Linear programs for entanglement and key distribution in the quantum internet

Bäuml, Stefan; Azuma, Koji; Kato, Go; Elkouss, David

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Check for updates

Linear programs for entanglement and key distribution in the quantum internet

Stefan Bäuml[1,2,3,4 ✉], Koji Azuma[3,4], Go Kato[4,5] & David Elkouss[1]

Quantum networks will allow to implement communication tasks beyond the reach of their classical counterparts. A pressing and necessary issue for the design of quantum network protocols is the quantification of the rates at which these tasks can be performed. Here, we propose a simple recipe that yields efficiently computable lower and upper bounds on the maximum achievable rates. For this we make use of the max-flow min-cut theorem and its generalization to multi-commodity flows to obtain linear programs. We exemplify our recipe deriving the linear programs for bipartite settings, settings where multiple pairs of users obtain entanglement in parallel as well as multipartite settings, covering almost all known situations. We also make use of a generalization of the concept of paths between user pairs in a network to Steiner trees spanning a group of users wishing to establish Greenberger-Horne-Zeilinger states.

[1] QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, Netherlands. [2] ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, Av. Carl Friedrich Gauss 3, 08860 Castelldefels, Barcelona, Spain. [3] NTT Basic Research Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan. [4] NTT Research Center for Theoretical Quantum Physics, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi 243-0198, Japan. [5] NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi 243-0198, Japan. ✉email: stefan.baeuml@icfo.eu

Quantum entanglement allows for the implementation of communication tasks not possible by classical means. The most prominent examples are quantum key distribution and quantum teleportation between two parties[1–3], but there is a host of other tasks also involving more than two parties[4]. An example of a protocol using multipartite entanglement is quantum conference key agreement[5], where multiple parties who trust each other need to establish a common key. Another example is quantum secret sharing[6], where multiple parties who do not trust each other wish to encrypt a message in such a way that it can only be decrypted if all parties cooperate. Multipartite entanglement can also be used for the synchronization of a network of clocks[7] and plays an important role in quantum computing[8]. Quantum networks allow for the distribution of entanglement as a resource for such tasks among parties, which could, in principle, be spread out across different continents in an efficient manner. Whereas small-scale quantum networks can be designed in such a way that they perform optimally in distributing a particular resource to a particular set of users, a future quantum version of the internet will most likely grow to have a complex structure and involve a number of user pairs, or groups, requiring entangled resources for different tasks in parallel.

Recently, in light of the experimental promise of short-term quantum network deployment, the community has begun to devote attention to communication problems for networks of noisy quantum channels and their general structures. Arguably, the most important one is the computation of the maximum rates at which the different tasks can be performed. Given that, even in the case of point-to-point links, entanglement makes the characterization of capacities notably more complicated than its classical counterpart, with phenomena such as superactivation[9], it was unclear how much it would be possible to borrow from the theory of classical networks. Besides, the usage of a quantum channel is much more expensive than that of its classical counterpart. This motivates the introduction of different capacities which account for resources in different ways. The results of refs. [10,11] introduced the quantum problem and successfully established upper and lower bounds on a capacity of a quantum network which quantifies the maximum size of bipartite maximally entangled states (for quantum teleportation) or private states (for quantum key distribution) per network use, as a generalization of the fundamental/established notion of classical network capacity[12]. These upper and lower bounds coincide when the network is composed only of a very relevant class of quantum channels, called distillable channels. The results of refs. [13–15] derive analogous bounds, alternatively defining the capacity of a quantum network per total number of channel uses (related with a cost) or per time, rather than per network use, for generality. In any case, rather surprisingly, the series of fundamental works[10,11,13–15] have shown that these capacity of quantum networks for bipartite communication behave similar to that of classical networks. The distribution of bi- and multipartite entanglement in quantum networks has been in considered in a number of other works, including refs. [16–22]. These works differ from refs. [10,11,13–15] in that they are not concerned with networks of general noisy channels.

Namely, given a network of quantum noisy channels and bounds on their capacities satisfying certain properties, one can conceptually construct a classical version of the quantum network where each quantum channel is replaced by a perfect classical channel with a capacity given by the bound on the quantum channel capacity. Then, by considering cuts between two nodes in the induced 'classical' network, it is possible to obtain upper and lower bounds on a capacity of the network for distributing private keys or entanglement between two clients. The same techniques have found application for many user pairs[10,23,24] and for the

distribution of multipartite entanglement among multiple users[24,25]. While the early work has laid down extremely useful techniques to characterize quantum network capacities, it has either not focused on their computation[13,24] or left open the computability of several of the scenarios considered[11]. However, this is rather important in practice, in the sense that the quantum network will be required to serve entanglement resources quickly according to the requests of clients, and, in so doing, efficient estimation of the quantum network capacities is a necessary basis for choosing a proper subnetwork to accomplish that. The goal of this paper is to provide a simple recipe to find such efficiently computable bounds for quantum network capacities.

In this paper, using the approach taken in refs. [13–15,24], i.e., defining a network capacity as a rate per the total number of channel uses or per time, we introduce or generalize the capacities for private or quantum communication in the following scenarios: bipartite communication, concurrent communication between multiple user pairs with the objective of (1) maximizing the sum of rates achieved by the user pairs or (2) maximizing the worst-case rate that can be achieved by any pair, as well as multipartite state sharing where the goal is either to distribute Greenberger–Horne–Zeilinger (GHZ) or multipartite private states[5] for a group of network users. We then provide linear-program lower and upper bounds on the all these capacities. The size of the linear programs (LPs) scales polynomially in the parameters of the network, making it computable in polynomial time by interior point algorithms[26]. A central tool deriving upper bounds in the case of multiple user pairs are approximate min-cut max-flow theorems for multi-commodity flows[27–29]. Up to a factor of the logarithmic order of the number of user pairs, these results link quantities that occur in the known upper bounds[24], such as the minimum cut ratio (i.e., the smallest ratio of the capacity of a cut and the demand across the cut) and the minimum capacity multicut (i.e., the smallest capacity set of edges whose removal disconnects all user pairs), both of which are NP-hard problems to calculate in general graphs[27,30], to multi-commodity flow maximizations that can be computed by LPs. A challenge we address in this work is to find protocols that can achieve the upper bounds. In the bipartite case, protocols involving distillation of Bell pairs across all edges of a network, and entanglement swapping along paths have been used to provide lower bounds on the network capacities[10,11,14]. Using such simple routing methods, it was shown in refs. [10,11,14] that the bipartite upper bounds can be achieved for networks consisting of a wide class of channels, known as distillable channels[31], which include erasure channels, dephasing channels, bosonic quantum amplifier channels, and lossy optical channels. Here, we extend the bipartite protocol presented in ref. [14] to the case of many user pairs and to the distribution of GHZ states among a set of users. We do so by considering edge-disjoint Steiner trees spanning the set of users.

Results

Our base setup is a network of nodes connected by noisy quantum channels (see Fig. 1). The nodes act either as end users or as repeater stations and have the ability to store and process quantum information locally. In addition, all nodes are connected by classical lines of communication, which can be used freely.

We are interested in the possibilities and limitations of quantum networks for different communication tasks and usage scenarios. Fortunately, most tasks of interest can be rephrased as the distribution of an entangled target state among users of the quantum network[32]. Here, we consider the distribution of a bipartite entangled target state between a pair of users, of multiple bipartite entangled target states between multiple pairs of users in parallel as well as of a multipartite entangled target state among a
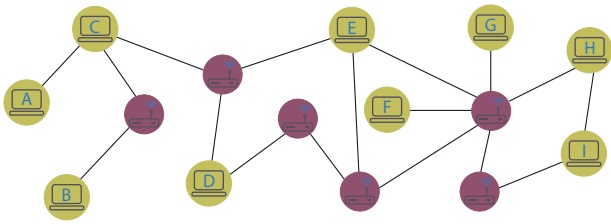
Fig. 1 Example of a quantum network: it consists of quantum channels, repeater stations and end users A, B,..., I. In such a network there are many possible communication tasks. Some examples are the distribution of private states, Bell states and Greenberger–Horne–Zeilinger (GHZ) states. The first two are bipartite tasks. We study the implementation of these tasks between a single pair of users, for instance *A* and *I*, and between multi-pairs of users in parallel, for instance *A* and *I*, *C* and *D*, and *F* and *H*. The last task, the distribution of GHZ states, is a multipartite user scenario, for instance *A*, *B*, *E*, *G*, and *I* could distill a five-partite GHZ state.

group consisting of more than two users. The distribution of these states is known to be equivalent to the problems of quantum information transmission, private classical communication, quantum key distribution, and quantum conference key agreement among others.

As we are interested in emergent, organically grown quantum networks, such as the classical internet, we do not make any assumptions on the structure of the network except that it can be described by a finite directed graph. Let the quantum network be given by the directed graph $G = (V, E)$, where V denotes the set of the finite vertices and E the set of the finite directed edges, which represent quantum channels. Each directed edge $e \in E$ has tail $v \in V$ and head $w \in V$. We also denote e by vw. We can also assign a nonnegative edge capacity (In order to avoid confusion, we will use the terms 'edge capacity' when referring to edges and 'capacity', when referring to quantum or private capacities of a channel or the entire network.) $c(e)$ to every edge. Each edge vw corresponds to a channel $\mathcal{N}^e = \mathcal{N}^{vw}$ with input in v and output in w. We assume that each vertex has the capability to store and process quantum information locally and that all vertices are connected by public lines of classical communication, the use of both of which is considered to be a free resource. Let us assume there is a subset $U \subset V$ of the vertices, the users who wish to establish a target state containing the desired resource, whereas the remaining vertices serve as repeater stations. In the following section we will elaborate on the exact form of the target state.

We assume that initially there is no entanglement between any of the vertices. A target state can be distributed by means of an adaptive protocol, consisting of local operations and classical communication (LOCC) among the vertices in the network interleaved by channel uses[10,11,13]. In this work we are not concerned with the inner workings of the protocol but describe a protocol only by the total number of channel uses, and usage frequencies of each channel. We describe a protocol as follows: Given upper bounds n_e on the average of the number of uses of each channel $\mathcal{N}^e$, we define a set of usage frequencies $\{p_e\}_{e \in E}$ of each channel $\mathcal{N}^e$ as $p_e := n_e/n (\geq 0)$. Here n can be regarded as time or n with $\sum_{e \in E} p_e = 1$ can be considered to be an upper bound on the average of total channel uses (see ref. [14]). Further, we introduce an error parameter ϵ such that after the final round of LOCC a state ϵ-close in trace distance to the target state is obtained. Depending on the user scenario, the target state can be a maximally entangled state, a tensor product of multiple maximally entangled states between multiple pairs of users or a GHZ state. By average we mean that parameters of a protocol are averaged over all possible LOCC outcomes. We call such a

protocol an $(n, \epsilon, \{p_e\}_{e \in E})$ adaptive protocol. In the asymptotic limit where $n \to \infty$ it then holds $n_e \to \infty$ for edge e with $p_e > 0$ while $\{p_e\}_{e \in E}$ remains fixed[14].

Note that whereas quantum channels are directed, the direction does not play a role when we use them to distribute entanglement under the free use of (two-way) classical communication. For example, once a channel has been used to distribute a Bell state, which is invariant under permutations of nodes across the channel. This motivates the introduction of an undirected graph $G' = (V, E')$, where E' is obtained from E as follows: for any edge $vw \in E$ with $wv \in E$, the directed edges vw and wv are replaced by single undirected edge $\{vw\}$ (or, equivalently $\{wv\}$) with $c'(\{vw\}) = c(vw) + c(wv)$, while, for any edge $vw \in E$ with $wv \notin E$, the directed edge vw is replaced by undirected edge $\{vw\}$ with $c'(\{vw\}) = c(vw)$. For more details about our notations see Supplementary Note 1.

Let us also note that whereas it is common from a quantum information theory point of view to allow for free LOCC operations, there are practical challenges to implement quantum memories with long storage times. By a slight abuse of our notation, however, it is possible to include such effects into our scenario, as well. Namely one could divide a vertex into a pre- and post storage vertex and add an additional noisy channel describing the noisy quantum memory (for instance, see ref. [13]).

Bipartite user scenario. In this section we obtain linear-program upper and lower bounds on the entanglement and key generation capacities of a network for bipartite scenarios. While some of the discussion have been made implicitly in earlier results[10,11,14], it is worth giving an explicit formulation here, given its relevance. It will also serve as a good starting point to demonstrate our method and introduce some notation. Let us suppose that the set of users only contains two vertices, $s \in E$, a.k.a Alice, and $t \in E$, also known as Bob. A possible target state could be a maximally entangled state $|\Phi^d\rangle_{M_s M_t} = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |ii\rangle_{M_s M_t}$ with $\log d$ ebits. We also use the notation $\Phi^d_{M_s M_t} = |\Phi^d\rangle\langle\Phi^d|_{M_s M_t}$. In the case of $d = 2$, this state is called a Bell state. The target state could also be a general private state[33,34], which is of the form $\gamma^d_{K_s K_t S_s S_t} = U^{\text{twist}} |\Phi^d\rangle\langle\Phi^d|_{K_s K_t} \otimes \sigma_{S_s S_t} U^{\text{twist}\dagger}$, where $\sigma_{S_s S_t}$ is an arbitrary state and $U^{\text{twist}} = \sum_{ik} |ik\rangle\langle ik|_{K_s K_t} \otimes U^{(ik)}_{S_s S_t}$ is a controlled unitary that 'twists' the entanglement in the subsystem $K_s K_t$ to a more involved form also including the subsystem $S_s S_t$. It has been shown that, by measuring the 'key part' $K_s K_t$, while keeping the 'shield part' $S_s S_t$ away from an eavesdropper Eve, $\log d$ bits of a private key can be obtained. The number of ebits or private bits is treated as the figure of merit.

We can now define a quantum network capacity $\mathcal{Q}_{\{p_e\}_{e \in E}}(G, \{\mathcal{N}^e\}_{e \in E})$ per time [$\mathcal{Q}(G, \{\mathcal{N}^e\}_{e \in E})$ per total channel use] as the largest rate $\langle \log d^{(k)} \rangle_k / n$ achievable by an adaptive $(n, \epsilon, \{p_e\}_{e \in E})$ protocol such that after n uses the finally obtained state $\rho^{(n,k)}_{M_s M_t}$ is ϵ-close to $\Phi^{d^{(k)}}_{M_s M_t}$, in the limit $n \to \infty$ and $\epsilon \to 0$ [maximized over all user frequencies $p_e \geq 0$ such that $\sum_e p_e = 1$]. Here k is a vector keeping the track of outcomes of the LOCC rounds and the notation $\langle \cdots \rangle_k$ corresponds to averaging over all LOCC outcomes. Similarly, we define a private network capacity $\mathcal{P}_{\{p_e\}_{e \in E}}(G, \{\mathcal{N}^e\}_{e \in E})$ per time [$\mathcal{P}(G, \{\mathcal{N}^e\}_{e \in E})$ per total channel use] as the largest rate $\langle \log d^{(k)} \rangle_k / n$ achievable by an adaptive $(n, \epsilon, \{p_e\}_{e \in E})$ protocol such that after n uses the state $\rho^{(n,k)}_{K_s K_t S_s S_t}$ is ϵ-close to $\gamma^{d^{(k)}}_{K_s K_t S_s S_t}$, in the limit $n \to \infty$ and $\epsilon \to 0$ [maximized over all user frequencies $p_e \geq 0$ such that $\sum_e p_e = 1$].
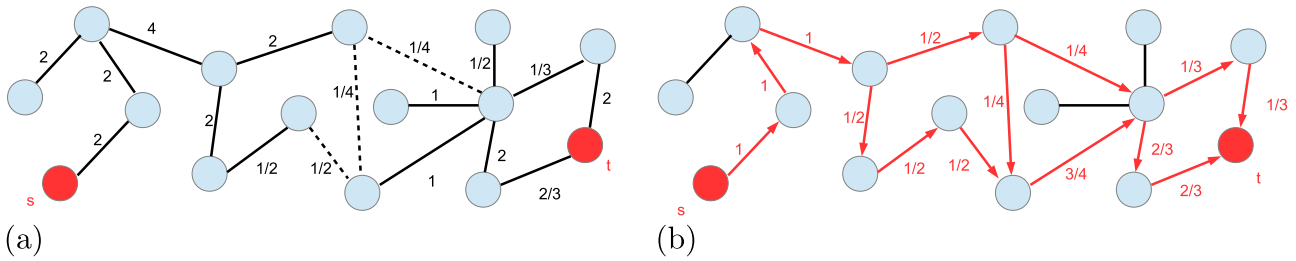
Fig. 2 Example of the max-flow min-cut theorem. a Min-cut: example of an undirected graph with a source-sink pair (in red). The labels of the edges denote their edge capacities. The edges in dashed lines represent a source-sink cut, i.e., their removal completely disconnects the source from the sink. The capacity of the cut is given by the sum over the edge capacities in the cut, in this case equal to 1, which is the minimum capacity of all source-sink cuts in this network. In other words, the min-cut is equal to 1. Note that the minimizing cut is not unique. **b** Max-flow: by the max-flow min-cut theorem the min-cut is equal to the maximum flow from the source to the sink. Here we have provided an example of a flow from the source to the sink. The labels denote the directed edge flows f_e. The flow from the source to the sink achieves the min-cut value of 1.

As the class of private states includes maximally entangled states, the private capacity is an upper bound on the quantum capacity[33,34]. Our main results in this section will be efficiently computable upper and lower bounds on the private and quantum capacities, respectively.

In a number of recent works upper bounds on private network capacities have been obtained[11,13,15]. The main idea behind those results is to assign nonnegative edge capacities to each edge e and to find the minimum edge capacity cut between s and t. By cut between s and t we mean a set of the edges whose removal disconnects s and t. The edge capacity of a cut can be defined as the sum of edge capacities of the edges in the cut. For details see the 'Methods' section. If the edge capacity $c(e)$ of an edge e is given by the usage frequency p_e of channel $\mathcal{N}^e$, multiplied by an entanglement measure $\mathcal{E}(\mathcal{N}^e)$ upper bounding the private capacity of $\mathcal{N}^e$, which is continuous near the target state and cannot be increased by amortization (see properties P1 and P2 of ref. [15] or Supplementary Note 2), the minimum edge capacity cut provides an upper bound on the private network capacity. Examples of suitable quantities $\mathcal{E}(\mathcal{N}^e)$ include the squashed entanglement E_{sq}[35], the max-relative entropy of entanglement $E_{\max}$[36] and, for a particular class of so-called Choi-stretchable channels/teleportation-simulable channels[31,37–39], the relative entropy of entanglement E_R[40] of the channel. If we know such quantities for all channels constituting the network, all that is left to do is finding the minimum edge capacity cut, which is a well-known problem in graph theory. However, it is not necessarily efficient to solve this optimization directly, because there is a case where we need to maximize further such a minimized edge capacity. For instance, it is not clear a priori how to maximize over channel frequencies the minimum capacity to find the capacity of the network per channel use. To tackle this issue, we resort to the duality of the problem.

In particular, using the max-flow min-cut theorem[41,42], we rephrase the problem of finding the minimum edge capacity cut as a network flow maximization problem in the undirected graph G'. Thanks to this, it becomes sufficient for us to consider maximization only, in every case. In a network flow maximization problem in an undirected graph, the idea is to assign a variable f_{vw} and f_{wv} to each undirected edge $\{vw\}$ which can take nonnegative values. f_{vw} is interpreted as an abstract flow of some commodity from vertex v to vertex w. As such, it has to fulfill the following constraint: interpreting the edge capacity $c'(\{vw\})$ of an undirected edge $\{vw\}$ as the capacity of its edge to transmit an abstract commodity, we require that the sum of edge flows f_{vw} and f_{wv} does not exceed the edge capacity $c'(\{vw\})$. We call this the edge capacity constraint.

Having defined a flow of an abstract commodity through an edge, the obvious next step is to consider a flow through the entire network. Namely, we mark two vertices, the source s and the sink t and define a flow from s to t as the sum of all 'outgoing' flows f_{sv}, where v is a vertex adjacent to s, such that for every edge the edge capacity constraint is fulfilled and that for every vertex $w \notin \{s, t\}$ the sum over v of 'incoming' edge flows f_{vw} is equal to the sum over v of 'outgoing' flows f_{wv}, where v are the vertices adjacent to w, which is known as flow conservation constraint. If the graph is undirected, the roles of the source and the sink can be exchanged, without changing the value of the flow. As both the edge capacity and the flow conservation constraint are linear, the maximization of the flow from s to t can be efficiently computed by means of linear programming[43]. The max-flow min-cut theorem now states that the minimum edge capacity cut that separates s and t is equal to the maximum flow from s to t. Figure 2 illustrates this with an example. For detailed definitions of cuts, flows and the max-flow min-cut theorem see the 'Methods' section.

Precisely, we use the max-flow min-cut theorem to transform the min-cut upper bounds on the private network capacity given in ref. [15] into an efficiently computable LP. To do so we define directed edge capacities $c(e) = p_e \mathcal{E}(\mathcal{N}^e)$, for every directed edge $e \in E$. Thus, entanglement takes the role of the abstract commodity considered above.

The interpretation of entanglement as a commodity raises the question if there exists a protocol that can distribute entanglement in a way similar to the flow of a commodity through a network. Ideally, one could construct such a protocol using the edge flows obtained in the flow maximization. To some extend this can be achieved by a quantum routing protocol, such as the aggregated repeater protocol introduced in ref. [14]. The aggregated repeater protocol consists of two steps: first each channel is used to distribute Bell states at a rate $p_e R^{\leftrightarrow}(\mathcal{N}^e)$ such that $R^{\leftrightarrow}(\mathcal{N}^e)$ reaches the quantum capacity $Q^{\leftrightarrow}(\mathcal{N}^e)$ in the asymptotic limit of many channel uses. This results in a network of Bell states, which can be described by an indirected multigraph, where each edge corresponds to one Bell pair. The second step of the protocol is to find edge-disjoint paths from Alice to Bob and connect them by means of entanglement swapping. The number of Bell pairs is thus equal to the number of edge-disjoint paths between Alice and Bob in the Bell network. Hence, in order to obtain a lower bound on the capacity, one would have to find the number of edge-disjoint paths in the multigraph corresponding to the Bell state network.

Finding the maximum number of edge-disjoint paths between s and t in a multigraph is the same as maximizing the flow in a

graph where the edge capacities are given by the number of parallel edges in the multigraph, however with the additional constraint that each edge flow takes integer values[44]. Such an integer flow maximization can no longer be formulated as a LP. Physically, the integer constraint corresponds to the fact that there is no such thing as 'half a Bell state'. Hence, if a LP provides an edge flow of 0.5 for some edge and 1 for another, we cannot translate this into a protocol distributing half a Bell sate over the first edge and one Bell state over the other. What we can do, however, is to multiply all edge flows obtained in the optimization by a factor of 2, and distribute one Bell state along the edge where we have obtained flow 0.5 and two Bell states along the edge where we have obtained 1. For rational edge flows obtained in the optimization and finite graphs, we can always find a large enough number to multiply the edge flows with to obtain integer values associated with each edge that can be translated into a number of Bell pairs distributed along this edge. If the edge flows obtained are real numbers, we can approximate them by rational numbers with arbitrary accuracy. This allows us to compute lower bounds on the quantum network capacities by means of maximizing the flow over a network with edge capacities given by $c(e) = p_e Q^{\hookrightarrow}(\mathcal{N}^e)$, providing us with a lower bounds that can be efficiently computed by linear programming. Finally, we can include an optimization over usage frequencies p_e into both LPs, providing us with:

Theorem 1 *For a network described by a finite directed graph G and an undirected graph G' as defined above, the private and quantum network capacities per total channel use, $\mathcal{P}(G, \{\mathcal{N}^e\}_{e\in E})$ and $\mathcal{Q}(G, \{\mathcal{N}^e\}_{e\in E})$, satisfy*

$$\bar{f}_{\max}^{s\to t}(G', \{Q^{\hookrightarrow}(\mathcal{N}^e)\}_{e\in E}) \leq \mathcal{Q}(G, \{\mathcal{N}^e\}_{e\in E}) \leq \mathcal{P}(G, \{\mathcal{N}^e\}_{e\in E}) \leq \bar{f}_{\max}^{s\to t}(G', \{\mathcal{E}(\mathcal{N}^e)\}_{e\in E}),$$
(1)

where $\bar{f}_{\max}^{s\to t}$ is given by the LP Eq. ([12](#)) in the 'Methods' section.

Further, $\mathcal{E}$ can be chosen to be the squashed entanglement E_{sq}, the max-relative entropy of entanglement $E_{\max}$ and, for Choi-stretchable channels, the relative entropy of entanglement E_R.

For the proof see Supplementary Note 2. As described in the 'Methods' section, the LPs scale polynomially with the size of the network.

Note that for a subset of Choi-stretchable channels, known as distillable channels, which include erasure channels, dephasing channels, bosonic quantum amplifier channels, and lossy optical channels, the relative entropy of entanglement of the channel $\mathcal{N}^e$ (and its Choi state σ^e) is equal to the two-way classical assisted quantum capacity[31], $E_R(\mathcal{N}^e) = E_R(\sigma^e) = Q^{\hookrightarrow}(\mathcal{N}^e)$. Hence the bounds in Theorem 1 become tight.

Multiple pairs of users. We now move on to the scenario of multiple pairs of users $(s_1, t_1),\ldots, (s_r, t_r)$ who wish to establish maximally entangled states or private states concurrently, i.e., we have target states of the form $\bigotimes_{i=1}^r \Phi_{M_{s_i} M_{t_i}}^{d_i}$ or $\bigotimes_{i=1}^r \gamma_{K_{s_i} K_{t_i} S_{s_i} S_{t_i}}^{d_i}$. This would be a typical scenario in a future 'quantum internet', where a number of user pairs might wish to perform QKD in parallel. In contrast to the bipartite scenario discussed in the previous section, where the goal is to simply optimize the rate at which entanglement is distributed between a user pair, there are a number of different figures of merit in the multi-pair scenario. We define the following three figures of merit: (1) a total multi-pair quantum (private) network capacity $\mathcal{Q}^{\text{total}}(G, \{\mathcal{N}^e\}_{e\in E})$ per total channel use [$\mathcal{Q}^{\text{total}}_{\{p_e\}_{e\in E}}(G, \{\mathcal{N}^e\}_{e\in E})$ per time] ($\mathcal{P}^{\text{total}}(G, \{\mathcal{N}^e\}_{e\in E})$ per total channel use [$\mathcal{P}^{\text{total}}_{\{p_e\}_{e\in E}}(G, \{\mathcal{N}^e\}_{e\in E})$ per time]), defined as the largest sum, over all user pairs, of the entanglement distribution rates achievable by an adaptive $(n, \epsilon, \{p_e\}_{e\in E})$ protocol such that after n uses we are ϵ-close to the target state, again taking the limit $n \to \infty$ and $\epsilon \to 0$ [and maximizing over all user frequencies $p_e \geq 0$ such that $\sum_e p_e = 1$]. Whereas maximizing the sum of rates is a good approach when the goal is to distribute as much entanglement as possible, it has the drawback that the protocol can be unfair in the sense that some pairs might get more entanglement than others, while some might not get anything at all. This drawback can be overcome by using our second figure of merit: (2) a worst-case multi-pair quantum (private) network capacity $\mathcal{Q}^{\text{worst}}(G, \{\mathcal{N}^e\}_{e\in E})$ per total channel use [$\mathcal{Q}^{\text{worst}}_{\{p_e\}_{e\in E}}(G, \{\mathcal{N}^e\}_{e\in E})$ per time] ($\mathcal{P}^{\text{worst}}(G, \{\mathcal{N}^e\}_{e\in E})$ per total channel use [$\mathcal{P}^{\text{worst}}_{\{p_e\}_{e\in E}}(G, \{\mathcal{N}^e\}_{e\in E})$ per time]), i.e., the least entanglement distribution rate that can be achieved by any user pair [by maximizing over all user frequencies $p_e \geq 0$ with $\sum_e p_e = 1$]. This approach is good in a scenario where the goal is to distribute entanglement in a fair way, in the sense that the amount of entanglement that each user pair obtains is maximized. Finally, we consider (3) the case where we assign weight q_i to each user pair (s_i, t_i). This approach can be used if user pairs are given different priorities. We call the corresponding figure of merit weighted multi-pair quantum (private) network capacity $\mathcal{Q}^{q_1,\ldots,q_r}(G, \{\mathcal{N}^e\}_{e\in E})$ per total channel use [$\mathcal{Q}^{q_1,\ldots,q_r}_{\{p_e\}_{e\in E}}(G, \{\mathcal{N}^e\}_{e\in E})$ per time] ($\mathcal{P}^{q_1,\ldots,q_r}(G, \{\mathcal{N}^e\}_{e\in E})$ per total channel use [$\mathcal{P}^{q_1,\ldots,q_r}_{\{p_e\}_{e\in E}}(G, \{\mathcal{N}^e\}_{e\in E})$ per time]) and define it as the largest achievable weighted sum of rates [with maximization over all user frequencies $p_e \geq 0$ with $\sum_e p_e = 1$]. We will now present our results for the total and worst-case scenario. For bounds on the weighted multi-pair network capacities see Supplementary Note 3.

Let us begin with scenario (1). As in the bipartite case, we can assign edge capacity $c(e) = p_e E_{sq}(\mathcal{N}^e)$ to each edge e in the graph corresponding to the network. From ref. [24] we can obtain upper bounds on the total multi-pair private network capacity which are given the minimum capacity multicut. A multicut is defined as a set of edges whose removal disconnects all pairs. The capacity of a multicut is defined by summing over the edge capacities of all edges in the multicut. Whereas this is a straightforward generalization of the problem of finding the minimum capacity cut that separates a single pair, there is no exact generalization of the max-flow min-cut theorem to multicuts. In fact, finding the minimum multicut in a general graph has been shown to be NP-hard[30].

It is however possible to upper bound the minimum multicut by means of a total multi-commodity flow optimization, also known as total multi-commodity flow, up to a factor $g_t(r)$ of order $\mathcal{O}(\log r)$[29]. A multi-commodity flow is a generalization of a flow to more than one source-sink pair, each exchanging a separate abstract 'commodity'. In order to maximize the total multi-commodity flow one introduces separate edge flow variables $f_e^{(i)}$ for each commodity i as well as each edge e and maximizes the sum of flows from s_i to t_i over all commodities $i \in \{1,\ldots, r\}$. In the optimization, one requires that for each commodity i the flow is conserved in all edges except at the corresponding source s_i and sink t_i, resulting in r separate flow conservation constraints. Thus, it is ensured that for each commodity the net flow leaving the source will reach the corresponding sink. A multi-commodity flow is concurrent if all commodities can be distributed in parallel without exceeding the edge capacities in any edge. In order to ensure this, one adds the constraint that for each undirected edge $\{vw\}$ the sum of flows of all commodities passing through the edge, $\sum_{i=1}^r (f_{vw}^{(i)} + f_{wv}^{(i)})$ does not exceed the edge capacity $c'(\{vw\})$. For details on multicuts and multi-commodity flows and the gaps that separate them see the 'Methods' section.
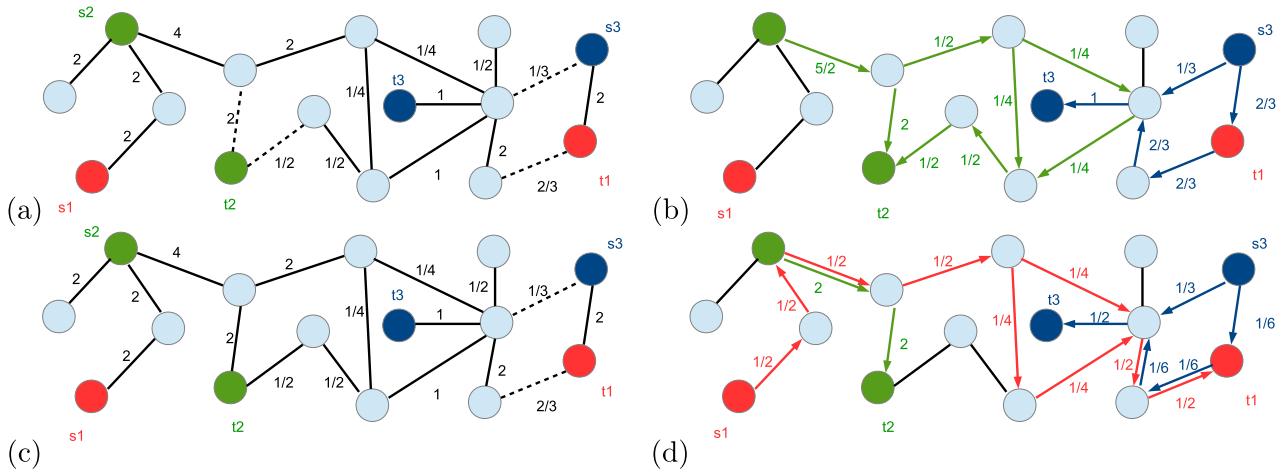
Fig. 3 Example of a multi-user scenario with source-sink pairs, denoted by the pairs of red, green and blue vertices, respectively. a A multicut (dashed edges) that separates all three source-sink pairs. The capacity of the multicut is equal to 7/2, which is the minimum value possible in this network. **b** A concurrent multi-commodity flow instance, with values 0, 5/2 and 1 for the red, green and blue pairs, respectively. The total multi-commodity flow is hence equal to minimum multicut capacity, however at the price that there is no flow for the red pair. **c** Same network and same edge capacities as in (**a**), with an example of a bipartite cut, denoted in dashed lines, with capacity 1 that separates two source-sink pairs, the red and the blue ones. Hence its cut ratio is given by 1/2, which is also the minimum cut ratio in this graph. **d** Example of a corresponding multi-commodity flow instance, with concurrent flows of values 1/2, 2 and 1/2 for the red, green and blue source-sink pairs. Hence, the worst-case multi-commodity flow is equal to 1/2, in this case matching the minimum cut ratio. Whereas the flows sum up to 3, which is less than the sum of flows in (**b**), this multi-commodity flow instance is fairer than the one in (**b**) as it also provides a flow for the red user pair.

Figure 3a shows an example of a minimum multicut separating all three source-sink pairs. Figure 3b shows a corresponding concurrent multi-commodity flow. The value of the minimum multicut is 3.5, which is equal to the sum of the three source-sink flows. So in this simple example there is no gap.

Applying the aggregated repeater protocol[14] to multiple user pairs and using the same reasoning as in the bipartite case, we also obtain lower bounds in terms of the maximum concurrent multi-commodity flows, providing us with the following efficiently computable bounds:

Theorem 2 *In a network described by a graph G with associated undirected graph G' and a scenario of r user pairs $(s_1, t_1),..., (s_r, t_r)$, the total multi-pair quantum and private network capacities per total channel use, $\mathcal{Q}^{\mathrm{total}}(G, \{\mathcal{N}^e\}_{e \in E})$ and $\mathcal{P}^{\mathrm{total}}(G, \{\mathcal{N}^e\}_{e \in E})$, satisfy*

$$\bar{f}_{\max}^{\mathrm{total}}(G', \{Q^{\leftrightarrow}(\mathcal{N}^e)\}_{e \in E}) \leq \mathcal{Q}^{\mathrm{total}}(G, \{\mathcal{N}^e\}_{e \in E}) \leq \mathcal{P}^{\mathrm{total}}(G, \{\mathcal{N}^e\}_{e \in E})$$
$$\leq g_t(r)\bar{f}_{\max}^{\mathrm{total}}(G', \{E_{\mathrm{sq}}(\mathcal{N}^e)\}_{e \in E}), \qquad (2)$$

where $\bar{f}_{\max}^{\mathrm{total}}$ is given by the polynomial sized LP Eq. (19) presented in the 'Methods' section and $g_t(r)$ is of order $\mathcal{O}(\log r)$ as described in ref.[29].

For the proof see Supplementary Note 3.

Let us now move on to scenario (2). Let us, again, describe the network by a capacitated graph with edge capacities $c(e) = p_e \mathcal{E}(\mathcal{N}^e)$, where $\mathcal{E}(\mathcal{N}^e)$ can be chosen to be the squashed entanglement E_{sq}, the max-relative entropy of entanglement $E_{\max}$ and, for Choi-stretchable channels, the relative entropy of entanglement E_R of the channel. Using the results of refs. [15,24], it is possible to show that the worst-case multi-pair private network capacity is upper bounded by the so-called minimum cut ratio with unit demands of the capacitated graph. Given a (bipartite) cut, which separates the set of vertices into two subsets, the cut ratio is defined as its capacity of the cut, i.e., the sum over edge capacities of the edges, divided by the demand across the cut, in this case the number of pairs separated by the cut. The minimum cut ratio is obtained by a minimization over all bipartite cuts. See Fig. 3c for an example of a minimum cut ratio.

As for the minimum multicut discussed above, the computation of the minimum cut ratio is an NP-hard problem in general graphs[27].

Whereas, as in the case of multicuts, there is no exact version of the max-flow min-cut theorem for the minimum cut ratio, there is again a connection to concurrent multi-commodity flows up to a factor $g_w(r)$, which can be of order up to $\mathcal{O}(\log r)$[27]. Namely, it has been shown that the minimum cut ratio is upper bounded by $g_w(r)$ times what we call the maximum worst-case multi-commodity flow, also known as maximum concurrent multi-commodity flow, which corresponds to the maximum flow that can be achieved by any of the commodities concurrently, with respect to the same edge capacity and flow conservation constraints as in the case of the total multi-commodity flow, discussed previously. Figure 3d contains an example of a maximum worst-case multi-commodity flow that achieves the cut ratio in Fig. 3c. Note that this flow is different from the one achieving the minimum multicut in Fig. 3b. In particular, it is 'fairer' in the sense that it also provides a flow for the red user pair (s_1, t_1). See the 'Methods' section for a detailed definition of the minimum cut ratio, the worst-case multi-commodity flow and the gap that separates them.

As in the previous scenarios, we can obtain a lower bound by application of the aggregated repeater protocol[14] to multiple user pairs and include an optimization over usage frequencies, resulting in the following result:

Theorem 3 *In a network described by a graph G with associated undirected graph G' and a scenario of r user pairs $(s_1, t_1),..., (s_r, t_r)$, the worst-case multi-pair quantum and private network capacities per total channel use, $\mathcal{Q}^{\mathrm{worst}}(G, \{\mathcal{N}^e\}_{e \in E})$ and $\mathcal{P}^{\mathrm{worst}}(G, \{\mathcal{N}^e\}_{e \in E})$, satisfy*

$$\bar{f}_{\max}^{\mathrm{worst}}(G', \{Q^{\leftrightarrow}(\mathcal{N}^e)\}_{e \in E}) \leq \mathcal{Q}^{\mathrm{worst}}(G, \{\mathcal{N}^e\}_{e \in E})$$
$$\leq \mathcal{P}^{\mathrm{worst}}(G, \{\mathcal{N}^e\}_{e \in E}) \leq g_w(r)\bar{f}_{\max}^{\mathrm{worst}}(G', \{\mathcal{E}(\mathcal{N}^e)\}_{e \in E}), \qquad (3)$$

where $\bar{f}_{\max}^{\mathrm{worst}}$ is given by the polynomially sized LP Eq. (21) presented in the 'Methods' section. Further $g_w(r)$ is the flow-cut gap described in the 'Methods' section. $\mathcal{E}$ can be chosen to be the squashed entanglement E_{sq}, the max-relative entropy of
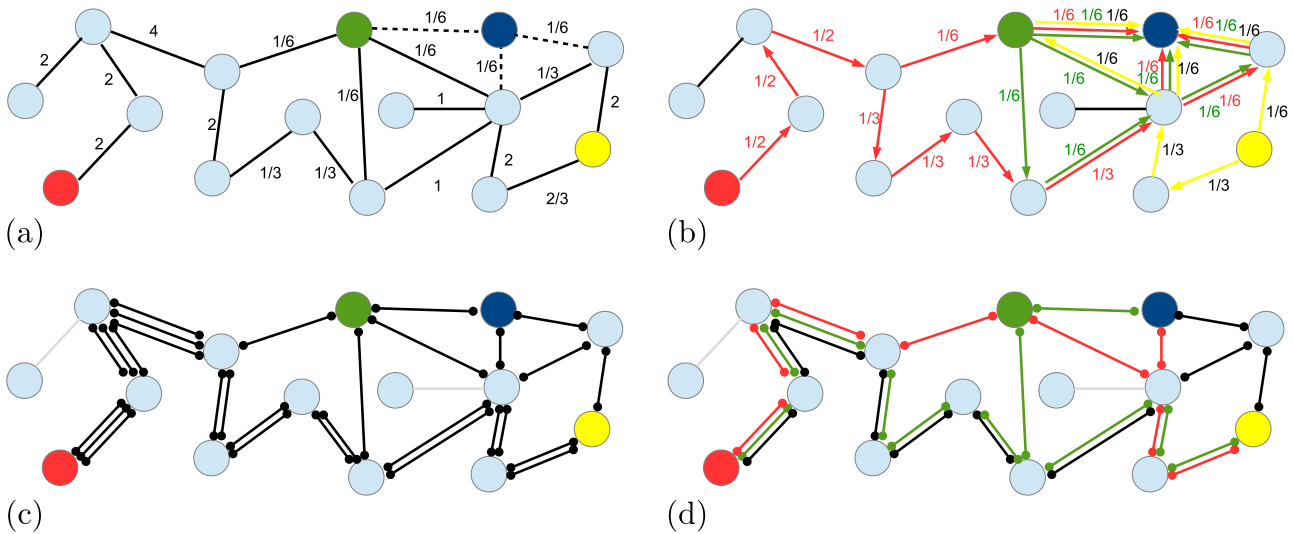
Fig. 4 Example of a setting where a group of four user nodes (red, green, blue, and yellow) wishes to establish a Greenberger–Horne–Zeilinger (GHZ) state. a The graph G' with labeled edge capacities. The dashed edges correspond to a minimum Steiner cut with respect to the set of the four users, i.e., it is a smallest capacity cut that separates at least one pair of vertices in the set. In this case it separates the red-blue, green-blue, and yellow-blue pairs and has capacity 1/2. In other words, the set of users is 1/2-connected. **b** Part of a flow instance corresponding to linear program (LP) given by Eqs. (32)–(36). Here the flows from the red, green, and yellow vertices to the blue vertex are shown in red, green, and yellow, respectively. For simplicity, flows between other nodes are not shown in this picture. The directed edge flows correspond to the variables $f_{vw}^{(ij)}$ of the LP given by Eqs. (32)–(36). By providing flows of value of at least 1/2 between all pairs in the set of users, the LP shows that the set of users is 1/2-connected. **c** Aggregated repeater protocol: assuming that the edges in (**a**) correspond to quantum channels (of some direction) and their capacities to non-asymptotic quantum capacities, one could, by using each channel (at most) six times, create a network of Bell states that is described by a three-connected undirected multigraph. **d** Steiner Trees: in our example the multigraph contains two edge-disjoint Steiner trees, depicted in red and green. The Bell pairs forming the Steiner trees can then be connected by means of a generalized entangled swapping protocol to form two qubit GHZ states among the four users.

---

entanglement E_{max} *and, for Choi-stretchable channels, the relative entropy of entanglement* E_R.

For the proof, see Supplementary Note 3. As a proof of principle demonstration, we have numerically computed the worst-case and total multi-commodity flows for an example network. See Supplementary Note 5 for details and plots.

Multipartite target states. In this section we present our results on the distribution of multipartite entanglement. Let us consider a set of disjoint users $S = \{s_1,…,s_l\}$, who wish to establish a multipartite target state, such as a GHZ state[45] $|\Phi^{GHZ,d}\rangle_{M_{s_1}…M_{s_l}} = \frac{1}{\sqrt{d}}\sum_{i=0}^{d-1} |i\rangle_{M_{s_1}} \otimes \cdots \otimes |i\rangle_{M_{s_l}}$ or a multipartite private state[5], $\gamma_{K_{s_1}S_{s_1}…K_{s_l}S_{s_l}}^d = U^{twist} |\Phi^{GHZ,d}\rangle\langle\Phi^{GHZ,d}|_{K_{s_1}…K_{s_l}} \otimes \sigma_{S_{s_1}…S_{s_l}} U^{twist\dagger}$, where $\sigma_{S_{s_1}…S_{s_l}}$ is an arbitrary state and $U^{twist} = \sum_{i_1,…,i_l} |i_1, \ … \ , i_l\rangle\langle i_1, \ … \ ,i_l|_{K_{s_1}…K_{s_l}} \otimes U_{S_{s_1}…S_{s_l}}^{(i_1,…,i_l)}$ is a controlled unitary operation. The corresponding multipartite quantum and private network capacities Q^S and $\mathcal{P}^S$ are defined analogously to the bipartite case.

As a consequence of ref. [24], the private capacity is upper bounded by the connectivity of the set S of user vertices in the graph capacitated by $p_e E_{sq}(\mathcal{N}^e)$. By connectivity of the set S we mean the minimum edge capacity cut that separates any two vertices s_i and s_j ($i \neq j$) in S. Such a cut is also known as minimum S-cut or minimum Steiner cut with respect to S. See Fig. 4a for an example of a minimum Steiner cut with respect to the set of the red, green, blue, and yellow nodes. The computation of the connectivity consists of a minimization of all possible disjoint vertex pairs within S as well as a cut minimization. See Eq. (31) in the 'Methods' section. Applying the max-flow min-cut theorem for every possible pair s_i and s_j ($i \neq j$) in S, we can transform the computation of the connectivity of S into another LP that upper

bounds the multipartite network private capacity. See Fig. 4b for an example.

Finding a lower bound on the multipartite network quantum capacity, i.e., the maximum rate at which we can distribute a GHZ state among S, is slightly more involved than in the previously considered scenarios. As we did in all previous scenarios, we begin by performing an aggregated repeater protocol to create a network of Bell states that can be described by an undirected multigraph. See Fig. 4c for an example. Whereas it is possible to create a GHZ state locally in one of the nodes and use chains of Bell pairs to teleport the respective subsystems of the GHZ state to all other nodes in S, it is easy to find a network where this is not the optimal strategy. Instead, the idea is to generalize the concept of paths linking two nodes to Steiner trees spanning the set S of users. In an undirected multigraph, a Steiner tree spanning S, or short S-tree, is an acyclic subgraph that connects all nodes in S. See Fig. 4d for an example of two edge-disjoint Steiner trees spanning the set of the red, green, blue, and yellow nodes. See the 'Methods' section for more information on Steiner trees.

A Steiner tree spanning S in the network of (qubit) Bell states can be transformed into a (qubit) GHZ state among all nodes in S by means of a protocol introduced in ref. [20], which can be seen as a generalization of entanglement swapping. Hence, the number of (qubit) GHZ states obtainable from the Bell state network is equal to the number of edge-disjoint Steiner trees spanning S. Computing this number is a referred to as a Steiner tree packing, which is another NP-complete problem[46]. However, the number of edge-disjoint Steiner trees in a multigraph can be lower bounded by its S-connectivity up to constant factor 1/2 and an additive constant[47–49]. Combining this with the max-flow min-cut theorem, this allows us to derive a linear-program lower bound on the multipartite quantum network capacity. Hence we obtain the following:

Theorem 4 *In a network described by a graph G with associated undirected graph G' and a scenario of a set $S = \{s_1, ..., s_r\}$ of users, the quantum and private network capacities per total channel use, $\mathcal{Q}^S(G, \{\mathcal{N}^e\}_{e \in E})$ and $\mathcal{P}^S(G, \{\mathcal{N}^e\}_{e \in E})$, satisfy*

$$\frac{1}{2}\bar{f}_{\max}^S(G', \{Q^{\rightarrow}(\mathcal{N}^e)\}_{e \in E}) \leq \mathcal{Q}^S(G, \{\mathcal{N}^e\}_{e \in E})$$
$$\leq \mathcal{P}^S(G, \{\mathcal{N}^e\}_{e \in E}) \leq \bar{f}_{\max}^S(G', \{E_{sq}(\mathcal{N}^e)\}_{e \in E}), \quad (4)$$

where $\bar{f}_{\max}^S$ is given by the polynomially sized LP Eq. (37) presented in the 'Methods' section.

For the proof see Supplementary Note 4.

Discussion

We have provided linear-program upper and lower bounds on the entanglement and key generation capacities in quantum networks for various user scenarios. We have done so by reducing the corresponding network-routing problems to flow optimizations, which can be written as LPs. The user scenarios we have considered are the distribution of Bell or private states between a single pair of users, the parallel distribution of such states between multi-pairs of users and the distribution of GHZ or multipartite private states among a group of multiple users. The size of the LPs scales polynomially in the parameters of the networks, and hence the LPs can be computed in polynomial time. In order to perform the LPs, upper and lower bounds on the two-way assisted private or quantum capacities of all the channels constituting the network have to be provided as input parameters. Thus the problem of bounding capacities for the entire network is reduced to bounding capacities of single channels, as well as performing an LP which scales polynomially in the network parameters.

For a large class of practical channels, including erasure channels, dephasing channels, bosonic quantum amplifier channels, and lossy optical channels, tight bounds can be obtained in the bipartite case. In the multi-pair case, however, there still remains a gap of order up to $\log r^*$ between the upper and lower bounds. This gap, also known as flow-cut gap, is due to the lack of an exact max-flow min-cut theorem for multi-commodity flows. From a complexity theory standpoint, the flow-cut gap separates the NP-hard problem of determining the minimum cut ratio from the problem of finding the maximum concurrent multi-commodity flow, which can be done in polynomial time[27]. From a network theoretic view the gap also leaves room for a possible advantage of network coding over network routing in undirected networks, which is still an open problem[50,51]. Another gap, of value 1/2, occurs between our upper and lower bounds in the multi-pair case. As in the multiple-pair case, this gap is significant in terms of computational complexity, as it separates our polynomial LP from the problem of Steiner tree packing, which is NP-complete[46].

While our LPs cover an important set of user scenarios and tasks, we believe that our recipe will find broader use. In the bipartite case, we could assign costs to the links and consider the problem of minimizing the total cost for a given set of user demands[52]. In the multipartite case, we could apply it to the distribution of multipartite entanglement between multiple groups of users, for which one could leverage results connecting the minimum ratio Steiner cut problem and the Steiner multicut problem with concurrent Steiner flows[53]. As another example, beyond network capacities, many algorithms for graph clustering and community detection in complex networks rely on the sparsest cut of graph[54,55]. This quantity is bounded from below by the uniform multi-commodity flow problem, which is an instance of our multi-pair entanglement distribution maximizing the worst-case multi-commodity flow, and from above by the same quantity multiplied by a value that scales logarithmically

with the number of nodes in the network. Hence, the direct solution of this instance could be used to solve the analogous problem in complex networks where the links are evaluated for their capability to transmit quantum information or private classical information. Although we have focused on a LP to bound capacities, rather than actual rates in practical scenarios with other imperfections, such as storage limitation or overheads, we believe that our program could be the basis to develop an algorithm to treat such practical scenarios as well.

Methods

Bipartite user scenario. In this section we will explicitly define all quantities that occur in our result for bipartite user scenarios, Theorem 1, and briefly review the main ingredient in its proof, the max-flow min-cut theorem. Let us begin with the definition of the capacities: the quantum and private network capacities per total channel use that occur in Theorem 1 are defined as

$$\mathcal{Q}(G, \{\mathcal{N}^e\}_{e \in E}) = \max_{p_e \geq 0, \sum_e p_e = 1} \lim_{\epsilon \to 0} \lim_{n \to \infty} \sup_{\Lambda} \left\{ \frac{\langle \log d^{(k)} \rangle_k}{n} : \left\| \rho_{M,M_t}^{(n,k)} - \Phi^{d^{(k)}}{}_{M,M_t} \right\|_1 \leq \epsilon \right\}, \quad (5)$$

$$\mathcal{P}(G, \{\mathcal{N}^e\}_{e \in E}) = \max_{p_e \geq 0, \sum_e p_e = 1} \lim_{\epsilon \to 0} \lim_{n \to \infty} \sup_{\Lambda} \left\{ \frac{\langle \log d^{(k)} \rangle_k}{n} : \left\| \rho_{K,K_t S,S_t}^{(n,k)} - \gamma_{K,K_t S,S_t}^{d^{(k)}} \right\|_1 \leq \epsilon \right\}, \quad (6)$$

where the suprema are over all adaptive $(n, \epsilon, \{p_e\}_{e \in E})$ protocols Λ. Further $k = (k_1, ..., k_{n+1})$ is a vector keeping the track of outcomes of the $n + 1$ LOCC rounds in Λ, the averaging, denoted by the parenthesis $\langle ... \rangle_k$, is over all those outcomes and $\rho_{M,M_t}^{(n,k)}$ is the final state of Λ for given outcomes k.

Let us discuss the difference between the above quantities and network capacities introduced in refs. [10,11], which consider rates per network use. There are two strategies considered in refs. [10,11], sequential (or single path) routing and multi-path routing. Both strategies are adaptive in the same sense as defined above, i.e., the channel uses are interleaved by LOCC operations among all nodes, the number of LOCC rounds being equal to the total number of channel uses.

In the case of sequential (or single path) routing, one use of the network involves usage of channels along a single path from Alice to Bob. The path, and its length, can change with every use of the network. This strategy could correspond to the external provider offering a path for the users (similar to the paradigm of circuit switching networks[56]) instead of allowing the users to precisely determine the usage frequencies of each channel.

In the case of multi-path routing, a flooding strategy is applied, where during each use of the network each channel is used exactly once. Hence the total number of channel uses is given by $|E|$ times the number of network uses. As shown in refs. [10,11], there are examples of networks, such as the so-called diamond network, for which such a strategy provides an advantage over single-path routing. The multi-path scenario could correspond to a private quantum network where the users are willing to use the whole of their resources each clock cycle to implement the desired communication task.

In the present paper, an alternative approach is taken. Instead of considering rates per use of the network, we consider rates per the total number of channel uses and per time. By setting our usage frequencies p_e constant for all nodes $e \in E$ in the network, we can incorporate the flooding strategy used in the multi-path routing scenario of refs. [10,11]. Hence, although phrased with the channel use metric, our results also can be used for the network use metric generalizing the original results in refs. [10,11] to multipartite settings. There is, however, no direct relation between our capacities and the single-path capacities. In fact they can differ by a factor $\mathcal{O}(|E|)$, which is the order of the number of vertices in the network, as shown in Fig. 5.

We will now introduce the LP that provides upper and lower bounds on the capacities Eqs. (5) and (6), respectively. Let us consider undirected graph $G' = (V, E')$, as defined at the beginning of the 'Result' section, with edge capacities $c'(\{vw\})$ for all $\{vw\} \in E'$. We assume that we have two special nodes $s, t \in V$, which we call the source and the sink. As the entanglement across an edge can be used in both directions, we assign two edge flows $f_{wv} \geq 0$ and $f_{vw} \geq 0$ to each edge $\{wv\} \in E'$, where f_{wv} corresponds to a flow from w to v and f_{vw} to a flow in the opposite direction.

The goal is now to maximize the flow from s to t over the graph G'. In order to be a feasible flow, it should not exceed the capacity of each edge. Namely, for each edge $\{vw\}$ we need

$$f_{wv} + f_{vw} \leq c'(\{wv\}). \quad (7)$$

We also need that for each edge $w \neq s, t$

$$\sum_{v:\{vw\} \in E'} f_{vw} = \sum_{v:\{vw\} \in E'} f_{wv}, \quad (8)$$

which is known as flow conservation. By this flow conservation the flow from s to t
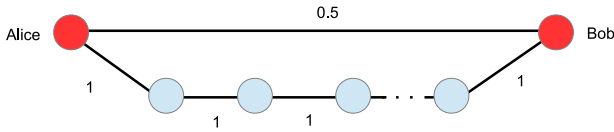
Fig. 5 Per network use versus per channel use capacities: simple example of a network, where our capacity Eq. (5) can differ arbitrarily from the single-path capacity introduced in refs. [10,11]. The numbers refer to capacities of the single channel. When the goal is to maximize the transmission per total number of channel uses, the upper route is preferable. It can achieve a transmission of 0.5 using a single channel, i.e., a rate per channel use of 0.5. The lower route can achieve a transmission of 1 using n (greater than two) channels, i.e., a rate per channel use of $1/n$. When the goal is to maximize the transmission per uses of the network over a single path as in refs. [10,11], the lower route is preferable as it can achieve a transmission of 1 per use of the network, whereas the upper route can achieve 0.5.

is equal to the flow leaving the source minus the flow entering the source,

$$f^{s\to t} = \sum_{v:\{sv\}\in E'} (f_{sv} - f_{vs}). \tag{9}$$

In order to obtain the maximum flow from s to t over the graph G', we need to maximize Eq. (9) over edge flows with respect to constraints Eqs. (7) and (8), which is a LP:

$$f^{s\to t}_{\max}(G', \{c'(\{wv\})\}_{\{wv\}\in E'}) = \max \sum_{v:\{sv\}\in E'} (f_{sv} - f_{vs})$$
$$\forall \{vw\} \in E': f_{wv} + f_{vw} \le c'(\{wv\})$$
$$\forall \{vw\} \in E': f_{wv}, f_{vw} \ge 0$$
$$\forall w \in V: w \ne s, t, \sum_{v:\{vw\}\in E'} f_{vw} = \sum_{v:\{vw\}\in E'} f_{wv}. \tag{10}$$

In Theorem 1, we set the capacities $c'(\{wv\})$ to

$$c'_C(\{vw\}, p_{wv}, p_{vw}) = p_{wv}C(\mathcal{N}^{wv}) + p_{vw}C(\mathcal{N}^{vw}), \tag{11}$$

where $C = Q^{\leftrightarrow}$ for the lower bound and $C = \mathcal{E}$ for the upper bound, respectively. In the case where $wv \in E$ but $vw \notin E$, we set $p_{vw}C(\mathcal{N}^{vw}) = 0$. Further we add an optimization over the usage frequencies

$$\bar{f}^{s\to t}_{\max}(G', \{C(\mathcal{N}^e)\}_{e\in E}) = \max_{0\le p_e\le 1, \sum_e p_e=1} f^{s\to t}_{\max}(G', \{c'_C(\{vw\}, p_{wv}, p_{vw})\}_{\{vw\}\in E'}). \tag{12}$$

In the following we will make use of the max-flow min-cut theorem: Given a subset $V' \subset V$ we define a cut of G' as the set

$$\partial(V') := \{\{vw\} \in E' : v \in V', w \in V \setminus V'\}. \tag{13}$$

If, for given vertices s and t, and a set $V_{s;t} \subset V$, $s \in V_{s;t}$ and $t \in V \setminus V_{s;t}$ we call $\partial(V_{s;t})$ an st-cut. Let us note that the first and second indices in the subscript of $V_{s;t}$ have different meanings. The minimum st-cut of G' is defined as

$$\min_{V_{s;t}} \sum_{\{vw\}\in\partial(V_{s;t})} c'(\{vw\}), \tag{14}$$

where the minimization is over all $V_{s;t} \subset V$ such that $s \in V_{s;t}$ and $t \in V \setminus V_{s;t}$. By the max-flow min-cut theorem[41,57] it holds

$$f^{s\to t}_{\max}(G', \{c'(\{wv\})\}_{\{wv\}\in E'}) = \min_{V_{s;t}} \sum_{\{vw\}\in\partial(V_{s;t})} c'(\{vw\}). \tag{15}$$

See Fig. 2 for an example illustrating the connection between cuts and flows.

Multiple pairs of users. In this section we will explicitly define all quantities that occur in our results for multiple pairs of users, Theorems 2 and 3. We also briefly introduce multi-commodity flows and the corresponding generalizations of the max-flow min-cut theorem, which are used in the proofs of Theorems 2 and 3.

We begin by defining a total multi-pair quantum network capacity and a worst-case multi-pair quantum network capacity per total channel use respectively as:

$$Q^{\text{total}}(G, \{\mathcal{N}^e\}_{e\in E}) = \max_{\substack{p_e\ge 0 \\ \sum_e p_e=1}} \lim_{\epsilon\leftarrow 0} \lim_{n\to\infty} \sup_\Lambda \left\{ \frac{\sum_{i=1}^r \langle\log d_i^{(k)}\rangle_k}{n} : \left\|\rho^{(n,k)}_{M_{s_1}M_{t_1}\cdots M_{s_r}M_{t_r}} - \bigotimes_{i=1}^r \Phi^{d_i^{(k)}}_{M_{s_i}M_{t_i}}\right\|_1 \le \epsilon \right\}, \tag{16}$$

$$Q^{\text{worst}}(G, \{\mathcal{N}^e\}_{e\in E}) = \max_{\substack{p_e\ge 0 \\ \sum_e p_e=1}} \lim_{\epsilon\leftarrow 0} \lim_{n\to\infty} \min_{i\in\{1\cdots r\}} \sup_\Lambda \left\{ \frac{\langle\log d_i^{(k)}\rangle_k}{n} : \left\|\rho^{(n,k)}_{M_{s_1}M_{t_1}\cdots M_{s_r}M_{t_r}} - \bigotimes_{i=1}^r \Phi^{d_i^{(k)}}_{M_{s_i}M_{t_i}}\right\|_1 \le \epsilon \right\}, \tag{17}$$

where the supremum is over all adaptive $(n, \epsilon, \{p_e\}_{e\in E})$ protocols Λ and $k = (k_1,\ldots, k_{m+1})$ is a vector of outcomes of the $m+1$ LOCC rounds in Λ, the averaging is over all those outcomes and $\rho^{(n, k)}$ is the final state of Λ for given outcomes k. The corresponding private capacities $\mathcal{P}^{\text{total}}(G, \{\mathcal{N}^e\}_{e\in E})$ and $\mathcal{P}^{\text{worst}}(G, \{\mathcal{N}^e\}_{e\in E})$ are defined by replacing $\Phi^{d_i^{(k)}}_{M_{s_i}M_{t_i}}$ by $\gamma^{d_i^{(k)}}_{K_{s_i}S_{s_i}K_{t_i}S_{t_i}}$ in Eqs. (16) and (17).

The bounds on Eqs. (16) and (17) given in Theorems 2 and 3, respectively are in terms of multi-commodity flow optimizations, which we will introduce in this section. A flow instance involving multiple sources and sinks $s_1,\ldots, s_r$ and $t_1, \ldots, t_r$ is known as a multi-commodity flow, each flow $f^{(i)}$ from s_i to t_i being considered to be a separate commodity. The maximum total multi-commodity flow is then obtained by maximizing the sum over all single-commodity flows. Generalizing LP Eq. (10) accordingly, we obtain the following LP:

$$f^{\text{total}}_{\max}(G', \{c'(\{vw\})\}_{\{vw\}\in E'}) = \max \sum_{i=1}^r \sum_{v:\{s_i v\}\in E'} \left(f^{(i)}_{s_i v} - f^{(i)}_{vs_i}\right)$$
$$\forall\{vw\}\in E': \sum_{i=1}^r \left(f^{(i)}_{vw} + f^{(i)}_{wv}\right) \le c'(\{vw\})$$
$$\forall\{vw\}\in E', \forall i: f^{(i)}_{vw}, f^{(i)}_{wv} \ge 0 \tag{18}$$
$$\forall i, \forall w\in V, w\ne s_i, t_i: \sum_{v:\{vw\}\in E'} \left(f^{(i)}_{vw} - f^{(i)}_{wv}\right) = 0.$$

Again, we can define

$$\bar{f}^{\text{total}}_{\max}(G', \{C(\mathcal{N}^e)\}_{e\in E}) = \max_{0\le p_e\le 1, \sum_e p_e=1} f^{\text{total}}_{\max}(G', \{c'_C(\{vw\}, p_{wv}, p_{vw})\}_{\{vw\}\in E'}). \tag{19}$$

Further, the maximum worst-case multi-commodity flow is obtained by adding additional variable f and maximizing f, while demanding that every single-commodity flow is greater or equal to f. Hence f corresponds to the least flow between any source-sink pair. This provides us with the following LP[58]:

$$f^{\text{worst}}_{\max}(G', \{c'(\{vw\})\}_{\{vw\}\in E'}) = \max f$$
$$\forall i: f - \sum_{v:\{s_i v\}\in E'} \left(f^{(i)}_{s_i v} - f^{(i)}_{vs_i}\right) \le 0$$
$$\forall\{vw\}\in E': \sum_{i=1}^r \left(f^{(i)}_{vw} + f^{(i)}_{wv}\right) \le c'(\{vw\})$$
$$\forall\{vw\}\in E', \forall i: f^{(i)}_{vw}, f^{(i)}_{wv} \ge 0 \tag{20}$$
$$\forall i, \forall w\in V, w\ne s_i, t_i: \sum_{v:\{vw\}\in E'} \left(f^{(i)}_{vw} - f^{(i)}_{wv}\right) = 0.$$

Again, we can define

$$\bar{f}^{\text{worst}}_{\max}(G', \{C(\mathcal{N}^e)\}_{e\in E}) = \max_{0\le p_e\le 1, \sum_e p_e=1} f^{\text{worst}}_{\max}(G', \{c'_C(\{vw\}, p_{wv}, p_{vw})\}_{\{vw\}\in E'}). \tag{21}$$

Next, we will consider generalizations of the max-flow min-cut theorem to multiple source-sink pairs: given source-sink pairs $(s_1, t_1),\ldots,(s_r, t_r)$, one can define a multicut $\{S\}\leftrightarrow\{T\}$ as a set of edges in E' whose removal disconnects all source-sink pairs and the capacity of a multicut as the sum over the capacity of its edges $\{S\}\leftrightarrow\{T\}$, namely

$$c'(\{S\}\leftrightarrow\{T\}) = \sum_{\{vw\}\in\{S\}\leftrightarrow\{T\}} c'(\{vw\}). \tag{22}$$

Whereas there is no known exact max-flow minimum cut-ratio theorem in the case of multiple flows, there exists a relation between the minimum multicut and the maximum total multi-commodity flow up to a factor $g_t(r)$ that scales as $\mathcal{O}(\log r)$[29]. Namely it holds

$$f^{\text{total}}_{\max}(G', \{c'(\{vw\})\}_{\{vw\}\in E}) \le \min_{\{S\}\leftrightarrow\{T\}} c'(\{S\}\leftrightarrow\{T\}) \le g_t(r) f^{\text{total}}_{\max}(G', \{c'(\{vw\})\}_{\{vw\}\in E}), \tag{23}$$

An example of the relation Eq. (23) is given in Fig. 3a, b. In the example $g_t(r) = 1$. In the case of the maximum worst-case multi-commodity flow there exists a similar

relation with the minimum cut ratio, which is defined as

$$R_{\min}\left(G', \{c'(\{vw\})\}_{\{vw\}\in E'}\right) = \min_{V'\subset V} \frac{\sum_{\{vw\}\in\partial V'} c'(\{vw\})}{d(\partial(V'))}, \tag{24}$$

where the minimization is over (bipartite) cuts $\partial V'$ and

$$d(\partial(V')) = |\{i : (s_i \in V', t_i \in V\setminus V') \vee (t_i \in V', s_i \in V\setminus V')\}| \tag{25}$$

describes the demand across a cut $\partial V'$. Note that in the case of only one source-sink pair the minimum cut ratio Eq. (24) reduces to the min-cut Eq. (14). Whereas there is no known exact max-flow minimum cut-ratio theorem in the case of multiple flows, there is a relation up to some factor $g_w(r)$[59],

$$f_{\max}^{\text{worst}}\left(G', \{c'(\{vw\})\}_{\{vw\}\in E'}\right) \leq R_{\min}\left(G', \{c'(\{vw\})\}_{\{vw\}\in E'}\right) \leq g_w(r) f_{\max}^{\text{worst}}\left(G', \{c'(\{vw\})\}_{\{vw\}\in E'}\right). \tag{26}$$

An example of the relation Eq. (26) is given in Fig. 3c, d. In the example $g_w(r) = 1$. The gap $g_w(r)$ is known as the flow-cut gap. In ref. [59] it has been shown to be of $\mathcal{O}(\log |E|)$. This was then improved to $\mathcal{O}(\log r)$, where r is the number of source-sink pairs, in refs. [27,60]. In the case of overlapping source and sink vertices, i.e., $s_i = s_j$, $s_i = t_j$, $t_i = s_j$, or $t_i = t_j$ for some $i \neq j$, the flow-cut gap has further been improved to $\mathcal{O}(\log r^*)$, where r^* is the size of the smallest set of vertices that contains at least one of such s_i or t_i for all $i = 1,\ldots, r$[28]. For a number of particular classes of graphs, it has been shown that the flow-cut gap can even be of $\mathcal{O}(1)$[61–65].

Multipartite target states. In this section we will explicitly define all quantities that occur in our result for multipartite target states, Theorem 4. We also briefly introduce the concept of Steiner cuts and Steiner trees, which are used in the proof of Theorem 4.

Again, we begin with the definition of the capacities: given a set $S \subset V$ of users that wish to establish a GHZ or multipartite private state, the multipartite quantum, and private network capacities are defined as:

$$\mathcal{Q}^S\left(G, \{\mathcal{N}^e\}_{e\in E}\right) = \max_{p_e\geq 0, \sum_e p_e=1} \lim_{\epsilon\to 0} \lim_{n\to\infty} \sup_\Lambda \left\{ \frac{\langle\log d^{(k)}\rangle_k}{n} : \left\|\rho_{M_{s_1}\ldots M_{s_j}}^{(n,k)} - \Phi_{M_{s_1}\ldots M_{s_j}}^{d^{(k)}}\right\|_1 \leq \epsilon \right\}, \tag{27}$$

$$\mathcal{P}^S\left(G, \{\mathcal{N}^e\}_{e\in E}\right) = \max_{p_e\geq 0, \sum_e p_e=1} \lim_{\epsilon\to 0} \lim_{n\to\infty} \sup_\Lambda \left\{ \frac{\langle\log d^{(k)}\rangle_k}{n} : \left\|\rho_{K_{s_1}S_{s_1}\ldots K_{s_j}S_{s_j}}^{(n,k)} - \gamma_{K_{s_1}S_{s_1}\ldots K_{s_j}S_{s_j}}^{d^{(k)}}\right\|_1 \leq \epsilon \right\}, \tag{28}$$

where the suprema are over all adaptive $(n, \epsilon, \{p_e\}_{e\in E})$ protocols Λ. As the class of multipartite private states includes GHZ states, the multipartite private capacity is an upper bound on the multipartite quantum capacity.

Let us now introduce the concept of Steiner cuts and Steiner trees: for a subset $S \subset V$ of vertices in G' we define a Steiner cut with respect to S, in short S-cut, as a cut $\partial(V_S)$ with respect to a set $V_S \subset V$ such that there is at least one pair of vertices $s_i, s_j \in S$ with $s_i \in V_S$ and $s_j \in V\setminus V_S$. When considering a minimization of the capacity over all S-cuts, we can divide the minimization into a minimization over pairs of vertices in S and a minimization over cuts separating the pairs,

$$\min_{V_S} \sum_{\{vw\}\in\partial(V_S)} c'(\{vw\}) = \min_{s_i,s_j\in S, s_i\neq s_j} \min_{V_{s_i:s_j}} \sum_{\{vw\}\in\partial(V_{s_i:s_j})} c'(\{vw\}), \tag{29}$$

where $\min_{V_S}$ is a minimization over all $V_S \subset V$ such that there is at least one pair of vertices $s_i, s_j \in S$ with $s_i \in V_S$ and $s_j \in V\setminus V_S$. Further $\min_{V_{s_i:s_j}}$ is a minimization over all $V_{s_i:s_j} \subset V$ such that $s_i \in V_{s_i:s_j}$ and $s_j \in V \setminus V_{s_i:s_j}$. Note that, as $\min_{s_i,s_j\in S, s_i\neq s_j} \min_{V_{s_i:s_j}} \sum_{\{vw\}\in\partial(V_{s_i:s_j})} c'(\{vw\})$ does not depend on the order, we can, without loss of generality restrict to disjoint s_i and s_j with $j > i$, reducing the number of resources needed in the outer minimization. We can then apply the max-flow min-cut theorem Eq. (15) to the inner minimization,

$$\min_{\substack{s_i,s_j\in S, s_i\neq s_j \\ j>i}} \min_{V_{s_i:s_j}} \sum_{\{vw\}\in\partial(V_{s_i:s_j})} c'(\{vw\}) = \min_{\substack{s_i,s_j\in S, s_i\neq s_j \\ j>i}} f_{\max}^{s_i\to s_j}\left(G', \{c'(\{wv\})\}_{\{wv\}\in E'}\right), \tag{30}$$

where $f_{\max}^{s_i\to s_j}\left(G', \{c'(\{wv\})\}_{\{wv\}\in E'}\right)$ is given by LP Eq. (10). As there are finitely many disjoint s_i, s_j-pairs in S, we could solve $f_{\max}^{s_i\to s_j}\left(G', \{c'(\{wv\})\}_{\{wv\}\in E'}\right)$ for every pair and then find the smallest solution. A more efficient way is to introduce flow variables $f_e^{(ij)}$ for every disjoint s_i, s_j-pair (and every edge) and maximize a slack variable $\tilde{f}$, while requiring the flow value for every s_i, s_j-pair to be greater or equal than $\tilde{f}$ and all other constraints of LP Eq. (10) to be fulfilled for every disjoint s_i, s_j-pair:

$$\min_{\substack{s_i,s_j\in S, s_i\neq s_j \\ j>i}} f_{\max}^{s_i\to s_j}\left(G', \{c'(\{wv\})\}_{\{wv\}\in E'}\right) = f_{\max}^S\left(G', \{c'(\{vw\})\}_{\{vw\}\in E'}\right), \tag{31}$$

where

$$f_{\max}^S\left(G', \{c'(\{vw\})\}_{\{vw\}\in E'}\right) = \max \tilde{f} \tag{32}$$

$$\forall i,j>i : \tilde{f} - \sum_{v:\{s_i v\}\in E'} \left(f_{s_i v}^{(ij)} - f_{v s_i}^{(ij)}\right) \leq 0 \tag{33}$$

$$\forall i,j>i, \{vw\} \in E' : f_{vw}^{(ij)} + f_{wv}^{(ij)} \leq c'(\{vw\}) \tag{34}$$

$$\forall i,j>i, \forall\{vw\} \in E' : f_{vw}^{(ij)}, f_{wv}^{(ij)} \geq 0 \tag{35}$$

$$\forall i,j>i, \forall w \in V, w\neq s_i, s_j : \sum_{v:\{vw\}\in E'} \left(f_{vw}^{(ij)} - f_{wv}^{(ij)}\right) = 0. \tag{36}$$

Adding a maximization over usage frequencies, we obtain

$$\vec{f}_{\max}^S(G', \{C(\mathcal{N}^e)\}_{e\in E}) = \max_{0\leq p_e\leq 1, \sum_e p_e=1} f_{\max}^S(G', \{c_C'(\{vw\}, p_{wv}, p_{vw})\}_{\{vw\}\in E'}). \tag{37}$$

It will be convenient to introduce an undirected multigraph $G''_{\lfloor c'\rfloor}$, by replacing each edge $\{vw\} \in E'$ with $\lfloor c'(\{vw\})\rfloor$ identical edges with unit-capacity connecting v and w. An S-cut in an undirected unit-capacity multigraph G'' is defined as a set of edges whose removal disconnects at least two vertices in S. The size $\lambda_S(G'')$ of the minimum S-cut in G'' is called the S-connectivity of G''.

In G'' we can also define a Steiner tree spanning S, in short S-tree, as a subgraph of G'' that contains all vertices in S and is a tree, i.e., does not contain any cycles. If S only consists of two vertices, we call an S-tree a path. We call two Steiner trees edge-disjoint, if they do not contain a common edge. The problem of finding the number $t_S(G'')$ of edge-disjoint Steiner trees in a general undirected multigraph is NP-complete[46]. However, there is a connection between S-connectivity and the number of edge-disjoint S-trees in an undirected unit-capacity multigraph[47–49]:

$$t_S(G'') \geq \lfloor g_1\lambda_S(G'')\rfloor - g_2. \tag{38}$$

In ref. [47] it has been conjectured that Eq. (38) holds for $g_1 = \frac{1}{2}$ and $g_2 = 0$. In ref. [48] it has been shown that the relation holds for $g_1 = \frac{1}{26}$ and $g_2 = 0$, whereas the authors of ref. [49] show that it holds for $g_1 = \frac{1}{2}$ and $g_2 = \frac{|V\setminus S|}{2} + 1$, which is finite in the graphs we are considering.

On complexity. Let us briefly discuss the computational complexity of our LPs Eqs. (12), (19), (21), and (37). Using interior point methods, e.g., ref. [26], a LP in standard form

$$\begin{aligned} \min \quad & c^T x \\ & Ax = b, x \geq 0, \end{aligned} \tag{39}$$

where $c, x \in \mathbb{R}^N$, $b \in \mathbb{R}^M$, and $A \in \mathbb{R}^{M\times N}$, can be solved using $\mathcal{O}(\sqrt{N}L)$ iterations and $\mathcal{O}(N^3 L)$ total arithmetic operations. Here L is the size of the problem data, A, b, c, which scales as $\mathcal{O}(MN + M + N)$[66]. If we assume A to be of full rank, it holds $M \leq N$, and hence, L scales as $\mathcal{O}(N^2)$. Using slack variables[26], all inequality constraints in our LPs can be converted into equality constraints. Linear equality constraints can be easily written in the form $Ax = b$. Hence N can be obtained by adding the number of variables and the number of inequality constraints in our LPs.

For LP Eq. (12) we have $N = 3|E'| + |E|$. LP Eq. (19) has $2r|E'| + |E|$ variables and $|E'| + |E|$ inequality constraints. Thus $N = (2r + 1)|E'| + 2|E|$ for LP Eq. (19). LP Eq. (21) has $2r|E'| + |E| + 1$ variables and $|E'| + |E| + r$ inequality constraints. Thus $N = (2r + 1)|E'| + 2|E| + 1 + r$ for LP Eq. (21). LP Eq. (37) has $2\binom{|S|}{2}|E'| + |E| + 1$ variables and $\binom{|S|}{2}|E'| + |E| + \binom{|S|}{2}$ inequality constraints. Thus $N = 3\binom{|S|}{2}|E'| + 2|E| + 1 + \binom{|S|}{2}$ for LP Eq. (37). Hence, all our LPs, the number of iterations as well as the number of total arithmetic operations scale polynomially with the size of the network.

References

1. Ekert, A. K. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* **67**, 661–663 (1991).
2. Bennett, C. H., Brassard, G. & Mermin, N. D. Quantum cryptography without Bell's theorem. *Phys. Rev. Lett.* **68**, 557 (1992).
3. Bennett, C. H. et al. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **70**, 1895 (1993).

4. Wehner, S., Elkouss, D. & Hanson, R. Quantum internet: a vision for the road ahead. *Science* **362**, eaam9288 (2018).

5. Augusiak, R. & Horodecki, P. Multipartite secret key distillation and bound entanglement. *Phys. Rev. A* **80**, 042307 (2009).

6. Hillery, M., Bužek, V. & Berthiaume, A. Quantum secret sharing. *Phys. Rev. A* **59**, 1829 (1999).

7. Komar, P. et al. A quantum network of clocks. *Nat. Phy.* **10**, 582–587 (2014).

8. Raussendorf, R., Browne, D. E. & Briegel, H. J. Measurement-based quantum computation on cluster states. *Phys. Rev. A* **68**, 022312 (2003).

9. Smith, G. & Yard, J. Quantum communication with zero-capacity channels. *Science* **321**, 1812–1815 (2008).

10. Pirandola, S. Capacities of repeater-assisted quantum communications. *arXiv preprint arXiv:1601.00966* (2016).

11. Pirandola, S. End-to-end capacities of a quantum communication network. *Commun. Phys.* **2**, 51 (2019).

12. El Gamal, A. & Kim, Y.-H. *Network Information Theory* (Cambridge University Press, 2011).

13. Azuma, K., Mizutani, A. & Lo, H.-K. Fundamental rate-loss tradeoff for the quantum internet. *Nat. Commun.* **7**, 13523 (2016).

14. Azuma, K. & Kato, G. Aggregating quantum repeaters for the quantum internet. *Phys. Rev. A* **96**, 032332 (2017).

15. Rigovacca, L. et al. Versatile relative entropy bounds for quantum networks. *N. J. Phys.* **20**, 013033 (2018).

16. Van Meter, R. & Touch, J. Designing quantum repeater networks. *IEEE Commun. Mag.* **51**, 64–71 (2013).

17. Van Meter, R., Satoh, T., Ladd, T. D., Munro, W. J. & Nemoto, K. Path selection for quantum repeater networks. *Netw. Sci.* **3**, 82–95 (2013).

18. Epping, M., Kampermann, H. & Bruß, D. Robust entanglement distribution via quantum network coding. *N. J. Phys.* **18**, 103052 (2016).

19. Epping, M., Kampermann, H. & Bruß, D. Large-scale quantum networks based on graphs. *N. J. Phys.* **18**, 053036 (2016).

20. Wallnöfer, J., Zwerger, M., Muschik, C., Sangouard, N. & Dür, W. Two-dimensional quantum repeaters. *Phys. Rev. A* **94**, 052307 (2016).

21. Hahn, F., Pappa, A. & Eisert, J. Quantum network routing and local complementation. *npj Quantum Inf.* **5**, 1–7 (2019).

22. Chakraborty, K., Rozpedek, F., Dahlberg, A. & Wehner, S. Distributed routing in a quantum internet. Preprint at https://arxiv.org/abs/1907.11630 (2019).

23. Pirandola, S. Bounds for multi-end communication over quantum networks. *Quantum Sci. Technol.* **4**, 045006 (2019).

24. Bäuml, S. & Azuma, K. Fundamental limitation on quantum broadcast networks. *Quantum Sci. Technol.* **2**, 024004 (2017).

25. Yamasaki, H., Soeda, A. & Murao, M. Graph-associated entanglement cost of a multipartite state in exact and finite-block-length approximate constructions. *Phys. Rev. A* **96**, 032330 (2017).

26. Ye, Y. An. O. An O(n^3 L) potential reduction algorithm for linear programming. *Math. Program.* **50**, 239–258 (1991).

27. Aumann, Y. & Rabani, Y. An O(log k) approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.* **27**, 291–301 (1998).

28. Günlük, O. A new min-cut max-flow ratio for multicommodity flows. *SIAM J Discret. Math.* **21**, 1–15 (2007).

29. Garg, N., Vazirani, V. V. & Yannakakis, M. Approximate max-flow min-(multi) cut theorems and their applications. *SIAM J. Comput.* **25**, 235–251 (1996).

30. Garg, N., Vazirani, V. V. & Yannakakis, M. Primal-dual approximation algorithms for integral flow and multicut in trees. *Algorithmica* **18**, 3–20 (1997).

31. Pirandola, S., Laurenza, R., Ottaviani, C. & Banchi, L. Fundamental limits of repeaterless quantum communications. *Nat. Commun.* **8**, 15043 (2017).

32. Wilde, M. M. *Quantum Information Theory* (Cambridge University Press, 2013).

33. Horodecki, K., Horodecki, M., Horodecki, P. & Oppenheim, J. Secure key from bound entanglement. *Phys. Rev. Lett.* **94**, 160502 (2005).

34. Horodecki, K., Horodecki, M., Horodecki, P. & Oppenheim, J. General paradigm for distilling classical key from quantum states. *IEEE Trans. Inf. Theory* **55**, 1898–1929 (2009).

35. Christandl, M. & Winter, A. Squashed entanglement: An additive entanglement measure. *J. Math. Phys.* **45**, 829–840 (2004).

36. Datta, N. Min-and max-relative entropies and a new entanglement monotone. *IEEE Trans. Inf. Theory* **55**, 2816–2826 (2009).

37. Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. Mixed-state entanglement and quantum error correction. *Phys. Rev. A* **54**, 3824–3851 (1996).

38. Horodecki, M., Horodecki, P. & Horodecki, R. General teleportation channel, singlet fraction, and quasidistillation. *Phys. Rev. A* **60**, 1888–1898 (1999).

39. Mueller-Hermes, A. Transposition in Quantum Information Theory. Master's thesis, Technical University of Munich (2012).

40. Vedral, V., Plenio, M. B., Rippin, M. A. & Knight, P. L. Quantifying entanglement. *Phys. Rev. Lett.* **78**, 2275 (1997).

41. Elias, P., Feinstein, A. & Shannon, C. A note on the maximum flow through a network. *IRE Trans. Inf. Theory* **2**, 117–119 (1956).

42. Ford, L. R. & Fulkerson, D. R. Maximal flow through a network. *Can. J. Math.* **8**, 399–404 (1956).

43. Murty, K. G. *Linear Programming* (Springer, 1983).

44. Nishizeki, T. Planar graph problems. In *Computational Graph Theory*, 53–68 (Springer, 1990).

45. Greenberger, D. M., Horne, M. A. & Zeilinger, A. Going beyond Bell's theorem. In *Bell's Theorem, Quantum Theory and Conceptions of the Universe* 69–72 (Springer, 1989).

46. Cheriyan, J. & Salavatipour, M. R. Hardness and approximation results for packing steiner trees. *Algorithmica* **45**, 21–43 (2006).

47. Kriesell, M. Edge-disjoint trees containing some given vertices in a graph. *J. Comb. Theory Ser. B* **88**, 53–65 (2003).

48. Lau, L. C. An approximate max-Steiner-tree-packing min-Steiner-cut theorem. In *Proc 45th Annual IEEE Symposium on Foundations of Computer Science* 61–70 (IEEE, 2004).

49. Petingi, L. & Talafha, M. Packing the steiner trees of a graph. *Networks* **54**, 90–94 (2009).

50. Li, Z. & Li, B. Network coding: the case of multiple unicast sessions. In *Allerton Conference on Communications*, Vol. 16, 8 (IEEE, Piscataway, New Jersey, 2004).

51. Harvey, N. J., Kleinberg, R. D. & Lehman, A. R. Comparing network coding with multicommodity flow for the k-pairs communication problem (2004).

52. Ford Jr, L. R. & Fulkerson, D. R. *Flows in networks (part III)* (Princeton University Press, 2015).

53. Klein, P. N., Plotkin, S. A., Rao, S. & Tardos, E. Approximation algorithms for steiner and directed multicuts. *J. Algorithms* **22**, 241–269 (1997).

54. Kannan, R., Vempala, S. & Vetta, A. On clusterings: Good, bad and spectral. *J. ACM* **51**, 497–515 (2004).

55. Schaeffer, S. E. Graph clustering. *Comput. Sci. Rev.* **1**, 27–64 (2007).

56. Leon-Garcia, A. & Widjaja, I. *Communication Networks* Ch. 4 (McGraw-Hill, Inc., 2003).

57. Dantzig, G. & Fulkerson, D. On the Max Flow Min Cut Theorem of Networks. Tech. Rep., Rand Corp Santa Monica, CA (1955).

58. Shahrokhi, F. & Matula, D. W. The maximum concurrent flow problem. *J. ACM* **37**, 318–334 (1990).

59. Leighton, T. & Rao, S. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *J. ACM* **46**, 787–832 (1999).

60. Linial, N., London, E. & Rabinovich, Y. The geometry of graphs and some of its algorithmic applications. *Combinatorica* **15**, 215–245 (1995).

61. Gupta, A., Newman, I., Rabinovich, Y. & Sinclair, A. Cuts, trees and l 1-embeddings of graphs. *Combinatorica* **24**, 233–269 (2004).

62. Chekuri, C., Gupta, A., Newman, I., Rabinovich, Y. & Sinclair, A. Embedding k-outerplanar graphs into l1. *SIAM J. Discret. Math.* **20**, 119–136 (2006).

63. Lee, J. R. & Sidiropoulos, A. On the geometry of graphs with a forbidden minor. In *Proc. 41st Annual ACM Symposium on Theory of Computing*, 245–254 (ACM, 2009).

64. Chakrabarti, A., Fleischer, L. & Weibel, C. When the cut condition is enough: A complete characterization for multiflow problems in series-parallel networks. In *Proc. 44th Annual ACM Symposium on Theory of Computing*, 19–26 (ACM, 2012).

65. Salmasi, A., Sidiropoulos, A. & Sridhar, V. On constant multi-commodity flow-cut gaps for directed minor-free graphs. Preprint at: https://arxiv.org/abs/1711.01370 (2017).

66. Wright, S. J. *Primal-dual Interior-point Methods* (Siam, 1997).

Acknowledgements

Author contributions

D.E. conceived the project, S.B. worked out the technical derivations with the help of K.A., D.E., and G.K. S.B. wrote the paper with the help of K.A. and D.E.

11

Competing interests

All authors declare that they have no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s42005-020-0318-2.

Correspondence and requests for materials should be addressed to S.B.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.