The Streaming Capacity of Sparsely-Connected P2P Systems with Distributed Control

Can Zhao and Xiaojun Lin
School of Electrical and Computer Engineering,
Purdue University, West Lafayette
Email: {zhao43,linx}@purdue.edu

Chuan Wu
Department of Computer Science,
The University of Hong Kong
Email: cwu@cs.hku.hk

Abstract

Peer-to-Peer (P2P) streaming technologies can take advantage of the upload capacity of clients, and hence can scale to large content distribution networks with lower cost. A fundamental question for P2P streaming systems is the maximum streaming rate that all users can sustain. Prior works have studied the optimal streaming rate for a complete network, where every peer is assumed to communicate with all other peers. This is however an impractical assumption in real systems. In this paper, we are interested in the achievable streaming rate when each peer can only connect to a small number of neighbors. We show that even with a random peer-selection algorithm and uniform rate allocation, as long as each peer maintains $\Omega(\log N)$ downstream neighbors, where N is the total number of peers in the system, the system can asymptotically achieve a streaming rate that is close to the optimal streaming rate of a complete network. We then extend our analysis to multi-channel P2P networks, and we study the scenario where "helpers" from channels with excessive upload capacity can help peers in channels with insufficient upload capacity. We show that by letting each peer select $\Omega(\log N)$ neighbors randomly from either the peers in the same channel or from the helpers, we can achieve a close-to-optimal streaming capacity-region. Simulation results are provided to verify our analysis.

1 Introduction

With the proliferation of high-speed broadband services, the demand for rich multimedia content over the internet, in particular high-quality video delivery over the Internet, has kept increasing. Streaming video directly from the server will require a large amount of upload bandwidth at the server, which can be costly. The service quality can also be poor when the clients are far away from the server. In addition, it may be difficult for the server bandwidth to keep up when the demand is exceeding high. There have been different approaches to off-load traffic from the server, using either CDN (content distribution network) or P2P (peer-to-peer) technologies. Deploying a large CDN can introduce a high fixed cost. In contrast, P2P technologies are particularly attractive because they take advantage of the upload bandwidth of the clients, which does not incur additional cost to the video service provider. Several well-known commercial P2P live streaming systems have been successfully deployed, include CoolStreaming [1], PPLIVE [2], TVAnts [3], UUSEE [4], PPStream [5]. A typical P2P live streaming system can now offer thousands of TV channels or movies for viewing, and may serve hundreds of thousands of users simultaneously [4].

In contrast to the practical success of these P2P live streaming systems, the theoretical understanding of the performance of P2P live streaming seems to be lagging behind, which may impede further improvement of P2P live streaming. For example, a basic question for a P2P live streaming system is that of its streaming capacity, i.e., what is the maximum streaming rate that all users can sustain? This question has been studied under the assumption of a complete network, where each peer can connect to all other peers simultaneously. Under this assumption, the maximum streaming capacity has been found in [6], and both centralized and distributed rate allocation algorithms to achieve this maximum streaming capacity have been developed [6–9]. However, the assumption of a complete network is impractical for any large-scale P2P streaming systems. In a real P2P streaming system, typically each peer is only given a small list of other peers (which we refers to as neighbors) chosen from the entire population, and each peer can only connect to this subset of neighboring peers (neighbors may not be close in terms of physical distance). The number of neighboring peers is often much smaller than the total population, in order to limit the control overhead.

When each peer only has a small number of neighbors, the P2P network can be modeled as an incomplete graph with node-degree constraints. In this case, the streaming capacity of P2P systems becomes more complicated to characterize. Liu et al. [10] investigate the case when the number of down-stream peers in a single sub-stream tree is bounded. However, the number of neighbors that each peer could have over all sub-streams can still be very large (in the worse case it can be connected to all the other peers simultaneously). Some approximated and centralized solutions to solve the optimal streaming capacity problem on a given incomplete network has been proposed in [11]. However, for large-scale P2P streaming systems, such a centralized approach will be difficult to scale. Liu et al. [12] proposed a Cluster-Tree algorithm to construct a topology subject to a bounded node-degree constraint, which could achieve a streaming rate that is close to the optimal streaming capacity of a complete network. This result gives us hope that, even with node-degree constraints, a P2P network may achieve almost the same streaming rate as that of a complete network. However, the Cluster-Tree algorithm is not a completely de-centralized algorithm because it requires the tracker (a central entity) to apply the Bubble algorithm at the cluster level. The Bubble algorithm is a centralized algorithm. Some other works such as SplitStream [13] and Chinasaw [14] have also studied the problem of how to improve the streaming capacity when there is a node-degree constraint. However, these works did not provide theoretical results on the achievable streaming rate. To the best of our knowledge, we have not been aware of a fully distributed algorithm in the literature that can achieve close-to-optimal P2P streaming capacity on incomplete networks.

All of the above works are for single-channel P2P systems. Today's P2P systems typically serve a large number of TV channels and movies at the same time. In most P2P streaming systems, peers exchange data only with other peers that are viewing the same channel. Hence, peers from different channels are isolated from each other. Recently, Wu et al. [15, 16] show that by allowing peers to exchange data with other peers that are not even viewing the same channel, the overall performance of a multi-channel system can be improved. Such cross-channel peer exchange is particularly helpful for channels that do not have enough upload capacity, and hence they need the upload capacity of peers from other channels to improve their streaming rate. [15,16] have proposed a View-Upload Decoupling (VUD) algorithm that sets up a semi-permanent distribution group of peers, who are not necessarily the peers interested in viewing a channel, to help distribute the content of the channel. Although the VUD algorithm has been shown to improve the multi-channel streaming capacity, it is again a centralized algorithm and it assumes that all peers can connect to all other peers simultaneously, which is impractical for real systems.

In this paper, we are interested in the following question: without centralized control, how many neighbors does a peer in a large P2P network need to be maintained in order to achieve a streaming capacity that is close to the optimal streaming capacity of an otherwise complete network? Further, can we develop fully-distributed algorithms for peer-selection and rate-allocation to achieve the close-to-optimal streaming capacity? This paper provides some interesting and positive answers to these questions. First, we show that, if each peer has $\Omega(\log N)$ neighbors, where N is the total number of peers in the system, close-to-optimal streaming rate can be achieved with probability approaching 1 as N goes to infinity. Further, in order to achieve, this goal, each peer only needs to choose $\Omega(\log N)$ downstream neighbors uniformly and randomly from the entire population, and simply allocates its upload capacity evenly among all downstream peers. Only the server needs a slightly different peer-selection policy (see Section 2.2 for details).

We also extend our analysis to multi-channel systems, and allow peers from those channels with abundant upload capacity to help other channels with insufficient upload capacity. Again, we show that by using a simple and distributed algorithm where each peer randomly selects a small number of neighbors from peers belonging to the same channel and from the helper peers from other channels, a close-to-optimal streaming capacity region for multi-channel systems can be achieved with high probability. Hence, our results indicate that the benefit of VUD can be retained in a distributed manner without the complete network assumption.

The results that we obtain have a similar flavor as scaling-law results in wireless ad hoc networks [17]. Although such results only hold when the size of the network N is large, they do provide important insights into the dynamics of the system. For example, our analysis indicates that, with a random peer selection, the last hop will most likely be the bottle-neck for the streaming capacity. This insight suggests that we could focus on balancing the capacity at the *last* hop when designing new distributed resource allocation algorithms for P2P streaming systems. As an initial application of this insight, we show with an example that, by slightly adjusting the uniform rate-allocation strategy, we can indeed improve the probability of attaining the near-optimal streaming rate empirically. Hence, we believe that the insights from these results can be very helpful for designing more efficient control algorithms for P2P streaming.

2 Single-Channel P2P Networks

In this section, we will show that even without centralized control, $\Omega(\log N)$ neighbors is sufficient for large single-channel P2P streaming networks. Specifically, we will show that just by letting each peer select its $\Omega(\log N)$ neighbors randomly, the close-to-optimal streaming rate could be achieved will high probability when the network size N is large.

2.1 System Model

We consider a peer-to-peer live streaming network with N peers and one source s. In the rest of the paper, we will use the terms "source" and "server" interchangeably. Similarly, we will use the terms "peer", "node" and "user" interchangeably. Denote the set of all peers and the source as V (thus, $|V| = N + 1$). We assume that the source has a video file with infinite size to be streamed to all peers and it has a fixed upload capacity u_s. Denote the upload capacity of peer i as U_i, which is a random variables defined as follows: each peer has an upload capacity of $U_i = u$ with probability p and an upload capacity of $U_i = 0$ with probability $1 - p$, i.i.d. across peers. Although this is a somewhat simplified ON-OFF model, we believe that the insights obtained from this model can also be generalized to other models for the distribution of the upload capacity. We assume that $u_s \geq u$. Like other works [6, 11, 12, 18], we assume that the download capacity and the core network capacity are sufficiently large, and hence the only capacity constraints are on the upload capacity. Each peer $i \in V \backslash \{s\}$ has a fixed set $\mathcal{E}_i$ of M downstream neighbors. Similarly, the source has a set $\mathcal{E}_s$ of M downstream peers. We can then model the P2P network as a directed and capacitated random graph [19]. If $j \in \mathcal{E}_i$, assign an directed edge (i, j) from i to j. Let the set of all edges be E. Note that there may be multiple peers that have a common downstream neighbor. Define C_{ij} and C_{sj} be the streaming rate from peer i and source s, respectively, to peer j.

The value of $\mathcal{E}_i$, $\mathcal{E}_s$, C_{ij} and C_{sj} depend on the peer-selection and rate-allocation algorithm. Given such an algorithm, we can define the "streaming capacity" of the system as the maximum rate that the source could distribute the streaming content to all peers. For example, for a complete network, we have $\mathcal{E}_i = V \backslash \{i, s\}$ and $\mathcal{E}_s = V \backslash \{s\}$. [6] shows that the optimal streaming capacity is on average

$$C_f = \min \left\{ u_s, \frac{u_s + \sum_{i \in V} \mathbf{E}[U_i]}{N} \right\}, \tag{1}$$

and can be achieved by setting $C_{ij} = U_i/(N-1)$ and $C_{sj} = U_s/N$ for all i, j. For our ON-OFF model of upload capacity, this optimal streaming capacity is equal to $C_f = \min \left\{ u_s, \frac{u_s}{N} + up \right\}$. However, as we discussed in the introduction, the assumption of a complete network is impractical. In this paper, we are interested in the streaming capacity of an incomplete network, which can be calculated by the minimum cuts. Specifically note that for a given user t, a cut that separates s and t is defined by dividing the peers in V into a set V_n of size $(n+1)$ that contains the server, and the complementary set V_n^c of size $(N-n)$ that contains the peer t, i.e.,

$$s \in V_n, |V_n| = n+1, t \in V_n^c \text{ and } |V_n^c| = N-n.$$

The capacity of the cut C_n is defined as $C_n = \sum_{i \in V_n} \sum_{j \in V_n^c} C_{ij}$. See fig 1 for illustration.

Let $C_{\min}(s \to t)$ denote the minimum-cut capacity, which is the minimum capacity of all cuts that separate the source s and the destination t. It is well-known that this min-cut capacity is equal to the maximum rate from s to t. Let $C_{\min-\min}(s \to \mathcal{T})$ denote the min-min-cut which is the minimum cut of all individual min-cut capacities from the source to each destination t within a set $\mathcal{T}$, i.e.,

$$C_{\min-\min}(s \to \mathcal{T}) = \min_{t \in \mathcal{T}} C_{\min}(s \to t).$$

The streaming capacity of the network is then equal to $C_{\min-\min}(s \to V \backslash \{s\})$ [20]. Note that given the graph and the capacity of each edge, this streaming capacity can be achieved with simple transmission schemes, e.g., with network coding [21, 22] or with a latest-useful-chunk policy [7]. However, it

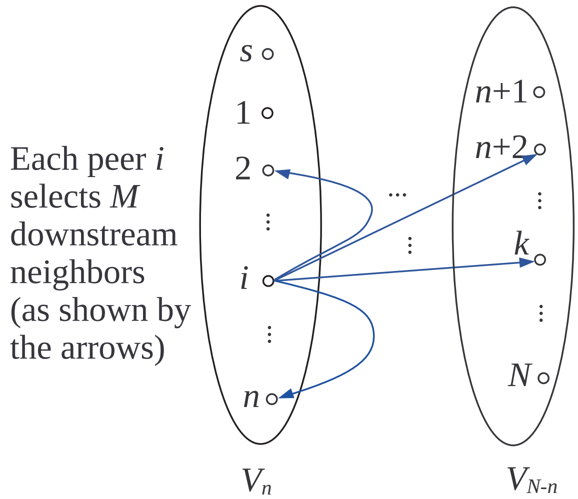

Figure 1: Illustration of the neighbor selection and a cut

may required global knowledge and centralized control in order to optimally construct the network graph and allocate the upload capacity. A natural question is then the following: without centralized control, can the streaming capacity over an incomplete network approach the optimal streaming capacity C_f of a complete network? In the next subsection we will provide a simple and distributed peer-selection and rate-allocation algorithm that can achieve this with high probability when the network size is large.

2.2 Algorithms

We will now give explicit description of our simple control algorithm. First, we use a random peer-selection algorithm. Specifically, each peer will randomly select M downstream neighbors uniformly from all other peers. On the other hand, the server will select M downstream neighbors uniformly and randomly among the ON peers. Second, we use a uniform rate-allocation algorithm, i.e., each peer i simply divides its upload capacity equally among all of its downstream neighbors in $\mathcal{E}_i$. Therefore, each peer in the set $\mathcal{E}_i$ will receive a streaming rate U_i/M from peer i. Similarly, each downstream peer of the server receives U_s/M from the server. Under the above scheme, the link capacity C_{ij} is given by

$$C_{ij} = \begin{cases} U_i/M, & \text{if } j \in \mathcal{E}_i, i \neq s \\ U_s/M, & \text{if } j \in \mathcal{E}_s, i = s \\ 0, & \text{otherwise.} \end{cases}$$

Note that since $\mathcal{E}_i$ and $\mathcal{E}_s$ are chosen randomly, C_{ij}'s will also be random variables. We define another import parameter for the total capacity that each peer i directly receives from its upstream neighbors, which is given by $C_i^R = \sum_{j \in V} C_{ji}$. We will see that this value is the main factor that determines the streaming capacity from the source to each node.

Note that the above algorithm is a very simple mesh-based algorithm with the following advantages:

- Simplicity - The random peer selection and uniform rate allocation are easy to implement.
- Robustness - If some peer leaves the system, only the upstream neighbors of that peer need to re-select another downstream neighbor. It is not necessary to reconstruct the whole topology. Further, when a peer switches ON or OFF, its set of downstream neighbors does not need to change.
- Low signaling overhead - Only the server need to know which peers are ON. The tracker does not need to update the upload capacity of peers to any other peer.

Somewhat surprisingly, we will show that, as long as $M = \Omega(\log N)$, the algorithm will achieve close-to-optimal streaming capacity, with probability approaching 1 as $N \to \infty$ (Theorem 1).

Remark: Note that the server will only choose ON peers as its downstream neighbors. This is essential for achieving the close-to-optimal streaming capacity. To see this, note that the optimal streaming capacity C_f of a complete network is also constrained by the server capacity (see Equation (1)). If the server had used a substantial fraction of its upload capacity to serve OFF peers, intuitively the rest of the peers would then suffer a lower streaming rate. With the same intuition, one would think that the peers directly connected to the server also need to be careful in choosing their downstream neighbors. However, this turns out to be unnecessary. For our main result (Theorem 1) to hold, no other peers (except the server) are required to differentiate their downstream neighbors. As readers will see, this is because those cuts with V_n only containing the downstream neighbors of s will play a small role in the overall probability of attaining the close-to-optimal streaming capacity.

We also note that the above algorithm uses the "push" model, where upstream peers choose downstream neighbors. An alternate model is the "pull" model, where downstream peers choose upstream neighbors. Note that both models create a mesh-topology, and there is considerable symmetry between the two models. We use the push model in this paper because it is easier to analysis, although we believe that the main results of the paper can also be generalized to the pull model, which we leave as future work.

2.3 Main Result

Theorem 1. *For any $\epsilon \in (0,1)$ and $d > 1$, there exists α and N_0 such that for any $M = \alpha \log(N)$ and $N > N_0$ the probability for the min-min-cut under the algorithm in Section 2.2 to be smaller than $(1-\epsilon)C_f$ is bounded by*

$$\mathbf{P}\left(C_{\min-\min}(s \to V) \leq (1-\epsilon)C_f\right) \leq O\left(\frac{1}{N^{2d-1}}\right).$$

Recall that the min-min-cut is equal to the streaming rate to all peers. Hence, Theorem 1 shows that as long as the number of downstream neighbors M is $\Omega(\log N)$, for any $\epsilon \in (0,1)$ the streaming rate of our algorithm will be larger than $(1-\epsilon)$ times the optimal streaming capacity with probability approaching 1 as the network size N increases.

2.4 Proof of Theorem 1

We first find the min-cut for any fixed peer t. We will use a similar approach as the one in [19]. We will show that the probability for the capacity of a cut to be smaller than $(1-\epsilon)$ times its mean is very small as N becomes large. Then, we will take the union bound over all cuts and show that overall probability is also very small. However, the techniques in [19] do not directly apply to our model due to the following two reasons. First, due to ON-OFF model, there are fewer "ON" peers and hence the probability for each cut to fall below its expected value will be larger than the case when all peers' upload capacity is the same. However, there are still the same number of cuts we need to account for, which may cause the union bound in [19] to diverge. Second, the link capacity C_{ij} in [19] is assumed to be independent across j, which is not the case in our model. To address the first difficulty, we will first consider the subgraph that only contains the ON users, and hence the number of cuts is also reduced correspondingly. To address the second difficulty, we will show that the joint distribution of C_{ij} can be approximated by i.i.d. random variables, which significantly simplifies the analysis.

We first introduce the following general relationship between the min-cut from the server s to the peer t in a random graph G and the min-cut from the server s to the peer t in the any subgraph H_t of G that contains s and t.

Proposition 2. *Let G be a random graph defined on some probability space Ω that has a fixed source s and a fixed destination t. Let H_t be another random graph defined on the same probability space such that $H_t(\omega) \subseteq G(\omega)$ for all $\omega \in \Omega$ and H_t contains s and t. Then for any given positive value C, the following holds,*

$$\mathbf{P}\left(C_{\min,G}(s \to t) \leq C\right) \leq \mathbf{P}\left(C_{\min,H_t}(s \to t) \leq C\right). \tag{2}$$

where $C_{\min,G}(s \to t)$ is the min-cut in G from s to t, and $C_{\min,H_t}(s \to t)$ is the min-cut in H_t from s to t.

Proof. Let $\mathcal{A} = \{G(\omega) : C_{\min,G(\omega)}(s \to t) \leq C\}$ and $\mathcal{B} = \{\omega : C_{\min,H_t(\omega)}(s \to t) \leq C\}$. For any $\omega \in \mathcal{A}$, the min-cut from s to t in the graph $G(\omega)$ is less than C. Since H_t is a subgraph of $G(\omega)$, the min-cut from s to t in $H_t(\omega)$ is smaller than the min-cut in $G(\omega)$, i.e.,

$$C_{\min,H_t(\omega)}(s \to t) \leq C_{\min,G(\omega)}(s \to t) \leq C.$$

Hence, $\omega \in \mathcal{B}$. We then have $\mathcal{A} \subseteq \mathcal{B}$ and (2) holds consequently. □

Proposition 2 is intuitive because every cut in $G(\omega)$ has a larger capacity than the corresponding cut in the subgraph $H_t(\omega)$. For a given destination t, let $H_t(W, F)$ be the subgraph of $G(V, E)$ such that W contains the peer t, the server and all of the nodes whose channel condition is ON, and $F \subset E$ is those edges between nodes in W. The capacity of the edges in F is the same as the capacity of the edges in E. Proposition 2 allows us to focus on the subnetwork H_t instead of the entire network G. Assume that there are Y ON peers in the network excluding peer t, and thus $|W| = Y + 2$. Clearly, Y is a random variable with binomial distribution with parameter $N-1$ and p. For ease of exposition, we assume that Y is fixed during the following discussion for one given cut, and we will consider the randomness of Y later when we take the union bound over all cuts. We define a cut on H_t by dividing the peers in W into a set W_m of size $m+1$ that contains the server, and the complementary set W_m^c of size $Y-m+1$ that contains peer t. The capacity of the cut D_m is then given by

$$D_m = \sum_{k \in W_m^c} C_{si} + \sum_{i \in W_m} \sum_{k \in W_m^c} C_{ik}.$$

Note that for each peer $i \in W_m$ (and $i \neq s$), we have $\sum_{k \in W_m^c} C_{ik} = L_i u/M$, where L_i is the number of downstream neighbors of peer i that are in the set W_m^c. Note that the value of L_i must satisfy $\max\{0, M-(N-Y+m-2)\} \leq L_i \leq \min\{M, Y-m+1\}$. Since downstream neighbors of peer i are uniformly chosen from other peers, we have,

$$\mathbf{P}\left(\sum_{k \in W_m^c} C_{ik} = l \cdot \frac{u}{M}\right) = \frac{\binom{Y-m+1}{l}\binom{N-Y+m-2}{M-l}}{\binom{N-1}{M}}.$$

This is the probability that l out of M downstream neighbors of peer i are in W_m^c (of size $Y-m+1$) and $M-l$ of them are in the set W_m. The distribution of L_i is known as a hypergeometric distribution with expectation $\frac{(Y-m+1)M}{N-1}$ [23, p167]. We can get a similar expression for the source s, i.e.,

$$\mathbf{P}\left(\sum_{i \in W_m^c} C_{si} = l \cdot \frac{u_s}{M}\right) = \begin{cases} \frac{\binom{Y-m}{l}\binom{m}{M-l}}{\binom{Y}{M}} & \text{if } t \text{ is OFF,} \\ \frac{\binom{Y-m+1}{l}\binom{m}{M-l}}{\binom{Y+1}{M}} & \text{if } t \text{ is ON.} \end{cases}$$

$$\mathbf{E}\left[\sum_{i \in W_m^c} C_{si}\right] = \begin{cases} \frac{u_s(Y-m)}{Y} & \text{if } t \text{ is OFF,} \\ \frac{u_s(Y+1-m)}{Y+1} & \text{if } t \text{ is ON.} \end{cases}.$$

Hence, we obtain the expectation of D_m as

$$\mathbf{E}\left[D_m\right] = \mathbf{E}\left[\sum_{k \in W_m^c} C_{si}\right] + \sum_{i \in W_m} \mathbf{E}\left[\sum_{k \in W_m^c} C_{ik}\right]$$
$$= \begin{cases} \frac{u_s(Y-m)}{Y} + \frac{u}{N-1} m(Y-m+1) & \text{if } t \text{ is OFF,} \\ \frac{u_s(Y+1-m)}{Y+1} + \frac{u}{N-1} m(Y-m+1) & \text{if } t \text{ is ON.} \end{cases} \quad (3)$$

Next, we are interested in the probability that $D_m \geq (1-\epsilon)\mathbf{E}[D_m]$ for all m for a given constant $\epsilon \in (0,1)$. In other words, this is the probability that the min-cut value is no less than $(1-\epsilon)$ times its average. For all m, it is not hard to see

$$\mathbf{E}[D_m] \geq \min\{\mathbf{E}[D_0], \mathbf{E}[D_Y]\} = \min\left\{u_s, \frac{u_s}{Y} + \frac{Y}{N-1}u\right\}.$$

If we have $Y \geq (1-\epsilon)p(N-1)$, we will get

$$\mathbf{E}[D_m] \geq (1-\epsilon)\min\left\{u_s, \frac{u_s}{N} + pu\right\}.$$

Recall that $C_f = \min\{u_s, \frac{u_s}{N} + pu\}$ is the optimal streaming capacity assuming a complete network [6]. Hence, $D_m \geq (1-\epsilon)\mathbf{E}[D_m]$ will then imply that $D_m \geq (1-\epsilon)^2 C_f$. In other words, the probability that $D_m \geq (1-\epsilon)\mathbf{E}[D_m]$ for all m will become a lower bound for the probability that the min-cut is no less than $(1-\epsilon)^2 C_f$. In the following, we will derive $\mathbf{P}(D_m \geq (1-\epsilon)\mathbf{E}[D_m])$. We will first find a bound on the moment generating function for D_m and take advantage of the Chernoff bound to obtain a good estimate of the above probability. Towards this end, we have the following Proposition. Before we go further, we need to address the second difficulty we mentioned above. To remove the coupling, we need to introduce the notion of negatively related for Bernoulli random variables [24, 25].

Definition 3. *The Bernoulli random variables $I_i, i = 1, ..., n$ are said to be negatively related if for each $i \leq n$ there exists random variables J_{ij}, such that the distribution of the random vector $[J_{i1}, J_{i2}, ..., J_{in}]$ is equal to the distribution of the random vector $[I_1, I_2, ..., I_n]$ given that $I_i = 1$, and $J_{ij} \leq I_j$ for $j \neq i$.*

For negatively related random variables, the following theorem holds (Theorem 4 in [25]).

Theorem 4. *Suppose I_i's are negatively related Bernoulli random variables with identical distribution, $i = 1, 2, ..., n$. Let $\tilde{I}_i$, $i = 1, 2, ..., n$ be i.i.d. random variables, where $\tilde{I}_i$ has the same distribution as I_i for all i. Then for any real t, we have*

$$\mathbf{E}\left[e^{t\sum_{i=1}^{n} I_i}\right] \leq \mathbf{E}\left[e^{t\sum_{i=1}^{n} \tilde{I}_i}\right].$$

Roughly speaking, for negatively related Bernoulli random variables, conditioned on the event that one of them is 1, the others are more likely to become small. Correspondingly, conditioned on the event that one of them is 0, the others are more likely to become large. Therefore, when t is positive, the moment generating function is mainly determined by the probability of the sum of all indicator random variables achieving the larger value. The sum of negatively related random variables is less likely to achieve larger value and hence the moment generation function is smaller. For negative t, the moment generating function is mainly determined by the probability of the sum of all indicator random variables achieving the smaller value. The sum of negatively related random variables is also less likely to achieve smaller value and hence the moment generation function is smaller.

One can show that hypergeometric random variables can be viewed as the sum of negatively related Bernoulli random variables (See Example 1 in [25]). We first construct I_i by choosing M neighbors out of $N-1$ peers. For each peer i on the right, let $I_i = 1$ if peer i is chosen as a neighbor, and let $I_i = 0$ otherwise (Note that I_i is not defined for peers on the left). We can then construct J_{ij} as the following. First, set $J_{ij} = I_j$ for all j. Then if $J_{ii} = 0$, in order to make $J_{ii} = 1$, we choose one neighbor k randomly (either from the left or the right), and exchange that neighbor with peer i. If k was on the left, we then let $J_{ii} = 1$. If k was on the right, we then let $J_{ii} = 1$ and $J_{ik} = 0$. Clearly, $\mathbf{J}_i$ has the same distribution as $\mathbf{I}$ given that $I_i = 1$. However, by our construction $J_{ij} \le I_j$ for all $j \neq i$. Hence, $I_i, i = 1, ..., M$ are negatively related. We can now bound the moment generation function of $\sum_{k \in W_m^d} C_{ik}$ by the moment generating functions of the sum of i.i.d. random variables. Towards this end, we have the following Proposition.

Proposition 5. *For any given cut V_k and V_k^c of a network $G(V,E)$, let $\tilde{W}_1$ and $\tilde{W}_2$ be subsets of V_k and V_k^c, respectively. Assume that $|\tilde{W}_1| = q \le k+1$ and $|\tilde{W}_2| = r \le N-k$. Let the upload capacity of each peer $i \in \tilde{W}_1$ be u. For each peer in $\tilde{W}_1$, it chooses M downstream neighbors uniformly and randomly from a given subset $\tilde{V}$ of V that is a superset of $\tilde{W}_2$. Let $\tilde{N} = |\tilde{V}|$. Then the moment generating function of $\sum_{i \in \tilde{W}_1} \sum_{j \in \tilde{W}_2} C_{ij}$ satisfy*

$$\mathbf{E}\left[e^{-\theta \sum_{i \in \tilde{W}_1} \sum_{j \in \tilde{W}_2} C_{ij}}\right] \le \exp\left[Mq\frac{r}{\tilde{N}}\left(e^{-\theta\frac{u}{M}} - 1\right)\right]. \quad (4)$$

Proof. We can write $\sum_{j \in \tilde{W}_2} C_{ij} = L_i \cdot \frac{u}{M}$, where L_i is the number of downstream neighbors of peer i in $\tilde{W}_2$. As mentioned above, peer i select M

downstream neighbors from $\tilde{N}$ different peers. Consider all the potential downstream neighbors $j \in \tilde{W}_2$. Let I_{ij} be the indicator function of the event that peer j is a downstream neighbors of i. Clearly, I_{ij} has a Bernoulli distribution with parameter $M/\tilde{N}$. Moreover, the number of downstream neighbors in $\tilde{W}_2$ would be equal to the summation of all the I_{ij}'s over j, i.e., $L_i = \sum_{j \in \tilde{W}_2} I_{ij}$ and follows a hypergeometric distribution. According to Theorem 4 in [25], if $\tilde{I}_{ij}$, $j \in \tilde{W}_2$ are i.i.d. Bernoulli random variables such that $\tilde{I}_{ij}$ has the same marginal distribution as I_{ij}, we will have, for any real t

$$\mathbf{E}\left[e^{t\sum_{j\in\tilde{W}_2} I_{ij}}\right] \leq \mathbf{E}\left[e^{t\sum_{j\in\tilde{W}_2} \tilde{I}_{ij}}\right]. \tag{5}$$

This means that we could use the moment generating function of a binomial random variable, which is the summation of i.i.d. Bernoulli random variables, to bound the moment generating function of the hypergeometric random variable. Letting $t = -\theta$, we then have, for each $i \in \tilde{W}_1$

$$\begin{aligned}
\mathbf{E}\left[e^{-\theta\sum_{j\in\tilde{W}_2} C_{ij}}\right] =& \mathbf{E}\left[\left(e^{-\theta\sum_{j\in\tilde{W}_2} I_{ij}}\right)^{\frac{u}{M}}\right] \\
\leq& \mathbf{E}\left[\left(e^{-\theta\sum_{j\in\tilde{W}_2} \tilde{I}_{ij}}\right)^{\frac{u}{M}}\right] \\
=& \left(\mathbf{E}\left[e^{-\theta\frac{u}{M}\tilde{I}_{ij}}\right]\right)^r \\
=& \left(1 - \frac{M}{\tilde{N}} + \frac{M}{\tilde{N}} e^{-\theta\frac{u}{M}}\right)^r
\end{aligned} \tag{6}$$

Note that,

$$\begin{aligned}
1 - \frac{M}{\tilde{N}} + \frac{M}{\tilde{N}} e^{-\theta\frac{u}{M}} =& 1 - \frac{M}{\tilde{N}}\left(1 - e^{-\theta\frac{u}{M}}\right) \\
\leq& \exp\left[\frac{M}{\tilde{N}}\left(e^{-\theta\frac{u}{M}} - 1\right)\right].
\end{aligned} \tag{7}$$

where the last inequality is due to $0 \leq \frac{M}{\tilde{N}}\left(1 - e^{-\theta\frac{u}{M}}\right) \leq 1$, and $1 - x \leq e^{-x}$ when $0 \leq x \leq 1$. Therefore, substituting (7) into (6) yields

$$\mathbf{E}\left[e^{-\theta\sum_{j\in\tilde{W}_2} C_{ij}}\right] \leq \exp\left[r\frac{M}{\tilde{N}}\left(e^{-\theta\frac{u}{M}} - 1\right)\right].$$

For different peers in $\tilde{W}_1$, they will select their downstream neighbors independently. Hence, $\sum_{j\in\tilde{W}_2} C_{ij}$ are independent across i. Therefore,

$$\begin{aligned}\mathbf{E}\left[e^{-\theta\sum_{i\in\tilde{W}_1}\sum_{j\in\tilde{W}_2} C_{ij}}\right] &= \left(\mathbf{E}\left[e^{-\theta\sum_{j\in\tilde{W}_2} C_{ij}}\right]\right)^q \\ &\leq \exp\left[\frac{Mrq}{\tilde{N}}\left(e^{-\theta\frac{u}{M}}-1\right)\right].\end{aligned}$$

□

Proposition 5 combined with the Chernoff bound will be frequently used to estimate the probability for a cut to "fail", i.e., when the capacity of a cut is less than $(1-\epsilon)$ its expected capacity. We have the following result for cuts W_m in H_t under the assumption of ON-OFF upload capacity.

Lemma 6. *Let $\epsilon \in (0,1)$. Given that the total number of ON peers in the entire network $Y = y$, the probability that the capacity of the cut D_m is less than $(1-\epsilon)\mathbf{E}[D_m]$ can be bounded by the following,*

$$\begin{aligned}&\mathbf{P}(D_m \leq (1-\epsilon)\mathbf{E}[D_m]|Y=y) \\ \leq&\exp\left[-\left(Mm\frac{y-m+1}{N-1}+M\frac{y-m}{y}\right)\frac{u}{u_s}\frac{\epsilon^2}{2}\right].\end{aligned}$$

Proof. By Chernoff bounds, we have for $\theta > 0$

$$\begin{aligned}&\mathbf{P}(D_m \leq (1-\epsilon)\mathbf{E}[D_m]|Y=y, t \text{ is ON}) \\ \leq&\frac{\mathbf{E}\left[e^{-\theta D_m}|Y=y\right]}{e^{-(1-\epsilon)\theta\mathbf{E}[D_m|Y=y]}} \\ \leq&\mathbf{E}\left[e^{-\theta\sum_{j=1}^{m}\sum_{i=m+1}^{y+1} C_{ji}}\right] e^{\theta(1-\epsilon)m(y-m+1)\frac{u}{N-1}}\,\mathbf{E}\left[e^{-\theta\sum_{i=m+1}^{y+1} C_{si}}\right] e^{\theta(1-\epsilon)(y-m+1)\frac{u_s}{y+1}} \\ =&e^{\phi(\theta)+\phi_s(\theta)},\end{aligned} \tag{8}$$

where

$$\phi(\theta) = \log\mathbf{E}\left[e^{-\theta\sum_{j=1}^{m}\sum_{i=m+1}^{y+1} C_{ji}}\right] + \theta(1-\epsilon)m(y-m+1)\frac{u}{N-1}; \tag{9}$$

$$\phi_s(\theta) = \log\mathbf{E}\left[e^{-\theta\sum_{i=m+1}^{y+1} C_{si}}\right] + \theta(1-\epsilon)(y-m+1)\frac{u_s}{y+1}. \tag{10}$$

Now we apply Proposition 5. Recall that we define a cut on H_t by dividing peers into sets W_m and W_m^c. We could also view W_m and W_m^c as subsets of some cut V_k and V_k^c of network G. We need to exclude the server from W_m since it has a different upload capacity. For each peer in $W_m \backslash s$, it will choose M downstream neighbors randomly from the entire network. Hence, $\tilde{V} = V$. According to proposition 5, we have $q = |W_m \backslash s| = m$, $r = |W_m^c| = y - m + 1$ and $|\tilde{V}| = N$. Therefore, using (4), we have,

$$\begin{aligned}
\phi(\theta) \leq & \log\left\{\exp\left[Mm\frac{y+1-m}{N-1}\left(e^{-\theta\frac{u}{M}} - 1\right)\right]\right\} \\
& + \theta(1-\epsilon)(y+1-m)\frac{u}{N-1} \\
= & M\frac{y+1-m}{N-1}\left(e^{-\theta\frac{u}{M}} - 1\right) + \theta(1-\epsilon)(y+1-m)\frac{u}{N-1} \\
= & \frac{1}{N-1}\left[M\left(e^{-\theta\frac{u}{M}} - 1\right) + \theta(1-\epsilon)u\right](y+1-m).
\end{aligned}$$

Note that the server only choose neighbors from the $y+1$ ON peers, $|\tilde{V}| = y+1$. Using similar techniques, for the server, we can bound $\phi_s(\theta)$ by

$$\phi_s(\theta) \leq \frac{1}{y+1}\left[M\left(e^{-\theta\frac{u_s}{M}} - 1\right) + \theta(1-\epsilon)u_s\right](y+1-m).$$

Define

$$\begin{aligned}
\tilde{\phi}(\theta) &\triangleq M\left(e^{-\theta\frac{u}{M}} - 1\right) + \theta(1-\epsilon)u; \\
\tilde{\phi}_s(\theta) &\triangleq M\left(e^{-\theta\frac{u_s}{M}} - 1\right) + \theta(1-\epsilon)u_s.
\end{aligned}$$

The $\phi(\cdot)$ and $\phi_s(\cdot)$ can be written as

$$\begin{aligned}
\phi(\theta) &\leq \frac{1}{N-1}\tilde{\phi}(\theta)m(y+1-m); \\
\phi_s(\theta) &\leq \frac{1}{y+1}\tilde{\phi}_s(\theta)(y+1-m).
\end{aligned}$$

Let $\tilde{\phi}_{\min}$ and $\tilde{\phi}_{s,\min}$ be the minimum of $\tilde{\phi}(\theta)$ and $\tilde{\phi}_s(\theta)$ respectively, over $\theta > 0$. It is easy to see $\tilde{\phi}_{\min} = \tilde{\phi}_{s,\min} < 0$. Also since $\tilde{\phi}$ and $\tilde{\phi}_s$ is convex on $\theta > 0$, these minimum is attainable. Let $\theta_{\min}$ and $\theta_{s,\min}$ be the minimizer respectively. We must have

$$\tilde{\phi}_s(\theta_{s,\min}) = \tilde{\phi}_{s,\min} = \tilde{\phi}_{\min} \leq \tilde{\phi}(\theta_{s,\min}). \tag{11}$$

One can show that

$$\theta_{s,\min} = -\frac{M}{u_s}\log(1-\epsilon). \tag{12}$$

Note that for $0 < a < 1$ and $0 \le x \le 1$, we have $(1-x)^a \le 1-ax$ since $(1-x)^a$ is concave and its derivative at 0 is $-a$. Moreover, for $0 \le x \le 1$, one can see that $(1-x)\log(1-x) \ge x^2/2-x$ by checking $\frac{\mathrm{d}}{\mathrm{d}x}(1-x)\log(1-x)-(x^2/2-x) = -\log(1-x)-x \ge 0$ and $(1-x)\log(1-x) = x^2/2-x$ when $x=0$. Then, substituting (12) into (11) and using the above relationship, we have

$$\begin{aligned}
\tilde{\phi}(\theta_{s,\min}) =& M\left[(1-\epsilon)^{\frac{u}{u_s}}-1\right] - M\frac{u}{u_s}(1-\epsilon)\log(1-\epsilon)\\
=& M\left\{\left[(1-\epsilon)^{\frac{u}{u_s}}-1\right] - \frac{u}{u_s}(1-\epsilon)\log(1-\epsilon)\right\}\\
\le& M\left[1-\frac{u}{u_s}\epsilon - 1 - \frac{u}{u_s}\left(\frac{\epsilon^2}{2}-\epsilon\right)\right]\\
=& -M\frac{u}{u_s}\frac{\epsilon^2}{2}.
\end{aligned}$$

We then have

$$\begin{aligned}
& m\phi(\theta_{s,\min}) + \phi_s(\theta_{s,\min})\\
\le& \frac{1}{N-1}\tilde{\phi}(\theta_{s,\min})m(y+1-m) + \frac{1}{y+1}\tilde{\phi}_s(\theta_{s,\min})(y+1-m)\\
\le& \frac{1}{N-1}\tilde{\phi}(\theta_{s,\min})m(y+1-m) + \frac{1}{y+1}\tilde{\phi}(\theta_{s,\min})(y+1-m)\\
\le& -\left(m\frac{y+1-m}{N-1} + \frac{y+1-m}{y+1}\right)M\frac{u}{u_s}\frac{\epsilon^2}{2}.
\end{aligned}$$

Since (8) holds for any $\theta > 0$, letting $\theta = \theta_{s,\min}$ yields

$$\begin{aligned}
& \mathbf{P}(D_m \le (1-\epsilon)\mathbf{E}[D_m] | Y = y, t \text{ is ON})\\
\le& \exp(m\phi(\theta_{s,\min}) + \phi_s(\theta_{s,\min}))\\
\le& \exp\left[-\left(Mm\frac{y+1-m}{N-1} + M\frac{y+1-m}{y+1}\right)\frac{u}{u_s}\frac{\epsilon^2}{2}\right].
\end{aligned}$$

Similarly, one can show that if the destination t is OFF, we have

$$\begin{aligned}
& \mathbf{P}(D_m \le (1-\epsilon)\mathbf{E}[D_m] | Y = y, t \text{ is OFF})\\
\le& \exp\left[-\left(Mm\frac{y+1-m}{N-1} + M\frac{y-m}{y}\right)\frac{u}{u_s}\frac{\epsilon^2}{2}\right].
\end{aligned}$$

Since $\frac{y+1-m}{y+1} \geq \frac{y-m}{y}$, we have

$$\begin{aligned}&\mathbf{P}(D_m \leq (1-\epsilon)\mathbf{E}[D_m]|Y=y, t \text{ is ON})\\ \leq &\mathbf{P}(D_m \leq (1-\epsilon)\mathbf{E}[D_m]|Y=y, t \text{ is OFF})\end{aligned}$$

Hence,

$$\begin{aligned}&\mathbf{P}(D_m \leq (1-\epsilon)\mathbf{E}[D_m]|Y=y)\\ \leq &\mathbf{P}(D_m \leq (1-\epsilon)\mathbf{E}[D_m]|Y=y, t \text{ is OFF})\\ \leq &\exp\left[-\left(Mm\frac{y+1-m}{N-1}+M\frac{y-m}{y}\right)\frac{u}{u_s}\frac{\epsilon^2}{2}\right].\end{aligned}$$

□

Lemma 6 gives us an upper bound on the probability that the capacity D_m of a cut W_m is less than $1-\epsilon$ times its mean conditioned on the event that total number of ON peers $Y=y$. Note that $Mm\frac{y-m+1}{N}$ is the average number of edges from peers in W_m to peers in W_m^c, while $M\frac{y-m}{y}$ is a lower bound on the average number of edges from the server to peers in W_m^c. Hence, the upper bound in Lemma 6 decreases exponentially if the average number of edges increases. Furthermore, since the average number of edges is proportional to M, the upper bound also decrease exponentially if M increases. The following lemma then bounds the effect of all cuts separating s and t. Note that for each value of m, there are $\binom{Y}{m}$ possible cuts W_m. Due to symmetry, the capacity of all $\binom{Y}{m}$ cuts has the same distribution.

Lemma 7. *Define $\tilde{\mathcal{B}}_m$ to be the event $\{D_m \leq (1-\epsilon)C_f$ for any cut W_m among the $\binom{Y}{m}$ cuts $\}$. The probability of the union of all $\tilde{\mathcal{B}}_m$'s can be bounded by*

$$\mathbf{P}\left(\bigcup_{m=0}^{Y}\tilde{\mathcal{B}}_m\right) \leq O(\exp(-\epsilon'^2p^2N)) + \beta^{\gamma}\left[\left(1+p\beta^{\frac{\gamma}{2}}\right)^{N-1}\right]. \tag{13}$$

More specifically, we can separate the union bound into to two parts:

$$\mathbf{P}\left(\bigcup_{m=0}^{Y-1}\tilde{\mathcal{B}}_m\right) \leq O(\exp(-\epsilon'^2p^2N)) \tag{14}$$

$$+\beta^{\gamma}\left[\left(1+p\beta^{\frac{\gamma}{2}}\right)^{N-1}-1\right], \tag{15}$$

$$\mathbf{P}\left(\tilde{\mathcal{B}}_Y\right) \leq O(\exp(-\epsilon'^2p^2N)) + \beta^{\gamma}. \tag{16}$$

where $\epsilon' = 1 - \sqrt{1-\epsilon}$, $\gamma = (1-\epsilon')p$ *and* $\beta \triangleq \exp(-M\frac{u}{u_s}\frac{\epsilon'^2}{2})$.

Proof. Choose a constant $\gamma = (1-\epsilon')p$. We then have

$$\begin{aligned}
&\mathbf{P}\left(\bigcup_{m=0}^{Y-1} \tilde{\mathcal{B}}_m\right) \\
\leq &\sum_{y=0}^{\lceil \gamma N \rceil - 1} \binom{N-1}{y-1} p^y (1-p)^{N-1-y} \mathbf{P}\left(\bigcup_{m=0}^{y} \tilde{\mathcal{B}}_m \middle| Y = y\right) \\
&\quad + \sum_{y=\lceil \gamma N \rceil}^{N-1} \binom{N-1}{y-1} p^y (1-p)^{N-1-y} \mathbf{P}\left(\bigcup_{m=0}^{y} \tilde{\mathcal{B}}_m \middle| Y = y\right).
\end{aligned}$$

The first term satisfies,

$$\begin{aligned}
&\sum_{y=0}^{\lceil \gamma N \rceil - 1} \binom{N-1}{y-1} p^y (1-p)^{N-1-y} \mathbf{P}\left(\bigcup_{m=0}^{y} \tilde{\mathcal{B}}_m \middle| Y = y\right) \\
\leq &\mathbf{P}(Y < \lceil \gamma N \rceil - 1) \\
\leq &\exp\left(-2\frac{(p(N-1) - (\lceil \gamma N \rceil - 1))^2}{N-1}\right) = O(\exp(-\epsilon'^2 p^2 N)),
\end{aligned}$$

where the last inequality follows from Hoeffding's inequality [26]. For the second term, recall that $D_m \geq \sqrt{1-\epsilon}\mathbf{E}[D_m]$ implies $D_m \geq (1-\epsilon)C_f$. For $m = 0$, $D_m = u_s \geq C_f$. Therefore the probability $\mathbf{P}(\tilde{\mathcal{B}}_0 | Y = y)$ is always 0.

We can then take the summation from $m = 1$. We have,

$$\begin{aligned}
&\mathbf{P}\left(\bigcup_{m=0}^{y-1} \tilde{\mathcal{B}}_m \middle| Y = y\right) \\
=&\sum_{m=0}^{y-1} \binom{y}{m} \mathbf{P}\left(D_m \leq \sqrt{(1-\epsilon)}\mathbf{E}[D_m] \middle| Y = y\right) \\
\leq&\sum_{m=1}^{y-1} \binom{y}{m} e^{-\left(Mm\frac{y-m+1}{N-1} + M\frac{y-m}{y}\right)\frac{u}{u_s}\frac{\epsilon'^2}{2}} \\
\leq&\sum_{m=1}^{y-1} \binom{y}{m} e^{-\left(Mm\frac{y-m+1}{N-1} + M\gamma\frac{y-m}{y}\right)\frac{u}{u_s}\frac{\epsilon'^2}{2}} \\
\leq&e^{-\gamma M\frac{u}{u_s}\frac{\epsilon'^2}{2}} \\
&\quad \times \sum_{m=1}^{y-1} \binom{y}{m} e^{-\left(Mm\gamma\frac{y-m+1}{y} + M\gamma\frac{-m}{y}\right)\frac{u}{u_s}\frac{\epsilon'^2}{2}}
\end{aligned} \tag{17}$$

Let $\beta \triangleq \exp(-M\frac{u}{u_s}\frac{\epsilon'^2}{2})$. Then,

$$(17) = \beta^\gamma \sum_{m=1}^{y-1} \binom{y}{m} \exp\left[-Mm\gamma\frac{u}{u_s}\frac{\epsilon'^2}{2}\frac{y-m+1-1}{y}\right] \tag{18}$$

$$\begin{aligned}
=&\beta^\gamma \sum_{m=1}^{y-1} \binom{y}{m} \beta^{m\gamma\frac{y-m}{y}} \\
=&\beta^\gamma \left[\sum_{m=1}^{\lfloor y/2 \rfloor} \binom{y}{m} \beta^{m\gamma\frac{y-m}{y}} + \sum_{m=\lfloor y/2 \rfloor+1}^{y-1} \binom{y}{m} \beta^{m\gamma\frac{y-m}{y}}\right] \\
\leq&\beta^\gamma \left[\sum_{m=1}^{\lfloor y/2 \rfloor} \binom{y}{m} \beta^{\frac{m\gamma}{2}} + \sum_{m=\lfloor y/2 \rfloor+1}^{y-1} \binom{y}{m} \beta^{\frac{(y-m)\gamma}{2}}\right] \\
\leq&\beta^\gamma \left[\left(1+\beta^{\frac{\gamma}{2}}\right)^y - 1 + \left(\beta^{\frac{\gamma}{2}}+1\right)^y - 1\right] \\
=&2\beta^\gamma \left[\left(1+\beta^{\frac{\gamma}{2}}\right)^y - 1\right].
\end{aligned} \tag{19}$$

We then have

$$\sum_{y=\lceil \gamma N \rceil}^{N-1} \binom{N-1}{y} p^y (1-p)^{N-1-y} \mathbf{P}\left(\bigcup_{m=0}^{y-1} \tilde{\mathcal{B}}_m \middle| Y = y \right)$$
$$\leq \sum_{y=\lceil \gamma N \rceil}^{N-1} \binom{N-1}{y} p^y (1-p)^{N-1-y}$$
$$\times 2\beta^{\gamma} \left[\left(1+\beta^{\frac{\gamma}{2}}\right)^y - 1 \right]$$
$$\leq \sum_{y=0}^{N-1} \binom{N-1}{y} 2\beta^{\gamma} \left(p(1+\beta^{\frac{\gamma}{2}})\right)^y (1-p)^{N-1-y}$$
$$= 2\beta^{\gamma} \left[\left(1+p\beta^{\frac{\gamma}{2}}\right)^{N-1} - 1 \right].$$

Then, plugging in the value of β will yields (15). For $m = y$, we have

$$\mathbf{P}\left(\tilde{\mathcal{B}}_y \middle| Y = y \right)$$
$$= \mathbf{P}\left(D_y \leq \sqrt{(1-\epsilon)} \mathbf{E}[D_y] \middle| Y = y \right)$$
$$\leq e^{-\left(My\frac{1}{N-1}\right)\frac{u}{u_s}\frac{\epsilon'^2}{2}}$$
$$= \beta^{\gamma}$$

(16) then follows trivially. □

We could now prove Theorem 1.

Proof of Theorem 1. According to Proposition 2 and Lemma 7, for any peer t, the minimum cut from the source s to t can be bounded by

$$\mathbf{P}\left(C_{\min}(s \to t) \leq (1-\epsilon) C_f \right)$$
$$\leq \mathbf{P}\left(C_{\min, H_t}(s \to t) \leq (1-\epsilon) C_f \right)$$
$$= \mathbf{P}\left(\bigcup_{m=0}^{Y} \tilde{\mathcal{B}}_m \right)$$
$$\leq \mathbf{P}\left(\bigcup_{m=0}^{Y} \{ D_m \leq (1-\epsilon') \mathbf{E}[D_m] \} \right)$$
$$\leq 2\beta^{\gamma} \left(1+p\beta^{\frac{\gamma}{2}}\right)^{N-1} + O(\exp(-\epsilon^2 p^2 N)).$$

Note that by assumption $M = \alpha \log(N)$. For any $\epsilon > 0$, and $\epsilon' = 1 - \sqrt{1-\epsilon}$, choose a sufficiently large α such that

$$\epsilon' = \sqrt{\frac{4d}{\alpha\gamma}\frac{u_s}{u}}.$$

We then have, for large N

$$\beta^\gamma = \exp(-M\gamma\frac{u}{u_s}\frac{\epsilon'^2}{2}) = \exp(-2d\log(N)) = 1/N^{2d}.$$

Hence, the minimum cut statistics,

$$\begin{aligned}&\mathbf{P}\left(C_{\min}(s \to t) \le (1-\epsilon')C_f\right)\\ \le&\frac{1}{N^{2d}}2\left(1 + pO(\frac{1}{N^d})\right)^{N-1} = O\left(\frac{1}{N^{2d}}\right).\end{aligned}$$

Thus, the min-min cut will satisfy

$$\begin{aligned}&\mathbf{P}\left(C_{\min-\min} \le (1-\epsilon)C_f\right)\\ \le&\sum_{t=1}^{N}\mathbf{P}\left(C_{\min}(s \to t) \le (1-\epsilon)C_f\right)\\ \le&O\left(\frac{1}{N^{2d}}\right)\cdot N = O\left(\frac{1}{N^{2d-1}}\right).\end{aligned}$$

□

We remark on several implications of Theorem 1. First, Theorem 1 not only shows that pure random selection is sufficient to achieve close-to-optimal streaming capacity as long as each peer has $\Omega(\log N)$ downstream neighbors, it also reveals important insights on the significance of different types of cuts. To see this, note that if we choose α as in the proof such that $\beta^\gamma = O(1/N^{2d})$, we have

$$\begin{aligned}&\mathbf{P}\left(\bigcup_{m=0}^{Y-1}\tilde{\mathcal{B}}_m\right) \le 2\beta^\gamma\left[\left(1 + p\beta^{\frac{\gamma}{2}}\right)^{N-1} - 1\right]\\ =&O(1/N^{2d})O(e^{1/N^{d-1}} - 1) = o(1/N^{2d}).\end{aligned}$$

On the other hand, we have $\mathbf{P}\left(\tilde{\mathcal{B}}_Y\right) = O(1/N^{2d})$. Hence, the probability that the last cut (the W_Y and W_Y^c cut) fails is much larger than the probability that any other cuts fails. Thus, for each peer t, the min-cut from the source to t is mainly determined by C_t^R (recall that C_t^R is the total capacity received by peer t directly from its upstream neighbors, which is also the capacity of the last cut).

The above insight suggests that, if we want to design improved distributed control algorithms for P2P streaming systems, we may want to focus on improving the capacity C_t^R at the last hop. Note that one of the main reasons for C_t^R to fall below its mean value is the imbalance of C_t^R across t. More specifically, some peers t may have a larger number of upstream peers, and hence have a larger-than-average value of C_t^R, while other peers may have a smaller-than-average value of C_t^R. Such imbalance will lead to an increase in the probability that some peers have low streaming rates. Based on this intuition, we can use the following slightly-modified algorithm. Suppose that a peer already receives enough capacity from its direct upstream neighbors (i.e., $C_t^R > C_f$), it is very likely that this peer will also have a min-cut from the source that is larger than C_f. We can then take away some upstream neighbors from this peer and allocate them to other peers. Intuitively, this modification will help to balance the values of C_t^R. Simulation results shows that this "adaptive" algorithm indeed reduces the "failure" probability compared to the pure random algorithm when the network size is the same.

Theorem 1 also reveals important relationship between the number of neighbors required and key system parameters. For example, if we require a better performance (smaller ϵ or larger d) or have fewer ON peers (smaller p), the number of downstream neighbors needed by each peer will increase. Specifically, according to the proof, we need $\alpha \geq \frac{4du_s}{\gamma u \epsilon'^2}$. If we require higher streaming rate or faster convergence rate, i.e., ϵ is smaller (consequently ϵ' is smaller) or d is larger, we will need a larger α. If the probability that a peer is ON is reduced, i.e., p is reduced, we will also need a larger α.

3 Multi-Channel P2P Networks

In this section, we will extended our analysis to a multi-channel network containing J different channels. We are interested in the scenarios where the upload capacity from one channel can be used to "help" the other channel.

For single-channel networks, the streaming capacity of the network is a real number. However, for a multi-channel network, the streaming rate requirements of different channels can be different. Let R_j be the streaming rate requirement of channel j, $j = 1, 2, ..., J$. There is clearly a tradeoff between the values of R_j in different channels, i.e., with finite upload capacity, increasing R_j for one channel j must be at the cost of reducing R_k of another channel k. To capture this tradeoff, we define the capacity region Λ as the set of streaming rate vectors $\mathbf{R} = [R_1, R_2, ..., R_J]^T$ such that whenever $\mathbf{R} \in \Lambda$, each user in the network will receive enough capacity to view its own channel of interest with high probability. Intuitively, if the upload capacity of the users and the server is the only bottleneck in the network, the best we can do is to support those rate vectors $\mathbf{R}$ such that the summation of all the demands is equal to all the supply. Hence, the largest possible capacity region will be no larger than

$$\Lambda_m = \left\{ \mathbf{R} \middle| \sum_{j=1}^{J} N_j R_j \leq \sum_{i \in V} \mathbf{E}[U_i], \sum_{j=1}^{J} R_j \leq u_s \right\}, \tag{20}$$

where N_j is the number of peers that are viewing channel j. With this largest possible capacity region in mind, we are going to present our multi-channel algorithm and we will show that our algorithm could achieve the following close-to-optimal capacity region with high probability when $N \to \infty$:

$$(1 - \epsilon)\Lambda_m = \{(1 - \epsilon)\mathbf{R} | \mathbf{R} \in \Lambda_m\}.$$

3.1 System Model

We consider a multi-channel P2P networks with N peers, one source s and J different channels. We will reuse the same notations as in the single-channel section. Let $\mathcal{J} = \{1, 2, ..., J\}$ denote the set of all the channels. We have $|\mathcal{J}| = J$. Denote the set of all peers that are viewing channel j as $\mathcal{N}_j$ and we have $|\mathcal{N}_j| = N_j$. We assume that $N_j = p_j N$, where p_j is a fixed constant and represents the fraction of peers interested in viewing channel j. We refer to the peers in $\mathcal{N}_j$ as the *normal peers* of channel j. Let the set of all peers (including the source) be V. Obviously, $V = \left(\bigcup_{j \in \mathcal{J}} \mathcal{N}_j\right) \cup \{s\}$ and $|V| = N + 1 = \sum_{j \in \mathcal{J}} N_j + 1$. Assume that the server allocates $u_{s,j}$ amount of capacity to channel j and $\sum_{j \in \mathcal{J}} u_{s,j} = u_s$. We also assume that each peer has an ON-OFF upload capacity, i.e., U_i is i.i.d. with the distribution

$\mathbf{P}(U_i = u) = p$ and $\mathbf{P}(U_i = 0) = 1 - p$. Each node (including the server) will still have M downstream neighbors. We will describe how these neighbors are chosen later.

In multi-channel P2P networks, the popularity N_j and streaming rate requirement R_j can vary from channel to channel. If we let the peers in the same channel to form a sub-network, for some channel j the streaming rate requirement R_j may exceed the maximum streaming capacity, while for other channels the upload capacity of the peers is more than needed. More specifically, for any $\epsilon \in (0, 1)$, let

$$\mathcal{I}_\epsilon = \{j \in \mathcal{J} | R_j > (1-\epsilon)up + u_{s,j}/N_j, \}$$
$$\mathcal{S}_\epsilon = \{j \in \mathcal{J} | R_j \leq (1-\epsilon)up + u_{s,j}/N_j.\}$$

We call the channels in $\mathcal{I}_\epsilon$ insufficient channels, and the channels in $\mathcal{S}_\epsilon$ sufficient channels. With a multi-channel network, we can let some peers in the sufficient channels help another insufficient channel, and both channels may be able to achieve their own streaming rate requirements [16]. We call these peers that are helping the other channel j the "helpers" for channel j. A helper will not allocate its upload capacity to its own channel but will contribute to the channel it is helping. Let the number of helpers for channel j be H_j. For convenience, we allow H_j be positive, negative or 0. $H_j > 0$ means that there are H_j helpers that are helping to distribute the content of an insufficient channel j. $H_j < 0$ means that a sufficient channel j is providing $|H_j|$ helpers to assist other channels. We emphasize that a helper from a sufficient channel j to help an insufficient channel k still needs to receive a streaming rate R_j for its interested content of channel j.

3.2 Algorithm

We will now present the scheme for allocating the capacity of the server for each channel and for choosing the number of helpers that each channel have. Fix $\epsilon \in (0, 1)$. In the following discussions, we assume that $\mathbf{R} \in (1-\epsilon)\Lambda_m$. The server will allocate the capacity for channel j according to the following equation

$$u_{s,j} = R_j/(1-\epsilon). \tag{21}$$

Each channel will determine the number of helpers it needs by the following equation

$$H_j = \left\lfloor \frac{N_j R_j}{(1-\epsilon)u} - \frac{u_{s,j}}{u} - pN_j \right\rfloor. \tag{22}$$

We require that $\sum_{j\in\mathcal{J}} H_j \leq 0$, i.e., the total number of helpers provided by the sufficient channels must be no smaller than the total number of helpers demanded by insufficient channels. We can check that (22) satisfies this condition (see Lemma 8 for details). Note that according to (22), a channel with higher streaming rate requirement will require more helpers. Let $K = \max_{j\in J} \frac{R_j}{(1-\epsilon)u} - p$. Note that for any insufficient channel j, we must have $H_j \leq KN_j$. For an insufficient channel j, denote the set of all helpers that are helping channel j as $\mathcal{H}_j$. Recall that each ON peer can have M downstream neighbors, which correspond to M links, each with capacity u/M. Each peer i in $\mathcal{N}_j$ will reserve K downstream links (out of a total of M links) to allow helpers to connect to peer i. Each helper i will choose one connections among $K\mathcal{N}_j$ reserved links uniformly randomly, and the owner of that reserved link becomes an upstream neighbor of that helper. In other words, each helper will try to find one and only one upstream neighbor in $\mathcal{N}_j$ and no peers support more than K downstream helpers. Each peer in $\mathcal{N}_j$ will choose $M - K$ downstream neighbors from $\mathcal{N}_j$ and each helper will also choose M downstream neighbors from $\mathcal{N}_j$. Note that there will be no connection between helpers, which avoids loops among helpers. Our multi-channel algorithm preserves the desirable features of our single-channel algorithms. The random peer-selection is simple, robust, and mesh-based.

3.3 Performance Analysis

Next we will provide the analysis of the capacity region of our algorithm. We next consider the asymptotic behavior of the system as $N \to \infty$ and $N_j = p_j N$ for some fixed value of $p_j, j = 1, ..., J$. We will show that as long as $M = \Omega(\log N)$, with high probability our algorithm can achieve the capacity region of $(1-\epsilon)\Lambda_m$, where Λ_m is the optimal capacity region given by (20). We start with a result of the performance bound on each channel, and then use that result to analyze the capacity region.

For each channel j, denote the set containing the server, the peers in $\mathcal{N}_j$ and the helpers in $\mathcal{H}_j$ as $\mathcal{V}_j$, i.e., $\mathcal{V}_j = \{s\} \cup \mathcal{N}_j \cup \mathcal{H}_j$. Let the subnetwork that contains all the nodes in $\mathcal{V}_j$ and the links between them as $\mathcal{G}_j$. We have on average pN_j native ON peers and H_j helpers that contribute their bandwidth for uploading. There are N_j peers that require the full streaming rate. Therefore, even under a complete-network assumption, the maximum streaming rate for each channel j will be

$$C_{f,j} \triangleq \min \left\{ u_{s,j}, \frac{pN_j + H_j}{N_j} u + \frac{u_{s,j}}{N_j} \right\}. \tag{23}$$

Note that if $N_j = p_j N$, under the assumption of ON-OFF upload capacity, the set of Λ_m in (20) can be written as

$$\Lambda_m = \left\{ \mathbf{R} \left| \sum_{j=1}^{J} p_j R_j \le up + \frac{u_s}{N}, \sum_{j=1}^{J} R_j \le u_s \right. \right\},$$

which is independent of N. We will adopt this definition of Λ_m in the rest of the paper. Assume that $\mathbf{R} \in (1-\epsilon)\Lambda_m$.

Similar to the single-channel P2P network, our algorithm could achieve a close-to-optimal streaming rate for each channel. We first introduce the following lemma which reveals some important properties of our algorithm.

Lemma 8. *Assume that $N_j = p_j N$. Given any $\epsilon > 0$ and $\mathbf{R} \in (1-\epsilon)\Lambda_m$, let $u_{s,j}$ be given by* (21) *and H_j be chosen as* (22)*. Then for any ϵ' such that $\epsilon < \epsilon' < 1$, there exist N_0 and $0 < \eta < 1$ such that if $N \ge N_0$ we have*

1) $H_j = \Theta(N)$ or 0; (24)

2) If $H_j < 0$ for all $N > N_0$, we have $|H_j| \le \eta p N_j$; (25)

3) and $R_j \le (1-\epsilon')C_{f,j}$; (26)

4) $\sum_{j \in \mathcal{J}} H_j \le 0.$ (27)

This lemma tell us: 1) The number of helpers H_j of either sufficient channel or insufficient channel has the same order as N, unless this channel cannot provide any helper and do not require any helper from other channels. 2) For a sufficient channel j, the number of helpers that it provides is less than pN_j, which is the expected number of ON peers in channel j. Consequently, with high probability, for any sufficient channel, there are enough ON peers to become helpers. 3) For any $\epsilon' > \epsilon$, there exists a sufficiently large N such that the target streaming rate R_j is less than $(1-\epsilon')$ times the maximum streaming capacity of channel j. 4) The sum of H_j is no greater than 0, i.e., the total number of helpers required by insufficient channel is less than the total number of helpers provided by sufficient channel.

Then the follow theorem holds for each channel under our multi-channel algorithm.

Theorem 9. *Fix $\epsilon \in (0,1)$. Assume that p_j, R_j are given and $H_j = \Theta(N)$ for large N. In addition, assume that for any j such that $H_j < 0$, there exist $\eta < 1$ such that $|H_j| \leq \eta p N_j$. Then for any channel j, any $\epsilon' > \epsilon$ and $d > 1$, there exists α_j and $N_{0,j}$ such that for any $M = \alpha_j log(N)$ and $N > N_{0,j}$ the probability for the min-min cut of channel j to be smaller than or equal to $(1-\epsilon')C_{f,j}$ is bounded by*

$$\mathbf{P}\left(C_{\min-\min}(s \to \mathcal{N}_j) \leq (1-\epsilon')C_{f,j}\right) \leq O\left(\frac{1}{N^{2d-1}}\right).$$

We will prove this Lemma 8 and Theorem 9 in the appendix. Theorem 9 provides a performance bound for each channel. The result is similar in flavor to the single-channel case. The choice of α_j is also very similar (see the remark at the end of Appendix B), i.e., the higher the streaming rate is (smaller ϵ) and the smaller the ON probability p is, the larger α_j is required to achieve faster convergence rate (larger d). According to Lemma 8, with the choice of H_j in (22), let ϵ' be a small constant such that $0 < \epsilon' < \epsilon$. For any $j \in \mathcal{J}$, $(1-\epsilon')C_{f,j} \geq R_j$, we can then conclude that

$$\begin{aligned} &\mathbf{P}\left(C_{\min-\min}(s \to \mathcal{N}_j) \leq R_j\right) \\ \leq &\mathbf{P}\left(C_{\min-\min}(s \to \mathcal{N}_j) \leq (1-\epsilon')C_{f,j}\right) \leq O\left(\frac{1}{N^{2d-1}}\right). \end{aligned}$$

With Theorem 9 and Lemma 8, we are able to show the capacity region of our algorithm. Theorem 10 summarizes the final result on the capacity region of our algorithm.

Theorem 10. *For any $\epsilon \in (0,1)$, $d > 1$ and $\mathbf{R} \in (1-\epsilon)\Lambda_m$, choose H_j as (22) for $j = 1, 2, ..., J$. There exists α and N_0 such that for any $M = \alpha log(N)$ and $N > N_0$, the following holds*

$$\mathbf{P}\left(C_{\min-\min}(s \to \mathcal{N}_j) \leq R_j, \text{ for some } j\right) \leq O\left(\frac{1}{N^{2d-1}}\right).$$

We omit the proof of this theorem since it follows trivially from Theorem 9 and Lemma 8. We see that $\Omega(\log N)$ neighbors are again sufficient for achieving a close-to-optimal streaming capacity with high probability when $N \to \infty$.

4 Simulation

In this section we provide simulation results to verify our analytical results in previous sections. We first simulate a single-channel P2P network with $N = 10000$ peers and one server. Each user has a ON-OFF upload capacity with ON probability $p = 0.5$. When a user is ON, it will contribute $u = 10$ amount of upload capacity. The server has a capacity of $u_s = 20$. The optimal streaming capacity would be $C_f = 5.002$. We vary the number of downstream neighbors of each user from $10 \log N = 90$ to $70 \log N = 630$, which corresponds to between 0.9% and 6.3% of the total number of pees N. For each choice of the number of downstream neighbors, we generate the network 100 times. During each iteration all users select their downstream neighbors randomly as described in section 2.2, and we use the algorithm in [27] to find the min-min cut from the source to all the users and compare it with $(1-\epsilon)C_f$, where $\epsilon = 0.3$. We count the number of times that the min-min cut of the network is larger than $(1-\epsilon)C_f$ and plot the probability for that to happen as the number of downstream neighbors of each peer varies. We also simulate the adaptive algorithm described at the end of section 2.4 where each peer only selects those peers who have not received enough capacity as its downstream neighbors. The result is shown as fig 2. We see that using pure random selection, when the number of downstream neighbors of each peer is more than $40 \log N = 360$ (3.6% of N), the probability that the system could sustain a streaming rate higher than 70% of the optimal streaming capacity is greater than 0.9. For the adaptive algorithm, the same performance is achieved when each peer only have $30 \log N = 270$ (2.7% of N) downstream neighbors. We see that by improving the capacity of the last cut, the performance of the system is also improved. We caution, however, that when the peer-selection is significantly different from the baseline random algorithm in Section 2.2, Theorem 1 will no longer apply. Hence, it remains an open question as to how to design hybrid schemes that adaptively improve the capacity in the last hop, while retaining the robustness of a random peer-selection scheme at the same time. We leave this question for future work.

Next we simulate a multi-channel P2P network with $N = 10000$ peers and 2 channels. We use the same setting as the single-channel simulation for the upload capacity for all ON peers, the capacity of the server, and the probability for a peer to be ON. We set $N_1 = 4000$ and $N_2 = 6000$. We choose a streaming rate vector $\tilde{\mathbf{R}} = [15/2, 10/3]^T$ in Λ_m and let our target streaming rate vector be $\mathbf{R} = 0.7\tilde{\mathbf{R}}$ (i.e., $\epsilon = 0.3$). Channel 1 will become an

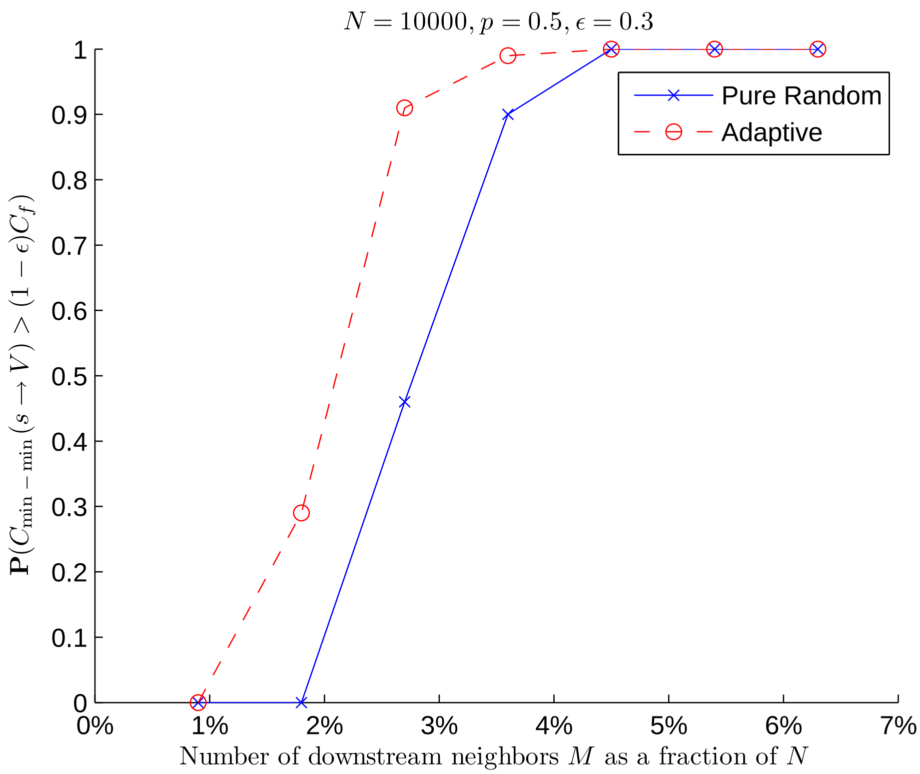

Figure 2: Single-Channel: The Success Probability versus The Number of Downstream Neighbors

insufficient channel and channel 2 will be a sufficient channel. For channel 1, we need 1000 helpers and channel 2 could provide 1000 helpers. In this case, each normal peer in channel 1 needs to reserve 1 link for helpers. We plot the probability that the streaming rate of each channel j is greater than its target streaming rates R_j and the probability that both channels 1 and 2 sustain a streaming rate greater than their corresponding target streaming rate R_1 and R_2, respectively, as the number of downstream neighbors of each peer varies. The result is shown in fig 3. We see that the performance of sufficient channel is worse than the insufficient channel. The reason is that there is only 6000 peers in the sufficient channel, and on average there are 3000 ON peers. However, 1000 of the ON peers is helping the insufficient channel. Hence, there are only 2000 ON peers left in the sufficient channel, which is equivalent to having an ON probability of 1/3. Hence, the network size and the ON probability of the sufficient channel are both smaller. We will then need a larger number of downstream neighbors to achieve the same success probability.

5 Conclusion

In this paper, we study the streaming capacity of sparsely-connected P2P networks. We show that even with a random peer-selection algorithm and uniform rate allocation, as long as each peer maintains $\Omega(\log N)$ downstream neighbors, the system can achieve close-to-optimal streaming capacity with high probability when the network size is large. We then extend our analysis to multi-channel P2P networks, and we let "helpers" from channels with excessive upload capacity to help peers in channels with insufficient upload capacity. We show again that we can achieve a close-to-optimal streaming capacity-region by letting each peer uniformly randomly select $\Omega(\log N)$ neighbors from either the peers in the same channel or from the helpers.

These results provide important new insights on the streaming capacity of large P2P network with sparse topology. In future work, we plan to study how to improve the peer-selection and rate-allocation algorithm to further optimize the streaming capacity. We note that although our analytical results show that having $\Omega(\log N)$ neighbors is sufficient to achieve close-to-optimal streaming capacity with high probability, our simulation results indicate that the actual number of peers required can still be fairly large. A natural next step is to improve the constant in front of the $\Omega(\log N)$ result

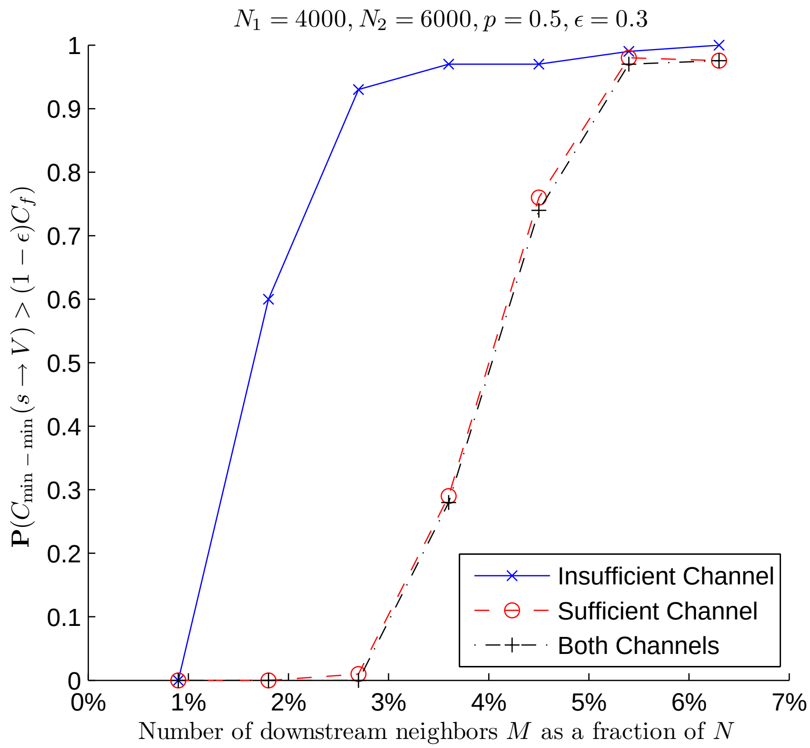

Figure 3: Multi-Channel: The Success Probability versus The Number of Downstream Neighbors

and still retain the simplicity and robustness of a random selection scheme. Our analysis provides an important insight that the capacity of the last cut (i.e., the capacity from direct upstream neighbors) is often the bottleneck. Our simulation results demonstrate that, by slightly modifying the control at the last hop, the performance of the system can indeed be improved. We envision that hybrid schemes that both balance the capacity at the last hop and exploit some level of random peer-selection may be able to achieve the best tradeoff between performance and complexity.

A Proof of Lemma 8

Proof. 1) It is clear from (22) that H_j and N_j have the same order unless $H_j = 0$. Hence, either $H_j = 0$ or

$$H_j = \Theta(N_j) = \Theta(N).$$

2) Note that $u_{s,j} = R_j/(1-\epsilon)$, we have

$$\begin{aligned} |H_j| &= \left| \left\lfloor \frac{N_j R_j}{(1-\epsilon)u} - \frac{u_{s,j}}{u} - pN_j \right\rfloor \right| \\ &\leq \left| \left\lfloor \frac{(N_j - 1)R_j}{(1-\epsilon)u} - pN_j \right\rfloor \right|. \end{aligned}$$

If $H_j < 0$, we have

$$\frac{(N_j - 1)R_j}{(1-\epsilon)u} - pN_j < 0.$$

Thus,

$$|H_j| \leq pN_j - \frac{(N_j - 1)R_j}{(1-\epsilon)u} + 1.$$

We then have

$$\begin{aligned} \limsup_{N\to\infty} \frac{|H_j|}{N_j} &= \limsup_{N_j\to\infty} \frac{|H_j|}{N_j} \\ &\leq p - \frac{R_j}{(1-\epsilon)u}. \end{aligned}$$

Let $\eta' \triangleq 1 - \frac{R_j}{(1-\epsilon)up}$. Since $\frac{R_j}{(1-\epsilon)up} > 0$, we have $\eta' < 1$. On the other hand, because

$$\limsup_{N\to\infty} \frac{|H_j|}{N_j} \leq p\eta', \tag{28}$$

and $\frac{|H_j|}{N_j} > 0, p > 0$, we have $\eta' > 0$. Now, choose η such $\eta' < \eta < 1$, according to (28), we have for large enough N

$$|H_j| \leq p\eta N_j.$$

3) From (22) we have

$$\begin{aligned}
&\frac{pN_j + H_j}{N_j}u + \frac{u_{s,j}}{N_j} \\
\geq &pu + \frac{u_{s,j}}{N_j} + \frac{u}{N_j}\left(\frac{N_jR_j}{(1-\epsilon)u} - \frac{u_{s,j}}{u} - pN_j - 1\right) \\
= &pu + \frac{u_{s,j}}{N_j} + \frac{R_j}{(1-\epsilon)} - \frac{u_{s,j}}{N_j} - pu - \frac{u}{N_j} \\
= &\frac{R_j}{(1-\epsilon)} - \frac{u}{N_j}.
\end{aligned}$$

Choose $\epsilon' < \epsilon$. We then have,

$$\begin{aligned}
&R_j \leq (1-\epsilon)\left(\frac{pN_j + H_j}{N_j}u + \frac{u_{s,j}}{N_j} + \frac{u}{N_j}\right) \\
&= (1-\epsilon')\left(\frac{pN_j + H_j}{N_j}u + \frac{u_{s,j}}{N_j}\right) + (\epsilon' - \epsilon)\left(\frac{pN_j + H_j}{N_j}u + \frac{u_{s,j}}{N_j}\right) + (1-\epsilon)\frac{u}{N_j}
\end{aligned}$$

Note that $\epsilon' - \epsilon < 0$. For large enough N, we have

$$-(1-\epsilon)\frac{u}{N_j} + (\epsilon' - \epsilon)\left(\frac{pN_j + H_j}{N_j}u + \frac{u_{s,j}}{N_j}\right) \leq 0.$$

Hence,

$$R_j \leq (1-\epsilon')\left(\frac{pN_j + H_j}{N_j}u + \frac{u_{s,j}}{N_j}\right). \tag{29}$$

Combining (21) and (29) yields (26).

4) To show (27), note that $\mathbf{R} \in (1-\epsilon)\Lambda_m$, we have

$$\sum_{j\in\mathcal{J}} \frac{N_jR_j}{\leq}(1-\epsilon)(upN + u_s).$$

Hence,

$$\begin{aligned}\sum_{j\in\mathcal{J}} H_j \leq & \sum_{j\in\mathcal{J}}\left(\frac{N_jR_j}{(1-\epsilon)u}-\frac{u_{s,j}}{u}-pN_j\right)\\ = & \sum_{j\in\mathcal{J}}\frac{N_jR_j}{(1-\epsilon)u}-\frac{u_s}{u}-pN\\ \leq & pN+\frac{u_s}{u}-\frac{u_s}{u}-pN(\text{using })\\ = & 0.\end{aligned}$$

Therefore, (27) holds. □

B Proof of Theorem 9

For any insufficient channel $j \in \mathcal{I}$, consider the subnetwork $\mathcal{G}_j$. We will use a similar technique as we did for the single-channel model. We first focus on the min-cut from the server s to any given destination t. Denote the further subnetwork of $\mathcal{G}_j$ that contains all the ON peers in $\mathcal{N}_j$, the destination t and the links between them as $\mathcal{G}_j^{ON,t}$, i.e., $\mathcal{G}_j^{ON,t}$ will contains the server, the peers in $\mathcal{H}_j$ and the ON peers in $\mathcal{N}_j$ and the links between them. Let the set of ON peers in $\mathcal{N}_j$ be $\mathcal{N}_j^{ON}$. Let the number of peers in $\mathcal{N}_j^{ON}$ excluding the server and the destination be Y_j. According to Proposition 2, the min-cut of $\mathcal{G}_j$ statistically dominates by the min-cut of $\mathcal{G}_j^{ON,t}$. We can then study $\mathcal{G}_j^{ON,t}$ instead of $\mathcal{G}_j$.

We will estimate the "failure" probability for each individual cut using Chernoff bound and take the union bound of them. However, we need to categorize cuts more carefully because now we have two different kinds of peers: the normal peer in channel i and the helpers borrowed from other channels. We define a cut of $\mathcal{G}_j^{ON,t}$ by dividing peers into a set $\mathcal{V}_j^{n,h}$ that contains the server, n ON peers from the set $\mathcal{N}_j$ and h helpers from the set $\mathcal{H}_j$, and the complementray set $\mathcal{V}_j^{n,h^c}$ that contains the destination, the remaining $Y_j - n$ ON peers and $H_j - h$ helpers. In the following, by "the left set" we mean the set $\mathcal{V}_j^{n,h}$ and by "the right set" we mean the set $\mathcal{V}_j^{n,h^c}$. For each pair of (n, h) there are totally $\binom{Y_j}{n}\binom{H_j}{h}$ such cuts. Define

$$\mathcal{N}_j^{n,h} \triangleq \mathcal{V}_j^{n,h} \cap \mathcal{N}_j, \mathcal{H}_j^{n,h} \triangleq \mathcal{V}_j^{n,h} \cap \mathcal{H}_j.$$

$\mathcal{N}_j^{n,h}$ is the set of normal ON peers of channel j that are in the left set and $\mathcal{H}_j^{n,h}$ is the set of helpers of channel j that are in the left set. Define the set $\mathcal{N}_j^{n,h^c}$ and $\mathcal{H}_j^{n,h^c}$ as the complementray set of $\mathcal{N}_j^{n,h}$ and $\mathcal{H}_j^{n,h}$ in $\mathcal{N}_j^{ON}$ and $\mathcal{H}_j$, respectively. Recall that $C_{i,k}$ is the link capacity between node i and node k. Let $C_j^{n,h}$ be the capacity of the cut, which is given by

$$C_j^{n,h} = \sum_{i \in \mathcal{V}_j^{n,h}} \sum_{k \in \mathcal{V}_j^{n,h^c}} C_{i,k}.$$

The capacity of a cut comes from four different kinds of contribution including (i) the contribution from the server to the normal peers in the right set, (ii) the contribution from the normal peers in the left set to the normal peers in the right set, (iii) the contribution from the normal peers in the left set to the helpers in the right set and (iv) the contribution from the helpers in the left set to the normal peers in the right set. Note that there is no contribution from helpers to other helpers since there are no connections between them. Denote these four parts of contribution as $C_{j,s\to N}^{n,h}$, $C_{j,N\to N}^{n}$, $C_{j,N\to H}^{n,h}$ and $C_{j,H\to N}^{n,h}$ (we omit the superscript h in $C_{j,N\to N}^{n}$ since it does not depend on the number of helpers on the left set h). More specifically,

$$C_{j,s\to N}^{n,h} = \sum_{k \in \mathcal{N}_j^{n,h^c}} C_{s,k},$$

$$C_{j,N\to N}^{n} = \sum_{i \in \mathcal{N}_j^{n,h}} \sum_{k \in \mathcal{N}_j^{n,h^c}} C_{i,k},$$

$$C_{j,N\to H}^{n,h} = \sum_{i \in \mathcal{N}_j^{n,h}} \sum_{k \in \mathcal{H}_j^{n,h^c}} C_{i,k},$$

$$C_{j,H\to N}^{n,h} = \sum_{i \in \mathcal{H}_j^{n,h}} \sum_{k \in \mathcal{N}_j^{n,h^c}} C_{i,k}.$$

The capacity of a cut can then be written as

$$C_j^{n,h} = C_{j,s\to N}^{n,h} + C_{j,N\to N}^{n} + C_{j,N\to H}^{n,h} + C_{j,H\to N}^{n,h}.$$

Note that for each normal peer, there are K reserved links for helpers, and therefore the upload capacity of each normal ON peer that can be utilized for other normal peers will be given by $u' = \frac{M'}{M}u$, where $M' = M - K$. The contribution of a normal peer in the left set to that in the right set normalized by the per link capacity u/M still has a hypergeometric distribution. So are the contribution from the server to the normal peers in the right set and the contribution from the helpers in the left set to the normal peers in the right set. The expectations of different parts conditioned on Y_j can be written as

$$\begin{aligned}
\mathbf{E}\left[C_{j,s\to N}^{n,h}|Y_j, t \text{ is ON}\right] &= (Y_j - n + 1)\frac{u_{s,j}}{Y_j+1}, \\
\mathbf{E}\left[C_{j,s\to N}^{n,h}|Y_j, t \text{ is OFF}\right] &= (Y_j - n)\frac{u_{s,j}}{Y_j}, \\
\mathbf{E}\left[C_{j,N\to N}^{n}|Y_j\right] &= (Y_j - n + 1)n\frac{u'}{N_j - 1}, \\
\mathbf{E}\left[C_{j,H\to N}^{n,h}|Y_j\right] &= (Y_j - n + 1)h\frac{u}{N_j}.
\end{aligned}$$

However, $C_{j,N\to H}^{n,h}$ has a different distribution. For each individual $i \in \mathcal{N}_j, k \in \mathcal{H}_j$, C_{ik} will be a Bernoulli random variable with parameter $1/N_j$ times the per link capacity $\frac{u}{M}$. Note however that such C_{ik} are not independent across i or k. Nonetheless, we can show that

$$\mathbf{E}\left[C_{j,N\to H}^{n,h}|Y_j\right] = \frac{n(H_j - h)}{Y_j M}u.$$

Now define

$$C_j'^{n,h} = C_{j,s\to N}^{n,h} + C_{j,N\to N}^{n} + C_{j,H\to N}^{n,h},$$

and let

$$\begin{aligned}
\bar{C}_j'^{n,h} =& \mathbf{E}\left[C_j'^{n,h}\middle| Y_j\right] \\
=& \mathbf{E}\left[C_{j,s\to N}^{n,h}|Y_j\right] + \mathbf{E}\left[C_{j,N\to N}^{n}|Y_j\right] + \mathbf{E}\left[C_{j,H\to N}^{n,h}|Y_j\right].
\end{aligned}$$

In addition, define the contribution from all the helpers to the normal peers on the right side as $C_{j,AH\to N}^{n}$, i.e.,

$$C_{j,AH\to N}^{n} = \sum_{k\in\mathcal{H}_j}\sum_{i\in\mathcal{N}_j^{n,h^c}} C_{i,k}.$$

We have

$$\mathbf{E}\left[C_{j,AH\to N}^{n}\right] = (Y_j - n)\frac{H_j}{N_j}u.$$

We have the following lemma.

Lemma 11. *For insufficient channel j and any $\epsilon > 0$, assume that $M = \Omega(\log N)$, $N_j = \Theta(N)$ and $H_j = \Theta(N)$. For any $\epsilon'' \in (0,1)$ there exists N_0 such that for $N > N_0$ if $\gamma \in [p(1-\epsilon'')^{\frac{1}{3}}, p)$ and $Y_j \geq \gamma N_j$, we have the following results for $0 \leq n \leq Y_j$ and $0 \leq h \leq H_j$*

1) When $n \leq \delta_1 Y_j$, and $0 < \delta_1 < 1$, we have

$$(1-\epsilon'')C_{f,j} \leq \bar{C}_j^{\prime n,h}; \tag{30}$$

2) When $Y_j - (K+1)/p \geq n \geq \delta_1 Y_j$ for some constants $\delta_1 > 0$, we have

$$(1-\epsilon'')C_{f,j} \leq \mathbf{E}\left[C^n_{j,N\to N}|Y_j\right]; \tag{31}$$

3) When $n \geq Y_j - (K+1)/p$, we have

$$(1-\epsilon'')^2 C_{f,j} \leq \mathbf{E}(1-\epsilon'')\left[C^n_{j,N\to N}|Y_j\right] + C^{n,h}_{w,j}, \tag{32}$$

where

$$C^{n,h}_{w,j} = \left[(1-\epsilon'')\mathbf{E}\left[C^n_{j,AH\to N}\right] - (H_j-h)(Y_j-n+1)\frac{u}{M}\right]^+ + \left[(H_j-h)u/M - K(Y_j-n+1)u/M\right]^+.$$

($C^{n,h}_{w,j}$ is the expectation of the worst case value of $C^{n,h}_{j,N\to H} + C^{n,h}_{j,H\to N}$ in these condition, which will be described later.)

Proof. We first consider part 1). Note that $N_j = \Theta(N)$. We can assume that $\beta_1 N \leq N_j \leq \beta_2 N$. We consider two different ranges of n. i) Suppose $n \leq \delta_3 \log(N)$ for some constants δ_3. We have

$$\begin{aligned}
\liminf_{N\to\infty} \bar{C}_j^{\prime n,h} &\geq \liminf_{N\to\infty} \mathbf{E}\left[C^{n,h}_{j,s\to N}|Y_j\right] \\
&\geq \liminf_{N\to\infty} (Y_j - n)\frac{u_{s,j}}{Y_j} \\
&\geq \liminf_{N\to\infty} \left(1 - \frac{\delta_3 \log(N)}{\gamma\beta_1 N}\right) u_{s,j} \\
&= u_{s,j} \geq C_{f,j}.
\end{aligned}$$

Therefore, (30) holds for large enough n.

ii) Suppose $\delta_3 \log(N) \le n \le \delta_1 Y_j$ for some constants δ_3. Then

$$\begin{aligned}
\liminf_{N\to\infty} \bar{C}_j^{\prime n,h} \geq & \liminf_{N\to\infty} \mathbf{E}\left[C^n_{j,N\to N}|Y_j\right] \\
\geq & \liminf_{N\to\infty} (Y_j - n + 1)n\frac{u'}{N_j - 1} \\
\geq & \liminf_{N\to\infty} (\gamma\beta_1(1-\delta_1)N - 1)\delta_3\log(N)\frac{u'}{\beta_2 N - 1} \\
= & \liminf_{N\to\infty} \frac{\gamma\beta_1(1-\delta_1)}{\beta_2}\delta_3\log(N)u' \\
= & +\infty.
\end{aligned}$$

Hence, (30) holds for large enough n. According to i) and ii), result 1) holds consequently.

We now consider 2). Note that

$$\mathbf{E}\left[C^n_{j,N\to N}|Y_j\right] = (Y_j - n + 1)n\frac{u'}{N_j - 1}$$

is a quadratic function of n. Therefore, for $Y_j - (K+1)/p \geq n \geq \delta_1 Y_j$,

$$\begin{aligned}
& \mathbf{E}\left[C^n_{j,N\to N}|Y_j\right] \\
= & (Y_j - n + 1)n\frac{u'}{N_j - 1} \\
\geq & \min\{(1+(K+1)/p)(Y_j - (K+1)/p), \delta_1 Y_j(Y_j - \delta_1 Y_j)\}\frac{u'}{N_j - 1}.
\end{aligned}$$

Note that $\delta_1 Y_j(Y_j - \delta_1 Y_j) = \Theta(N^2)$, $(1+(K+1)/p)(Y_j-(K+1)/p) = \Theta(N)$. For large N, we will have

$$\delta_1 N(Y_j - \delta_1 N) >> (1 + (K+1)/p)(Y_j - (K+1)/p).$$

Hence, for $Y_j - (K+1)/p \geq n \geq \delta_1 N$, we have

$$\mathbf{E}\left[C^n_{j,N\to N}|Y_j\right] > (1 + (K+1)/p)(Y_j - (K+1)/p).$$

If $Y_j \geq \gamma N_j$, we have

$$(1+(K+1)/p)(Y_j-(K+1)/p)\frac{u'}{N_j - 1} \geq (1+(K+1)/p)(\gamma N_j-(K+1)/p)\frac{u'}{N_j - 1}.$$

Hence,

$$\liminf_{N\to\infty} \mathbf{E}\left[C^n_{j,N\to N}|Y_j\right] \geq (1+(K+1)/p)\gamma u'.$$

For any ϵ'', if N is large enough such that $N_j = p_j N$ and $M = \alpha\log(N)$ are large enough, we have

$$\mathbf{E}\left[C^n_{j,N\to N}|Y_j\right] \geq (1+(K+1)/p)(1-\epsilon'')^{\frac{1}{3}}\gamma u',$$

and

$$u' = \frac{M-K}{M} \geq (1-\epsilon'')^{\frac{1}{3}}u$$

Choose γ such that $\gamma \geq p(1-\epsilon'')^{\frac{1}{3}}$. We then have

$$\begin{aligned}&\mathbf{E}\left[C^n_{j,N\to N}|Y_j\right]\\ \geq&(1+(K+1)/p)(1-\epsilon'')^{\frac{1}{3}}p(1-\epsilon'')^{\frac{1}{3}}(1-\epsilon'')^{\frac{1}{3}}u\\ =&(1-\epsilon'')(up+(K+1)u).\end{aligned}$$

Note that $H_j \leq KN_j$, for large enough N

$$\begin{aligned}C_{f,j} \leq&\frac{pN_j+H_j}{N_j}u + \frac{u_{s,j}}{N_j}\\ \leq&up+Ku+\frac{u_{s,j}}{N_j}\\ \leq&up+(K+1)u\end{aligned}$$

Therefore, for large enough N,

$$(1-\epsilon'')C_{f,j} \leq (1-\epsilon''')up + (K+1)u \leq \mathbf{E}\left[C^n_{j,N\to N}|Y_j\right].$$

For 3), we have $Y_j - (K+1)/p \leq n$. Then if $Y_j \geq \gamma N_j$

$$\begin{aligned}(Y_j-n+1)n\frac{u'}{N_j-1} \geq&Y_j\frac{u'}{N_j-1}\\ \geq&\gamma\frac{M-K}{M}u\end{aligned}$$

For any ϵ'', we can choose γ such that $\gamma \geq p\sqrt{1-\epsilon}$ and M large enough such that $\frac{M-K}{M} \geq \sqrt{1-\epsilon''}$. Then we will get

$$(Y_j-n+1)n\frac{u'}{N_j-1} \geq (1-\epsilon'')up. \tag{33}$$

Next we will show that for large enough N,

$$(1-\epsilon'')^2\frac{H_j}{N_j}u \le C_{w,j}^{n,h}.$$

First, if $H_j - h \le K(Y_j - n+1)$, we have $(H_j-h)u/M - K(Y_j-n+1)u/M \le 0$, and

$$\begin{aligned}\limsup_{N\to\infty}(H_j-h)(Y_j-n+1)\frac{u}{M} \le& \limsup_{N\to\infty} K(Y_j-n+1)^2\frac{u}{M}\\ \le& \limsup_{N\to\infty} K((K+1)/p+1)^2\frac{u}{\alpha\log(N)}\\ =&0.\end{aligned}$$

On the other hand, since $(H_j-h)(Y_j-n+1)\frac{u}{M} \ge 0$, we have

$$\liminf_{N\to\infty}(H_j-h)(Y_j-n+1)\frac{u}{M} \ge 0.$$

Hence,

$$\lim_{N\to\infty}(H_j-h)(Y_j-n+1)\frac{u}{M} = 0.$$

Therefore, for large enough N, $(H_j-h)(Y_j-n+1)\frac{u}{M}$ will be smaller than $(1-\epsilon'')\frac{(Y_j-n+1)H_j}{N_j}u$, whose limit is at least $\lim_{N\to\infty}(1-\epsilon'')\frac{H_j}{N_j}$ when N approaches infinity, i.e.,

$$(1-\epsilon'')\mathbf{E}\left[C_{j,AH\to N}^{n}\right] = (1-\epsilon'')\frac{(Y_j-n+1)H_j}{N_j}u \ge (H_j-h)(Y_j-n+1)\frac{u}{M}.$$

Thus

$$\begin{aligned}C_{w,j}^{n,h} &= \left[(1-\epsilon'')\mathbf{E}\left[C_{j,AH\to N}^{m}\right] - (H_j-h)(Y_j-n+1)\frac{u}{M}\right]^+\\ &+ \left[(H_j-h)u/M - K(Y_j-n+1)u/M\right]^+\\ &= (1-\epsilon'')\frac{(Y_j-n+1)H_j}{N_j}u - (H_j-h)(Y_j-n+1)\frac{u}{M} + 0\\ &\ge (1-\epsilon'')\frac{H_j}{N_j}u - (H_j-h)(Y_j-n+1)\frac{u}{M}\\ &\stackrel{N\to\infty}{\longrightarrow} (1-\epsilon'')\frac{H_j}{N_j}u.\end{aligned}$$

Second, if $K(Y_j - n + 1) \leq H_j - h \leq (1-\epsilon'')\frac{H_j}{N_j}M$, we have

$$(H_j - h)\frac{u}{M} - K(Y_j - n + 1)\frac{u}{M} \geq K(Y_j - n + 1)\frac{u}{M} - K(Y_j - n + 1)\frac{u}{M} = 0.$$

and

$$\begin{aligned}
&(1-\epsilon'')\frac{(Y_j - n + 1)H_j}{N_j}u - (H_j - h)(Y_j - n + 1)\frac{u}{M} \\
\geq &(1-\epsilon'')\frac{(Y_j - n + 1)H_j}{N_j}u - \frac{H_j}{N_j}M(Y_j - n + 1)\frac{u}{M} \\
= &0.
\end{aligned}$$

Hence,

$$\begin{aligned}
C_{w,j}^{n,h} = &(1-\epsilon'')\frac{(Y_j - n + 1)H_j}{N_j}u - (H_j - h)(Y_j - n + 1)\frac{u}{M} + (H_j - h)\frac{u}{M} \\
&- K(Y_j - n + 1)\frac{u}{M} \\
= &(1-\epsilon'')\frac{(Y_j - n + 1)H_j}{N_j}u - (H_j - h)(Y_j - n)\frac{u}{M} - K(Y_j - n + 1)\frac{u}{M} \\
\geq &(1-\epsilon'')\frac{(Y_j - n + 1)H_j}{N_j}u - \frac{H_j}{N_j}M(Y_j - n)\frac{u}{M} - K(Y_j - n + 1)\frac{u}{M} \\
= &(1-\epsilon'')\frac{H_j}{N_j}u - K(Y_j - n + 1)\frac{u}{M} \\
\xrightarrow{N\to\infty} &(1-\epsilon'')\frac{H_j}{N_j}u.
\end{aligned}$$

Finally, if $H_j - h \geq (1-\epsilon'')\frac{H_j}{N_j}M$, we have

$$(H_j - h)\frac{u}{M} - K(Y_j - n + 1)\frac{u}{M} \geq \frac{H_j}{N_j}M\frac{u}{M} - K(Y_j - n + 1)\frac{u}{M} \geq 0,$$

and

$$\begin{aligned}
&(1-\epsilon'')\frac{(Y_j - n + 1)H_j}{N_j}u - (H_j - h)(Y_j - n + 1)\frac{u}{M} \\
\leq &(1-\epsilon'')\frac{(Y_j - n + 1)H_j}{N_j}u - \frac{H_j}{N_j}M(Y_j - n + 1)\frac{u}{M} \\
= &0.
\end{aligned}$$

Thus,

$$\begin{aligned} C_{w,j}^{n,h} =&0 + (H_j - h)\frac{u}{M} - K(Y_j - n + 1)\frac{u}{M} \\ \geq& \frac{H_j}{N_j} M \frac{u}{M} - K(Y_j - n + 1)\frac{u}{M} \\ \xrightarrow{N\to\infty}& \frac{H_j}{N_j} u. \end{aligned}$$

Combine the three different ranges of h together, we have

$$\liminf_{N\to\infty} C_{w,j}^{n,h} = (1-\epsilon'') \lim_{N\to\infty} \frac{H_j}{N_j} u.$$

Thus for large enough N,

$$C_{w,j}^{n,h} \geq (1-\epsilon'')^2 \frac{H_j}{N_j} u.$$

Consequently, part 3) of the lemma holds.

In conclusion, for any $\epsilon'' \in (0,1)$, and any γ such that $\gamma \geq p(1-\epsilon'')^{\frac{1}{3}}$ and $\gamma \geq p\sqrt{1-\epsilon''}$, all the three parts hold. Obviously, $p > p(1-\epsilon'')^{\frac{1}{3}} > p\sqrt{1-\epsilon''}$. The lemma holds if we choose $\gamma \geq p(1-\epsilon'')^{\frac{1}{3}}$ for any given $\epsilon'' \in (0,1)$.

□

Part 1) Lemma 11 basically says that when there are few normal peers on the left the contribution from the peers on the left to the normal peers on the left will be big enough (larger than $C_{f,j}$). Part 2) Lemma 11 says that when there are many normal peers on the left but there are still a few normal peers on the right, the contribution from the normal peers on the left to the helpers on the right will dominant. Part 3) of Lemma 11 says that when almost all the normal peers are on the left, we need to consider all the three parts of the contribution except the contribution of the server. Lemma 11 allows us to analyze the "failure" probability of a cut based on how many peers that each side of the cut has, and also allows us to compare the capacity of the cut to the mean capacity instead of $C_{f,j}$. Law of large numbers then says the "failure" probability will converge to 0 if we compare the cut capacity with its mean.

However, the distribution of $C_{j,N\to H}^{n,h}$ is different from the other three parts of the contribution. It is not the sum of hypergeometric random variables.

But its characteristic function can still be bounded by the characteristic function of the sum of independent Bernoulli random variables. Note that for each $k \in \mathcal{H}_j^{n,h^c}$, $I_k = \sum_{i \in \mathcal{N}_j^{n,h}} C_{i,k} M/u$ is a Bernoulli random variables with parameter $\frac{n}{Y_j}$. We can show that I_k's are negatively related. To see this, construct J_{kl} as the following. First set $J_{kl} = I_l$ for any l. Then if $J_{kk} \neq 1$, choose one out of all reserved links from the left normal nodes randomly and exchange it with the reserved links that peer k is connecting to. It is clear that $\mathbf{J}_k$ has the same distribution as $\mathbf{I}$ given $I_k = 1$. If the chosen reserved link connects to a helper k' on the right, then after the exchanging, we have $J_{kk} = 1$ and $J_{kk'} = 0$. If the chosen reserved link connects to a helper on the left, then we have $J_{kk} = 1$ and the other $J_{kl}, l \neq k$ remain the same. Under either case, we have $J_{kl} \leq I_l$, $l \neq k$. We then can apply Theorem 4 and bound the moment generating function of $C_{j,N\to H}^{n,h}$ as the sum of i.i.d. Bernoulli random variables with parameter $\frac{n}{Y_j}$.

Now we consider sufficient channel j. We denote the cut by $\mathcal{V}_j^n$ and $\mathcal{V}_j^{nc}$ since there are no helpers involved. Let the capacity of the cut be C_j^n, which is given by

$$C_j^n = \sum_{i \in \mathcal{V}_j^n} \sum_{k \in \mathcal{V}_j^{nc}} C_{i,k}.$$

Assuming that $Y_j + H_j > 0$, the conditional expectation given the number of ON peers Y_j can be calculated easily

$$\bar{C}_j^n \triangleq \mathbf{E}\left[C_j^n | Y_j\right] = \begin{cases} \frac{u_{s,j}(Y_j+H_j-n)}{Y_j+H_j} + \frac{u}{N_j-1} n(Y_j + H_j - n) & \text{if } t \text{ is OFF,} \\ \frac{u_{s,j}(Y_j+H_j+1-n)}{Y_j+H_j+1} + \frac{u}{N_j-1} n(Y_j + H_j - n) & \text{if } t \text{ is ON.} \end{cases}.$$

We have corresponding result for sufficient channels regarding the mean of the cut capacity.

Lemma 12. *Given ϵ, choose H_j's according to (22). For sufficient channel j, assume there exist $\eta < 1$ such that $|H_j| \leq \eta p N_j$. For any $0 \leq n \leq Y_j + H_j$, and $\epsilon'' \in (0,1)$ if $\gamma' \in [(1-\epsilon''+\eta\epsilon'')p, p)$ and $Y_j > \gamma' N_j$ we have*

$$Y_j + H_j \geq (1-\epsilon'')(1-\eta)pN_j, \tag{34}$$

and,

$$(1-\epsilon'')C_{f,j} \leq \bar{C}_j^n. \tag{35}$$

Proof. It is not hard to see

$$\begin{aligned}\bar{C}_j^n &\leq \min\left\{C_{f,j}^0, C_{f,j}^{Y_j+H_j}\right\} \\ &\leq \min\left\{u_{s,j}, \frac{Y_j+H_j}{N_j}u + \frac{u_{s,j}}{N_j}\right\}.\end{aligned}$$

We have

$$\begin{aligned}&(1-\epsilon'')C_{f,j} \leq \bar{C}_j^n \\ \Leftrightarrow &(1-\epsilon'')\min\left\{u_{s,j}, \frac{pN_j+H_j}{N_j}u + \frac{u_{s,j}}{N_j}\right\} \leq \min\left\{u_{s,j}, \frac{Y_j+H_j}{N_j}u + \frac{u_{s,j}}{N_j}\right\} \\ \Leftrightarrow &(1-\epsilon'')\frac{pN_j+H_j}{N_j}u \leq \frac{Y_j+H_j}{N_j}u \\ \Leftrightarrow &(1-\epsilon'')(pN_j+H_j) \leq Y_j+H_j\end{aligned}$$

Hence, for (35) to holds, it is sufficient to show that there exist $\gamma' < p$ such that

$$\gamma' N_j + H_j \geq (1-\epsilon'')(pN_j+H_j),$$

which is equivalent to

$$\gamma' \geq (1-\epsilon)p - \epsilon''\frac{H_j}{N_j}. \tag{36}$$

Note that from Lemma 8 2), there exists $\eta < 1$ such that $-\frac{H_j}{N_j} \leq \eta p$, we have,

$$\begin{aligned}&(1-\epsilon'')p - \epsilon\frac{H_j}{N_j} \\ \leq &(1-\epsilon'')p + \eta p\epsilon'' \\ = &(1-\epsilon''+\eta\epsilon'')p \\ < &p.\end{aligned}$$

Then we can choose any γ' between $[(1-\epsilon''+\eta\epsilon'')p, p)$ as a constant and the (35) holds. Substitute γ' into (36) will yields (34). □

Lemma 12 says that if Y_j is close to or larger than its mean value pN_j, the capacity of each cut is greater than the $C_{f,j}$, which is the minimum over the mean value of any cut capacity. Now we can prove Theorem 9.

Proof of Theorem 9. The proof is divided into two main parts.

A. Insufficient channels

For given ϵ', let $\epsilon'' = 1 - \sqrt{(1-\epsilon')}$, i.e., $(1-\epsilon'')^2 = 1-\epsilon'$. For insufficient channel, define $\tilde{\mathcal{B}}_j^{n,h}$ to be the event $\{C_j^{n,h} \leq (1-\epsilon')C_{f,j}$ for any cut among the $\binom{Y_j}{n}\binom{H_j}{h}$ cuts $\}$. Let $0 < \delta_1 < 1$ be a constant. We will first assume that $Y_j = y_j \geq \gamma N_j$ for the γ as in Lemma 11 during the following discussion, and consider the randomness of Y_j at the end. We will consider three different ranges of n and h.

1) When $n \leq \delta_1 y_j$, is satisfied. From part 1) of Lemma 11 we will only need to worry about the $C_j^{'n,h}$. Using Chernoff bound, for any $\theta > 0$ we have

$$\begin{aligned}
&\mathbf{P}\left(C_j^{'n,h} \leq (1-\epsilon')C_{f,j}\right)\\
\leq &\mathbf{P}\left(C_j^{'n,h} \leq (1-\epsilon'')\mathbf{E}\left[C_j^{'n,h}\middle| Y_j = y_j\right]\right)\\
\leq &\frac{\mathbf{E}\left[\exp(-\theta C_j^{'n,h})\middle| Y_j = y_j\right]}{\exp(-\theta(1-\epsilon'')\mathbf{E}\left[C_j^{'n,h}\middle| Y_j = y_j\right])}\\
= &\frac{\mathbf{E}\left[e^{-\theta C_{j,s\to N}^{n,h}}\middle| Y_j = y_j\right]\mathbf{E}\left[e^{-\theta C_{j,N\to N}^{n}}\middle| Y_j = y_j\right]\mathbf{E}\left[e^{-\theta C_{j,H\to N}^{n,h}}\middle| Y_j = y_j\right]}{e^{-\theta(1-\epsilon'')\left(\mathbf{E}\left[C_{j,s\to N}^{n,h}\middle|Y_j=y_j\right]+\mathbf{E}\left[C_{j,N\to H}^{n,h}\middle|Y_j=y_j\right]+\mathbf{E}\left[C_{j,H\to N}^{n,h}\middle|Y_j=y_j\right]\right)}}.
\end{aligned}$$

We can use the same method as we did in Lemma 6 and we will have

$$\begin{aligned}
&\mathbf{P}\left(C_j^{'n,h} \leq (1-\epsilon'')\mathbf{E}\left[C_j^{'n,h}\middle| Y_j = y_j\right] |Y_j = y_j\right)\\
\leq &\exp\left(-\frac{u}{u_{s,j}}\frac{\epsilon''^2}{2}M'\left(\frac{y_i - n}{y_i} + \frac{(n+h)(y_i-n)}{N_i}\right)\right)
\end{aligned}$$

Since $n \leq \delta_1 y_j$, we have $y_i - n \geq (1-\delta_1)y_j \geq \gamma(1-\delta_1)N_j$. Letting $\beta =$

$e^{-\frac{u}{u_{s,j}}\frac{\epsilon''^2}{2}M'}$, then

$$\begin{aligned}
&\mathbf{P}\left(\bigcup_{n=0}^{\delta_1 y_j}\bigcup_{h=0}^{H_j}\tilde{\mathcal{B}}_j^{n,h}|Y_j=y_j\right)\\
\le&\sum_{n=0}^{\delta_1 y_j}\sum_{h=0}^{H_i}\binom{y_i}{n}\binom{H_i}{h}\mathbf{P}\left(C_j'^{n,h}\le(1-\epsilon'')\mathbf{E}\left[C_j'^{n,h}\middle|Y_j=y_j\right]|Y_j=y_j\right)\\
\le&\sum_{n=0}^{\delta_1 y_j}\sum_{h=0}^{H_i}\binom{y_i}{n}\binom{H_i}{h}\exp\left(-\frac{u}{u_{s,j}}\frac{\epsilon''^2}{2}M'\left(\frac{y_i-n}{y_i}+\frac{(n+h)(y_i-n)}{N_i}\right)\right)\\
\le&\beta\sum_{n=0}^{\delta_1 y_j}\sum_{h=0}^{H_i}\binom{y_i}{n}\binom{H_i}{h}\exp\left(-\frac{u}{u_{s,j}}\frac{\epsilon''^2}{2}M'\left(\frac{-n}{y_i}+\delta_1\gamma(n+h)\right)\right)\\
\le&\beta\sum_{n=0}^{y_j}\binom{y_i}{n}\beta^{-\frac{n}{y_i}+\delta_1\gamma n}\sum_{h=0}^{H_i}\binom{H_i}{h}\beta^{\delta_1\gamma h}\\
=&\beta(1+\beta^{\delta_1\gamma-1/y_i})^{y_i}(1+\beta^{\delta_1\gamma})^{H_i}.
\end{aligned}\tag{37}$$

2) When $y_j-(K+1)/p\ge n\ge\delta_1 y_j$, from part 2) Lemma 11 we can found N large enough and $y_j\ge\gamma N_j$ such that $(1-\epsilon'')C_{j,f}\le\mathbf{E}\left[C_{j,N\to N}^n|Y_j=y_j\right]$. Hence, we only need to consider $C_{j,N\to N}^n$. As result, for these cuts the capacity between normal peers will be big enough, and we do not need to consider the contribution from the helpers. Let us focus on the subnetwork only contains the ON normal peers and the cuts of the subnetwork that divide all the ON normal peers to two sets: one on the left with n peers and the other on the right with y_j-n peers. Define $\tilde{\mathcal{B}}_j^n$ to be the event $\{C_{j,N\to N}^n\le(1-\epsilon'')\left[C_{j,N\to N}^n|Y_j=y_j\right]$ for any cut among the $\binom{y_j}{n}$ cuts $\}$. For a given n such that $y_j-(K+1)/p\ge n\ge\delta_1 y_j$, if a cut in this subnetworks has enough capacity then any cuts in the subnetwork contains both ON normal peers and helpers that divides the normal peers the same will also has enough capacity.

More specifically, for a fixed $n : y_j - (K+1)/p \geq n \geq \delta_1 y_j$, we have

$$\begin{aligned}
&\left(\tilde{\mathcal{B}}_j^n\right)^c \\
\triangleq&\{C_{j,N\to N}^n \geq (1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^n | Y_j = y_j\right] \text{ for all the } \binom{y_j}{n} \text{ cuts}\} \\
\subset&\{C_{j,N\to N}^n \geq (1-\epsilon'')^2 C_{f,j} \text{ for all the } \binom{y_j}{n}\binom{H_j}{h} \text{ cuts in } \mathcal{G}_j^{ON,t} \text{ for all } h\} \\
\subset&\{C_j^{n,h} \geq (1-\epsilon') C_{f,j} \text{ for all the } \binom{y_j}{n}\binom{H_j}{h} \text{ cuts in } \mathcal{G}_j^{ON,t} \text{ for all } h\} \\
=&\bigcap_{h=0}^{H_j} \left(\tilde{\mathcal{B}}_j^{n,h}\right)^c
\end{aligned}$$

Therefore, for $y_j - (K+1)/p \geq n \geq \delta_1 y_j$

$$\bigcup_{h=0}^{H_j} \tilde{\mathcal{B}}_j^{n,h} \subset \tilde{\mathcal{B}}_j^n. \tag{38}$$

For each $n \geq \delta_1 y_j$

$$\begin{aligned}
&\mathbf{P}\left(C_{j,N\to N}^n \leq (1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^n | Y_j = y_j\right] | Y_j = y_j\right) \\
\leq&\frac{\mathbf{E}\left[\exp(-\theta C_{j,N\to N}^n) | Y_j = y_j\right]}{\exp(-\theta(1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^n | Y_j = y_j\right])}.
\end{aligned}$$

Using the same method as we did in Lemma 6 we can get

$$\begin{aligned}
&\mathbf{P}\left(C_{j,N\to N}^n \leq (1-\epsilon'')\left[C_{j,N\to N}^n\right] | Y_j = y_j\right) \\
\leq&\exp\left(-\frac{\epsilon''^2}{2}\frac{Mn(y_i - n + 1)}{N_i - 1}\right).
\end{aligned}$$

then

$$
\begin{aligned}
&\mathbf{P}\left(\bigcup_{h=0}^{H_j} \tilde{\mathcal{B}}_j^{n,h} | Y_j = y_j\right) \\
\leq &\mathbf{P}\left(\tilde{\mathcal{B}}_j^{n} | Y_j = y_j\right) \\
\leq &\binom{y_i}{n} \mathbf{P}\left(C_{j,N\to N}^n \leq (1-\epsilon'') \left[C_{j,N\to N}^n\right] | Y_j = y_j\right) \\
\leq &\binom{y_i}{n} \exp\left(-\frac{\epsilon''^2}{2} \frac{Mn(y_i - n + 1)}{N_i - 1}\right)
\end{aligned}
$$

Note that for $n \geq \delta_1 y_j$ and $y_j \geq \gamma N_j$, we have

$$
\begin{aligned}
\frac{Mn(y_i - n + 1)}{N_i - 1} &= M\frac{y_i - n + 1}{N_i - 1} + M\frac{(n-1)(y_i - n + 1)}{N_i - 1} \\
&\geq M\frac{\gamma N_i - n + 1}{N_i - 1} + M\frac{(n-1)(y_i - n + 1)}{N_i - 1} \\
&\geq \gamma M + M\frac{-n + 1}{N_i - 1} + M\frac{(n-1)(y_i - n + 1)}{N_i - 1} \\
&\geq \gamma M + M\frac{(n-1)(y_i - n)}{N_i - 1} \\
&\geq \gamma M + \gamma\delta_1 (y_i - n) M
\end{aligned}
$$

Therefore, letting $\beta' = e^{-M\frac{\epsilon''2}{2}}$, for $n \geq \delta_1 y_j$ and $y_j \geq \gamma N_j$, we have

$$
\begin{aligned}
&\mathbf{P}\left(\tilde{\mathcal{B}}_j^{n} | Y_j = y_j\right) \\
\leq &\binom{y_i}{n} \exp\left(-\frac{\epsilon''^2}{2} \frac{Mn(y_i - n + 1)}{N_i - 1}\right) \\
\leq &\binom{y_i}{n} \exp\left(-\frac{\epsilon''^2}{2}\gamma M - \frac{\epsilon''^2}{2}\gamma\delta_1 (y_i - n) M\right) \\
\leq &\beta'^{\gamma} \binom{y_i}{n} \beta'^{\delta_1 \gamma (y_i - n)}.
\end{aligned} \tag{39}
$$

3) Then consider the case when $y_j - (K+1)/p \leq n$. There will be at most $(K+1)/p + 1$ normal peers on the right. Now consider $C_{j,H\to N}^{n,h}$, which is the contribution from all the helpers on the left to the normal peers on the right.

For the cut that all helpers is on the left, we have $C^n_{j,AH\to N} = C^{n,H_j}_{j,H\to N}$ (recall that $C^n_{j,AH\to N}$ is the contribution from all the helpers to the normal peers on the right side). Now we move the helpers to the right one by one. Note that for each helper moved to the right, it takes away at most $(y_j - n + 1)u/M$ of capacity from $C^n_{j,AH\to N}$. We then have

$$C^{n,h}_{j,H\to N} \geq \left[C^n_{j,AH\to N} - (H_j - h)(y_j - n + 1)\frac{u}{M}\right]^+,$$

where $[\cdot]^+$ denotes the projection to $[0, +\infty)$.

Now consider $C^{n,h}_{j,N\to H}$. Since each helpers will have one and only one upstream neighbors from all the normal peers, we have $C^{n,h}_{j,N\to H}$ is at most $(H_j - h)u/M$, which is achieved when all the upstream neighbors of the helpers on the right are on the left. However, for each normal peers on the right, it will have at most K helpers as downstream neighbors. Therefore, there are at most $K(y_j - n + 1)$ helpers will select their upstream neighbors on the right. Hence,

$$C^{n,h}_{j,N\to H} \geq [(H_j - h)u/M - K(y_j - n + 1)u/M]^+.$$

We get for any h

$$\begin{aligned} C^{n,h}_{j,H\to N} + C^{n,h}_{j,N\to H} \geq & \left[C^n_{j,AH\to N} - (H_j - h)(y_j - n + 1)\frac{u}{M}\right]^+ \\ & + [(H_j - h)u/M - K(y_j - n + 1)u/M]^+. \end{aligned}$$

Recall that

$$\begin{aligned} C^{n,h}_{w,j} =& \mathbf{E}\Bigg[(1-\epsilon'')\left[C^n_{j,AH\to N} - (H_j - h)(Y_j - n + 1)\frac{u}{M}\right]^+ \\ & + [(H_j - h)u/M - K(Y_j - n + 1)u/M]^+ |Y_j = y_j\Bigg]. \end{aligned}$$

According to part 3) of Lemma 11, for $y_j \geq \gamma N_j$, we have

$$(1-\epsilon'')^2 C_{f,j} \leq (1-\epsilon'')\mathbf{E}\left[C^n_{j,N\to N}|Y_j = y_j\right] + C^{n,h}_{w,j}.$$

Note that, for any small h such that $(1-\epsilon'')\mathbf{E}\left[C^n_{j,AH\to N}|Y_j = y_j\right] \leq (H_j -$

$h)(Y_j - n + 1)\frac{u}{M}$, we have

$$
\begin{aligned}
&C_j^{n,h} \leq (1-\epsilon')C_{f,j} \\
\Rightarrow &C_{j,N\to N}^{n} + C_{j,H\to N}^{n,h} + C_{j,N\to H}^{n,h} \leq (1-\epsilon'')^2 C_{f,j} \\
\Rightarrow &C_{j,N\to N}^{n} + C_{j,H\to N}^{n,h} + C_{j,N\to H}^{n,h} \leq (1-\epsilon'')\left[\mathbf{E}\left[C_{j,N\to N}^{n}|Y_j = y_j\right] + C_{w,j}^{n,h}\right]. \\
\Rightarrow &C_{j,N\to N}^{n} + [(H_j - h)u/M - K(y_j - n + 1)u/M]^+ \\
&\qquad + \left[C_{j,AH\to N}^{n} - (H_j - h)(Y_j - n + 1)\frac{u}{M}\right]^+ \\
&\leq (1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^{n}|Y_j = y_j\right] + [(H_j - h)u/M - K(y_j - n + 1)u/M]^+ \\
&\qquad + \left[(1-\epsilon'')\mathbf{E}\left[C_{j,AH\to N}^{n}|Y_j = y_j\right] - (H_j - h)(Y_j - n + 1)\frac{u}{M}\right]^+ \\
\Leftrightarrow &C_{j,N\to N}^{n} + \left[C_{j,AH\to N}^{n} - (H_j - h)(Y_j - n + 1)\frac{u}{M}\right]^+ \\
&\leq (1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^{n}|Y_j = y_j\right] \\
&\qquad + \left[(1-\epsilon'')\mathbf{E}\left[C_{j,AH\to N}^{n}|Y_j = y_j\right] - (H_j - h)(Y_j - n + 1)\frac{u}{M}\right]^+ \\
\Leftrightarrow &C_{j,N\to N}^{n} + \left[C_{j,AH\to N}^{n} - (H_j - h)(Y_j - n + 1)\frac{u}{M}\right]^+ \\
&\leq (1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^{n}|Y_j = y_j\right] \\
\Rightarrow &C_{j,N\to N}^{n} \leq (1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^{n}|Y_j = y_j\right].
\end{aligned}
$$

For any large h such that $(1-\epsilon'')\mathbf{E}\left[C_{j,AH\to N}^{n}|Y_j = y_j\right] > (H_j - h)(Y_j - n + 1)\frac{u}{M}$, we have

$$
\begin{aligned}
&C_j^{n,h} \le (1-\epsilon')C_{f,j} \\
\Rightarrow &C_{j,N\to N}^n + \left[C_{j,AH\to N}^n - (H_j-h)(Y_j-n+1)\frac{u}{M}\right]^+ \\
&\le (1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^n | Y_j = y_j\right] \\
&\qquad + \left[(1-\epsilon'')\mathbf{E}\left[C_{j,AH\to N}^n | Y_j = y_j\right] - (H_j-h)(Y_j-n+1)\frac{u}{M}\right]^+ \\
\Leftrightarrow &C_{j,N\to N}^n + \left[C_{j,AH\to N}^n - (H_j-h)(Y_j-n+1)\frac{u}{M}\right]^+ \\
&\le (1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^n | Y_j = y_j\right] \\
&\qquad + (1-\epsilon'')\mathbf{E}\left[C_{j,AH\to N}^n | Y_j = y_j\right] - (H_j-h)(Y_j-n+1)\frac{u}{M} \\
\Rightarrow &\qquad C_{j,N\to N}^n + C_{j,AH\to N}^n \\
&\le (1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^n | Y_j = y_j\right] + (1-\epsilon'')\mathbf{E}\left[C_{j,AH\to N}^n | Y_j = y_j\right].
\end{aligned}
$$

The last step holds because

$$
\begin{aligned}
C_{j,AH\to N}^n \le (H_j-h)(Y_j-n+1)\frac{u}{M} \\
+ \left[C_{j,AH\to N}^n - (H_j-h)(Y_j-n+1)\frac{u}{M}\right]^+.
\end{aligned}
$$

We can then conclude that

$$
\begin{aligned}
&C_j^{n,h} \le (1-\epsilon')C_{f,j} \text{ for any of the } \binom{h}{H_j} \text{ cut, any } h \\
\Rightarrow &C_{j,N\to N}^n \le (1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^n | Y_j = y_j\right] \\
&\text{or} \\
&\quad C_{j,N\to N}^n + C_{j,AH\to N}^n \\
&\le (1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^n | Y_j = y_j\right] + (1-\epsilon'')\mathbf{E}\left[C_{j,AH\to N}^n | Y_j = y_j\right]
\end{aligned}
$$

Now define $\tilde{\mathcal{C}}_j^n$ be the event that

$$
\begin{aligned}
&\{ \text{ for any of the } \binom{y_j}{n} \text{ cuts }, C_{j,N\to N}^n + C_{j,AH\to N}^n \\
\le &(1-\epsilon'')\mathbf{E}\left[C_{j,N\to N}^n | Y_j = y_j\right] + (1-\epsilon'')\mathbf{E}\left[C_{j,AH\to N}^n | Y_j = y_j\right]\}
\end{aligned}
$$

We have

$$
\begin{aligned}
&\mathbf{P}\left(\bigcup_{n=y_j-(K+1)/p}^{y_j}\ \bigcup_{h=0}^{H_j}\tilde{\mathcal{B}}_j^{n,h}|Y_j=y_j\right)\\
\leq&\mathbf{P}(C^n_{j,N\to N}\leq(1-\epsilon'')\mathbf{E}\left[C^n_{j,N\to N}|Y_j=y_j\right]\ \text{ for any cut, any } h\\
&\qquad\text{and any } n\geq y_j-(K+1)/p)\\
&\ +\mathbf{P}(C^n_{j,N\to N}+C^n_{j,AH\to N}\leq(1-\epsilon'')\mathbf{E}\left[C^n_{j,N\to N}|Y_j=y_j\right]\\
+&\ \ (1-\epsilon'')\mathbf{E}\left[C^n_{j,AH\to N}|Y_j=y_j\right]\ \text{ for any cut, any } n\geq y_j-(K+1)/p,\ \text{ any } h)\\
\leq&\mathbf{P}\left(\bigcup_{n=y_j-(K+1)/p}^{y_j}\tilde{\mathcal{B}}_j^{n}|Y_j=y_j\right)+\mathbf{P}\left(\bigcup_{n=y_j-(K+1)/p}^{y_j}\tilde{\mathcal{C}}_j^{n}\right). \qquad (40)
\end{aligned}
$$

For each cut, we have

$$
\begin{aligned}
&\mathbf{P}\left(C^n_{j,N\to N}+C^n_{j,AH\to N}\leq(1-\epsilon'')(\mathbf{E}\left[C^n_{j,N\to N}|Y_j=y_j\right]+\mathbf{E}\left[C^n_{j,AH\to N}|Y_j=y_j\right]\right)\\
&\leq\frac{\mathbf{E}\left[\exp(-\theta(C^n_{j,N\to N}+C^n_{j,AH\to N}))|Y_j=y_j\right]}{\exp(-\theta(1-\epsilon'')(\mathbf{E}\left[C^n_{j,N\to N}|Y_j=y_j\right]+\mathbf{E}\left[C^n_{j,AH\to N}|Y_j=y_j\right]))}.
\end{aligned}
$$

We can apply Lemma 5 to obtain a bound on the moment generating function of $C^n_{j,AH\to N}$, which is given by

$$
\mathbf{E}\left[\exp(-\theta(C^n_{j,N\to N}+C^n_{j,AH\to N}))|Y_j=y_j\right]\leq\exp\left[M(y_j-n+1)\frac{H_j+N_j}{N_j}(e^{-\theta\frac{u}{M}}-1)\right].
$$

Using the same method as we did in Lemma 6, one can show then that

$$
\begin{aligned}
&\mathbf{P}\left(C^n_{j,N\to N}+C^n_{j,AH\to N}\leq(1-\epsilon'')(\mathbf{E}\left[C^n_{j,N\to N}|Y_j=y_j\right]+\mathbf{E}\left[C^n_{j,AH\to N}|Y_j=y_j\right]\right)\\
&\leq\exp\left(-\frac{\epsilon''}{2}\frac{M(y_j-n+1)(H_j+N_j)}{N_j}\right).
\end{aligned}
$$

Therefore, for $n\geq y_j-(K+1)/p$, we have

$$
\begin{aligned}
&\mathbf{P}\left(\bigcup_{n=y_j-(K+1)/p}^{y_j}\tilde{\mathcal{C}}_j^{n}\right)\\
\leq&\binom{y_j}{n}\exp\left(-\frac{\epsilon''}{2}\frac{M(y_j-n+1)(H_j+N_j)}{N_j}\right)\\
\leq&\beta'^{\gamma\frac{H_j+N_j}{N_j}}\binom{y_j}{n}\beta'^{(y_j-n)\frac{H_j+N_j}{N_j}}. \qquad (41)
\end{aligned}
$$

We are now ready to prove Theorem 9. Note that (39) holds for all $n \geq \delta_1 y_j$. We now combine 2) and 3) together and obtain the following from (38) (39), (40) and (41).

$$
\begin{aligned}
&\mathbf{P}\left(\bigcup_{n=\delta_1 y_j}^{y_j} \bigcup_{h=0}^{H_j} \tilde{\mathcal{B}}_j^{n,h} | Y_j = y_j\right) \\
=&\mathbf{P}\left(\bigcup_{n=\delta_1 y_j}^{y_j-(K+1)/p} \bigcup_{h=0}^{H_j} \tilde{\mathcal{B}}_j^{n,h} | Y_j = y_j\right) + \mathbf{P}\left(\bigcup_{n=y_j-(K+1)/p}^{y_j} \bigcup_{h=0}^{H_j} \tilde{\mathcal{B}}_j^{n,h} | Y_j = y_j\right) \\
\leq&\mathbf{P}\left(\bigcup_{n=\delta_1 y_j}^{y_j-(K+1)/p} \tilde{\mathcal{B}}_j^{n} | Y_j = y_j\right) + \mathbf{P}\left(\bigcup_{n=y_j-(K+1)/p}^{y_j} \tilde{\mathcal{B}}_j^{n} | Y_j = y_j\right) \\
&\quad + \mathbf{P}\left(\bigcup_{n=y_j-(K+1)/p}^{y_j} \tilde{\mathcal{C}}_j^{n}\right) \\
\leq& \sum_{n=\delta_1 y_j}^{y_j} \beta'^{\gamma} \binom{y_i}{n} e^{-\frac{\epsilon'}{2} M \delta_1 \gamma (y_i - n)} + \sum_{n=y_j-(K+1)/p}^{y_j} \beta'^{\gamma \frac{H_j+N_j}{N_j}} \binom{y_j}{n} \beta'^{(y_j-n)\frac{H_j+N_j}{N_j}} \\
=\leq& \beta'^{\gamma}(1+\beta'^{-\delta_1\gamma})^{y_j} + \beta'^{\gamma \frac{H_j+N_j}{N_j}} (1 + \beta'^{\delta_1 \gamma \frac{H_j+N_j}{N_j}})^{y_j}. \qquad (42)
\end{aligned}
$$

Now combine (37) and (42) we have

$$
\begin{aligned}
&\sum_{y_j=\lceil \gamma N_j \rceil}^{N_j-1} \binom{N_j-1}{y_j} p_j^y (1-p)^{N_j-1-y_j} \mathbf{P}\left(\bigcup_{n=0}^{y_j} \bigcup_{h=0}^{H_j} \tilde{\mathcal{B}}_j^{n,h} \middle| Y_j = y_j\right) \\
\leq& \sum_{y_j=\lceil \gamma N_j \rceil}^{N_j-1} \binom{N_j-1}{y_j} p_j^y (1-p)^{N_j-1-y_j} \Bigg[\beta^{\gamma}(1+\beta^{\delta_1\gamma - 1/y_i})^{y_i}(1+\beta^{\delta_1\gamma})^{H_i} \\
&\quad + \beta'^{\gamma}(1+\beta'^{\delta_1\gamma})^{y_j} + \beta'^{\gamma \frac{H_j+N_j}{N_j}} (1 + \beta'^{\delta_1 \gamma \frac{H_j+N_j}{N_j}})^{y_j}\Bigg] \\
\leq& \beta^{\gamma}(1+p\beta^{\delta_1\gamma - 1/\gamma N_j})^{N_j}(1+\beta^{\delta_1\gamma})^{KN_j} \\
&\quad + +\beta'^{\gamma}(1+p\beta'^{\delta_1\gamma})^{N_j} + \beta'^{\gamma \frac{H_j+N_j}{N_j}} (1 + p\beta'^{\delta_1 \gamma \frac{H_j+N_j}{N_j}})^{N_j}
\end{aligned}
$$

Recall that $M = \alpha_j \log(n)$, choose α_j, δ_1 large enough such that for $d > 1$

$$\alpha_j \geq d \frac{4u_{s,j}}{\delta_1 \gamma \epsilon''^2 u}.$$

We will have

$$\beta' = \frac{1}{N^{\frac{2du_{s,j}}{\gamma\delta_1 u}}},$$
$$\beta = \frac{1}{N^{\frac{2d}{\gamma\delta_1}}}.$$

Consequently

$$\beta^\gamma = \frac{1}{N^{\frac{2d}{\delta_1}}} \leq O\left(\frac{1}{N^{2d}}\right),$$
$$\beta'^\gamma = \frac{1}{N^{\frac{2du_{s,j}}{\delta_1 u}}} \leq O\left(\frac{1}{N^{2d}}\right),$$
$$\beta'^{\gamma\frac{H_j}{N_j}} = \frac{1}{N^{\frac{2du_{s,j}\frac{H_j+N_j}{N_j}}{\delta_1 u}}} \leq O\left(\frac{1}{N^{2d}}\right).$$

Moreover,

$$\begin{aligned}&(1 + p\beta^{\delta_1\gamma - 1/\gamma N_i})^{N_i}\\ =&(1 + pO(\frac{1}{N^{2d}}))^{N_i}\\ =&O(1),\end{aligned}$$

and similarly

$$(1 + p\beta'^{\delta_1\gamma})^{N_i} = O(1),$$
$$(1 + p\beta'^{\delta_1\frac{H_j+N_j}{N_j}\gamma})^{N_i} = O(1).$$

Therefore,

$$\sum_{y_j=\lceil\gamma N_j\rceil}^{N_j-1}\binom{N_j-1}{y_j}p_j^y(1-p)^{N_j-1-y_j}\mathbf{P}\left(\bigcup_{n=0}^{y_j}\bigcup_{h=0}^{H_j}\tilde{\mathcal{B}}_j^{n,h}\middle| Y_j=y_j\right)$$
$$\leq O\left(\frac{1}{N^{2d}}\right)O(1)+O\left(\frac{1}{N^{2d}}\right)O(1)+O\left(\frac{1}{N^{2d}}\right)O(1)$$
$$\leq O\left(\frac{1}{N^{2d}}\right).$$

Finally,

$$\mathbf{P}\left(C_{\min-\min}(s\to\mathcal{N}_j)\leq(1-\epsilon')\left\{u_{s,j},\frac{pN_j+H_j}{N_j}u+\frac{u_{s,j}}{N_j}\right\}\right)$$
$$\leq N_j\mathbf{P}\left(C_{\min}(s\to t)\leq(1-\epsilon')\left\{u_{s,j},\frac{pN_j+H_j}{N_j}u+\frac{u_{s,j}}{N_j}\right\}\right)$$
$$\leq N_j\mathbf{P}\left(\bigcup_{n=0}^{y_j}\bigcup_{h=0}^{H_j}\tilde{\mathcal{B}}_j^{n,h}\right)$$
$$\leq N_j\mathbf{P}(Y_j\leq\gamma N_j)$$
$$+N_j\sum_{y_j=\lceil\gamma N_j\rceil}^{N_j-1}\binom{N_j-1}{y_j}p_j^y(1-p)^{N_j-1-y_j}\mathbf{P}\left(\bigcup_{n=0}^{y_j}\bigcup_{h=0}^{H_j}\tilde{\mathcal{B}}_j^{n,h}\middle| Y_j=y_j\right)$$
$$\leq N_jO(\exp(-(1-\gamma)N_j))+N_jO\left(\frac{1}{N^2d}\right)$$
$$=O\left(\frac{1}{N^{2d-1}}\right)$$

B. Sufficient Channel

For sufficient channel j, if Y_j is given, then it is equivalent to a single channel with N_j peers, $\tilde{Y}_j=Y_j+H_j$ of which is ON. Then the result obtained for single model when the number of ON peers is given still holds, i.e., Lemma 6 and (19) still holds. Let the ϵ'' in Lemma 12 be $\epsilon''=1-\sqrt{1-\epsilon'}$ and choose the corresponding γ'. We have $\tilde{Y}_j\geq -H_j$. The probability that $Y_j\leq\gamma' N_j$ is of order $O(\exp-(1-\gamma')^2N_j)$.

Define $\tilde{\mathcal{B}}_j^n$ to be the event $\{C_j^n\leq(1-\epsilon')C_{f,j}$ for any cut among the $\binom{Y_j+H_j}{n}$ cuts $\}$. We have $\{C_j^n\leq(1-\epsilon')C_{f,j}\}$ implies $\{C_j^n\leq\sqrt{1-\epsilon'}\bar{C}_j^n\}$

given that $Y_j \geq \gamma' N_j$ since $\sqrt{1-\epsilon'} C_{f,j} \leq \bar{C}_j^n$. Let $\tilde{\gamma}' = \gamma' + \eta p$ so that $\gamma' \geq (1-\epsilon')(1-\eta)p$. Since $\tilde{Y}_j \geq \tilde{\gamma}' N_j$, letting $\beta'' = e^{-M \frac{u_{s,j}}{u} \frac{\epsilon''}{2}}$, from (19) we have

$$
\begin{aligned}
&\mathbf{P}\left(\bigcup_{m=0}^{y_j} \tilde{\mathcal{B}}_j^n \middle| \tilde{Y}_j = y_j \right) \\
\leq & 2\beta''^{\tilde{\gamma}'} \left(1 + \beta''^{\frac{\tilde{\gamma}'}{2}}\right)^{y_j}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\sum_{y_j=\lceil \tilde{\gamma}' N \rceil}^{N_j+H_j} \binom{N_j+H_j}{y_j} p_j^y (1-p)^{N_j+H_j-1-y_j} \mathbf{P}\left(\bigcup_{m=0}^{y_j} \tilde{\mathcal{B}}_m \middle| \tilde{Y}_j = y_j \right) \\
\leq & \sum_{y_j=\lceil \tilde{\gamma}' N \rceil}^{N_j+H_j-1} \binom{N_j+H_j-1}{y} p^y (1-p)^{N_j+H_j-1-y_j} \\
& \times 2\beta''^{\tilde{\gamma}'} \left(1 + \beta''^{\frac{\tilde{\gamma}'}{2}}\right)^{y} \\
\leq & \sum_{y_0=0}^{N_j+H_j-1} \binom{N_j+H_j-1}{y_j} 2\beta''^{\tilde{\gamma}'} \left(p(1 + \beta''^{\frac{\tilde{\gamma}'}{2}})\right)_j^y (1-p)^{N_j+H_j-1-y_j} \\
= & 2\beta''^{\tilde{\gamma}'} \left(1 + p\beta''^{\frac{\tilde{\gamma}'}{2}}\right)^{N_j+H_j-1}.
\end{aligned}
$$

Then using exactly the same method as in Theorem 1, if $M = \alpha_j \log(N)$ where $\alpha_j \geq \frac{4du_{s,j}}{\tilde{\gamma}' u \epsilon''^2}$, we have

$$
\begin{aligned}
&\mathbf{P}\left(C_{\min-\min}(s \to \mathcal{N}_j) \leq (1-\epsilon') \left\{ u_{s,j}, \frac{pN_j + H_j}{N_j} u + \frac{u_{s,j}}{N_j} \right\} \right) \\
\leq & O\left(\frac{1}{N^{2d-1}} \right).
\end{aligned}
$$

□

Remark: For insufficient channel, we need

$$
\alpha_j \geq d \frac{4u_{s,j}}{\delta_1 \gamma \epsilon''^2 u},
$$

where γ can be chosen close to p and δ_1 can chosen close to 1 according to Lemma 11. For sufficient channel, we need

$$\alpha_j \geq \frac{4du_{s,j}}{\tilde{\gamma}' u \epsilon''^2},$$

where $\tilde{\gamma}'$ can be chosen close to p according to Lemma 12. Therefore, for any channel j, α_j is close to

$$\frac{4du_{s,j}}{pu\epsilon''^2}.$$

Similar to the single-channel model, we observe that if we require larger streaming capacity or faster convergence rate, i.e., ϵ is smaller (consequently ϵ' and ϵ'' is smaller) or d is larger, we will need a larger α_j. If the probability that a peer is ON is reduced, i.e., p is reduced, we will also need a larger α_j.

References

[1] X. Zhang, J. Liu, B. Li, and Y.-S. Yum, "Coolstreaming/donet: a data-driven overlay network for peer-to-peer live media streaming," in *Proceedings IEEE INFOCOM*, vol. 3, 2005, pp. 2102 – 2111.

[2] X. Hei, C. Liang, J. Liang, Y. Liu, and K. Ross, "A measurement study of a large-scale p2p iptv system," *IEEE Transactions on Multimedia*, vol. 9, no. 8, pp. 1672 –1687, Dec. 2007.

[3] T. Silerston and O. Fourmaux, "Measuring p2p iptv systems," in *Proceedings of NOSSDAV'07*, June 2007.

[4] C. Wu, B. Li, and S. Zhao, "Exploring large-scale peer-to-peer live streaming topologies," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 4, no. 3, pp. 1–23, 2008.

[5] W. Liang, J. Bi, R. Wu, Z. Li, and C. Li, "On characterizing ppstream: Measurement and analysis of p2p iptv under large-scale broadcasting," in *IEEE GLOBECOM 2009*, Nov. 2009, pp. 1 –6.

[6] R. Kumar, Y. Liu, and K. Ross, "Stochastic fluid theory for p2p streaming systems," in *IEEE INFOCOM*, May 2007, pp. 919 –927.

[7] L. Massoulié and A. Twigg, "Rate-optimal schemes for peer-to-peer live streaming," *Perform. Eval.*, vol. 65, no. 11-12, pp. 804–822, 2008.

[8] C. Feng and B. Li, "On large-scale peer-to-peer streaming systems with network coding," in *Proceeding of the 16th ACM international conference on Multimedia*, Vancouver, British Columbia, Canada, 2008, pp. 269–278.

[9] T. Bonald, L. Massoulié, F. Mathieu, D. Perino, and A. Twigg, "Epidemic live streaming: optimal performance trade-offs," in *ACM SIGMETRICS '08*, 2008, pp. 325–336.

[10] S. Liu, R. Zhang-Shen, W. Jiang, J. Rexford, and M. Chiang, "Performance bounds for peer-assisted live streaming," in *ACM SIGMETRICS '08*, 2008, pp. 313–324.

[11] S. Sengupta, S. Liu, M. Chen, M. Chiang, J. Li, and P. A. Chou, "Streaming capacity in peer-to-peer networks with topology constraints," *submitted to IEEE Transaction on Information Theory*, 2009.

[12] S. Liu, M. Chen, S. Sengputa, M. Chiang, J. Li, and P. A. Chou, "P2p streaming capacity under node degree bound," in *Proc. IEEE ICDCS*, Genoa, Italy, June 2010.

[13] M. Castro, P. Druschel, A.-M. Kermarrec, A. Nandi, A. Rowstron, and A. Singh, "Splitstream: high-bandwidth multicast in cooperative environments," in *SOSP '03*, Bolton Landing, NY, 2003, pp. 298–313.

[14] V. Pai, K. Kumar, K. Tamilmani, V. Sambamurthy, A. E. Mohr, and E. E. Mohr, "Chainsaw: Eliminating trees from overlay multicast," in *IPTPS*, 2005, pp. 127–140.

[15] D. Wu, C. Liang, Y. Liu, and K. Ross, "View-upload decoupling: A redesign of multi-channel p2p video systems," in *IEEE INFOCOM*, 2009, pp. 2726 –2730.

[16] D. Wu, Y. Liu, and K. Ross, "Queuing network models for multi-channel p2p live streaming systems," in *IEEE INFOCOM*, 2009, pp. 73 –81.

[17] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388 –404, Mar 2000.

[18] L. Massoulié, A. Twigg, C. Gkantsidis, and P. Rodriguez, "Randomized decentralized broadcasting algorithms," in *Proceedings of IEEE INFOCOM*, 2007, pp. 1073–1081.

[19] A. Ramamoorthy, J. Shi, and R. Wesel, "On the capacity of network coding for random networks," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2878 –2885, Aug. 2005.

[20] J. Edmonds, "Edge-disjoint branchings," *Combinatorial Algorithms, ed. R. Rustin*, pp. 91–96, 1973.

[21] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204 –1216, Jul 2000.

[22] S.-Y. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371 –381, Feb. 2003.

[23] S. Ross, *A first course in probability*, 5th ed. Upper Saddle River, NJ: Prentice-Hall, 1998.

[24] A. D. Barbour, L. Holst, and S. Janson, *Poisson Approximation*. Oxford University Press, 1992.

[25] S. Janson, "Large deviation inequalities for sums of indicator variables," Uppsala Universitet, Tech. Rep., 1994.

[26] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963.

[27] J. Hao and J. B. Orlin, "A faster algorithm for finding the minimum cut in a graph," in *SODA '92: Proceedings of the third annual ACM-SIAM symposium on Discrete algorithms*, Orlando, Florida, 1992, pp. 165–174.