can be encoded in linear constraints between the corresponding matrix elements. These constraints are compatible with semidefinite programming, so one can again use efficient solving algorithms.

As an illustration, consider the bilocal scenario of Figure 2b, with Alice and Charlie performing two dichotomic measurements each, $\{A_0, A_1\}$ and $\{C_0, C_1\}$, respectively. A possible extended moment matrix is that generated by the set of operators $\{\mathbb{1}, A_0 A_1, C_0 C_1, \langle A_0 A_1 \rangle \mathbb{1}\}$, reading

$$
\Gamma = \begin{matrix} & \\ \mathbb{1} & \\ (A_0 A_1)^\dagger & \\ (C_0 C_1)^\dagger & \\ \langle A_0 A_1 \rangle^* \mathbb{1} & \end{matrix}
\begin{pmatrix}
\overset{\mathbb{1}}{1} & \overset{A_0 A_1}{v_1} & \overset{C_0 C_1}{v_2} & \overset{\langle A_0 A_1 \rangle \mathbb{1}}{v_3} \\
& 1 & v_4 & v_5 \\
& & 1 & v_6 \\
& & & v_7
\end{pmatrix}.
\tag{32}
$$

Per the original NPA prescription [92], every cell in the moment matrix above contains the expectation value of the operator characterising the row and the operator characterising the column under an unknown quantum state, i.e., $\Gamma_{R,C} = \mathrm{Tr}\left(\rho R^\dagger C\right) = \langle R^\dagger C \rangle_\rho$. Note that the last row and column are the only ones different from standard NPA, as they are indexed by the operator $\langle A_0 A_1 \rangle \mathbb{1}$, the identity operator multiplied by the scalar $\langle A_0 A_1 \rangle_\rho$.

It is easy to see that, in the matrix above, $v_1 = \langle A_0 A_1 \rangle_\rho$ and $v_3 = \langle \langle A_0 A_1 \rangle_\rho \mathbb{1} \rangle_\rho = \langle A_0 A_1 \rangle_\rho$, and that $v_5 = \langle \langle A_0 A_1 \rangle_\rho (A_0 A_1)^\dagger \rangle_\rho = |\langle A_0 A_1 \rangle_\rho|^2$ and $v_7 = \langle \langle A_0 A_1 \rangle_\rho \langle A_0 A_1 \rangle_\rho^* \mathbb{1} \rangle_\rho = |\langle A_0 A_1 \rangle_\rho|^2$. Therefore, two immediate linear constraints, that are not imposed by the

network structure but rather by scalar extension, are $v_1 = v_3$ and $v_5 = v_7$. More importantly, the structure underlying the bilocal scenario imposes that expectations involving only operators of Alice and Charlie factorise. In Eq. (32), the factorisation $\langle A_0 A_1 C_0 C_1 \rangle_\rho = \langle A_0 A_1 \rangle_\rho \langle C_0 C_1 \rangle_\rho$ is realised by the linear equality $v_4 = v_6$. Therefore, the factorisation can be imposed, via scalar extension, in a way amenable to semidefinite programming. Note that the example above is just an illustration. When applying scalar extension to a given problem, a general procedure consists in extending the corresponding moment matrix with as many columns as the smallest number of factors that allow to set equality constraints for all elements in the moment matrix that must factorise (see, e.g., Refs. [115, 116]).

While conceptually simple, scalar extension has proved itself useful in the analysis of correlations in networks with conditional independences. Indeed, small levels of scalar extension are capable (see Refs. [116, Chapter 5] and [117]) of identifying as incompatible with a bilocal model mostly all the distributions that are known to be incompatible due to the version of the BRGP inequality of Eq. (11) based on separable measurements. Yet, it remains unknown whether, upon the choice of a sufficiently large extension, one is able to discard all distributions incompatible with a network. Recent results in noncommutative polynomial optimisation [118] hint to a positive answer.

Scalar extension is employed in Ref. [115], in combination with semidefinite relaxations of sets of local-variable correlations [119], to demonstrate that the ability of measurement devices to create nonlocal correlations can be activated. In a tripartite grey-box scenario where the devices held by the parties are known to have non-unity detection efficiencies, it is known that detection efficiencies below $\eta = 2/3$ do not allow to certify the presence of standard Bell nonlocality in the two-input, two-output scenario[15]. However, by using scalar extension, Ref. [115] proves that violations of bilocality can be certified from the observed correlations for detection efficiencies down to $\eta = 0.5291$. Moreover, scalar extension has been employed in Ref. [31] for providing quantum bounds to bilocal Bell inequalities tailored for the Elegant Joint Measurement.

B. Inflation of networks

Inflation is a powerful concept that allows to set constraints on the correlations that can arise in any network. Its primary idea is to analyse the hypothetical situation where one has access to multiple copies of the sources and measurement devices that conform the network, and can arrange them in arbitrary configurations. By analysing the correlations that are

generated in these new configurations, the fact that the network elements are copies of those in the original network allows to translate conclusions obtained in the inflation networks back to the original one, providing a means of studying the correlations that are generated there. While this concept was briefly illustrated in Section II C 2, here we provide more details on how the concept of inflation materialises into systematic methods that constrain the classical and quantum correlations that can be generated in networks.

1. Classical inflation

Inflation was originally developed in Ref. [32] with the aim of solving the problem of causal compatibility: given a causal structure (of which the network scenarios discussed in this review are concrete instances) and a probability distribution over its visible nodes, determine whether the distribution can be generated in the structure. In the context of Bell nonlocality in networks, where the visible nodes of the structure represent the parties' outcomes after measuring their respective systems, this problem is equivalent to that of discerning whether an observed probability distribution admits a local model in the network, this is, if the given distribution admits a model of the type (5).

The method that Ref. [32] proposes for solving the causal compatibility problem was exemplified in Section II C 2 for the triangle network. We now discuss the general situation. One begins by assuming that the distribution under scrutiny, P_{obs}, can be generated in the network, i.e. that there exist random variables (corresponding to the sources) and deterministic functions of such variables (which correspond to the parties) that give rise to the target distribution. Then, one considers inflations of the network by adding copies of the sources and parties available, and connecting them in such a way that the each copy of a party is only connected to copies of the sources that connect to the original party in the original network.

The inflated network is analysed, and necessary constraints on compatible probability distributions are derived, which are later "translated" into constraints in the original network. This translation is possible because certain marginals of the distributions compatible with the inflated scenario, namely those which exactly reproduce (collections of) sub-networks of the original network, must exactly match (products of) marginals of the original distribution. For instance, in the so-called web inflation of the triangle scenario, illustrated in Figure 7, any marginal of the form $p_{\text{inf}}(a^{i,j}, b^{j,k}, c^{k,i})$ must match the original distribution under scrutiny, $p_{\text{obs}}(a, b, c)$, since the sub-network formed by parties $\{A^{i,j}, B^{j,k}, C^{k,i}\}$ and the local variables that connect to them, $\{\alpha^k, \beta^i, \gamma^j\}$, reproduce the original triangle scenario. Moreover, in addition to identifications with P_{obs}, the distributions that can be generated in the inflated network satisfy many other constraints that are derived from the intrinsic symmetries of the inflated network. This is the case of, for example, $p_{\text{inf}}(a^{1,1}, b^{1,1}, c^{1,2})$ in Figure 7. This marginal does not correspond to any marginal of the original distribution because the copies of the source β that

---

[15] The scenario exemplified in Ref. [115] contains one party that performs a single measurement, which effectively acts as a projection of the state shared by the remaining parties. Consequently, the detection efficiency required to demonstrate nonlocality is that required to demonstrate it with the projected states, which is given in Refs. [120, 121].
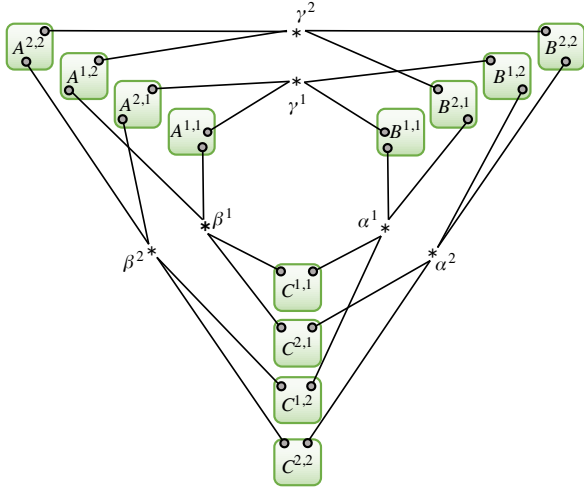
FIG. 7. The web inflation of the triangle network (for simplicity, the arrows that denote inputs and outputs are omitted). It is also regarded as the second step in the hierarchy of inflations defined in Ref. [122]. Note that copies of the individual random variables sent to the original parties are sent to various parties in the inflation. This is impossible in quantum mechanics due to the no-cloning theorem [123], and thus this inflation can be used to identify correlations that cannot be written as Eq. (22).

send systems to $A^{1,1}$ and $C^{1,2}$ are different. Nevertheless, it must coincide with all of the form $p_{\text{inf}}(a^{i,j}, b^{j,k}, c^{k,l \neq i})$ because all are created by the same sources and measurement devices, wired in the exact same way.

Finally, one needs to determine whether a P_{inf} satisfying all possible constraints of the type outlined above exists. Determining the existence of such a probability distribution can be formulated as a linear program, which can always be evaluated in an efficient manner, for instance by using the algorithms [107–109, 112, 114].

There are infinitely many ways in which a given network can be inflated, and finding just a single inflation for which there exists no distribution satisfying all the above constraints proves that the original distribution does not have a network local model. Moreover, it was later proved [122] that one can define a hierarchy of inflation tests that converges to the conclusive answer: if all the tests in such hierarchy are passed, the distribution is guaranteed to have a local model in the network. In addition, passing a particular test guarantees a maximum total variation distance between the distribution and the set of network local models. Thus, inflation is a powerful tool for determining that an observed probability distribution does not have a local model in a network.

2. Quantum inflation

It is not straightforward to generalise the inflation technique for the case when the sources distribute shares of joint quantum states. The reason for this is the fact that classical inflation crucially exploits the property that classical information

can be cloned, and this is impossible when the information is encoded in quantum systems [123]. In Ref. [67], this problem with no-cloning is circumvented by not copying the subsystems that the sources send to the parties as classical inflation does, but by considering copies of the sources themselves. In such a situation, the parties receive multiple copies of quantum states, and can choose on which copies they will perform their measurements. This scheme is illustrated in Figure 8. If a probability distribution can be realised in a given quantum network, then there exists a global quantum state on an inflated scenario and measurements whose statistics (i) are invariant under the permutation of Hilbert spaces corresponding to the different copies of a same source, and (ii) reproduce the original distribution when the parties make a choice of copies that reproduces the original network. One such possible state is the tensor product of copies of the sources, along with the original measurements applied on different subspaces as in Figure 8b. Finding this realisation in a given inflation is a task as hard as finding a realisation in the original scenario. However, the relaxed problem of finding a quantum state and measurements whose statistics are invariant under suitable permutations[16] can be solved via non-commuting polynomial optimisation [102]. The non-existence of a solution of the relaxed problem is thus a proof that the original distribution cannot be realised in the original quantum network.

Ref. [67] outlines with examples various families of problems for which quantum inflation can be used to provide a solution. These include the certification of a distribution being impossible to generate in a quantum network, the optimisation over the set of distributions that can be generated in a quantum network, the extraction of polynomial witnesses of incompatibility with quantum network models and, as a more practical application, the bounding of the information that a malicious party can obtain when eavesdropping on a quantum repeater. Moreover, quantum inflation is, at the time of writing, the only method for characterising quantum correlations in networks that is capable of analysing continuous-valued distributions.

The implementation of quantum inflation synthesises two different hierarchies: the one of inflations, and for each inflation, the non-commuting polynomial optimisation hierarchy used to determine whether a distribution admits such an inflation. There is evidence [116, Chapter 5] supporting that each hierarchy enforces constraints of different nature. Thus, in situations of limited computational resources, it is possible to prioritise one hierarchy over the other if prior knowledge or intuition about the distribution under scrutiny is available.

Quantum inflation, in the same way as the scalar extension method described in Section III A 3, can be modified [119] to characterise the sets of correlations that admit network local models. Indeed, by doing so quantum inflation generates semidefinite relaxations of the linear program that is defined by classical inflation at the same inflation level. This gives rise to an interesting characterization that, by combining

––––––––

[16] The relaxation is thus that in the inflation one does not require the tensor product structure of the state, only the permutation symmetries derived from it.
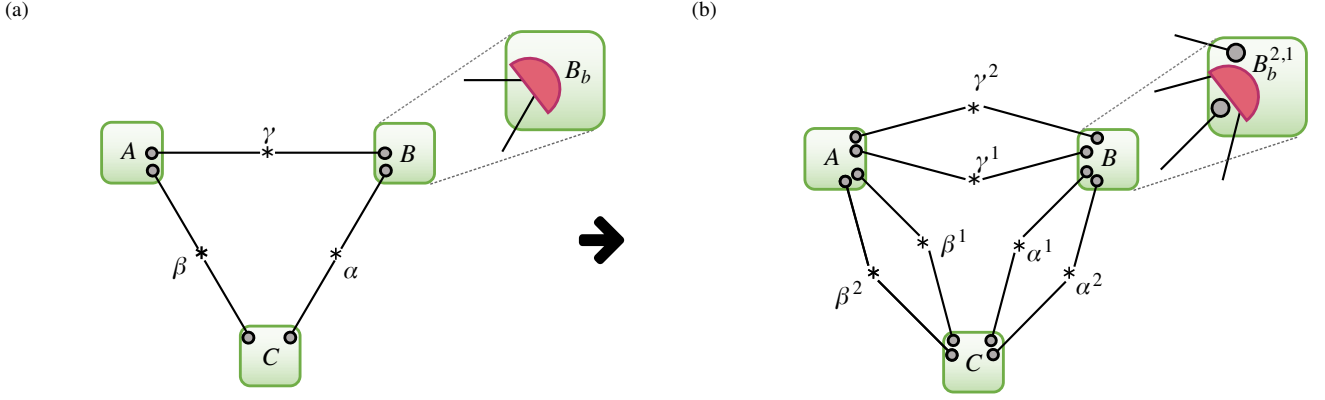
(a)

(b)

FIG. 8. Exemplification of quantum inflation in the triangle scenario (for simplicity, the arrows that denote inputs and outputs are omitted). (a) In the original scenario, the different outcomes b for Bob's measurement are produced by applying $\{B_b\}_b$, depicted as a semi-circle, on the shares α_B and γ_B. The outcomes of Alice's and Charlie's measurements are produced similarly. (b) In quantum inflation, one distributes n (in the case shown, $n = 2$) independent copies of the same states to the parties, which apply their measurements on different pairs of copies of the states they receive. In the case shown, the operators $\{B_b^{2,1}\}_b$ are applied on the shares α_B^2 and γ_B^1. In contrast with Figure 7, this inflation is not affected by no-cloning, and thus can be used to identify correlations that cannot be written as Eq. (6) in the triangle network.

and trading-off various relaxations, is capable of alleviating to some extent the computational load of analyzing network local models.

C. Rigidity-based arguments

Network nonlocality can also be proven from making use of the rigidity property of some classical strategies. In a first step, this property is employed to restrict the possible local explanations of a quantum distribution to a tiny family of local strategies. In a second step, this family is ruled out, exploiting its characteristics.

In Ref. [62], given a fixed network, a classical strategy associated to a correlation P_R is said to be rigid if it is the only classical strategy which obtains this correlation. In other words, rigidity can be thought of as a self-testing of classical strategies: the observation of the correlation P_R is sufficient to characterise the classical strategy which obtained it. This characterisation is defined up to a sole relabelling of the local variables and parties' outcome functions. Two families of strategies are shown to be rigid in some restricted classes of networks: token-counting and color-matching strategies. As will be seen in the following, this rigidity property can be used to show that certain quantum network correlations P admitting a rigid coarse-graining P_R are nonlocal.

Triangle nonlocality from three-token-counting rigidity: A primal example of proofs of nonlocality from rigidity arguments is the proof that the quantum-triangle distribution P_{RGB4}, introduced in Section II C 1 and explicitly given in Eq. (A3), does not admit a local model in the triangle network. To obtain it, each source distributes maximally entangled states $|\psi^+\rangle$, which can be interpreted as a unique token sent in superposition to the left ($|01\rangle$) and to the right ($|10\rangle$) parties it is connected to. Each party first measures in the same basis given by $\{\bar{0} := |00\rangle\langle00|, \bar{1} := |01\rangle\langle01| + |10\rangle\langle10|,$

$\bar{2} := |11\rangle\langle11|\}$, where $\bar{n}$ counts the number of received tokens, obtaining the correlation P_{count}. Whenever the outcome $\bar{1}$ is obtained (so one token is received), the party who obtained it performs a second, dichotomic measurement given by $\{|\bar{1}_0\rangle := c|01\rangle + s|10\rangle, |\bar{1}_1\rangle := s|01\rangle - c|10\rangle\}$.

The three-outcome distribution P_{count} can be obtained from a classical strategy. Indeed, since the operators $\{\bar{0}, \bar{1}, \bar{2}\}$ are diagonal in the token number basis, replacing the quantum sources by classical sources emitting the state $1/2(|10\rangle\langle10| + |01\rangle\langle01|)$ does not change P_{count} and provides a classical token-counting strategy. Moreover, the authors of Ref. [62] prove the rigidity of P_{count}, this is, they show that its observation self-tests the local variables and response functions in the triangle network. The proof is based on the fact that the total number of tokens seen by all parties is always three: in particular, if a change in the value of source α produces a change in the number of tokens detected by Bob, then it must be compensated in the number of tokens detected by Charlie in order to keep the total sum invariant (since Alice does not have access to α).

Hence, any hypothetical strategy simulating P_{RGB4} induces a classical strategy for P_{count}, which by rigidity must be the classical token-counting strategy outlined above. Therefore, in any hypothetical simulating strategy of P_{RGB4}, a new classical hidden variable can be introduced: the movement of each individual tokens. For instance, postselecting on the case where all parties obtain $\bar{1}$, the tokens of the source must be sent incoherently in the clockwise ($t = \circlearrowright$) or in the anti-clockwise ($t = \circlearrowleft$) direction, as shown in Figure 9. Introducing the (unknown) joint probability distribution of this hidden variable with the parties output

$$q(a, b, c, t) = \Pr(a, b, c, t|\text{all parties measure } \bar{1} \text{ token}), \tag{33}$$

it is possible to explicitly compute some of the marginals of $q(a, b, c, t)$, expressed in terms of the probability distribution
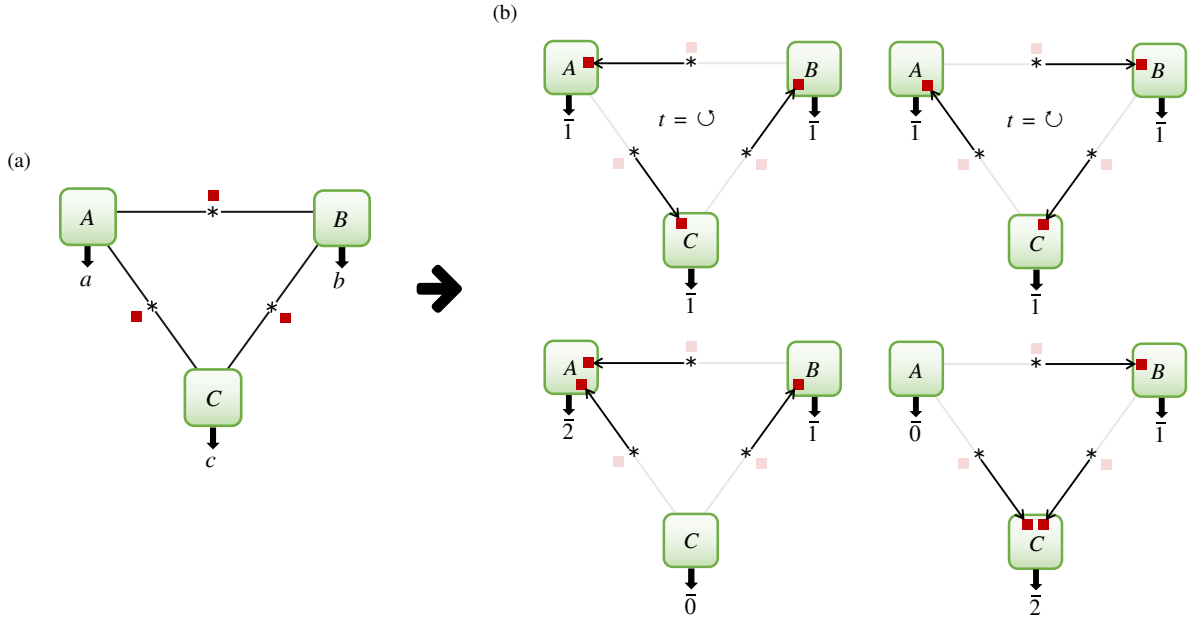
FIG. 9. Token counts in the uniform token-counting strategy in the triangle scenario. (a) Each source uniformly distributes a unique token (red square) to one of the two parties it is connected to, either in a quantum way, using states of the form $|\psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$, or a classical way, distributing states of the form $\frac{1}{2}(|10\rangle\langle10| + |01\rangle\langle01|)$. (b) Each party first measures the total number $\bar{0}, \bar{1}, \bar{2}$ of tokens received, giving rise to the distribution P_{count}, which is the same for classical and quantum strategies. P_{count} is a rigid distribution, so this is the unique classical strategy that is capable of generating it in the triangle network. In the quantum counting strategy, whenever $\bar{1}$ token is detected by a party, that party further measures $\{\bar{1}_0, \bar{1}_1\}$ to measure where the token comes from, obtaining the full correlation P. When all parties obtain $\bar{1}$, the token trajectories are undetermined between $t = \circlearrowleft$ and $t = \circlearrowright$ (top row). In all other cases, the token trajectories are fully determined by the token counts (bottom row).

P_{RGB4}. By proving that those marginals are incompatible with any valid probability distribution for some specific choices of the operators $\{\bar{1}_0, \bar{1}_1\}$, the authors of Ref. [62] demonstrate that P_{RGB4} does not admit a local model in the triangle.

Notice a key difference between classical and quantum token-counting strategies: in the quantum strategy, when all parties measure $\bar{1}$ token, the global state is a coherent superposition of the tokens turning clockwise and anti-clockwise. This is, once all parties obtain one token, the global state of the sources is left in the state:

$$\text{Classical token counting}: \frac{1}{2}\left(|\circlearrowright\rangle\langle\circlearrowright| + |\circlearrowleft\rangle\langle\circlearrowleft|\right)$$

$$\text{Quantum token counting}: \frac{1}{\sqrt{2}}\left(|\circlearrowright\rangle + |\circlearrowleft\rangle\right).$$

Hence, in the quantum case, one observes the creation of a global coherent superposition between all parties through local measurements in the network. This is the core feature that rigidity-based methods exploit, and what ultimately allows these methods to identify network nonlocality.

Network nonlocality from token-counting rigidity: The method described above can be extended to a large class of networks beyond the triangle. More precisely, a generic classical token-counting strategy involves sources distributing an arbitrary fixed number of tokens, and parties measuring the total number of tokens they receive, obtaining the token count distribution P_{count}. Such strategies are proven to be rigid in all networks where any pair of parties shares at most one source. Corresponding quantum token-counting strategies are obtained by allowing superposition in the tokens sent by the sources, and measuring the provenance of the tokens using entangled measurements. Ref. [62] generalises the initial triangle example outlined above, which was first introduced in Ref. [35], to prove the nonlocality of quantum token-counting distributions in the four-partite network where every pair of parties except $\{A_1, A_2\}$ share a bipartite source, and in all ring-type networks, where the every pair A_j, A_{j+1} shares a bipartite source and $A_{N+1} = A_1$.

Network nonlocality from color-matching rigidity: A second class of strategies is known to be rigid: these are known as the color-matching strategies. In a quantum color-matching strategy, each source picks a color from a pre-established set, in superposition. In the triangle network, the sources can for instance distribute the state $(|00\rangle + |11\rangle + |22\rangle)/\sqrt{3}$, where $0, 1, 2$ represent colors. A party first checks if all her neighbouring sources sent to her the same color, in which case she outputs that color. In the triangle, the measurement performed by each of the parties would correspond to rank one projectors over $\{|00\rangle, |11\rangle, |22\rangle\}$, with an extra no-color-match output, corresponding to the projector over the remaining dimension-six vector space. This gives the color-matching distribution P_{match}. At a second step, and in analogy with the procedure for token-counting, the no-color-match output can be refined by asking the parties to perform a second measurement using refined projectors to reveal information about the pattern of the received col-

ors. In the triangle, an example measurement would be $\{(|01\rangle \pm |10\rangle)/\sqrt{2}, (|02\rangle \pm |20\rangle)/\sqrt{2}, (|12\rangle \pm |21\rangle)/\sqrt{2}\}$.

As in the token-counting case, P_{match} can be obtained from a classical color-matching strategy, in which there is no superposition of colors in the states or measurements. Using techniques based on Finner inequalities (which we describe in Section IV C), Ref. [62] proves color-matching strategies to be rigid in a number of networks, following the same procedure as for token-counting, just by replacing P_{count} with P_{match}. This allows to find network nonlocal quantum color-matching strategies in the four-partite network where $\{A_1, A_2, A_3\}$ and $\{A_1, A_2, A_4\}$ share each a tripartite source and $\{A_3, A_4\}$ share a bipartite source, and in all networks where every pair of parties shares a bipartite source. Moreover, Ref. [62] establishes a connection between network nonlocality and graph coloring, which is used to prove the nonlocality of quantum color-matching distributions in all networks where every party is connected to a different pair of sources and every possible pair of sources connects to a party. For this, the parties are asked to check that two neighbouring sources have distinct colors. When this is the case for all parties, it indicates a proper coloring of the graph of sources. These measurements project the joint state of all sources to a coherent mixture of all possible proper colorings, which then leads to network nonlocality.

D. Entropic characterisation

Network correlations can also be analysed in terms of the entropies of those variables that are jointly measurable in the network. This approach, which was first employed in the bipartite Bell scenario in the late 1980s [124], has been found useful in the study of network correlations for two main reasons. Firstly, variable entropies are insensitive to their cardinalities. In the analysis of correlations in networks, this implies that the conclusions derived from this approach hold independently of the number of outcomes. Secondly, as will be shown later on, some constraints imposed by the network that are nonlinear in the elements of the probability distribution become linear in the entropic picture, enabling the use of numerical tools that cannot be leveraged in the direct study of compatible correlations. The entropic approach to the study of correlations is reviewed in Ref. [125], covering not just network correlations but also correlations in more general scenarios. While referring the reader to Ref. [125] for more detailed discussions, we here survey the main ideas and advances concerning this approach in network nonlocality.

In the same way that the primary object in the probability picture is the set of (classical, quantum or theory-independent) distributions compatible with a network structure, the primary object in the entropic picture is the so-called entropy cone. For a fixed number of random variables n, the entropy cone is the closure of the set of all vectors $v \in \mathbb{R}^{2^n - 1}$ of the form

$$v = \big(H(X_1), H(X_2), \ldots, H(X_1 X_2), H(X_1 X_3),$$
$$\ldots, H(X_1 X_2 \ldots X_n)\big), \quad (34)$$

where H is an entropy function. In the traditional approach to the entropic analysis of causal structures, which assumes that all nodes in the structure are classical random variables, the entropy function used is the Shannon entropy, given by

$$H(\mathcal{X}) = - \sum_{x \in X} p(x) \log_2 p(x), \quad (35)$$

where $p(x)$ is the distribution of events for a subset of variables $\mathcal{X} \subset \{X_1, \ldots, X_n\}$ whose range is X. Other entropies can be used alternatively, such as the Rényi entropies [126] or the Tsallis entropies [127]. In particular, when one deals with random variables resulting from quantum measurements, a popular choice is the von Neumann entropy, $H(\mathcal{X}) = - \text{Tr}(\rho_{\mathcal{X}} \log \rho_{\mathcal{X}})$.

The boundary of the entropy cone may be described by potentially infinitely many linear inequalities, and thus it is difficult to characterise in general. This is the reason why outer approximations of the cone are used instead. The most common approximation is the Shannon cone, which is the polyhedral cone characterised by three basic constraints that every entropy function satisfies, namely positivity (the uncertainty about a variable is a non-negative quantity), monotonicity (the uncertainty about a set of variables is not smaller than the uncertainty about any subset of it), and submodularity (the conditional mutual information is positive)[17]. The Shannon cone coincides with the entropy cone only for $n \leq 3$. For $n > 3$, the Shannon cone is an outer approximation of the entropy cone. All those inequalities that describe the entropy cone but cannot be derived from the conditions above are called non-Shannon-type inequalities [129–133].

In order to deal with correlations in networks, the Shannon cone must be further constrained in order to account for the network structure. In what follows, we review how this is done for local and quantum correlations. A brief remark on the entropic analysis of more general correlations in networks is deferred to Section IV A.

1. Local correlations

In the case of characterising local models in networks, the condition of locality implies that all variables are jointly measurable, and thus one can define a probability distribution over all events [10]. Therefore, the corresponding Shannon cone can be built by imposing the Shannon constraints (recall, positivity, monotonicity, and submodularity) to all the possible subsets of variables in the network. Then, the constraints implied by the network structure are imposed by the entropic analogue of the compatibility condition of a probability distribution with a network. While in the probability picture compatibility was captured by Eq. (5), in the entropic picture this

---

[17] The corresponding mathematical formulations of these properties can be found, for instance, in [125, Section 2(a)(i)] or in [128, Appendix I].

notion is captured by

$$I(\lambda_j : \lambda_k) = 0,$$
$$I(A_j : A_k | \lambda_{i \to j,k}) = 0, \qquad (36)$$

where $I(X : Y | Z) = H(X | Z) + H(Y | Z) - H(X, Y | Z)$ is known as the conditional mutual information, λ denotes the sources, A denotes the parties, and $\lambda_{i \to j,k}$ is a source that is connected to parties A_j and A_k. The first equation captures the independence of different sources, while the second establishes that all correlations between two parties are established exclusively via the sources that connect to them. It is important to note that, while Eq. (5) gives rise to equality constraints that are polynomial in the elements of the probability distribution, Eq. (36) is linear in the variables' entropies. Indeed, the problem of whether a given entropy vector lies within the set defined by the Shannon cone and Eq. (36) can be cast as a linear program, which can be efficiently solved using numerical methods [107–109, 112, 114]. For example, the entropic constraint that imposes the independence of the sources in the bilocal scenario where Alice and Charlie can perform, respectively, m_A and m_C measurements is $H(A_1, \ldots, A_{m_A}, C_1, \ldots, C_{m_C}) = H(A_1, \ldots, A_{m_A}) + H(C_1, \ldots, C_{m_C})$ [134].

In order to obtain constraints on the entropies of the variables an experimenter has access to, one must marginalise the entropic cone over the variables that represent the sources of shared randomness. This procedure can be addressed with Fourier-Motzkin elimination [135, 136]. An explicit example of this procedure, that derives an entropic inequality in the triangle scenario, can be found in [27, Proof of Lemma 2.14]. Another example, in the bilocal scenario, is given in Ref. [128]. This is the inequality

$$H(A_0, C) \leq H(A_0, B) + H(C | A_1, B), \qquad (37)$$

where A_0 and A_1 are the random variables corresponding to the two possible measurements that Alice can choose to perform. This inequality is framed in a scenario where Alice has a choice of measurement while Bob and Charlie do not. While the distributions that arise in this scenario all admit a Bell-local model [recall Eq. (2)], Ref. [128] finds distributions that violate Eq. (37). As a final example, the entropic approach is used in Ref. [137] to analyse the freedom of choice loophole in the bipartite Bell scenario, by characterising the correlations compatible with the network that models an experiment where the loophole is present.

The fact that the entropic approach considers approximations of the entropic cone for $n > 3$, even with the use of non-Shannon constraints, prevents it of being capable of identifying all the observations that can be produced by network local models. Also, on a more fundamental perspective, the full probability distribution represents a larger amount of information than the variables' entropies. While it is possible to identify nonlocal network correlations using the entropic approach (an example of this is the case of the GHZ distribution of Eq. (25) violating [27, Lemma 2.14]), it is not clear that a precise characterisation of the corresponding sets of correlations can be achieved by constraining the variables' entropies.

For example, the W distribution of Eq. (26) has not been identified as incompatible with a local model in the triangle using entropic techniques [125], while other methods such as inflation [32] have been capable of doing so.

2. Quantum correlations

In contrast with local models, quantum correlations are characterised by the fact that there exists sets of observables that cannot be jointly measured. Examples of these sets are the set of different measurements performed by a same party, or the systems input to a measurement device and the variable representing the outcome of its measurement. Therefore, when using the entropic approach to characterise quantum correlations in networks, one cannot set the basic Shannon constraints on the joint entropies of any set of variables, but only on the entropies of jointly measurable ones [138], for which a joint quantum state can be defined. Moreover, in further analogy, the entropy function now corresponds to a quantum entropy, such as the von Neumann entropy. These new functions, however, do not always satisfy the monotonicity constraint. For the cases where monotonicity does not hold, a weaker notion is imposed [125, 130, 139], which instead ensures the non-negativity of the sum of conditional entropies.

Finally, in order to relate the entropies of different jointly measurable sets, one makes use of data-processing inequalities, that capture the notion that the information content of a quantum system cannot be increased by means of local operations. An example of this type of inequalities is the upper bound of the mutual information of two outcome variables by the mutual information of their parent quantum states. More precisely, in the example of the triangle scenario of Figure 2c, the data-processing inequalities would take the form $I(A : B) \leq I(\beta_A \gamma_A : \alpha_B \gamma_B)$, and analogously for the pairs $\{A, C\}$ and $\{B, C\}$.

The quantum entropic approach is developed in Ref. [138] and used to recover and generalise Information Causality, an information-theoretic principle satisfied by quantum mechanics [140]. In the same work, the quantum entropic approach is also applied to characterise the correlations achievable in the triangle scenario, observing that the entropic cones obtained by using just Shannon-type constraints is the same in the quantum and classical cases. The classical cone has been since then further shrunk through the use of non-Shannon inequalities [133], and initial steps have been taken in the case of the quantum cone [130].

IV. TOWARDS THEORY-INDEPENDENT NETWORK CORRELATIONS

In all the situations studied thus far, we have phrased compatibility relative to a network with the sources sending systems described by a particular theory, be it classical [where we looked for compatibility with Eq. (5)] or quantum [where the corresponding model is Eq. (6)]. A very interesting avenue is removing as many assumptions as possible in the the-
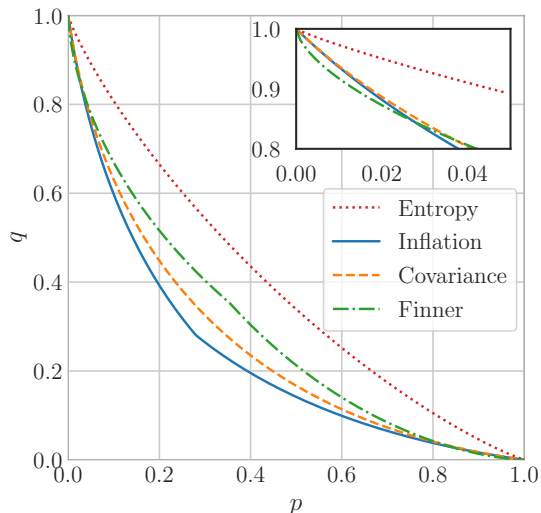
FIG. 10. Application of all the methods discussed in Section IV for deducing necessary conditions for the compatibility of a family of probability distributions with the triangle network. The considered family is $P_{p,q} \coloneqq p[000] + q[111] + \frac{1-p-q}{6}([001] + [010] + [011] + [100] + [101] + [110])$, with $p \geq 0$, $q \geq 0$, $p + q \leq 1$. The different curves denote the boundaries in the $p - q$ plane above which each method is capable of certifying that the distribution cannot be generated in the triangle scenario. The dotted red curve is implied by the entropic condition (38). The solid blue curve is implied by the inflation-based arguments behind the condition (39). The dashed orange curve is implied by the covariance matrix approach described in Section IV B (also analytically derived in [141, Figure 1]). The dot-dashed green curve is given by the Finner inequality (43). The inset depicts a zoom-in on the region $p \in [0, 0.05]$, where the boundaries illustrating the different methods intersect.

ory that describes the systems sent by the sources, leaving only those that are strictly necessary to have a reasonable notion of a system and transformations over it (see, for instance, Refs. [28, 29, 142, 143]), and thus exploring the most general correlations that can be achieved in a network and the constraints on compatible correlations that arise exclusively from the network structure. These correlations are the focus of this section. There are several approaches to the matter, and those that we review below are compared through application to a simple problem illustrated in Figure 10[18]. On one hand, in Section IV A we consider general techniques that have been described throughout the review, which now are applied without any regard to the particular theory that describes the states and measurement devices, only focusing on the constraints imposed by the network structure. On the other, in Sections

IV B and IV C we describe two interesting methods that, while originally tailored to be satisfied by network local models, they have been eventually proven to hold also by network quantum correlations and in several generalised operational-probabilistic theories.

A. Theory-independent network constraints

Classical sources, represented by local variables, and quantum sources, represented by quantum states, may be viewed as particular instances of more general, theory-independent, sources. The latters are restricted by the no-signaling principle: operations on one subsystem cannot instantaneously relay information to another subsystem. This principle stems from relativity and it is also the cornerstone behind the notion of distinct parties and is therefore the minimal requirement for a set of space-like separated observations to be compatible with a given network. This naturally leads to the question of how one can determine the most general, theory-independent, correlations relevant in networks and how to understand the role of the no-signaling principle.

1. Entropic approach

One approach, which extends ideas encountered previously in Section III D, is based on using entropies of the observed correlations in order to capture theory-independent constraints. Entropic formulations of such correlations are, however, not straightforward because there exists no equivalent of the Shannon and von Neumann entropies defined for theory-independent systems. However, a characterisation of the entropic cones corresponding to theory-independent correlations in several networks is developed in Ref. [128]. The entropy cone for a particular scenario is, in this case, the intersection of the Shannon cones (recall, characterised in Section III D) corresponding to every set of coexisting variables in the scenario. Since the independence of the sources implies linear constraints at the level of entropies, the relevant entropic characterisation is an intersection of the convex cone intersection of all Shannon cones (that imposes the standard no-signaling constraints) with a hyperplane (the linear constraints implied by the independence of the sources). Ref. [128] completely characterises this set in the bilocal scenario with binary inputs and finds that there exists correlations that violate the entropic bilocal Bell inequality (37). Furthermore, in the triangle network without inputs, Ref. [28] proves that the inequality

$$I(A : B) + I(A : C) \leq H(A), \tag{38}$$

initially proved in [27] to hold for correlations admitting triangle local models of the form of Eq. (22), is in fact a theory-independent inequality, and any correlation generated in the triangle will satisfy it irrespective of the nature of the sources and measurement devices. This inequality enables one to bound e.g. the noise to which the GHZ distribution must be

---

[18] It must be noted, however, that the general performance of each method in arbitrary networks is still not fully understood. Therefore, one should not be surprised to find that methods that seem to provide looser characterisations in Figure 10 happen to be tighter when applied to different distributions or different networks.

exposed in order to be compatible with the network structure (see Figure 10).

2. Inflation

Another approach to theory-independent correlations in networks continues the idea of inflation, which was discussed for classical and quantum systems in Section III B. A procedure for adapting the inflation technique to general correlations was already put forward in the original work [32]. Recall that the success of inflation for local correlations rests on the fact that classical information (such as a local variable) can be cloned and therefore several copies can be sent to different parties. Such cloning is not possible in generalised operational-probabilistic theories [142] (in fact, it is not possible already in quantum theory). One can consider, however, inflations that do not require information cloning. These are called non-fanout inflations. An example is the so-called cut inflation, already depicted in Figure 5b. Another one is illustrated in Figure 11, where the triangle network is inflated to a six-ring, and which contains the cut inflation upon marginalisation. Due to the absence of cloned subsystems, these non-fanout inflations cannot directly constrain classical correlations, but they can be used to deduce constraints on any distribution arising in the original triangle scenario. This is, one may deduce constraints implied by the network structure. In fact, the proof of the validity of Eq. (38) for theory-independent correlations compatible with the triangle is arguably centered around the cut inflation of Fig. 5b, before the formal concept of inflation was developed. In practice, however, the constraints based on inflation arguments quickly become computationally demanding when one considers large networks or many copies.

Non-fanout inflations have been explicitly studied in the context of extending the role of the no-signaling principle to network scenarios. Ref. [69] deduces no-signaling constraints in the triangle network when neither Alice, Bob and Charlie have an input. Naively, one might believe that no-signaling is not relevant in the absence of inputs, but this is indeed not true in network scenarios due to the independence of the sources. Even in network scenarios where the parties do not have a choice of inputs, the outcomes of one party should be insensitive to whatever the remaining parties do, including the particular arrangement of the topology of their part of the network. This is the core of the notion of no-signaling presented in Ref. [69]. For the case of binary outcomes, Ref. [69] analyses the six-ring inflation of the triangle network in terms of the one-, two- and three-body correlators between Alice, Bob and Charlie[19]. However, due to the six-ring structure effectively opening up the triangle network, the three-body correlator E_{ABC} does not appear in the inflation. It is shown that no-
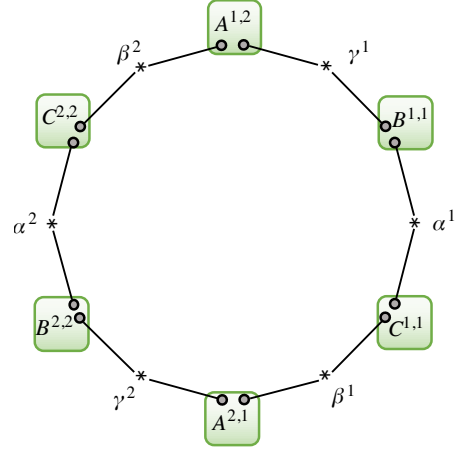

FIG. 11. Example of non-fanout inflation: six-ring inflation of the triangle scenario (the arrows denoting inputs and outputs have been removed for clarity). In this inflation, the sources and measurement devices are exact copies of those in the original triangle (see Figure 5a). In particular, the systems prepared by those sources are not cloned before being sent to the parties, in contrast with Figure 7. Therefore, this inflation only captures theory-independent constraints on correlations compatible with the triangle scenario.

signaling and independence implies the following constraint on the one- and two-body correlators in the triangle network,

$$
\begin{aligned}
& (1 + |E_A| + |E_B| + E_{AB})^2 \\
& + (1 + |E_A| + |E_C| + E_{AC})^2 \\
& + (1 + |E_B| + |E_C| + E_{BC})^2 \\
& \le 6 (1 + |E_A|) (1 + |E_B|) (1 + |E_C|).
\end{aligned}
\tag{39}
$$

In Figure 10 this criterion is illustrated for the noisy GHZ distribution and found to give stronger constraints than the entropic inequality (38). Notably, when the three one-body correlators vanish and the three two-body correlators are equal, the inequality can be saturated by wiring copies of a unique building block composed of a nonlocal source and a measurement device [144]. It is not yet known whether this building block can be realised in quantum theory.

In summary, theory-independent correlations in networks are strikingly different than in standard Bell scenarios. While the latter is characterised by a linear program [because, aside from positivity and normalisation, the only constraint to impose is no-signaling, which is linear in the probabilities, recall Eq. (1)], currently the former is only partially characterised through a linear program e.g. via analysis of entropies or inflations. This leaves an important open problem in whether it is possible to obtain simple and computable necessary and sufficient conditions that characterise the most general correlations achievable in networks.

B. Covariance matrices

It is also possible to characterise network constraints through the analysis of the covariance matrix of the random

---

[19] These are written E_A, E_B, E_C, E_{AB}, E_{AC}, E_{BC}, E_{ABC}. For instance, we have $E_A = \sum a^{1,2} p(a^{1,2}, b^{1,1}, c^{1,1}, a^{2,1}, b^{2,2}, c^{2,2})$ when all outcomes take values ± 1, which due to the inflation constraints can be expressed equivalently as $E_A = \sum a^{2,1} p(a^{1,2}, b^{1,1}, c^{1,1}, a^{2,1}, b^{2,2}, c^{2,2})$.

variables that describe the parties' outcomes. This perspective, developed in Refs. [141, 145–147], has the benefit of being notably simple.

As an illustration in the case of networks without inputs, consider the correlation $p(a_1, \ldots, a_n)$. In the covariance matrix approach, each party is associated a vector space, and each of the parties' outputs is mapped to an element of the corresponding vector space, i.e. $a_j \mapsto \mathbf{v}^{(i)}(a_j)$ for every party i obtaining output a_j upon measurement. Thus, the final joint observation of outcomes $\bar{a} = (a_1, \ldots, a_n)$ is mapped to a vector in the direct sum of the parties' spaces, $\mathbf{v}(\bar{a}) = \mathbf{v}^{(1)}(a_1) \oplus \cdots \oplus \mathbf{v}^{(n)}(a_n)$. When one considers the covariance matrix of this vector under the correlation,

$$
\begin{aligned}
\mathrm{Cov}(\mathbf{v})_{i,j} &= \langle v_i v_j \rangle - \langle v_i \rangle \langle v_j \rangle \\
&= \sum_{a_1,\ldots,a_n} p(\bar{a}) v_i(\bar{a}) v_j(\bar{a}) \\
&\quad - \sum_{a_1,\ldots,a_n} p(\bar{a}) v_i(\bar{a}) \sum_{a_1,\ldots,a_n} p(\bar{a}) v_j(\bar{a}),
\end{aligned}
\tag{40}
$$

not only this matrix is positive semidefinite as any covariance matrix is, but Refs. [141, 145] prove that, in the case of local and quantum realisations, if the distribution is compatible with a network, then it decomposes in a sum of positive-semidefinite blocks that is dictated by the network structure. This is,

$$
\begin{aligned}
\exists C_i &= \Pi^{(i)} C_i \Pi^{(i)} \succeq 0 \text{ s.t.} \\
\mathrm{Cov}(\mathbf{v}) &= \sum_i^m C_i,
\end{aligned}
\tag{41}
$$

where $\Pi^{(i)}$ are the projectors onto the subspaces corresponding to the parties that are connected to the source i and m is the total number of sources in the network. Figure 12 depicts the decomposition above in the case of the triangle scenario without inputs. If such a decomposition does not exist, then the distribution cannot be generated in the network neither with local variables nor with quantum systems.

While this method uses distributions generated by parties that do not have a choice of input, Ref. [141] provides a generalisation to networks where the parties have input choices. In such a situation, a decomposition analogous to Eq. (41) still holds, that contains entries which cannot be computed directly from the probability distribution. However, finding such a decomposition can still be formulated as a semidefinite programming problem, so the covariance matrix approach maintains its numerical simplicity.

As remarked by Ref. [146], the decomposition (41) can be directly deduced from block coherence theory, which, given a quantum state and an orthonormal basis, aims at characterising the minimal number of the basis elements needed to decompose the state. This connection allows to directly solve the existence of such a decomposition in case of dichotomic measurements, by providing sufficient conditions for compatibility. In Refs. [141, 145, 146], the decomposition (41) is shown to hold in the restricted case of local and quantum realisations. Ref. [147] promotes it to a fully theory-independent
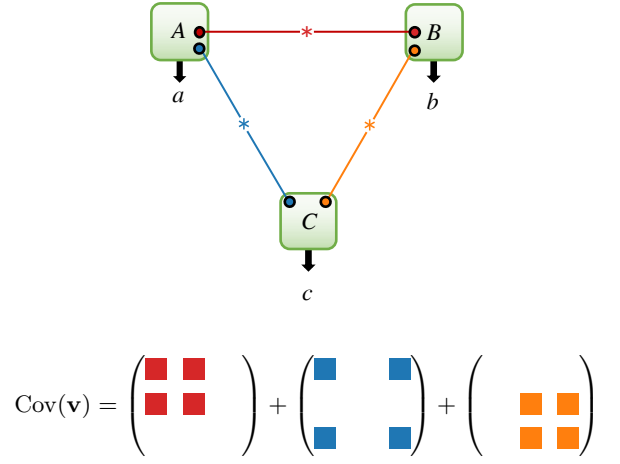

FIG. 12. Example of semidefinite decomposition of the covariance matrix according to the network topology, in the triangle scenario without inputs. For distributions which can be generated in the triangle scenario (regardless of the nature of the sources), the corresponding covariance matrix decomposes into the block structure depicted in the bottom, where every block is a positive-semidefinite matrix. This method holds for any network, the decomposition being dictated by the network structure.

network constraint. The proof is restricted to the case of all networks where any pair of parties shares at most one source. It is also restricted to the case of vectors $\mathbf{v}^{(i)}$ of dimension one. These allow to recover the standard formulation of the covariance matrix in Eq. (40), at the expense of providing a weaker characterisation of the sets of theory-independent correlations than what is theoretically possible. It first reformulates the decomposition (41), considering its dual cone in the set of positive matrices, which is characterised using quantum embezzlement concepts [148]. Then, generalising the connection to block coherence theory established in Ref. [146], it shows that this decomposition is a consequence of the positivity of the covariance matrices in the asymptotic non-fanout inflations of the networks.

C. Limits from Finner inequalities

Another method to determine whether a probability distribution can be generated in a particular network arises from the scrutiny of the problem under the perspective of pure probability theory. Indeed, the Finner inequalities [149] give rise, in the context of Bell nonlocality in networks, to necessary criteria for correlations to be compatible with a network. These constraints have been proven to apply to local, quantum, and correlations arising in some generalised probabilistic theories in large families of networks.

Finner inequalities are a generalisation of the Hölder inequality [150, 151] (which bounds the expectation of a product of functions $f_j : \mathbb{R} \mapsto \mathbb{R}$ by the product of individual expectations), that are easy to interpret as necessary conditions that correlations compatible with a network must sat-

isfy: Consider local stochastic post-processing functions $f_j : a_j \mapsto f_j(a_j) \in \mathbb{R}$ of the parties' outputs. Consider also a vector of weights $(\eta_1, \ldots, \eta_m)$ satisfying $\forall i, \sum_{j \,:\, i \to j} \eta_j \leq 1$, where the sum is over all parties connected to source i. Then, the Finner inequalities are constraints on the joint expectation value of $\prod_j f_j$ that read

$$\left\langle \prod_j f_j \right\rangle \leq \prod_j \left\langle f_j^{1/\eta_j} \right\rangle^{\eta_j}, \qquad (42)$$

where the averages are taken over the probability distribution under scrutiny. In the case where there is just one source, one recovers the original Finner inequality by forgetting the post-processing functions and directly taking $a_j = \bar{\lambda}_j$. For different networks the subsets of $(\eta_1, \ldots, \eta_m)$ whose sum is bounded are different, giving rise to distinct Finner inequalities. Moreover, for a given network, there are infinitely many Finner inequalities, corresponding to all suitable choices of weights. In certain cases, however, a finite collection of them provides a complete description of the space of allowed values. An example of such cases can be found in [23, Appendix D].

The Finner inequalities were introduced in the context of probability theory and graph theory. Their original proof [149], when interpreted in the context of nonlocality in networks, demonstrates that they hold for any local distribution compatible with the network it applies to. Beyond that, Ref. [23] proves that the Finner inequalities also hold in networks composed of bipartite quantum sources, and in the triangle scenario where the sources are no-signaling boxes[20] and the parties are only allowed to do wirings of their systems. It also provides an explicit example of one of these inequalities, namely

$$p(a, b, c) \leq \sqrt{p(a)p(b)p(c)}, \qquad (43)$$

which holds in the triangle scenario of Figure 2c and is used in Ref. [152] to analyse quantum networks obtained in the context of limited quantum resources.

Ref. [23] conjectures that the Finner inequalities hold for any network and any no-signalling theory. This conjecture is proved in Ref. [153], which also provides an explicit proof of the inequality $p(a, b, c) \leq p(a)^{1 - \frac{1}{m}} p(b)^{\frac{1}{m}} p(c)^{1 - \frac{1}{m}}$ holding in the bilocal scenario for any integer $m \geq 2$.

Finner inequalities are naturally formulated in networks where the parties do not have a choice of measurement performed to their respective systems. Nevertheless, there is a straightforward way of applying them in networks with inputs, by checking them for every possible choice of inputs. The question of whether this approach is the only possible one, or in the contrary new inequalities can be devised for networks with inputs, remains open.

V. EXPERIMENTS

Quantum nonlocality in networks is now not just a theoretical concept, but also a subject of experiments. This constitutes a natural continuation of decades of experimental investigation of quantum nonlocality in standard Bell scenarios (for a review of experiments in standard Bell nonlocality see, e.g., Refs. [2, 3, 154]). Its interest lies not only in testing the predictions of quantum theory in more sophisticated scenarios, but also in the key role attributed to the experimental realisation of entangled measurements and for the future development of quantum information protocols in networks, such as the quantum internet [6, 7]. A program for quantum nonlocality in networks constitutes a formidable challenge, which ultimately includes the implementation of synchronised, high-quality, independent sources distributing entangled states over large distances, and scalable entangled measurements of several distinct systems. In addition, quantum nonlocality in networks faces a challenge inherited from standard Bell experiments, namely the closing of experimental loopholes, commonly associated to technical imperfections in Bell experiments, which can be exploited to reproduce quantum correlations in (network) local models[21]. Furthermore, a pivotal challenge is the justification of the fundamental assumption of network Bell experiments, namely the independence of the sources. Increasing confidence in this assumption is closely related to increasingly sophisticated realisations of synchronised, physically distinct, quantum sources. In what follows, we first discuss entanglement swapping experiments in the context of quantum nonlocality and then direct experimental tests of network nonlocality.

A. Indirect tests: entanglement swapping and nonlocality

Entanglement swapping is a key primitive for quantum nonlocality in networks. Indeed, the simplest entanglement swapping scenario is identical to the bilocal scenario (see Figure 2b): two independent sources each distribute an entangled pair of quantum systems and by performing an entangled measurement on one share of each pair, the two remaining systems, which have never interacted, can become entangled [9]. The paradigmatic measurement for entanglement swapping is the Bell state measurement (12), i.e. the projection of two qubits in a basis of maximally entangled states. The state of the *a posteriori* entangled qubits (those held by Alice and Charlie) depends on the specific outcome of the Bell state measurement performed by Bob. By conditioning on this outcome, the swapped entanglement is said to be event-ready.

---

[20] No-signaling boxes are the building elements of the box-world generalised probabilistic theory [142] that generates stronger-than-quantum correlations [15].

[21] While the idea of nature conspiring against experimenters using seemingly contrived local models is arguably far-fetched, the practical importance of closing of such loopholes, beyond purely foundational considerations, becomes more evident in quantum information tasks such as random number generation and quantum key distribution which feature a natural notion of an adversary attempting to corrupt the protocol.

One can then attempt to use this entanglement (e.g. as a quality benchmark) to violate a Bell inequality. Naturally, since standard local models are strictly stronger than network local models, it follows that a violation of a standard Bell inequality based on event-ready entanglement constitutes an indirect demonstration of network nonlocality. Therefore, such indirect tests significantly pre-date the research program aimed at understanding Bell nonlocality in networks.

In photonics-based setups, which are natural candidates due to the possibility of realising significant distances, Bell state measurements [155] constitute a challenge. In fact, a complete Bell state measurement cannot be performed deterministically with linear optics [156] (moreover, the efficiency cannot exceed 50% [157]) unless auxiliary degrees of freedom are made available [158]. Nevertheles, it is possible to implement partially resolved Bell state measurements, i.e. measurements that only distinguish two of the four basis states, using linear optics without auxiliary degrees of freedom [159, 160]. Such a measurement was used to realise photonic entanglement swapping for the first time in 1998 [40]. However, in this proof-of-principle experiment, the visibility of the event-ready entanglement was insufficient to enable a violation of the CHSH inequality. This limitation was overcome in 2001 when Ref. [161] demonstrated event-ready entanglement of a visibility well above the required threshold. A few years later, such visibilities were achieved with photons separated over a substantial distance (over 2 km) [162]. In 2002, entanglement swapping with high visibility was demonstrated in a different way, by using only two independent photons with each one being a physical carrier of entanglement between two of its degrees of freedom [163]. Furthermore, entanglement swapping has also been demonstrated in continuous-variable systems [164, 165].

The first demonstration of nonlocality from event-ready entanglement was reported in Ref. [41] in 2001. In this experiment, both entangled photon pairs were generated with a spontaneous parametric down-conversion (SPDC) process from the same laser pulse, which led to correlations in the phases of the two entangled photon pairs. In 2006, this motivated a more faithful event-ready violation of the CHSH inequality which featured synchronised entanglement sources based on separate SPDC processes and separate (but optically dependent) lasers [166]. Soon afterwards, Refs. [167, 168] implemented entanglement swapping with fully independent sources, but the event-ready entanglement was not strong enough to violate the CHSH inequality. It is interesting note that the visibilities reported in Refs. [167, 168] would have been sufficient to violate the BRGP inequality (11), which was developed a few years later. Subsequently, in 2009, Ref. [169] reported on a stringent demonstration of nonlocality from event-ready entanglement with separate and independent sources. Notably, taking another route to entanglement swapping, Ref. [170] demonstrated a violation of the CHSH inequality via event-ready entanglement established through a probabilistically implemented complete Bell state measurement. Furthermore, event-ready entanglement and nonlocality have been demonstrated in hybrid systems featuring one source of discrete-variable single-photon entanglement and one source of continuous-variable entanglement [171, 172].

Photonic entanglement swapping has also been demonstrated beyond the simplest network scenario. Ref. [87] considered a four-party chain network involving three SPDC sources and performed two separate partial Bell state measurements in order to entangle two initially independent photons. Also, a four-party star network has been realised in which three sources emit two-photon entangled states: three initially independent photons are projected onto a GHZ state in order to remotely prepare the remaining three photons in a GHZ state [173]. In addition, entanglement swapping has been performed between two initially independent three-photon GHZ states [174].

Photons can also be employed as means for event-ready entanglement between two, initially independent, matter qubits. To this end, one can consider sources of atom-photon entanglement in which the photon acts as a flying qubit [175–177]. By subjecting the photons from two different sources to an entangled measurement, one obtains event-ready atom-atom entanglement with significant spatial separation. This was demonstrated in 2007 and 2008 in Refs. [178, 179]. In Ref. [178], using a photonic partial Bell state measurement, two ions separated by one meter were entangled, albeit at a low success probability and with a visibility insufficient to violate the CHSH inequality. The visibility of the event-ready entanglement was improved in 2008 and nonlocality was demonstrated with a high detection efficiency [180]. In 2012, event-ready atom-atom entanglement with a separation of 20 meters, as well as a demonstration of its nonlocality, was reported in Ref. [181]. Notably, loophole-free violations of the CHSH inequality have been reported based on the implementation of entanglement swapping in photon-electron spin systems [19] as well as in photon-atom systems [21]. In contrast, doing away with photons and instead considering ion-trap systems, fully deterministic entanglement swapping and a violation of the CHSH inequality was reported in Ref. [182]. In 2019, high-fidelity deterministic entanglement swapping was demonstrated on superconducting qubits [183].

B. Direct tests: violating network Bell inequalities

Direct experimental tests of network nonlocality, i.e. experiments that implement protocols tailored for violating the constraints of network local models, is a recent and growing program. The first proof-of-principle photonics experiments were reported in 2017 by Saunders et al. [44] and Carvacho et al. [184], both focused on the bilocal scenario. In 2019, a more stringent violation of bilocal Bell inequalities was demonstrated, which closed several loopholes [185]. The bilocal Bell inequalities discussed in Section II B have also been violated in quantum protocols without entangled measurements [188]. In 2020, violations based on separable measurements were demonstrated for star networks with up to four branch observers [186]. Departing from photonic implementations, Ref. [187] considered transmon quantum computers and star networks of up to five branch parties and demonstrated quantum correlations using fully entangled multi-qubit

Reference	Network	System	Inequality	Type of measurement	Independence of sources	LOC	DET	FOC
Saunders et al. [44]	Bilocal	Photons	[30]	Partial BSM	Partly	Open	Open	Open
Carvacho et al. [184]	Bilocal	Photons	[30]	Partial BSM	Assumed	Open	Open	Open
Sun et al. [185]	Bilocal	Photons	[30]	Partial BSM	Strict	Closed	Open	Closed
Poderini et al. [186]	Star $N \leq 4$	Photons	[33] & [186]	Separable	Strict	Open	Open	Open
Bäumer et al. [187]	Star $N \leq 5$	Transmons	[33] & [31]	Full BSM & EJM	Assumed	Open	Closed	Open

TABLE III. Summary of selected network nonlocality experiments. LOC, DET and FOC denote the locality, detection and freedom of choice loopholes, respectively.

measurements. We now proceed to discuss these experiments one by one. Their key features are listed in Table III.

Refs. [44, 184] report on bilocality experiments performed using polarisation qubits generated via SPDC sources. In the spirit of independent sources, two separate nonlinear crystals were used to generate the respective entangled pairs. However, in both experiments, the beams impinging on the respective crystals originate from the same laser. In the experiment of Saunders et al., this potential origin of correlation between the sources is addressed by using a quantum random number generator together with a phase shifter in order to give a random offset to the beam impinging on the second crystal. Both experiments use linear optics to violate the BRGP inequality (11), achieving $\mathcal{S}_{\mathrm{BRGP}} = 1.25 \pm 0.04$ [44] and $\mathcal{S}_{\mathrm{BRGP}} = 1.268 \pm 0.014$ [184], respectively. The implemented protocol is based on the complete Bell state measurement. However, due to the limitations of standard linear optics [189], the experiments actively exploit an assumption of the detected events constituting a fair sample of all de-facto events (the so-called detection loophole) in order to emulate this measurement by using several implementations of a partial Bell state measurement for entanglement swapping. Notably, Ref. [44] also demonstrates a violation of a modified version of the BRGP inequality, which is tailored for a partial Bell state measurement. This protocol was implemented without active exploitation of the detection loophole.

Sun et al. [185] take a stringent approach to realising the bilocal scenario. In this experiment, the independence of the sources is imposed using two separate, synchronised, SPDC sources with their relative phases randomised by periodically shifting the laser from spontaneous emission to stimulated emission. Furthermore, the experiment implements fast switching, which allows for spacelike separation between the relevant parts of the experiment (state emissions, generation of input and performing the measurements). This allows for the elimination of the so-called locality loophole (the possibility of exchanging a signal between events) as well as the freedom of choice loophole (the possibility of correlating the inputs of Alice and Charlie when they are generated within the same lightcone). It does, however, leave the detection loophole open. The experiment also captures the spirit of a quantum network since the sources and parties are all subject to significant separations: the smallest distance between the

parties and sources is roughly 100 meters. This experiment demonstrates a faithful violation of a a modified version of the BRGP inequality which is tailored for implementation with a partial Bell state measurement.

Recently, violations of network Bell inequalities have been demonstrated beyond the bilocal scenario. Poderini et al. [186] considered star networks with up to four branch parties, with each of the four bipartite sources of polarisation qubits being realised with independent lasers and SPDC processes, and parties and sources being separated by tens of meters. In contrast to the previous, this experiment is based on a separable measurement in the central node of the network. This considerably simplifies the implementation, as there is no need for intereference in the central node between the photons originating from the different sources. Instead, the central node performs synchronised single-photon measurements implemented, for instance, using a large coincidence window. Therefore, this may be viewed as a synchronisation of several independent standard Bell experiments. For star networks with two, three and four branch parties, the measured correlations were shown to violate a modified version of the star network Bell inequality (29) based on binary outputs [33]. Furthermore, motivated by device-independent considerations, Ref. [186] also considered the case of adding more inputs for the branch parties. In addition, violations of bilocality based on separable measurements have also been demonstrated in Ref. [188] without assuming a fixed reference frame between the involved parties. Furthermore, Ref. [190] considered the bilocal network but without assuming the underlying network structure, and performed two synchronised tests of the CHSH inequality in order to certify that the global state is a pair of singlets. The experiment performs the certification up to a global fidelity of 86.3%.

Network Bell experiments have also been performed beyond optical setups, specifically in transmon quantum computers. In 2020, Ref. [187] employed the publicly available, cloud-controlled IBM quantum computers [191] to implement star network Bell experiments. Since all the qubits are located on a single quantum processor, this experiment does not follow the spirit of sizable separation between parties in Bell experiments. In this sense, it may be considered a simulation of network Bell experiments. Furthermore, the device's architecture is trusted, meaning that the independence of the sources is

assumed. Nevertheless, Ref. [187] evidences that the assumption of independent sources is accurate on the level of the measured probabilities. The experiment targets a violation of the star network Bell inequality (29) by implementing a deterministic, fully entangled, measurement for entanglement swapping in the central node, i.e. a 2^m-outcome Bell state measurement of m qubits. Violations are reported for two, three, four and five branch parties. Furthermore, the same platform is used to implement the Elegant Joint Measurement (15) by realising a circuit, requiring three CNOT gates, proposed in Ref. [31]. This is used both to violate a bilocal Bell inequality and to implement the quantum protocol for the triangle network discussed in Section II C 1. However, it remains unclear whether the measured data violates triangle local models due to the lack of noise-robust analytical criteria for detecting the nonlocality of the distribution given in Eq. (A2). No postselection is made on the detections and therefore the nonlocality is not vulnerable to the detection loophole.

VI. FURTHER TOPICS AND RELATED AREAS

In this section we discuss additional topics in network nonlocality. This includes the current progress towards definitions of genuine nonlocality in networks, its application in device-independent quantum information processing protocols, its relationship to standard multipartite Bell nonlocality, and machine learning approaches to tackle problems in the area. Moreover, we also discuss topics closely related to network nonlocality, where there is an interesting exchange of ideas and concepts. These include the notion of entanglement in networks and the physics of more general causal structures.

A. Stronger notions of network nonlocality

As forefronted already by the discussion in Section II A, a central issue in the topic concerns distinguishing disguised forms of network nonlocality from those that may be viewed as genuine to the network structure on which they arise. This is a challenging matter. It involves both the development of physically insightful notions of network nonlocality that are stronger than the standard notion, and the technical development of the tools needed to explore its features and consequences. It appears reasonable to expect that for the purpose of capturing more genuine notions of network nonlocality a variety of different concepts may be concieved, each geared towards a particular aspect in which the standard definition proves itself unsatisfactory. This makes it plausible that a variety of stronger notions of network nonlocality may be of separate interest for investigation, as opposed to a unique notion of "genuine" network nonlocality. This, however, does not compromise the foundational and experimental motivation for understanding such matters.

Although the number is likely to grow, two conceptually different approaches have hitherto been taken to address this matter. Ref. [192] takes a theory-independent approach to networks and considers correlations of interest those which can be explained only if every source in a given network distributes systems that cannot be modeled by local variables. Such correlations are termed fully network nonlocal. Thus, if a distribution obtained on a given network can be modelled through at least one source admitting a local-variable description, it is not fully network nonlocal. This definition is convenient in terms of numerical characterisation because it enables one to describe the desired correlations by considering hybrid models of inflation, in which one source is local and all other sources are general nonlocal resources [32, 69]. On the conceptual level, detection of full network nonlocality guarantees that no source of a network distributes classical systems. Notably, adding more sources and parties that act classically, to a network that already displays full network nonlocality, causes the correlations in the expanded network to no longer be fully nonlocal. Thus, it apropriately addresses e.g. the Fritz example in the triangle (recall Section II A and Appendix A). Another interesting consequence of this concept is that no known quantum violation of the BRGP inequality implies full network nonlocality, i.e. the known violations can all be simulated in the bilocal scenario when one source is described by a local variable. In order to reveal full network nonlocality in quantum theory, concrete protocols based on i.a. the Elegant Joint Measurement and the three-qubit Bell state measurement have been proposed [192]. However, full network nonlocality does not in principle (although it does for many concrete quantum examples) prevent wirings of standard Bell inequalities to be considered fully network nonlocal. It can also be activated in a trivial manner, by grouping parties in a network to create a larger, single party. This property is known as instability under composition, which notably also applies to genuine multipartite nonlocality [24], genuine multipartite entanglement [193], and robust genuine multipartite entanglement [194].

A different, complementary, approach is proposed in Ref. [195]. Although in principle extendible also to theory-independent systems, the focus is on networks that are assumed to be governed by quantum theory. The authors label a correlation arising in a given network as genuine network nonlocal if it cannot be reproduced by wiring standard Bell nonlocal distributions according to the network structure. In other words, each source is considered to distribute a black box, and any distribution that cannot be obtained by classically processing the inputs and outputs to those boxes is considered to be genuinely network nonlocal. Notably, this does not necessarily imply a need for a measurement with entangled eigenvectors due to the inability of simulating all joint product measurements by means of local operations and classical communication [37]. For instance, genuine quantum network nonlocality apropriately addresses the Fritz distribution in the triangle and by construction also wirings of standard Bell experiments. By employing self-testing techniques, Ref. [195] proves that the correlation achieving the maximum known violation of the BRGP inequality is genuinely network nonlocal under the assumption of quantum theory. However, noise-robust proofs of genuine network nonlocality, as well as reasonably versatile methods to characterise the non-genuine set of correlations, are still lacking.

It is interesting to note that neither of the two above discussed concepts are special cases of the other. For instance, two wired PR boxes in a bilocal scenario manifest full network nonlocality but not genuine network nonlocality (when quantum theory is not assumed). Conversely, the largest known quantum violation of the BRGP inequality (11) is not fully nonlocal but is genuinely nonlocal in the quantum bilocal scenario. This makes their comparison less straightforward and further lends credit to the plausibility that many stronger, complementary, notions of network nonlocality are both interesting and relevant.

B. Device-independent considerations

Nonlocality is at the heart of the device-independent approach to quantum information science (see e.g. Ref. [196] for a short overview in the context of Bell scenarios). In this approach, deductions are made based on experiments subject to minimal assumptions, most notably without any characterisation of the internal workings of the operating devices [197–199]. This approach is naturally extended to network scenarios, but requires one to assume the independence of the sources. Notably, such independence cannot be guaranteed in a device-independent manner. Whereas source independence can be justified with increasing confidence by suitably adapting experimental setups (see Section V), it is, ultimately, an additional assumption. This warrants a distinction in network scenarios for device-independent protocols aiming to assert a property of nature and device-independent protocols aiming to perform a quantum information task. In the former, the independence of the sources is a reasonable assumption whereas in the latter, it is not far-fetched to consider a malicious party tampering with the sources in a coordinated manner, thus affecting their independence.

From another point of view, the assumption of independent soruces has, in certain contexts, a status similar to the freedom of choice assumption in standard Bell scenarios. For instance, in Section II A (see Figure 4), we saw that correlations violating the CHSH Bell inequality can be identified with a nonlocal quantum distribution in the four-party chain network arising from parties without inputs. This identification exchanges the freedom of choice assumption with the assumption of independent sources.

In the following we provide an overview of a few aspects where the device-independent framework and network nonlocality are combined.

Quantum key distribution: The device-independent framework, when applied in quantum key distribution allows two parties to extract a common key and guarantee its security by analysing the outcome statistics of a Bell-type experiment. It enables a high level of security which is relevant in view of increasingly sophisticated eavesdropping attacks [200–203]. Generally, an eavesdropper can correlate a system of their possession with the systems measured by the parties in order to gain information about their outcomes. Such eavesdropping can be detected by testing the nonlocal correlations of the honest participants. Specific Bell inequalities are used for bounding the information leaked to the eavesdropper or, alternatively, the secure key rate (see e.g. Refs. [198, 204–208]).

In Bell scenarios, the eavesdropper is assumed to hold a (potentially stronger-than-quantum) system correlated with the source that distributes systems to the honest parties. In network scenarios, this changes into independent systems that can each be correlated with different sources. Ref. [209] introduced an intermediate device-independent scenario by trusting a subset of sources that are not correlated with a single system held by an eavesdropper. For a chain network consisting of n sources, Ref. [209] finds a monogamy relation between the amount of information that the eavesdropper can obtain about the outcomes of the extremal parties and the violation of the BRGP inequality that these parties can achieve: if the violation is larger, the leaked information is smaller. Similar results were also proven to hold in star networks [209]. These results have been extended to general networks in Ref. [210] by using the network Bell inequality derived in Ref. [70] as the quantifier of monogamy. All these monogamy relations are generally conservative because of the assumptions of independence and that the eavesdroppers are allowed to manipulate stronger-than-quantum systems. In the case of the chain network, Ref. [67] strengthened the existing bounds by making the realistic assumption that the eavesdropper has access only to quantum systems. It is important to note that if all the independent sources are correlated by the eavesdropper, the quantum nonlocal correlations may be simulated by eavesdroppers with classical sources [209]. This stresses the pivotal role of the of the source independence assumption in device-independent protocols in networks [211].

Self-testing: Another relevant device-independent consideration is self-testing [212], which aims to characterise an unknown quantum device based only on some observed correlation, in a black-box scenario (see e.g. Ref. [213] for a review). In networks, it is possible to self-test entangled measurements [214, 215]. In particular, the Bell state measurement can be certified in the bilocal scenario (recall Figure 2b), asking the central party to perform a complete Bell state measurement and the two remaining parties to perform the standard CHSH measurements. This maximally violates four relabelings of the CHSH inequality. A certification of the central party performing an entangled measurement is obtained for a violation of the CHSH inequality larger than $\mathcal{S}_{\mathrm{CHSH}} \approx 2.69$. This protocol can be generalised to certifying the tilted Bell state measurement associated to four partially entangled two-qubit states as well as a measurement in a basis of GHZ states [215]. This protocol has been experimentally implemented in Ref. [216].

Self-testing has also been considered at the level of the physical theory itself. This aims at identifying a task in networks where the optimal performance singles out quantum theory as the only possible physical model (up to reasonable degrees of freedom). In standard Bell scenarios, no-signaling resources enable larger state spaces and therefore more general correlations than those of quantum theory. However, enlarged state spaces typically come with restricted notions of joint measurements [142]. Since networks endow joint mea-

surement with a pivotal role, it becomes interesting to ask whether they can be used to self-test quantum theory. Recently, such arguments have been used in the bilocal scenario with the same protocol involving a central Bell state measurement and the maximal violation of four relabelled CHSH inequalities. For this task, no generalised probabilistic theories performs better than quantum theory, and quantum theory outperforms many of them [217, 218].

The role of complex numbers: Complex numbers are essential for mathematical analysis. While some physical theories do not make use of complex numbers, others, such as electromagnetism, employ them as convenient, but in principle dispensible, tools for calculations. Similarly, complex numbers are highly useful in the standard Hilbert space formalism of quantum theory. However, it is unclear if they are truly necessary in the standard Hilbert space formalism or if it is possible to simulate all predictions of quantum theory using only real-valued Hilbert spaces. A partial answer is provided by the fact that the statistics of every quantum implementation of a Bell experiment can be simulated in another Bell experiments using a quantum model that only requires real Hilbert spaces [219, 220]. Such a simulation requires one to enlarge the Hilbert space by introducing ancillary real qubits; otherwise it cannot succeed [221–224]. However, the situation changes when one examines a network scenario. Assuming that two sources are independent, it is possible to device-independently certify the need for complex numbers in the standard Hilbert space formalism of quantum theory [225]. This result is shown in a scenario reminiscent of the bilocal scenario; it involves three parties and two independent, bipartite quantum sources. In contrast to the bilocal scenario, all three parties are allowed to share a local variable. By Bob performing a Bell state measurement, Alice and Charlie can test a version of the CHSH inequality [226] in order to device-independently assert the impossibility of a quantum simulation based on real numbers.

Ref. [82] shows that the nonlocality of a Bell-local state can be activated by distributing many copies of it in a star network, using an event-ready protocol which postselects on the outcome of an entanglement swapping measurement. The same idea is used in Ref. [83] to activate nonlocality for all one-way distillable states. Event-ready protocols were first introduced to prove nonlocality of all multipartite entangled pure states in Ref. [229]. These can be used in chain networks (recall Figure 6b) to identify unknown nonlocal resources using erasure channels. Furthermore, it is also possible to activate nonlocality by allowing the black boxes held by the parties to admit quantum inputs [230]. This latter work shows, additionally, that nonlocality is not an additive quantity. Activation of nonlocality has also been analysed through the verification of genuine multipartite nonlocality [24]. Ref. [231] shows that genuine multipartite nonlocality can be activated in networks that exclusively contain bipartite sources. The main concept in the proof of Ref. [231] is applying the lifting method of Ref. [232] for obtaining the boundary of the set of correlations in increasingly complex networks starting from inequalities for bipartite systems. Outside the context of activation of nonlocality, the approach of using new Bell inequalities with multiple inputs in order to verify the genuinely multipartite nonlocality of quantum networks is employed in Ref. [211] to analyse protocols in quantum key distribution, blind quantum computing, and quantum secret sharing. A third example has been demonstrated for the triangle network without inputs. Ref. [36] introduces a way of quantifying the amount of independence between sources, and investigates to what extent the independence assumption can be relaxed in order to demonstrate network nonlocality. By generalising Fritz's construction for disguising a bipartite Bell test in the triangle network [27], Ref. [36] shows that correlations incompatible with a local model in the triangle can be observed even for arbitrarily small levels of independence between the sources.

C. Relaxing source-independence

Quantum networks are present in many experimental implementations of Bell experiments [19] and generation of entangled states [85, 227, 228]. In these cases, the quantum network is the physical framework in which the quantum correlations are obtained. However, the theoretical analysis is performed by studying the generated correlations with respect to a single source of randomness being shared by all parties, this is, in the associated Bell scenario. In the following, we discuss a series of theoretical works in which the notion of a network is still present at the theoretical level, but the analysis of the correlations obtained is performed at the level of the corresponding Bell scenario. There are two different lines of research that do not use the source independence assumption. On one hand, several works studied how the nonlocality of an entangled state could be activated by distributing copies of the state in a quantum network. On the other hand, this framework has been employed in the analysis of multipartite nonlocal games.

The second approach originates from the analysis of quantum games [233]. A nonlocal game is described as spatially separated players who interact with a referee. Assuming that the players cannot communicate with each other, they are required to answer questions in a coordinated manner. Nonlocal correlations are used to achieve a quantum advantage over similar experiments where the parties share classical resources. Generally, any nonlocal game naturally corresponds to a standard Bell inequality, although the converse is not true [234]. As a natural extension, Ref. [235] considers nonlocal games in networks with bipartite sources, where sharing quantum states allows for larger winning probabilities than those achievable when the players share classical systems. The winning condition is consistent with the assumptions required for the derivation of the network Bell inequality of Ref. [70]. Thus, these games provide necessary and sufficient conditions for witnessing quantum networks consisting of bipartite sources.

D. Machine learning in network nonlocality

Machine learning, the discipline that analyses the automated extraction of relevant features from data, provides many important tools that are being increasingly used in many fields of physics (see e.g. the review [236] for a wide account of the different contexts where it is being used). In the context of nonlocality, the fact that knowing whether a given correlation in a given network admits a network local model is a computationally challenging but well-defined task makes it a suitable candidate for applying machine learning techniques. However, problems in quantum information theory in general, and quantum correlations in particular, often require rigorous proofs that falsify a hypothesised class of models. This requirement for rigorous proofs is typically at odds with the uncertainties inherent to established machine learning techniques.

This, however, has not prevented machine learning tools to be used in the study of network nonlocality. Initial attempts followed the traditional approach to regression via machine learning: a neural network being used to approximate a function that is difficult to compute otherwise. Ref. [237] applies this scheme in the bilocal scenario to compute quantifiers of non-compliance with the network structure (i.e., quantifiers of the distance between the bipartite probabilities of the events of the extremal parties and the product of their corresponding single-party marginals), training several networks whose results are combined at a second stage. This work shows that neural-network regression can lead to very accurate results, despite the fact of needing significant amounts of labeled data for training to be effective. A more interesting approach is that introduced by Ref. [65]. By using a neural-network-based parametrization of Eq. (5), the deep learning toolbox is used to find the parameters that minimise the distance of the corresponding model to a target distribution. Such distance being made zero is a proof[22] that the target distribution admits a local variable model in the network. However, the converse is not true: the failure in obtaining a vanishing distance does not necessarily imply that the distribution is nonlocal. It can instead be a consequence of i.e. insufficient expressiveness of the neural network architecture chosen or unsatisfactory parameter initialisation.

This construction is employed to address whether some interesting tripartite quantum distributions admit local models in the triangle network. Interestingly, when applied to the RGB4 distribution given in Eq. (A3), the machine learning approach fails to find local models in a larger parameter range than the analytical arguments of Refs. [35, 62] (recall also Section III C). Moreover, Ref. [65] also reports the failure of finding a local model for the triangle distribution in Eq. (A2), providing an estimate on the visibility and the detection efficiency required to demonstrate its (potential) nonlocality. These results support the conjectures of the nonlocality of those dis-

tributions, although formal proofs must yet be developed. Finally, this approach is used in Ref. [63] to estimate parameter ranges and physical requirements of experimental realisations in the triangle network.

It is far from clear whether approaches based on machine learning could, eventually, provide proofs of nonlocality. Approaches similar to the example above seem, *a priori*, much more promising. Machine learning is a powerful exploratory tool for building intuition on where to look for nonlocal behaviour, because it offers a great computational improvement compared to traditional methods which typically require navigating in non-convex, high-dimensional spaces. It is reasonable to expect that the major contributions of machine learning to research in network nonlocality do not directly deal with the development of proofs in an automated fashion, but rather provide useful heuristics for computer-assisted discovery [238, 239], whereby automated procedures suggest where the interesting phenomena may be, and provide intuition to the scientists who later develop the formal proofs. Indeed, this approach has already proven successful in neighboring fields, for instance in problems concerning experiment design [240–242].

E. Multipartite network entanglement

Entanglement [243–245] is at the heart of quantum nonlocality. Much like standard Bell nonlocality, the conventional approach to multipartite entanglement considers particles originating from the same source. However, investigating multipartite entanglement in network structures with independent sources leads to novel questions and challenges in entanglement theory.

The standard definition of genuine multipartite entanglement stems from the need to assess control over an ever-growing number of quantum systems, and requires that a quantum state cannot be written as a convex mixture of states that are separable under a bipartition of the parties [193]. This notion, however, presents some arguably undesired features when considering entanglement in networks. As shown in Ref. [246], sources capable of distributing bipartite entanglement can, by themselves, generate genuinely k-partite entangled states for any k. Moreover, there are multipartite entangled states, such as cluster states [247] or graph states [248], that can be naturally defined through local operations on shares of several entangled sources. Thus, parallel to the development of network Bell nonlocality, several works have proposed different ways of defining network entanglement.

On the one hand, Ref. [246] proposes a definition whereby a quantum state is genuinely network k-entangled if it cannot be produced by applying local, completely positive and trace-preserving linear maps over several $(k-1)$-partite entangled states distributed among the parties, even with the aid of global shared randomness. The simplest example is the state $|\psi\rangle_{ABC} = |\psi_1\rangle_{AB_1} \otimes |\psi_2\rangle_{B_2C}$, where $|\psi_1\rangle$ and $|\psi_2\rangle$ are bipartite entangled quantum states and system B is composed by two subsystems, B_1 and B_2. The state $|\psi\rangle$ is genuinely tripartite entangled per the standard definition, but only

---

[22] Note that these proofs are not immediately rigorous due to the fact that machine precision is finite for all practical purposes.

genuinely network 2-entangled. The definition provided by Ref. [246] is particularly suitable for numerical characterisation, since the constraints that are derived from it are amenable to semidefinite programming. In Ref. [246], several witnesses of genuine network entanglement are presented, that make use of the quantum inflation technique discussed in Section III B 2 to derive analytical and numerical bounds to the fidelity of multipartite states that can be generated in network structures.

The same problem is considered in Ref. [249] for the specific case of the triangle network. In contrast with Ref. [246], in this work the parties are only allowed to perform local unitary transformations in the states that they receive, and classical randomness is shared not by the parties but by the sources. The restrictions considered in Ref. [249] allow providing upper bounds to the dimension of the sources. Moreover, Ref. [249] gives necessary conditions that are satisfied by all tripartite correlations that can be generated in the triangle network and proves that there are no three-qubit genuinely multipartite entangled states which, even when embedded in larger-dimensional systems, can be prepared in the triangle network even when one consider sources of quantum states that can be classically correlated via global shared randomness.

The same setup (parties are only allowed to perform unitary operations on their states) is characterised in Ref. [250], which provides many witnesses of genuine multipartite network entanglement in terms of the fidelities with experimentally realisable states (e.g. the GHZ state, the W state [251], or Dicke states [252]), and computes several visibility thresholds of mixed states that certify genuine network entanglement. Furthermore, Ref. [250] proves that any $n \geq 3$-partite permutationally symmetric entangled pure state, and any $n \geq 3$-partite entangled pure state composed of qubits and qutrits, are genuinely network entangled when the parties are restricted to perform unitary transformations and they are allowed to share a source of classical randomness.

Quantum network entanglement has recently found use in the analysis of entanglement-based attacks to quantum bit commitment [253, 254]. Bit commitment protocols aim to commit secret bits between two parties, with the certification that neither the committer nor the receiver can forge their observations. The no-go theorems of Refs. [253, 254] forbid non-relativistic quantum bit commitment with single particles or bipartite entangled states. It is possible, however, to perform quantum bit commitment in relativistic settings [255, 256]. Exploiting network entanglement, Ref. [257] proposes an alternative protocol based on GHZ states, where one of the parties holds multiple shares of the state. Although a party may forge her operations by modifying the overall network to be the bilocal scenario (in Figure 2b), the fact that the works [246, 249, 250] find certificates of incompatibility of quantum states with networks allow the parties to certify that the state shared cannot be generated in the bilocal scenario, thus making the protocol resistant to dishonest parties.

F. Network-based genuinely multipartite nonlocality

In the previous section it was discussed how the traditional notion of genuinely multipartite entanglement suffered from issues when confronted with ideas brought by the consideration of network structures. The clearest example is the state $|\psi\rangle_{ABC} = |\psi_1\rangle_{AB_1} \otimes |\psi_2\rangle_{B_2C}$ in the bilocal scenario, which is genuinely tripartite entangled according to its traditional definition [193] if $|\psi_1\rangle$ and $|\psi_2\rangle$ are entangled states, yet is possible to generate by the exclusive use of bipartite sources. In that case, as described in Section VI E, the solution to these issues comes from taking an approach that uses networks as the basis to define non-genuine forms of multipartite entanglement.

When considering the concept of genuinely multipartite nonlocality, a similar situation arises. The traditional definition of genuinely multipartite nonlocality [24] is based on the notion of bipartitions of the set of parties: a correlation is genuinely multipartite nonlocal if and only if it cannot be written as a sum, over bipartitions of the set of parties, of convex combinations of distributions involving each of the subsets of the bipartition (see, for instance, [231, Eq. 1] for its mathematical expression). As in the case of genuinely multipartite entanglement, this notion includes distributions that confront the standard understanding of genuine multipartiteness: for instance, the tripartite distribution $p(a, b, c|x, y, z) = p_{\text{CHSH}}(a, b_1|x, y_1)p_{\text{CHSH}}(b_2, c|y_2, z)$, where p_{CHSH} is the bipartite distribution that achieves the maximum quantum violation of the CHSH inequality (recall Section I A), is genuinely multipartite nonlocal according to the traditional definition of Ref. [24], yet it is easy to generate by considering two separate Bell tests and grouping one party of each of the tests in order to define the central party.

The origin of this apparent discrepancy is the fact that the standard definition of genuinely multipartite nonlocality is formulated under the assumption that parties can use local operations and classical communication freely. This does not correspond to the situation assumed when dealing with nonlocality, where parties are spacelike separated and thus cannot communicate with each other until the experiment is completed. Motivated by this, Refs. [258, 259] consider instead a situation where local operations are still free to apply by the parties, but the assumption of free classical communication is substituted by free access to a global resource of shared classical randomness. This is, despite not being able to communicate with each other until the experiment is completed, the parties can agree on a common strategy prior to the experiment taking place. Assuming that the parties have access to a global source of shared classical randomness gives rise to a definition of genuinely multipartite nonlocality completely based in network notions, and that parallels that of Ref. [246] for the case of genuinely multipartite entanglement: according to Refs. [258, 259], an n-partite distribution is genuinely network k-nonlocal if it cannot be produced by applying local operations over several $(k-1)$-partite nonlocal resources distributed among the parties even with the aid of global shared randomness, and thus it is genuinely network nonlocal if it cannot be generated by the use of $(n-1)$-partite nonlocal re-

sources. Non-genuine models of nonlocal correlations are, under this definition, easy to characterise under the inflation paradigm, hence including any exotic physical theory that can extend quantum entanglement and the process of combining subsystems via entangled joint quantum measurements as resources for generating nonlocality.

Based on this new definition, Refs. [258, 259] provide several device-independent and noise-tolerant witnesses of genuine network nonlocality and analyse the genuinely multipartite nature of distributions generated by measurements performed on genuinely network entangled quantum states. In particular, they show that the n-partite GHZ state and the tripartite W state (see, e.g., Ref. [259] for the definition of these states) can give rise to distributions that are genuinely network n-nonlocal, thus proving that only boundlessly multipartite nonlocal theories[23] can explain all quantum correlations. Not only that, they suggest that an experimental device-independent proof of the genuinely tripartite character of nature is already feasible with current technology [260].

G. Bayesian causal inference and quantum causal models

This review focuses on the correlations that can be generated in networks, where the parties' outcomes depend solely on their choice of input and the systems they receive from independent sources. These are the fundamental properties that are captured by the network local and network quantum models of Eqs. (5) and (6). It is natural to further wonder about the correlations that can be achieved in more general scenarios, where these conditions are not met. A simple example of a scenario that is not a network is the so-called instrumental scenario depicted in Figure 13, where the outcome of Alice serves as input for Bob's measurement.

The field of causal inference (see Ref. [261] for an introduction in the context of correlations admitting models based on local variables) deals with the task of understanding the correlations that can be generated in these general scenarios. A major insight, that constitutes one of the building blocks of the field, is that the scenarios allowed by causality[24] can be captured in a type of graphs called directed acyclic graphs. This enables the use of graph theory to understand correlations and causal relations. In the language of directed acyclic graphs, network structures can be described as two-layer graphs where one layer of visible variables (corresponding to the parties' outcomes) is connected to both visible (denoting the input choices) and unobservable variables (that describe the sources) in another layer, and there are no intra-layer connections. However, it is possible to envision many

---

[23] These are the theories where, for any number of parties, there exist sources of multipartite states that cannot be simulated by combining sources that distribute states to fewer parties.

[24] Causality here is defined in terms of Reichenbachs' principle, which states that if a correlation between two events, A and B, exists, then either A is a cause of B, B is a cause of A, or A and B are both caused by a third event, C, that occurs prior to A and B [262].
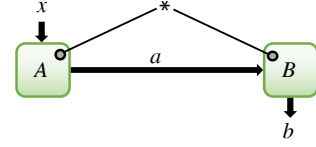
FIG. 13. The instrumental scenario, depicting Alice and Bob receiving systems from a shared source. This structure is not a network, because the outcome of Alice acts as an input for Bob. In the particular case of the instrumental scenario, the question of whether a probability distribution $p(a, b|x)$ can be generated in it can be mapped to an analysis of the bipartite Bell scenario [263–265], but this is not the general case in arbitrary causal structures.

phenomena of relevance that cannot be captured by these two-layer graphs, such as the signaling of an outcome described above, or the controlled preparation of the systems sent by a source,[25] to name a few, where causal inference techniques may prove themselves useful.

Causal inference benefits greatly from the developments in the study of nonlocality in networks, since many of the methods described in this review can be applied to the study not only of correlations in networks, but of correlations in general causal structures as well. The clearest example is the classical inflation technique described in Section III B 1, which gives asymptotically sufficient conditions for compatibility in any causal structure with just local variables [122]. Broadly, Ref. [122] (and Ref. [67] for the case of quantum variables) describes a series of steps that allow to reduce any causal structure and correlations generated in it to a network scenario and associated correlations that are dictated by those achievable in the original causal structure. Then, any method for characterising correlations in that network can be applied in order to characterise the correlations of the original causal structure. This procedure is exemplified in Ref. [67] by bounding the quantum average causal effect [266] of a visible variable on another in a triangle scenario with signaling. Also, the methods developed for the entropic characterisation of correlations can be readily applied to any causal structure, as described in Ref. [125], by properly enumerating its coexisting sets and their respective Shannon cones as described in Sections III D and IV A.

It must be noted, however, that classical causal inference is an established field that has developed its own analysis techniques [261, 267]. Moreover, since the goal of classical causal inference can be understood to be finding models based in local variables in arbitrary directed acyclic graphs, the techniques developed within its formalism have found application in the certification of nonlocality in causal structures that are not networks in the sense of this review [265, 268, 269]. The maturity of classical causal inference motivates the generalisation of its framework to assess causal relations between quantum systems. This is the effort of quantum causal mod-

---

[25] In the formalism of directed acyclic graphs, this is achieved by an observable variable connecting to an unobservable variable.

elling, which has since developed quantum generalisations of, for instance, Bayesian networks [28], Reichenbach's common cause principle [270], d-separation [271], the rules of do-calculus [272], and causal discovery algorithms [273].

VII. CONCLUSIONS AND OPEN PROBLEMS

Nonlocality in networks is a research program that conceptually, technically and practically takes the decades-old field of Bell nonlocality beyond the traditional Einstein-Podolsky-Rosen scenario. This review article surveys it up to the time of writing. It appears sensible to claim that this rapidly developing topic presently finds itself at a point where several basic methods and tools for its systematic analysis are being established, but many elementary questions remain wide open. While we have mentioned many concrete open problems along the way, we conclude this article with a discussion and highlight of some of the most interesting and relevant lines of future research in network nonlocality.

Genuine network nonlocality.— Despite the advances outlined in Section VI A, many questions remain to be solved regarding this aspect. How can one formalise and analyse the intuitive idea that some forms of network nonlocality are more genuine than others? This is a key question which reasonably should not be expected to boil down to a single concept, but may rather motivate a systematic classification of conceptually different forms of network nonlocality. To what extent are nonlocal resources required to be distributed in a network in order to explain quantum nonlocality? When does network nonlocality reduce to building blocks of standard Bell nonlocality (as discussed in Section II A)? What is the role of entanglement swapping in more genuine notions of quantum network nonlocality? Important steps in this direction include developing efficient methods to identify disguised Bell nonlocality in networks and understanding the relationship between network nonlocality and standard multipartite nonlocality.

Entangled measurements.— Standard quantum nonlocality is powered by entangled states. However, quantum nonlocality in networks endows also entangled measurements (measurements in an entangled basis) with a fundamental role. This motivates the development of entangled measurements, beyond the well-established Bell state measurement and the recent Elegant Joint Measurement, that possess conceptually or practically appealing properties. Are there new classes of entangled measurements that can reveal interesting features of quantum correlations in networks? How, and to what extent, can entangled measurements be viewed as resources in network nonlocality?

Quantum versus no-signaling.— In standard Bell scenarios, no-signaling correlations are sometimes called "post-quantum" or "supra-quantum" correlations, because they are strictly more powerful than quantum nonlocality. In networks, however, entangled measurements are brought to the forefront of the physical process. How does this impact the role and status of quantum theory in the broader landscape of no-signaling theories? How can we understand the counterpart of entanglement swapping in more general theories? Are there natural

counterparts to the Popescu-Rohrlich nonlocal box in network scenarios, and what can they teach us? Is it possible to find experiments in which quantum theory is singled out as the unique no-signaling theory for explaining predicted correlations?

Activation of nonlocality.— Networks are known to reveal the nonlocality of some states that do not violate any Bell inequality, even without the crucial assumption of independent sources (see Section VI C). It is therefore reasonably expected that networks with independent sources should be able to reveal such nonlocality in an even more powerful manner. However, there presently exists no example of using independent sources in a network to activate nonlocality in a way that cannot already be achieved in networks without assuming independent sources. Finding such an example, for instance based on the isotropic state (14), is an interesting problem with both fundamental and practical relevance. In a similar vein, given any entangled state, does there exist a network in which nonlocality can be detected if all sources emit one copy of the state?

Mixed resources.— This review, and indeed most research thus far, has largely focused on networks in which all sources are local, quantum or no-signaling, respectively. However, an interesting avenue of research concerns investigating the correlations that can arise in networks composed of sources described by different physical models. At a foundational level, this can shed light on the extent to which nonlocal resources are needed in a network in order to explain observed correlations. Do network nonlocal correlations truly exploit entanglement in all sources in the network in order to assert their nonlocality? This can also teach us about the relationship between different nonlocal theories in terms of studying their interplay on the level of correlations. Moreover, at a practical level, networks composed of both classical and quantum sources can find relevance in quantum information protocols for long-distance networks and allow to integrate natural assumptions such as the sources correlating via classical shared randomness.

Device-independence.— An interesting avenue is the exploration of network nonlocality as a phenomenon for powering device-independent quantum information protocols. Indeed, as discussed in Section VI B, initial steps in this direction have already been taken. Which tasks are natural to consider in a network scenario, as opposed to standard Bell scenarios? Do different networks enable qualitatively different protocols? Are there networks that support particularly well-performing protocols? What are the requirements for their implementation in relevant physical systems? More generally, what are the limits of device-independent inference in networks?

Justifying the independence of the sources.— The source-independence assumption is critical in network nonlocality. Contrary to the no-signaling assumption, it cannot be justified via reference to a fundamental principle similar to space-like separation. In particular, shared randomness can be simulated in synchronised protocols, for instance by independent parties separately measuring some global physical quantity (e.g. observing solar flare fluctuations). Hence, it is

a potential loophole in any experiment or device-independent protocol based on network nonlocal correlations. How critical is this loophole? Moreover, can the independence assumption be relaxed in some protocols, e.g. so that it is approximately or probabilistically satisfied? What are the best ways to experimentally justify the independence of the sources?

Exterior- and interior-point methods.— Developing methods for characterising the sets of local, quantum and no-signaling correlations in networks from both the interior and the exterior is a key challenge, since it allows to understand the ultimate limits of each theory. Many features of these methods are important to develop and improve upon. This includes which networks they apply to, which input/output scenarios to they apply to in a given network, how accurately can they bound the relevant correlations, and to what extent can they actually be computed with standard resources. While several such techniques have been discussed in this article, the need for powerful and practically useful methods, even in simple networks with a small number of inputs and outputs, remains central. In this line, an important aspect that remains open is to investigate how well the current techniques for bounding the correlation sets from the exterior characterise the sets of network correlations.

Bilocal and triangle networks.— Several concrete, yet elementary, questions lay open for the two smallest networks, namely the bilocal and the triangle scenarios. An interesting matter is to explore the geometry of the sets of local and quantum correlations in simple input/output scenarios in these networks. This may, e.g., enable the device-independent certification of entanglement swapping processes in the bilocal scenario. Furthermore, in the triangle network with no inputs and binary outputs, the conjecture that the local and quantum sets are identical remains open. Immediately related is the open problem of whether the set of triangle no-signaling correlations is equivalent to the quantum set. Another relevant open problem is to decide the triangle nonlocality of the highly symmetric quantum distribution, given in Eq. (A2), based on the Elegant Joint Measurement.

Better network Bell inequalities.— Many presently known network Bell inequalities ultimately rely on exploiting the core idea behind the BRGP bilocal Bell inequality. A natural challenge is to find qualitatively different, systematic, ways of constructing network Bell inequalities. Particularly relevant are constructions that are not based on standard Bell inequalities and that are tailored to quantum protocols based on entanglement swapping. Indeed, from the standpoint of present knowledge, this is relevant for every non-trivial network (see Section III A 2).

Experimental considerations.— Many fundamental aspects and technical challenges are relevant for experimental demonstrations and applications of network nonlocality. Perhaps most notably, an experiment that simultaneously closes the locality, fair sampling and freedom of choice loopholes is still missing. Moreover, none of the existing experiments go beyond the standard notion of network nonlocality; for instance, the network Bell inequality violations hitherto reported are insufficient for demonstrating stronger notions such as full network nonlocality. Furthermore, although several experiments have been performed in the bilocal scenario, no photonics experiment has thus far considered a different quantum network without resorting to wiring many standard violations of standard Bell inequalities. These more general networks may notably include multipartite sources, whose relevance in network nonlocality has yet not been demonstrated. Furthermore, a key challenge concerns the implementation and application of entangled measurements, in particular those that are different from a standard partial Bell state analyser. On this front, little is still known. These are important steps towards the long-term challenge of reliable and high-quality entanglement distribution and entanglement swapping in large-scale quantum networks.

Appendix A: Notable quantum distributions in the triangle network

In this appendix we compile a number of relevant distributions generated in the triangle scenario by means of quantum resources.

1. The Fritz distribution

The Fritz distribution [27] is the archetype example of a distribution that presents disguised Bell nonlocality. It is built from the distribution $p(a, b|x, y) = \frac{1}{4}[1 + \frac{1}{\sqrt{2}}(-1)^{a+b+xy}]$, which achieves the maximal quantum violation of the CHSH inequality in Eq. (3), by transforming the bipartite Bell network into the triangle network as described in Section II A. The Fritz distribution has no inputs and four outputs per party, and is given by

$$
\begin{aligned}
\frac{1}{32}(2+\sqrt{2}) =\ & p(0,0,0) = p(1,1,0) = p(0,2,1) = p(1,3,1) \\
=\ & p(2,0,2) = p(3,1,2) = p(2,3,3) = p(3,2,3), \\
\frac{1}{32}(2-\sqrt{2}) =\ & p(0,1,0) = p(1,0,0) = p(0,3,1) = p(1,2,1) \\
=\ & p(2,1,2) = p(3,0,2) = p(2,2,3) = p(3,3,3),
\end{aligned}
\tag{A1}
$$

with all probabilities not appearing above equal zero.

2. The triangle distribution based on the Elegant Joint Measurement

When each of the parties in the triangle scenario performs the Elegant Joint Measurement described in Eq. (15) in the respective shares of singlets $|\psi^-\rangle$ sent by the sources, a particularly simple distribution without inputs and four outputs is

generated. It reads

$$p(a = r, b = r, c = r) = \frac{25}{256} \qquad \text{for } r = 1, 2, 3, 4,$$

$$p(a = r, b = r, c = s) = \frac{1}{256} \qquad \text{for } r \neq s, \qquad \text{(A2)}$$

$$p(a = r, b = s, c = t) = \frac{5}{256} \qquad \text{for } r \neq s \neq t \neq r,$$

plus cyclic permutations of (a, b, c). A number of interesting properties about this distribution are collected in Ref. [64]. Despite many efforts, neither a local model nor a proof of network nonlocality has been found for this distribution.

3. The RGB4 distribution

The RGB4 distribution is the first triangle nonlocal distribution that is, arguably, conceptually different from the Fritz distribution. It was originally derived in Ref. [35] by having the parties perform the same four-outcome measurement on their respective shares of singlets $|\psi^-\rangle$ distributed by the sources, and it is given by

$$p(0, 1, 2) = p(1, 0, 3) = \frac{c^2}{8},$$

$$p(1, 0, 2) = p(0, 1, 3) = \frac{1 - c^2}{8},$$

$$p(a = b = c = 2, 3) = \frac{[(1 - c^2)^{3/2} + (-1)^a c^3]^2}{8}, \qquad \text{(A3)}$$

$$p(2, 2, 3) = \frac{[c^2 \sqrt{1 - c^2} - c(1 - c^2)]^2}{8},$$

$$p(2, 3, 3) = \frac{[c^2 \sqrt{1 - c^2} + c(1 - c^2)]^2}{8},$$

plus cyclic permutations, all the remaining probabilities being 0. The distribution depends on a parameter $c \in [-1, 1]$, and for values $0.785 \lesssim c^2 < 1$ it is known, using token-counting arguments (see Section III C), not to admit a local model in the triangle network.

[1] J. S. Bell, Physics Physique Fizika **1**, 195 (1964).
[2] M. Genovese, Phys. Rep. **413**, 319 (2005).
[3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).
[4] V. Scarani, *Bell nonlocality* (Oxford University Press, 2019).
[5] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
[6] H. J. Kimble, Nature **453**, 1023 (2008).
[7] S. Wehner, D. Elkouss, and R. Hanson, Science **362**, 6412 (2018).
[8] W. Kozlowski and S. Wehner, in *Proceedings of the Sixth Annual ACM International Conference on Nanoscale Computing and Communication*, NANOCOM '19 (Association for Computing Machinery, New York, NY, USA, 2019).
[9] M. Żukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, Phys. Rev. Lett. **71**, 4287 (1993).
[10] A. Fine, Phys. Rev. Lett. **48**, 291 (1982).
[11] D. Kaszlikowski, P. Gnaciński, M. Żukowski, W. Miklaszewski, and A. Zeilinger, Phys. Rev. Lett. **85**, 4418 (2000).
[12] M. Żukowski, D. Kaszlikowski, A. Baturo, and J.-Å. Larsson, Strengthening the Bell Theorem: conditions to falsify local realism in an experiment, arXiv:quant-ph/9910058.
[13] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[14] B. S. Cirel'son, Lett. Math. Phys. **4**, 93 (1980).
[15] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[16] S. J. Freedman and J. F. Clauser, Phys. Rev. Lett. **28**, 938 (1972).
[17] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. **49**, 1804 (1982).
[18] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-Å. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W.

Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger, Phys. Rev. Lett. **115**, 250401 (2015).

[19] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson, Nature **526**, 682–686 (2015).

[20] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam, Phys. Rev. Lett. **115**, 250402 (2015).

[21] W. Rosenfeld, D. Burchardt, R. Garthoff, K. Redeker, N. Ortegel, M. Rau, and H. Weinfurter, Phys. Rev. Lett. **119**, 010402 (2017).

[22] D. Rosset, N. Gisin, and E. Wolfe, Quantum Inf. Comput. **18**, 0910 (2018).

[23] M.-O. Renou, Y. Wang, S. Boreiri, S. Beigi, N. Gisin, and N. Brunner, Phys. Rev. Lett. **123**, 070403 (2019).

[24] G. Svetlichny, Phys. Rev. D **35**, 3066 (1987).

[25] T. Fritz, Commun. Math. Phys. **341**, 391 (2016).

[26] C. Branciard, N. Gisin, and S. Pironio, Phys. Rev. Lett. **104**, 170401 (2010).

[27] T. Fritz, New J. Phys. **14**, 103001 (2012).

[28] J. Henson, R. Lal, and M. F. Pusey, New J. Phys. **16**, 113043 (2014).

[29] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **80**, 022339 (2009).

[30] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, Phys. Rev. A **85**, 032119 (2012).

[31] A. Tavakoli, N. Gisin, and C. Branciard, Phys. Rev. Lett. **126**, 220401 (2021).

[32] E. Wolfe, R. W. Spekkens, and T. Fritz, J. Causal Inference **7**, 20170020 (2019).

[33] A. Tavakoli, P. Skrzypczyk, D. Cavalcanti, and A. Acín, Phys. Rev. A **90**, 062109 (2014).

[34] T. C. Fraser and E. Wolfe, Phys. Rev. A **98**, 022113 (2018).

[35] M.-O. Renou, E. Bäumer, S. Boreiri, N. Brunner, N. Gisin, and S. Beigi, Phys. Rev. Lett. **123**, 140401 (2019).

[36] I. Šupić, J.-D. Bancal, and N. Brunner, Phys. Rev. Lett. **125**, 240403 (2020).

[37] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **59**, 1070 (1999).

[38] M. Żukowski, A. Zeilinger, and H. Weinfurter, Ann. New York Acad. Sci **755**, 91 (1995).

[39] S. Bose, V. Vedral, and P. L. Knight, Phys. Rev. A **60**, 194 (1999).

[40] J.-W. Pan, D. Bouwmeester, H. Weinfurter, and A. Zeilinger, Phys. Rev. Lett. **80**, 3891 (1998).

[41] T. Jennewein, G. Weihs, J.-W. Pan, and A. Zeilinger, Phys. Rev. Lett. **88**, 017903 (2001).

[42] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

[43] C. H. Bennett and S. J. Wiesner, Phys. Rev. Lett. **69**, 2881 (1992).

[44] D. J. Saunders, A. J. Bennet, C. Branciard, and G. J. Pryde, Sci. Adv. **3**, e1602743 (2017).

[45] N. Gisin, Q. Mei, A. Tavakoli, M.-O. Renou, and N. Brunner, Phys. Rev. A **96**, 020304 (2017).

[46] F. Andreoli, G. Carvacho, L. Santodonato, R. Chaves, and F. Sciarrino, New J. Phys. **19**, 113020 (2017).

[47] R. Horodecki, P. Horodecki, and M. Horodecki, Phys. Lett. A **200**, 340 (1995).

[48] K. Mukherjee, B. Paul, and D. Sarkar, Quantum Inf. Process. **15**, 2895 (2016).

[49] A. Acín, T. Durt, N. Gisin, and J. I. Latorre, Phys. Rev. A **65**, 052325 (2002).

[50] T. Maudlin, PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association **1992**, 404 (1992).

[51] G. Brassard, R. Cleve, and A. Tapp, Phys. Rev. Lett. **83**, 1874 (1999).

[52] M. Steiner, Phys. Lett. A **270**, 239 (2000).

[53] D. Bacon and B. F. Toner, Phys. Rev. Lett. **90**, 157904 (2003).

[54] B. F. Toner and D. Bacon, Phys. Rev. Lett. **91**, 187904 (2003).

[55] J. Barrett and N. Gisin, Phys. Rev. Lett. **106**, 100406 (2011).

[56] C. Branciard, N. Brunner, H. Buhrman, R. Cleve, N. Gisin, S. Portmann, D. Rosset, and M. Szegedy, Phys. Rev. Lett. **109**, 100401 (2012).

[57] N. Gisin, Entropy **21**, 325 (2019).

[58] N. Gisin and S. Popescu, Phys. Rev. Lett. **83**, 432 (1999).

[59] J. Czartowski and K. Życzkowski, Quantum **5**, 442 (2021).

[60] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, J. Math. Phys. **45**, 2171 (2004).

[61] R. F. Werner, Phys. Rev. A **40**, 4277 (1989).

[62] M.-O. Renou and S. Beigi, Network nonlocality via rigidity of token-counting and color-matching, arXiv:2011.02769.

[63] P. Abiuso, T. Kriváchy, E.-C. Boghiu, M.-O. Renou, A. Pozas-Kerstjens, and A. Acín, Quantum networks reveal single-photon nonlocality, arXiv:2108.01726.

[64] N. Gisin, The Elegant joint quantum measurement and some conjectures about N-locality in the triangle and other configurations, arXiv:1708.05556.

[65] T. Kriváchy, Y. Cai, D. Cavalcanti, A. Tavakoli, N. Gisin, and N. Brunner, npj Quantum Inf. **6**, 70 (2020).

[66] B. Steudel and N. Ay, Entropy **17**, 2304 (2015).

[67] E. Wolfe, A. Pozas-Kerstjens, M. Grinberg, D. Rosset, A. Acín, and M. Navascués, Phys. Rev. X **11**, 021043 (2021).

[68] N. D. Mermin, Phys. Rev. Lett. **65**, 1838 (1990).

[69] N. Gisin, J.-D. Bancal, Y. Cai, P. Remy, A. Tavakoli, E. Zambrini Cruzeiro, S. Popescu, and N. Brunner, Nat. Commun. **11**, 2378 (2020).

[70] M.-X. Luo, Phys. Rev. Lett. **120**, 140402 (2018).

[71] D. M. Greenberger, M. A. Horne, and A. Zeilinger, Going Beyond Bell's Theorem, in: Bell's Theorem, Quantum Theory, and Conceptions of the Universe', M. Kafatos (Ed.), Kluwer, Dordrecht, 69-72 (1989), arXiv:0712.0921.

[72] L.-Y. Hsu and C.-H. Chen, Phys. Rev. Research **3**, 023139 (2021).

[73] D. Rosset, C. Branciard, T. J. Barnea, G. Pütz, N. Brunner, and N. Gisin, Phys. Rev. Lett. **116**, 010403 (2016).

[74] A. Tavakoli, Phys. Rev. A **93**, 030101 (2016).

[75] R. Chaves, Phys. Rev. Lett. **116**, 010402 (2016).

[76] C. Budroni and A. Cabello, J. Phys. A: Math. Theor. **45**, 385304 (2012).

[77] T. Fritz and R. Chaves, IEEE Trans. Inf. Theory **59**, 803 (2013).

[78] H. P. Williams, Am. Math. Mon. **93**, 681 (1986).

[79] A. Tavakoli, J. Phys. A: Math. Theor. **49**, 145304 (2016).

[80] A. Tavakoli, M.-O. Renou, N. Gisin, and N. Brunner, New J. Phys. **19**, 073003 (2017).

[81] R. Augusiak, M. Demianowicz, and A. Acín, J. Phys. A: Math. Theor. **47**, 424002 (2014).

[82] A. Sen(De), U. Sen, Č. Brukner, V. Bužek, and M. Żukowski, Phys. Rev. A **72**, 042310 (2005).

[83] D. Cavalcanti, M. L. Almeida, V. Scarani, and A. Acín, Nat. Commun. **2**, 184 (2011).

[84] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **81**, 5932 (1998).

[85] N. Sangouard, C. Simon, H. de Riedmatten, and N. Gisin, Rev. Mod. Phys. **83**, 33 (2011).

[86] K. Mukherjee, B. Paul, and D. Sarkar, Quantum Inf. Process. **14**, 2025 (2015).

[87] A. M. Goebel, C. Wagenknecht, Q. Zhang, Y.-A. Chen, K. Chen, J. Schmiedmayer, and J.-W. Pan, Phys. Rev. Lett. **101**, 080403 (2008).

[88] K. Mukherjee, B. Paul, and D. Sarkar, Phys. Rev. A **96**, 022103 (2017).

[89] K. Mukherjee, B. Paul, and A. Roy, Phys. Rev. A **101**, 032328 (2020).

[90] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, 2004).

[91] L. Vandenberghe and S. Boyd, SIAM Rev. **38**, 49 (1996).

[92] M. Navascués, S. Pironio, and A. Acín, Phys. Rev. Lett. **98**, 010401 (2007).

[93] M. Navascués, S. Pironio, and A. Acín, New J. Phys. **10**, 073013 (2008).

[94] M. Navascués and T. Vértesi, Phys. Rev. Lett. **115**, 020501 (2015).

[95] I. Kogias, P. Skrzypczyk, D. Cavalcanti, A. Acín, and G. Adesso, Phys. Rev. Lett. **115**, 210401 (2015).

[96] Y. Wang, I. W. Primaatmaja, E. Lavie, A. Varvitsiotis, and C. C. W. Lim, npj Quantum Inf. **5**, 17 (2019).

[97] J. Bowles, F. Baccari, and A. Salavrakos, Quantum **4**, 344 (2020).

[98] A. Tavakoli, E. Zambrini Cruzeiro, E. Woodhead, and S. Pironio, Informationally restricted correlations: a general framework for classical and quantum systems, arXiv:2007.16145.

[99] A. Tavakoli, E. Z. Cruzeiro, R. Uola, and A. A. Abbott, PRX Quantum **2**, 020334 (2021).

[100] A. Chaturvedi, M. Farkas, and V. J. Wright, Quantum **5**, 484 (2021).

[101] H. H. Jee, C. Sparaciari, O. Fawzi, and M. Berta, in *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 198, edited by N. Bansal, E. Merelli, and J. Worrell (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2021) pp. 82:1–82:20.

[102] S. Pironio, M. Navascués, and A. Acín, SIAM J. Optim. **20**, 2157 (2010).

[103] W. Slofstra, For. Math., Pi **7**, e1 (2019).

[104] Z. Ji, A. Natarajan, T. Vidick, J. Wright, and H. Yuen, MIP*=RE, arXiv:2001.04383.

[105] P. A. Parrilo, Math. Program. **96**, 293 (2003).

[106] M. F. Anjos and J. B. Lasserre, *Handbook on Semidefinite, Conic and Polynomial Optimization* (Springer, 2011).

[107] M. Grant and S. Boyd, *CVX: MATLAB Software for Disciplined Convex Programming, version 2.1* (2014).

[108] M. Grant and S. Boyd, in *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, edited by V. Blondel, S. Boyd, and H. Kimura (Springer-Verlag Limited, 2008) pp. 95–110.

[109] J. Löfberg, in *2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508)* (2004) pp. 284–289.

[110] P. Wittek, ACM Trans. Math. Softw. **41**, 21 (2015).

[111] G. Sagnol and M. Stahlberg, *PICOS: A Python interface to conic optimization solvers* (2020).

[112] MOSEK ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.* (2019).

[113] M. Yamashita, K. Fujisawa, M. Fukuda, K. Kobayashi, K. Nakta, and M. Nakata, in *Handbook on Semidefinite, Cone and Polynomial Optimization: Theory, Algorithms, Software and Applications*, edited by M. F. Anjos and J. B. Lasserre (Springer-Verlag Limited, 2011) Chap. 24, pp. 687–714.

[114] Gurobi Optimization, LLC, *Gurobi Optimizer Reference Manual* (2020).

[115] A. Pozas-Kerstjens, R. Rabelo, Ł. Rudnicki, R. Chaves, D. Cavalcanti, M. Navascués, and A. Acín, Phys. Rev. Lett. **123**, 140503 (2019).

[116] A. Pozas-Kerstjens, *Quantum information outside quantum information*, Ph.D. thesis (2019).

[117] A. Pozas-Kerstjens, Zenodo **2648405** (2019).

[118] I. Klep, V. Magron, and J. Volčič, Ann. Henri Poincaré (2021).

[119] G. ver Steeg and A. Galstyan, in *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, UAI'11 (AUAI Press, Arlington, Virginia, USA, 2011) p. 717–726.

[120] P. H. Eberhard, Phys. Rev. A **47**, R747 (1993).

[121] S. Massar and S. Pironio, Phys. Rev. A **68**, 062109 (2003).

[122] M. Navascués and E. Wolfe, J. Causal Inference **8**, 70 (2020).

[123] V. Scarani, S. Iblisdir, N. Gisin, and A. Acín, Rev. Mod. Phys. **77**, 1225 (2005).

[124] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **61**, 662 (1988).

[125] M. Weilenmann and R. Colbeck, Proc. R. Soc. A **473**, 20170483 (2017).

[126] A. Rényi, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (University of California Press, Berkeley, Calif., 1961) pp. 547–561.

[127] V. Vilasini and R. Colbeck, Phys. Rev. A **100**, 062108 (2019).

[128] R. Chaves and C. Budroni, Phys. Rev. Lett. **116**, 240501 (2016).

[129] Z. Zhang and R. W. Yeung, IEEE Trans. Inf. Theory **44**, 1440 (1998).

[130] N. Linden and A. Winter, Commun. Math. Phys. **259**, 129 (2005).

[131] R. Dougherty, C. Freiling, and K. Zeger, in *2006 IEEE International Symposium on Information Theory* (2006) pp. 233–236.

[132] R. Dougherty, C. Freiling, and K. Zeger, Non-Shannon information inequalities in four random variables, arXiv:1104.3602.

[133] M. Weilenmann and R. Colbeck, Quantum **2**, 57 (2018).

[134] R. Chaves and T. Fritz, Phys. Rev. A **85**, 032113 (2012).

[135] T. Fritz and R. Chaves, IEEE Trans. Inf. Theory **59**, 803 (2013).

[136] R. Chaves, L. Luft, T. O. Maciel, D. Gross, D. Janzing, and B. Schölkopf, in *Proc. of the 30th Conference on Uncertainty in Artificial Intelligence* (AUAI, 2014) pp. 112–121.

[137] R. Chaves, G. Moreno, E. Polino, D. Poderini, I. Agresti, A. Suprano, M. R. Barros, G. Carvacho, E. Wolfe, A. Canabarro, R. W. Spekkens, and F. Sciarrino, Causal networks and freedom of choice in Bell's theorem, arXiv:2105.05721.

[138] R. Chaves, C. Majenz, and D. Gross, Nat. Commun. **6**, 5766 (2015).

[139] E. H. Lieb and M. B. Ruskai, J. Math. Phys. **14**, 1938 (1973).

[140] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, Nature **461**, 1101 (2009).

[141] J. Åberg, R. Nery, C. Duarte, and R. Chaves, Phys. Rev. Lett. **125**, 110505 (2020).

[142] J. Barrett, Phys. Rev. A **75**, 032304 (2007).

[143] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **81**, 062348 (2010).

[144] J.-D. Bancal and N. Gisin, Non-local boxes for networks, arXiv:2102.03597.

[145] A. Kela, K. Von Prillwitz, J. Åberg, R. Chaves, and D. Gross, IEEE Trans. Inf. Theory **66**, 339 (2020).

[146] T. Kraft, C. Spee, X.-D. Yu, and O. Gühne, Phys. Rev. A **103**, 052405 (2021).

[147] S. Beigi and M.-O. Renou, Covariance decomposition as a universal limit on correlations in networks, arXiv:2103.14840.

[148] W. van Dam and P. Hayden, Phys. Rev. A **67**, 060302 (2003).

[149] H. Finner, Ann. Probab. **20**, 1893 (1992).

[150] L. J. Rogers, Messenger Math. New Ser. **XVII**, 145 (1888).

[151] O. Hölder, Nachr. Königl. Ges. Wiss. Georg-Augusts-Univ. **1889**, 38 (1889).

[152] M. K. Shukla, M. Huang, I. Chakrabarty, and J. Wu, Three party quantum networks created by quantum cloning, arXiv:2011.07554.

[153] M.-X. Luo, Network configuration theory for all networks, arXiv:2107.05846.

[154] J.-W. Pan, Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger, and M. Żukowski, Rev. Mod. Phys. **84**, 777 (2012).

[155] K. Mattle, H. Weinfurter, P. G. Kwiat, and A. Zeilinger, Phys. Rev. Lett. **76**, 4656 (1996).

[156] N. Lütkenhaus, J. Calsamiglia, and K.-A. Suominen, Phys. Rev. A **59**, 3295 (1999).

[157] J. Calsamiglia and N. Lütkenhaus, Appl. Phys. B **72**, 67 (2001).

[158] P. G. Kwiat and H. Weinfurter, Phys. Rev. A **58**, R2623 (1998).

[159] H. Weinfurter, EPL **25**, 559 (1994).

[160] S. L. Braunstein and A. Mann, Phys. Rev. A **51**, R1727 (1995).

[161] J.-W. Pan, M. Daniell, S. Gasparoni, G. Weihs, and A. Zeilinger, Phys. Rev. Lett. **86**, 4435 (2001).

[162] H. de Riedmatten, I. Marcikic, J. A. W. van Houwelingen, W. Tittel, H. Zbinden, and N. Gisin, Phys. Rev. A **71**, 050302 (2005).

[163] F. Sciarrino, E. Lombardi, G. Milani, and F. De Martini, Phys. Rev. A **66**, 024309 (2002).

[164] X. Jia, X. Su, Q. Pan, J. Gao, C. Xie, and K. Peng, Phys. Rev. Lett. **93**, 250503 (2004).

[165] N. Takei, H. Yonezawa, T. Aoki, and A. Furusawa, Phys. Rev. Lett. **94**, 220502 (2005).

[166] T. Yang, Q. Zhang, T.-Y. Chen, S. Lu, J. Yin, J.-W. Pan, Z.-Y. Wei, J.-R. Tian, and J. Zhang, Phys. Rev. Lett. **96**, 110501 (2006).

[167] R. Kaltenbaek, B. Blauensteiner, M. Żukowski, M. Aspelmeyer, and A. Zeilinger, Phys. Rev. Lett. **96**, 240502 (2006).

[168] M. Halder, A. Beveratos, N. Gisin, V. Scarani, C. Simon, and H. Zbinden, Nat. Phys. **3**, 692 (2007).

[169] R. Kaltenbaek, R. Prevedel, M. Aspelmeyer, and A. Zeilinger, Phys. Rev. A **79**, 040302 (2009).

[170] C. Schmid, N. Kiesel, U. K. Weber, R. Ursin, A. Zeilinger, and H. Weinfurter, New J. Phys. **11**, 033008 (2009).

[171] S. Takeda, M. Fuwa, P. van Loock, and A. Furusawa, Phys. Rev. Lett. **114**, 100501 (2015).

[172] G. Guccione, T. Darras, H. Le Jeannic, V. B. Verma, S. W. Nam, A. Cavaillès, and J. Laurat, Sci. Adv. **6**, eaba4508 (2020).

[173] C.-Y. Lu, T. Yang, and J.-W. Pan, Phys. Rev. Lett. **103**, 020501 (2009).

[174] X. Su, C. Tian, X. Deng, Q. Li, C. Xie, and K. Peng, Phys. Rev. Lett. **117**, 240503 (2016).

[175] B. B. Blinov, D. L. Moehring, L. Duan, and C. Monroe, Nature **428**, 153 (2004).

[176] D. L. Moehring, M. J. Madsen, B. B. Blinov, and C. Monroe, Phys. Rev. Lett. **93**, 090410 (2004).

[177] J. Volz, M. Weber, D. Schlenk, W. Rosenfeld, J. Vrana, K. Saucke, C. Kurtsiefer, and H. Weinfurter, Phys. Rev. Lett. **96**, 030404 (2006).

[178] D. L. Moehring, P. Maunz, S. Olmschenk, K. C. Younge, D. N. Matsukevich, L. Duan, and C. Monroe, Nature **449**, 68 (2007).

[179] Z.-S. Yuan, Y.-A. Chen, B. Zhao, S. Chen, J. Schmiedmayer, and J.-W. Pan, Nature **454**, 1098 (2008).

[180] D. N. Matsukevich, P. Maunz, D. L. Moehring, S. Olmschenk, and C. Monroe, Phys. Rev. Lett. **100**, 150404 (2008).

[181] J. Hofmann, M. Krug, N. Ortegel, L. Gérard, M. Weber, W. Rosenfeld, and H. Weinfurter, Science **337**, 72 (2012).

[182] M. Riebe, T. Monz, K. Kim, A. S. Villar, P. Schindler, M. Chwalla, M. Hennrich, and R. Blatt, Nat. Phys. **4**, 839 (2008).

[183] W. Ning, X.-J. Huang, P.-R. Han, H. Li, H. Deng, Z.-B. Yang, Z.-R. Zhong, Y. Xia, K. Xu, D. Zheng, and S.-B. Zheng, Phys. Rev. Lett. **123**, 060502 (2019).

[184] G. Carvacho, F. Andreoli, L. Santodonato, M. Bentivegna, R. Chaves, and F. Sciarrino, Nat. Commun. **8**, 14775 (2017).

[185] Q.-C. Sun, Y.-F. Jiang, B. Bai, W. Zhang, H. Li, X. Jiang, J. Zhang, L. You, X. Chen, Z. Wang, Q. Zhang, J. Fan, and J.-W. Pan, Nat. Photonics **13**, 687 (2019).

[186] D. Poderini, I. Agresti, G. Marchese, E. Polino, T. Giordani, A. Suprano, M. Valeri, G. Milani, N. Spagnolo, G. Carvacho, R. Chaves, and F. Sciarrino, Nat. Commun. **11**, 2467 (2020).

[187] E. Bäumer, N. Gisin, and A. Tavakoli, npj Quantum Inf. **7**, 117 (2021).

[188] F. Andreoli, G. Carvacho, L. Santodonato, M. Bentivegna, R. Chaves, and F. Sciarrino, Phys. Rev. A **95**, 062315 (2017).

[189] P. van Loock and N. Lütkenhaus, Phys. Rev. A **69**, 012302 (2004).

[190] I. Agresti, B. Polacchi, D. Poderini, E. Polino, A. Suprano, I. Šupić, J. Bowles, G. Carvacho, D. Cavalcanti, and F. Sciarrino, PRX Quantum **2**, 020346 (2021).

[191] IBM Quantum Experience, http://www.research.ibm.com/ibm-q/.

[192] A. Pozas-Kerstjens, N. Gisin, and A. Tavakoli, Full network nonlocality, arXiv:2105.09325.

[193] M. Seevinck and J. Uffink, Phys. Rev. A **65**, 012107 (2001).

[194] M.-X. Luo and S.-M. Fei, Robust multipartite entanglement without entanglement breaking, arXiv:2106.09240.

[195] I. Šupić, J.-D. Bancal, Y. Cai, and N. Brunner, Genuine network quantum nonlocality and self-testing, arXiv:2105.12341.

[196] S. Pironio, V. Scarani, and T. Vidick, New J. Phys. **18**, 100202 (2016).

[197] D. Mayers and A. Yao, in *Proceedings 39th Annual Symposium on Foundations of Computer Science (FOCS)* (1998) pp. 503–509.

[198] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).

[199] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).

[200] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Nat. Photonics **4** (2010).

[201] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002).

[202] F. Xu, X. Ma, Q. Zhang, H.-K. Lo, and J.-W. Pan, Rev. Mod. Phys. **92**, 025002 (2020).

[203] C. Portmann and R. Renner, Security in quantum cryptography, arXiv:2102.00021.

[204] E. P. Wigner, Am. J. Phys. **38**, 1005 (1970).

[205] S. L. Braunstein and C. M. Caves, Ann. Phys. (N.Y.) **202**, 22 (1990).

[206] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).

[207] L. Masanes, S. Pironio, and A. Acín, Nat. Commun. **2**, 238 (2011).

[208] U. Vazirani and T. Vidick, Phys. Rev. Lett. **113**, 140501 (2014).

[209] C. M. Lee and M. J. Hoban, Phys. Rev. Lett. **120**, 020504 (2018).

[210] M.-X. Luo, Phys. Rev. A **101**, 062317 (2020).

[211] M.-X. Luo, Fully device-independent model on quantum networks, arXiv:2106.15840.

[212] D. Mayers and A. Yao, Quantum Inf. Comput. **4**, 273 (2004).

[213] I. Šupić and J. Bowles, Quantum **4**, 337 (2020).

[214] J.-D. Bancal, N. Sangouard, and P. Sekatski, Phys. Rev. Lett. **121**, 250506 (2018).

[215] M.-O. Renou, J. Kaniewski, and N. Brunner, Phys. Rev. Lett. **121**, 250507 (2018).

[216] W.-H. Zhang, G. Chen, X.-X. Peng, X.-J. Ye, P. Yin, X.-Y. Xu, J.-S. Xu, C.-F. Li, and G.-C. Guo, Phys. Rev. Lett. **122**, 90402 (2019).

[217] M. Weilenmann and R. Colbeck, Phys. Rev. Lett. **125**, 060406 (2020).

[218] M. Weilenmann and R. Colbeck, Phys. Rev. A **102**, 022203 (2020).

[219] K. F. Pál and T. Vértesi, Phys. Rev. A **77**, 042105 (2008).

[220] M. McKague, M. Mosca, and N. Gisin, Phys. Rev. Lett. **102**, 020505 (2009).

[221] T. Vértesi and E. Bene, Phys. Rev. A **82**, 062115 (2010).

[222] O. Andersson, P. Badziąg, I. Bengtsson, I. Dumitru, and A. Cabello, Phys. Rev. A **96**, 032119 (2017).

[223] A. Tavakoli, M. Smania, T. Vértesi, N. Brunner, and M. Bourennane, Sci. Adv. **6**, aaw6664 (2020).

[224] M. Smania, P. Mironowicz, M. Nawareg, M. Pawłowski, A. Cabello, and M. Bourennane, Optica **7**, 123 (2020).

[225] M.-O. Renou, D. Trillo, M. Weilenmann, L. P. Thinh, A. Tavakoli, N. Gisin, A. Acín, and M. Navascués, Quantum physics needs complex numbers, arXiv:2101.10873.

[226] J. Bowles, I. Šupić, D. Cavalcanti, and A. Acín, Phys. Rev. A **98**, 042336 (2018).

[227] L.-M. Duan and C. Monroe, Rev. Mod. Phys. **82**, 1209 (2010).

[228] S. Ritter, C. Nölleke, C. Hahn, A. Reiserer, A. Neuzner, M. Uphoff, M. Mücke, E. Figueroa, J. Bochmann, and G. Rempe, Nature **484**, 195 (2012).

[229] S. Popescu and D. Rohrlich, Phys. Lett. A **166**, 293 (1992).

[230] M.-X. Luo, Phys. Rev. A **98**, 042317 (2018).

[231] P. Contreras-Tejada, C. Palazuelos, and J. I. de Vicente, Phys. Rev. Lett. **126**, 040501 (2021).

[232] S. Pironio, J. Math. Phys. **46**, 062112 (2005).

[233] N. Brunner and N. Linden, Nat. Commun. **4**, 2057 (2013).

[234] J. Silman, S. Machnes, and N. Aharon, Phys. Lett. A **372**, 3796 (2008).

[235] M.-X. Luo, npj Quantum Inf. **5**, 91 (2019).

[236] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Rev. Mod. Phys. **91**, 045002 (2019).

[237] A. Canabarro, S. Brito, and R. Chaves, Phys. Rev. Lett. **122**, 200401 (2019).

[238] S. Carter and M. Nielsen, Distill **9** (2017).

[239] P. Friederich, M. Krenn, I. Tamblyn, and A. Aspuru-Guzik, Mach. Learn.: Sci. Technol. **2**, 025027 (2021).

[240] M. Krenn, M. Malik, R. Fickler, R. Lapkiewicz, and A. Zeilinger, Phys. Rev. Lett. **116**, 090405 (2016).

[241] M. Krenn, J. Kottmann, N. Tischler, and A. Aspuru-Guzik, Phys. Rev. X **11**, 031044 (2021).

[242] A. A. Melnikov, P. Sekatski, and N. Sangouard, Phys. Rev. Lett. **125**, 160401 (2020).

[243] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).

[244] O. Gühne and G. Tóth, Phys. Rep. **474**, 1 (2009).

[245] N. Friis, G. Vitagliano, M. Malik, and M. Huber, Nat. Rev. Phys. **1**, 72 (2019).

[246] M. Navascués, E. Wolfe, D. Rosset, and A. Pozas-Kerstjens, Phys. Rev. Lett. **125**, 240505 (2020).

[247] H. J. Briegel and R. Raussendorf, Phys. Rev. Lett. **86**, 910 (2001).

[248] M. Hein, J. Eisert, and H. J. Briegel, Phys. Rev. A **69**, 062311 (2004).

[249] T. Kraft, S. Designolle, C. Ritz, N. Brunner, O. Gühne, and M. Huber, Phys. Rev. A **103**, L060401 (2021).

[250] M.-X. Luo, Adv. Quantum Technol. **4**, 2000123 (2021).

[251] W. Dür, G. Vidal, and J. I. Cirac, Phys. Rev. A **62**, 062314 (2000).

[252] G. Tóth, J. Opt. Soc. Am. B **24**, 275 (2007).

[253] D. Mayers, Phys. Rev. Lett. **78**, 3414 (1997).

[254] H.-K. Lo and H. F. Chau, Phys. Rev. Lett. **78**, 3410 (1997).

[255] A. Kent, Phys. Rev. Lett. **83**, 1447 (1999).

[256] A. Kent, New J. Phys. **13**, 113015 (2011).

[257] M.-X. Luo and X. Wang, Unconditionally secure quantum bit commitment: Revised, Cryptology ePrint Archive, Report 2020/1077 (2020).

[258] X. Coiteux-Roy, E. Wolfe, and M.-O. Renou, No bipartite-nonlocal causal theory can explain nature's correlations, arXiv:2105.09381.

[259] X. Coiteux-Roy, E. Wolfe, and M.-O. Renou, Any physical theory of nature must be boundlessly multipartite nonlocal, arXiv:2105.09380.

[260] D. R. Hamel, L. K. Shalm, H. Hübel, A. J. Miller, V. B. Verma, R. P. Mirin, S. Woo Nam, K. J. Resch, and T. Jennewein, Nat. Photonics **8**, 801 (2014).

[261] J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, 2009).

[262] H. Reichenbach and M. Reichenbach, *The Direction of Time* (University of California Press, 1991).

[263] I. Agresti, G. Carvacho, D. Poderini, L. Aolita, R. Chaves, and F. Sciarrino, Proc. MDPI AG **12**, 27 (2019).

[264] T. van Himbeeck, J. Bohr Brask, S. Pironio, R. Ramanathan, A. B. Sainz, and E. Wolfe, Quantum **3**, 186 (2019).

[265] M. Gachechiladze, N. Miklin, and R. Chaves, Phys. Rev. Lett. **125**, 230401 (2020).

[266] P. W. Holland, J. Am. Stat. Assoc. **81**, 945 (1986).

[267] I. Shpitser and J. Pearl, J. Mach. Learn. Res. **9**, 1941 (2008).

[268] R. Chaves, G. Carvacho, I. Agresti, V. Di Giulio, L. Aolita, S. Giacomini, and F. Sciarrino, Nat. Phys. **14**, 291 (2018).

[269] I. Agresti, D. Poderini, L. Guerini, M. Mancusi, G. Carvacho, L. Aolita, D. Cavalcanti, R. Chaves, and F. Sciarrino, Commun. Phys. **3**, 110 (2020).

[270] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, Phys. Rev. X **7**, 031021 (2017).

[271] J. Pienaar and Č. Brukner, New J. Phys. **17**, 073020 (2015).

[272] J. Barrett, R. Lorenz, and O. Oreshkov, Quantum causal models, arXiv:1906.10726.

[273] C. Giarmatzi and F. Costa, npj Quantum Inf. **4** (2018).