Neighbor selection for proportional fairness in P2P networks

Martín Zubeldía, Andrés Ferragut*, Fernando Paganini

Universidad ORT Uruguay, Montevideo, Uruguay

Abstract

This paper analyzes reciprocation strategies in peer-to-peer networks from the point of view of the resulting resource allocation. Our stated aim is to achieve through decentralized interactions a weighted proportionally fair allocation. We analyze the desirable properties of such allocation, as well as an ideal proportional reciprocity algorithm to achieve it, using tools of convex optimization. We then seek suitable approximations to the ideal allocation which impose practical constraints on the problem: numbers of open connections per peer, with transport layer-induced bandwidth sharing, and the need of random exploration of the peer-to-peer swarm. Our solution in terms of a Gibbs sampler dynamics characterized by a suitable energy function is implemented in simulation, comparing favorably with a number of alternatives.

Keywords: Peer-to-peer networks, Resource allocation, Distributed algorithms, Performance evaluation

1. Introduction

Peer-to-peer (P2P) file sharing networks constitute a popular alternative to distribute content over the Internet. The main principle behind these networks is that participating peers, which are downloading some content, also contribute their upload bandwidth to simultaneously serve other peers. Thus service capacity scales with demand, a valuable self-scaling property.

The allocation of such capacity is however a non-trivial issue, particularly when peers have non-homogeneous parameters. Assume that μ_i, $i = 1, \ldots, N$ are the maximum upload capacities of a set of peers, and that there are no other bottlenecks in the network. For *efficiency* purposes we want the entire $\sum_i \mu_i$ distributed among downloaders, but *fairness* is also important: if a specific peer is not getting a fair share it may have incentives to reduce its contribution μ_i.

These questions have been analyzed by many researchers in the past, often combined with efforts to characterize the behavior of prevailing P2P protocols such as BitTorrent [1]. In [2], the authors explore the design space of possible resource allocations, showing relevant tradeoffs between efficiency and fairness in the allocation rule. From an efficiency perspective, [3, 4, 5] seek to minimize the content distribution time, assuming a fixed number of participants in the network. However, when incentives come into play, minimizing the average download time may not be a good objective, since peers with above average bandwidth are equalized with the rest of the system. From the point of view of incentives, it is worth noting that BitTorrent has a *reciprocity* mechanism similar to *tit-for-tat* procedures from game theory. In this system, each peer allocates its uplink bandwidth to those from which it received the most in the previous exchanges, thus incentives for cooperation are provided (see [6]).

In this paper we adopt the position that, whenever feasible, a peer should receive from fellow peers as much as it gives (in line with the notion of fairness in [2]), because it provides the most direct, transparent incentives. If such perfect reciprocation is not feasible, the most natural approximation is a *proportional fairness* criterion (as advocated in [7] for Internet bandwidth sharing), but with *weights* corresponding to the bandwidths μ_i. This turns out to be equivalent to hav-

*Corresponding Author
Email address: ferragut@ort.edu.uy
Postal address: Cuareim 1451, 11100
Tel./Fax: (+598) 29021505 / 29081370

ing the download and upload rate vectors be close in the sense of Kullback-Leibler divergence [8].

The subject of proportional fairness in P2P reciprocity was first discussed in [9] and further developed in [6, 10]. A natural proposal in this regard is a decentralized *proportional reciprocity* algorithm, where peers allocate upload resources in proportion to bandwidth received. Mathematically this algorithm is very closely related to a matrix renormalization technique studied classically by Sinkhorn [11, 12], and by many others since (see e.g. [13]). From this literature convergence results for this iteration follow [10], with economic interpretations. Indeed this path is further extended in [14] to more general economic markets [15]. In this regard, a first objective of this paper is to provide a self-contained treatment of the properties of this ideal reciprocity mechanism in peer-to-peer networks, and its relationship with the desired proportional fairness, using techniques of convex optimization. While some of the results we present could be traced to this rich literature, our contribution is to give a detailed account of how these results are applicable to peer-to-peer systems while also highlighting additional facts about this powerful method. These aspects are covered in Section 2.

These ideal reciprocity schemes are not easily taken into practice, since they require a large number of open connections and fine-grained control of connection rates which in typical scenarios are governed by the transport layer, not under control of the application. To impose such practical design constraints (number of connections, transport layer bandwidth sharing) an *energy function* is introduced in Section 3, which is zero under ideal reciprocity, and which a practical scheme should try to minimize. Although optimizing this energy under the aforementioned limitations has combinatoric complexity, we identify cases where the minimum is indeed achievable. We further show that if each peer seeks to myopically reduce its portion of the energy, a tit-for-tat structure similar to BitTorrent's comes out naturally.

The final step is to devise an algorithm that introduces randomness in peer selection, to allow the system to explore the set of peers which will in practice vary with time. For this task we introduce in Section 4 a Gibbs sampler dynamics [16] designing a Markov process guided by a potential related to the energy function. In this regard, it is worth noting that [17, 18] have introduced this technique in the study of P2P systems. As we will explain, there are differences between the two proposals reflected in the potential function used and the optimization objective.

In Section 5, we analyze the convergence speeds of the proposed algorithms, establishing bounds on the mixing times of the underlying Markov processes. The resulting neighbor selection algorithm was also tested in simulation and compared to other existing protocols, such as standard BitTorrent, the PropShare algorithm from [6] as well as the proposal in [17], performing well against the alternatives in terms of reciprocity and fairness. These results are reported in Section 6.

Conclusions and given in Section 7, and some proofs are relayed to Appendices. Partial results leading up to this paper were presented in [19].

2. Proportional fairness in P2P systems

We begin by defining some notation. A set of N peers share information through a connectivity graph G: two peers are neighbors in this graph if they can exchange information. Let $A = (a_{ij})$ denote the adjacency matrix of this graph, which we assume *symmetric*, with $a_{ii} = 0$ for every i, and with no rows of zeros (each peer can exchange with at least one other). Assume also that every peer has an offered uplink capacity μ_i to share with other peers.

Define a *resource sharing* matrix $Z \in \mathbb{R}_+^{N\times N}$ in which z_{ij} represents the offered throughput from peer i to peer j. Z should satisfy the following restrictions:

$$z_{ij} \geq 0, \quad z_{ij} = 0 \text{ if } a_{ij} = 0, \tag{1a}$$

$$\sum_j z_{ij} = \mu_i \quad \forall i. \tag{1b}$$

This states that peers do not exchange information when not connected, and that each peer contributes its entire upload capacity. The second condition assumes there are no other network constraints (internal network capacity, download capacity), and therefore we are seeking an *efficient* allocation where all upload bandwidth is used. This is common in practice, where the main constraint is in the access capacity of peers.

The resource sharing matrix determines the amount of information received by each peer from fellow peers, given by the following expression:

$$r_j(Z) = \sum_i z_{ij} \quad \forall j. \tag{2}$$

Equivalently, in matrix form, equations (1) and (2) can be summarized as:

$$Z\mathbf{1} = \mu, \quad \mathbf{1}^T Z = r^T,$$

where $\mu, r, \mathbf{1} \in \mathbb{R}^N$ are interpreted as column vectors, the latter being the vector of ones, and ^{T} denotes transpose. The questions of interest are:

(i) What is a suitable resource allocation Z?

(ii) Can it be found through decentralized peer interactions?

2.1. Proportionally fair allocation

We now consider the first of the preceding issues by stating our allocation objective. As analyzed in [2], there is a fundamental tradeoff between efficiency and fairness. A maximum throughput allocation (i.e. one that minimizes the mean download time) can punish peers that contribute a high upload rate, giving an incentive to reduce the offered μ_i. The alternative advocated in [2] is an allocation in which each peer receives as much as it contributes, thus providing a transparent incentive for peers to fully cooperate. This fairness objective however is not always feasible within the constraints, so we choose to state our objective in terms of the following convex optimization problem:

Problem 1 (Proportionally fair allocation)**.** *Given A and μ, choose Z as a solution of*

$$\max_Z \sum_j \mu_j \log\left(r_j(Z)\right), \text{ subject to (1).}$$

The above criterion is an instance of (weighted) *proportional fairness*, extensively studied in the realm of Internet resource allocation [7]. Here, we choose each node's weight as its own contribution to the network, in order to reward those peers contributing more bandwidth[1].

An alternative way of expressing the above objective is to consider the *Kullback-Leibler divergence* (see e.g. [8]):

$$D(\mu||r) := \sum_j \mu_j \log\left(\frac{\mu_j}{r_j}\right). \tag{3}$$

As long as both vectors have the same sum (which happens in this case since $\sum_j r_j = \sum_{i,j} z_{ij} = \sum_i \mu_i$), the K-L divergence is always non-negative and only zero if $r = \mu$.

Our proportional allocation is equivalent to minimizing $D(\mu||r(Z))$ among feasible Z; thus, if feasible, each peer will receive as much as it contributes to the network, which creates a strong cooperation incentive. In case of infeasibility an approximation is sought in the sense of minimal K-L divergence between download and upload vectors.

[1] Proportional fairness can also be interpreted in terms of bargaining theory [20]. The weighted case corresponds to an asymmetric Nash solution [21] where the μ_i's represent bargaining powers.

Example 1. *Consider a full-mesh network where $a_{ij} = 1$ whenever $i \neq j$, and assume a descending order for upload capacities ($\mu_1 \geq \mu_2 \geq \ldots \mu_N$). In this case, a necessary and sufficient condition obtained in [9] for feasibility of $r(Z) = \mu$ is that $\mu_1 \leq \sum_{i=2}^N \mu_i$, i.e. no peer's capacity is greater than the sum of the remaining ones. This is arguably a typical scenario under a relatively homogeneous population of peers.*

We now characterize the proportionally fair allocation by writing the Lagrangian with respect to the upload capacity constraints,

$$L(Z, p) = \sum_j \mu_j \log\left(\sum_i z_{ij}\right) + \sum_{i=1}^N p_i \left(\mu_i - \sum_j z_{ij}\right). \tag{4}$$

At optimality, all multipliers (prices) p_i^* must be strictly positive. This is because the objective is strictly increasing in the free variables z_{ij}, which are independent per row (constraint); so if we were to relax (1b) to an inequality version, all constraints would be strictly active. Given such price vector $p^* = (p_i^*) > 0$, let us characterize the optimality conditions for $Z^* = \arg\max L(Z, p^*)$ over Z satisfying (1a). Noting that

$$\frac{\partial L}{\partial z_{ij}} = \frac{\mu_j}{r_j(Z)} - p_i^*,$$

we must have for any pair of neighbors (i, j):

$$\text{either } z_{ij}^* = 0, \quad \frac{\mu_j}{r_j^*} \leq p_i^*; \tag{5a}$$

$$\text{or } z_{ij}^* > 0, \quad \frac{\mu_j}{r_j^*} = p_i^*. \tag{5b}$$

Since for each j, $r_j^* = \sum_i z_{ij}^* > 0$, case (5b) must hold in at least one entry per column; at optimality peer j is served only from connected peers of minimum price. If ideal reciprocity $r = \mu$ is feasible, then all prices must be equal to one. More generally let

$$\pi_j^* := \frac{1}{\min\{p_i^* : a_{ij} = 1\}}, \tag{6}$$

then we have $r_j/\mu_j = \pi_j^*$, i.e. π_j^* determines the level of reciprocation that peer j is obtaining from the network.

Remark 1. *The objective of Problem 1 is not strictly concave in the variable Z, so it need not have a unique optimum. However it is strictly concave in $r(Z)$, so the optimum column sums r^* are uniquely determined. From* (5b) *it follows that optimal prices are uniquely determined as well.*

Example 2. *Return to the previous example but assume that* $\mu_1 > \sum_{i=2}^N \mu_i$*, so that it is infeasible to perfectly reciprocate. Define* $\kappa = \frac{\mu_1}{\sum_{i=2}^N \mu_i} > 1$*; it is easily checked that*

$$Z^* = \begin{bmatrix} 0 & \kappa\mu_2 & \dots & \kappa\mu_N \\ \mu_2 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mu_N & 0 & \dots & 0 \end{bmatrix}, \quad p^* = \begin{bmatrix} 1/\kappa \\ \kappa \\ \vdots \\ \kappa \end{bmatrix} \tag{7}$$

satisfies the conditions (5) *and thus is the proportionally fair solution. The best approximation to perfect reciprocation is to have peers* $2, \dots N$ *give the most possible rate to peer 1, who reciprocates with a factor* κ *of increase.*

Remark 2. *Note that in the example* $r_j^* = p_j^* \mu_j$ *for every* j*, or equivalently* $\pi_j^* = p_j^*$*. This implies from* (6) *that there is an* inverse *relation between a peer's own price and the prices of those peers with which it interacts at optimality. As we will see below in Proposition 2, this a general result and is a consequence of the uniqueness stated in Remark 1.*

What happens with the feasibility of perfect reciprocation when a graph structure A is given? This question amounts to finding a matrix Z of structure A whose rows and columns have sum μ, and as such has been studied in the literature of matrix renormalization since Sinkhorn [11, 12]. We extract from it (see [13] and references therein) the following characterization, which generalizes the case of the example.

Proposition 1. *Given an adjacency matrix* A *and a vector of capacities* μ*, there exists a matrix* Z *satisfying* (1) *and* $r(Z) = \mu$*, if and only if for every* zero minor *of* A*, i.e every* $(I, J) \subset \{1, \dots, N\}^2$ *such that* $a_{ij} = 0 \ \forall i \in I, j \in J$*, the following inequality holds:*

$$\mu_J := \sum_{j \in J} \mu_j \leq \sum_{i \notin I} \mu_i =: \mu_{I^c}. \tag{8}$$

In words: whenever a set of peers J is not allowed to receive data from another set I, J's total capacity should not exceed the total capacity of the complement of I. It is not difficult to see the necessity of the condition for perfect reciprocation. A proof of sufficiency based on optimization arguments is given in the Appendix.

2.2. Proportional reciprocity algorithm

We now turn to the question of achieving the proportionally fair allocation through a decentralized, iterative algorithm. At each iteration, each peer chooses how to split its bandwidth among neighbors, only using as information the rate from prior exchanges.

A natural proposal [6, 9, 10] is to reciprocate other peers in proportion to bandwidth received in the previous step, as follows:

$$z_{ij}(t+1) = \mu_i \frac{z_{ji}(t)}{r_i(t)}; \tag{9}$$

we call this the *proportional reciprocity* algorithm. It says that in the next step, peer i allocates to peer j a fraction of its bandwidth μ_i equal to the proportion of bandwidth received from peer j in the previous step.

In matrix form we have $Z(t+1) = \mathcal{R}[Z(t)]$ where the reciprocity mapping is defined by

$$\mathcal{R}[Z] := \operatorname{diag}\left(\frac{\mu_i}{r_i(Z)}\right) \cdot Z^T. \tag{10}$$

This is also equivalent to first re-normalizing the *columns* of Z to have sum μ, and subsequently transposing. Thus except for this transpose, the preceding iteration amounts to one step in the Sinkhorn algorithm [12] of alternating row and column renormalization, whose convergence has been extensively studied. This observation was already made in [10], and used to establish convergence of the even and odd subsequences of the above iteration, with a rather technical proof. In subsequent work [14], the author extended the above kind of proportional response dynamics to the more general context of Fisher market from economics [15]. Here the market has two sets of variables, bids and quantities, which are successively updated in what would correspond to two steps of the previous algorithm.

Our aim in this section is to extract from the above literature a streamlined account of the convergence properties of the proportional reciprocity iteration as applied to the peer-to-peer problem. The main result is Theorem 4 which states that the rate received by each peer converges to the optimum of Problem 1. In this sense, Theorem 4 is an adaptation of the main result in [14] to the case of bandwidth exchanges, and the proof is directly inspired by this paper.

We begin by characterizing the following property of the optima of Problem 1.

Proposition 2. *For a given* A *and* μ*, let* Z^* *denote any solution of Problem 1. Then* Z^* *is a fixed point of* $\mathcal{R}^2$*, square of the reciprocity iteration* (10)*. Furthermore* $Z^+ := \mathcal{R}[Z^*]$ *is also a solution of Problem 1.*

Proof. Let p^* be the unique dual optimal multiplier vector associated with Problem 1; (5) implies

$$\frac{\mu_j}{r_j^*} z_{ij}^* = p_i^* z_{ij}^* \quad \forall i, j. \tag{11}$$

Now according to (9) the left-hand side of (11) is z^+_{ji}, the (j,i)-entry of $Z^+ := \mathcal{R}(Z^*)$. Adding over j we obtain

$$r^+_i = \sum_j z^+_{ji} = p^*_i \sum_j z^*_{ij} = p^*_i \mu_i. \tag{12}$$

Now apply the reciprocity mapping again, $Z^{++} = \mathcal{R}[Z^+]$, writing

$$z^{++}_{ij} = \mu_i \frac{z^+_{ji}}{r^+_i} = \frac{z^+_{ji}}{p^*_i} = z^*_{ij}. \tag{13}$$

Here the second equality uses (12) and the last (11). Therefore $Z^{++} = Z^*$, a fixed point of $\mathcal{R}^2$.

Consider now a fixed index j, and the set

$$I_j = \left\{ i : z^+_{ji} > 0 \right\} = \left\{ i : z^*_{ij} > 0 \right\};$$

all $i \in I_j$ must have minimum price p^*_i among neighbors of j, so defining π^*_j as in (6) we conclude from (12) that

$$\pi^*_j = \frac{\mu_i}{r^+_i} \quad \forall i : z^+_{ji} > 0;$$
$$\pi^*_j \geq \frac{\mu_i}{r^+_i} \quad \forall i : z^+_{ji} = 0, a_{ji} = 1;$$

These are precisely the optimality conditions (5) for the pair (Z^+, π^*), therefore Z^+ is also an optimal point. Note finally that by uniqueness of optimal prices we must have $\pi^* = p^*$. □

Remark 3. *A consequence of the above is that $r^*_j = p_j \mu_j$ for every peer. At the same time recall from* (5b) *that $r^*_j = \mu_j \frac{1}{p_i}$ whenever $z^*_{ij} > 0$, so at optimality a peer can only receive/send rate to another of* inverse *price.*

In general, Z^+ need not be equal to Z^*, i.e. Z^* need not be a fixed point of the map $\mathcal{R}$ itself (in Example 2 it is actually the same point). However we have the following:

Corollary 3. *If Z^* is an optimum of Problem 1, and $Z^+ := \mathcal{R}[Z^*]$. Then $\tilde{Z} = \frac{Z^* + Z^+}{2}$ is another optimum and fixed point of $\mathcal{R}$. In the special case where $r = \mu$ is feasible, then there is always a symmetric optimal solution.*

Proof. Z^*, Z^+, both satisfy the conditions

$$Z\mathbf{1} = \mu, \quad \mathbf{1}^T Z = (r^*)^T$$

for the same (unique) optimal rate vector r^*. Hence by linearity $\tilde{Z}$ satisfies the same, and is thus an optimal allocation. Also (10) gives

$$\mathcal{R}[\tilde{Z}] = \operatorname{diag}\left(\frac{\mu_i}{r^*_i}\right) \cdot \tilde{Z}^T = \operatorname{diag}\left(\frac{\mu_i}{r^*_i}\right) \cdot \frac{[Z^*]^T + [Z^+]^T}{2}$$
$$= \frac{1}{2}\mathcal{R}[Z^*] + \frac{1}{2}\mathcal{R}[Z^+] = \frac{1}{2}(Z^+ + Z^{++}) = \tilde{Z}.$$

Finally note that for the case $r^* = \mu$, $Z^+ = [Z^*]^T$ therefore $\tilde{Z}$ is a symmetric matrix. □

The final result is that repeated application of the proportional reciprocity map converges to the set of optimal proportionally fair allocations, under the only requirement that the initial condition must enable all allowable exchange options.

Theorem 4. *Consider a trajectory of the proportional reciprocity iteration $Z(t+1) = \mathcal{R}[Z(t)]$ starting from an initial condition $Z(0)$ satisfying* (1)*, with also $z_{ij}(0) > 0$ whenever $a_{ij} = 1$. Then*

$$\lim_{k\to\infty} Z(2k) = Z^*, \quad \lim_{k\to\infty} Z(2k+1) = Z^+,$$

where both Z^* and Z^+ are optimal points of Problem 1. In particular, the rate sequence $r(Z(t))$ converges to the (unique) proportionally fair allocation.

Proof is given in the Appendix.

3. Approximating proportional reciprocity by neighbor selection

The proportional response is a decentralized algorithm that achieves our target bandwidth allocation in a P2P network. However, some features of this algorithm are not adapted to practical network scenarios.

A first difficulty is that in order to have a diverse set of exchange opportunities, peers are typically in contact with a moderately large number of neighbors. Assigning a positive rate $z_{ij} > 0$ to each such neighbor on a permanent basis requires maintaining a large number of open connections, with its associated overhead. Secondly, if the underlying transport protocol is TCP, the application layer does not have a simple way to control its rate to each of its neighbors. Instead, under normal circumstances (bottleneck in the upload) TCP will split the upload bandwidth uniformly between active outgoing connections[2]. Finally, we have given a completely deterministic algorithm in which exchange partners remain fixed throughout; as such it lacks the ability to explore the different peering options as the network evolves.

Before moving on to incorporate such restrictions in the analysis, we briefly review how things are handled in BitTorrent [1], the most popular P2P protocol. BitTorrent peers open a maximum amount (usually four) of connections to other peers. The neighbor selection algorithm has essentially two parts:

[2]The main source of (uncontrolled) differences between TCP rates would be round-trip-times; we will ignore this issue in what follows.

- The tit-for-tat part: each peer, every 10sec, chooses to upload to the three peers which, in the preceding 20sec, have given it the most data.
- The optimistic unchoke: each peer, every 30sec, opens a connection to a random peer for 30sec.

Thus at any time every peer is uploading to 4 peers at most, 3 of them chosen based on a ranking of the received bandwidths and the other one at random. We see that this protocol has incorporated the implementation restrictions discussed above (small number of connections, random exploration), and also *some* notion of reciprocity, but in principle not the proportional response we were seeking. In this regard, [6] argues that tit-for-tat behaves like a bandwidth auction, with weaker incentives to contribute than perfect reciprocity ($r = \mu$). Nevertheless such result has been approximately observed in empirical studies [22], which reveal that tit-for-tat tends to form *cliques* of peers with similar bandwidth parameters. Now the optimistic unchoke portion of the protocol, which is essentially egalitarian in its distribution, departs completely from proportionality and even introduces the possibility of free-riding by peers that live from these optimistic connections.

There is thus room left for exploring alternatives to the BitTorrent neighbor selection, that will more closely reflect our design objective of proportional allocation, within the practical constraints that have been identified. In this Section we will address two of these constraints:

1. Each peer can only open a maximum amount of N_0 connections.
2. The upload capacity μ_i of each peer is equally distributed between all outbound connections.

We postpone to the following section the issue of incorporating randomness in peer selection. We will also restrict our attention to the case of feasible perfect reciprocity. As argued before, this is not a very restrictive assumption for peers with similar capacities and degrees of connectivity:

Assumption 1. *The connectivity graph A and the upload bandwidths $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_N$ satisfy the conditions of Proposition 1. There are no other bottlenecks in the network.*

Under this assumption, we have seen in the previous Section that not only is perfect reciprocity $r = \mu$ feasible, but that it can be achieved with a *symmetric* matrix Z, as shown in Corollary 3. This means that there is balance of bandwidth not only in the global outcomes but also in peerwise interactions.

We now begin to incorporate the discrete restrictions imposed by the number N_0 of peer connections, and the equal bandwidth between them. At this point it is convenient to factor out the peer bandwidths and introduce a matrix X with coefficients in $\{0, \frac{1}{N_0}\}$ that stores the neighboring configurations in terms of the fractions x_{ij} of its own bandwidth that peer i allocates to each peer j. From it, the rate allocation can be obtained as

$$Z = \operatorname{diag}(\mu_i) X.$$

From a structural point of view, X must have the same hard zeros imposed by the connectivity matrix A. This is expressed by saying that X belongs to the following set:

$$\Lambda^S = \left\{ X \in \left\{0, \tfrac{1}{N_0}\right\}^{N^2} : x_{ij} = 0 \text{ if } a_{ij} = 0; \sum_{j \in S} x_{ij} = 1 \right\}.$$

3.1. Energy driven allocations

As a means to study the impact of discrete constraints on the desired reciprocity, we will introduce an *energy function* $\mathcal{E}(X)$, sum of squares of the peerwise discrepancies in exchange rates, as follows:

$$\mathcal{E}(X) = \frac{1}{2} \sum_{i,j} (\mu_i x_{ij} - \mu_j x_{ji})^2. \tag{14}$$

This energy is thus equal to $\frac{1}{2} \sum_{i,j} (z_{ij} - z_{ji})^2$, and has a minimal value of zero for symmetric allocations. At this point it may not be entirely clear why we define this new energy instead of just using the Kullback-Leibler divergence, but it will become clear in section 4.

We would like to minimize the energy $\mathcal{E}(X)$ over $X \in \Lambda^S$. This is a discrete optimization problem which has no explicit solution, but in certain cases we can identify interesting properties. One such case is where there are repeated values in the sequence of upload bandwidths $\{\mu_i\}$, of enough multiplicity with respect to the connectivity parameter N_0. We state the following result:

Proposition 5. *Suppose that N_0 is even. Divide the set of peers into K groups with the same upload bandwidth $\mu^{(k)}$ for each member of group k. If every group has $N_k > N_0$ peers, there exists at least one configuration X^* such that $\mathcal{E}(X^*) = 0$, resulting in the proportional allocation.*

Proof. As we have groups of peers with the same bandwidth, $\mathcal{E}(X^*) = 0$ holds for a configuration X^* where peers interact only within their group, provided each receives from N_0 others. Formally, for each group of N_k

peers we have to find a N_0-regular graph (undirected, where every node has N_0 neighbors). Fortunately, the existence of such graphs is a known result in graph theory when N_0 is even [23] (for instance, a solution is a so-called Cayley graph). As a result, every group of N_0-regular graphs would make the energy equal to 0 and thus yield a proportional allocation. □

Remark 4. *A N_0-regular graph is fundamentally different to the formation of* cliques *(complete subgraphs) which has been shown to be a property of the BitTorrent tit-for-tat mechanism [2]. A N_0-order clique has by definition $N_0 + 1$ nodes; so unless the cardinality of the repeated bandwidths happens to coincide with this value, the result will be different. In fact, an algorithm that forces cliques of fixed size can lead to severe loss in proportional reciprocity, as portrayed in the following example.*

Example 3. *Suppose that $N_0 = 4$ and $N = 15$, where seven peers have $\mu_i = 10$ and the other eight have $\mu_i = 1$. The method of Proposition 5 forms two regular graphs and achieves proportional reciprocity. If instead we form 3 cliques of size $N_0 + 1 = 5$, only two of these can involve homogeneous peers and deliver proportional reciprocity. The third clique will have two fast peers ($\mu_i = 10$) and three slow peers ($\mu_i = 1$), resulting in an allocation of $r = 3.25$ for the fast peers, and $r = 5.5$ for the slow ones. Not only is proportionality broken, but the fast peers are being penalized!*

One might think that having exact repetition of the upload bandwidths is a very special case. However, if peers can be grouped in classes with *approximately* equal bandwidth, we can bound the minimum energy as follows.

Proposition 6. *Suppose that N_0 is even. Divide the set of peers into K groups, where the bandwidths $\{\mu_i\}$ for peers in each group occupy an interval of length δ. If every group has $N_k > N_0$ peers, there exists at least one configuration X^* such that $\mathcal{E}(X^*) \leq \delta^2 \frac{N}{2N_0}$.*

Proof. Consider the same X^* constructed in Proposition 5. Write the total energy as $\mathcal{E}(X^*) = \sum_k \mathcal{E}_k(X^*)$, adding the energy contributions of each disconnected group. For group k we have $N_k N_0$ mutual connections, each with energy

$$\frac{1}{2}(\mu_i x_{ij} - \mu_j x_{ji})^2 \leq \frac{\delta^2}{2N_0^2}.$$

Therefore $\mathcal{E}_k(X^*) \leq \delta^2 \frac{N_k}{2N_0}$ and the result follows from $\sum_k N_k = N$. □

Thus suggests that grouping peers in subsets of similar bandwidth, of any size greater than N_0, is a good strategy to approximate the goal of proportional reciprocity. The size of the classes will be a function of the existing set of μ_i's; the flexibility of going beyond cliques of size $N_0 + 1$ can lead to significant improvements.

Remark 5. *We note, however, a limitation of seeking perfect reciprocity with the above discrete connections, which did not occur with optimal allocations Z^* of the previous section. In the discrete case we are favoring connectivity only between peers of the same (or similar) bandwidth, thus reducing the level of file sharing, which could be detrimental from the point of view of piece diversity. This issue will be later mitigated by adding randomness.*

3.2. Decentralized energy minimization and tit-for-tat

The question to ask at this point is: can the energy by minimized by a *decentralized* algorithm? Given the combinatoric nature of the problem we do not expect the global optimum to be computable, but a reasonable heuristic is to have each peer i choose its outgoing connections seeking to myopically reduce its own portion of the energy,

$$\mathcal{E}_i(X) := \sum_j (\mu_i x_{ij} - \mu_j x_{ji})^2.$$

In this minimization we assume given the rates x_{ji} received by peer i, and we introduce the notation $J^{in} = \{j : x_{ji} \neq 0\}$ for the set of peers from which peer i is currently receiving data. Let N^{in} be the cardinality of this set, and note that there are no a priori constraints on it, in principle $0 \leq N^{in} \leq N - 1$.

Since peer i will divide its bandwidth uniformly among its N_0 outgoing connections, the myopic optimization is just to choose the set $J^{out} = \{j : x_{ij} \neq 0\}$, of cardinality N_0, to minimize the energy portion $\mathcal{E}_i(X)$. The following proposition characterizes the optimal configuration.

Proposition 7. *Given a set J^{in} of peers uploading to i, a configuration X^* minimizes the local energy $\mathcal{E}_i(X)$ if and only if it solves*

$$\max_{J^{out}} \sum_{J^{in} \cap J^{out}} \mu_j. \tag{15}$$

Proof. For convenience we will denote by $\tilde{\mu}_j := \frac{\mu_j}{N_0}$, the fraction of bandwidth allocated in a single connection

from peer j. The local energy of a given configuration X can then be expressed as follows:

$$\mathcal{E}_i(X) = \sum_{j \in J^{in} \cap J^{out}} \left(\tilde{\mu}_i - \tilde{\mu}_j\right)^2 + \sum_{j \in J^{in} \setminus J^{out}} \tilde{\mu}_j^2 + \sum_{j \in J^{out} \setminus J^{in}} \tilde{\mu}_i^2.$$

Expanding the square $\left(\tilde{\mu}_i - \tilde{\mu}_j\right)^2 = \tilde{\mu}_i^2 + \tilde{\mu}_j^2 - 2\tilde{\mu}_i\tilde{\mu}_j$ and rearranging terms leads to the equivalent expression

$$\mathcal{E}_i(X) = \sum_{j \in J^{in}} \tilde{\mu}_j^2 + \sum_{j \in J^{out}} \tilde{\mu}_i^2 - 2 \sum_{j \in J^{in} \cap J^{out}} \tilde{\mu}_i\tilde{\mu}_j.$$

The first term above is given, and the second is fixed at $N_0\tilde{\mu}_i^2$ for all allowable configurations, so only the third term can be minimized by choice of J^{out}; noting that $\tilde{\mu}_i$ is fixed, and $\mu_j = N_0\tilde{\mu}_j$, we arrive at the equivalent maximization (15). □

To interpret the max-weight type condition (15), we distinguish two cases:

(i) $N^{in} \leq N_0$. In this case it is clearly optimal in (15) to cover the entire set J^{in} with J^{out}, assigning any extra elements arbitrarily.

(i) $N^{in} > N_0$. In this case only a portion of the μ_j can be included. The maximum weight is achieved by assigning J^{out} to the largest N_0 values of $\{\mu_j, j \in J^{in}\}$.

So we see that the local reciprocity energy is minimized by picking N_0 peers that are currently giving the most bandwidth to peer i, and assigning any extra slots arbitrarily. Interestingly, this corresponds exactly to the tit-for-tat part of the BitTorrent algorithm. Therefore, the myopic optimization of our energy cost is consistent with this widespread reciprocity mechanism.

What happens if we iterate on the above deterministic algorithm, each peer successively updating its configuration based on the tit-for-tat like reciprocity scheme? In general, it is difficult to characterize the behavior of such dynamics over a discrete set of configurations. The trajectory will depend on initial conditions, and there is no reason to expect the global energy-minimizing configuration will be found. For example, the initial file-exchange may break the graph into components, leaving some peers disconnected from their optimal neighbors; these will never be discovered by the above deterministic reciprocity. This suggests that a certain amount of random exploration is required. An additional argument for randomization is mentioned earlier in Remark 5. BitTorrent addresses this issue through the optimistic unchoke portion; however this egalitarian neighbor selection implies an important deviation from proportionality. An alternative is studied in the following section.

4. Incorporating randomness

From the preceding analysis, it is clear that randomness is needed in some form for a neighbor selection algorithm to work properly in P2P networks. In this section we will introduce a stochastic process over the configuration space of neighbor connections Λ^S that induces reciprocity in equilibrium. We would like the algorithm to be decentralized, so a natural set of conditions for the equilibrium distribution is the following:

- The connections of any peer, conditioned on the connections of its neighbors under the connectivity graph, should be independent from all the others (as each node should decide its connections only taking into account its neighbors).
- All configurations are possible a priori, so every one should have positive probability.
- The configurations where reciprocity is higher should have a higher probability.

We begin by introducing the following family of probability distributions:

$$\pi^T(X) = \frac{1}{Z_T} \exp\left(-\frac{\mathcal{E}(X)}{T}\right), \tag{16}$$

where $\mathcal{E}(X)$ is called the *energy function*, and the real parameter is T called the "temperature". Here Z_T is just a suitable normalization constant, often called *partition function* in statistical mechanics.

Note that the above distribution gives the highest probability to the low-energy states, provided the temperature is low. More precisely:

Proposition 8 ([16])**.** *Let $\{X_1^*, \ldots, X_K^*\}$ be the set of configurations that minimize the energy $\mathcal{E}(X)$, then as $T \to 0^+$ the distribution π^T converges to $\sum_{i=1}^K \frac{1}{K}\delta_{X_i^*}$, the uniform distribution on the optimal set.*

The energy function in (16) can be very general, but the most interesting case is when it is based on local interactions in the underlying connectivity graph. These are the so-called *Gibbs distributions* [16]. Let C be a complete subgraph or *clique* of the connectivity graph and $\mathcal{C}$ the set of all cliques. Let the energy take the form:

$$\mathcal{E}(x) = \sum_{C \in \mathcal{C}} V_C(x) \tag{17}$$

where $V_C(x)$ is *potential* associated to the *clique* C. In this case the Hammersley-Clifford equivalence Theorem [16] shows that every Gibbs distribution whose energy satisfies (17) produces a *Markov random field* in

the configuration space. This special structure ensures that the connections any peer chooses depend only on the connections of neighboring peers, the first assumption we imposed at the beginning of the section. Moreover, the equivalence theorem states that every Markov random field arises in this way, so there is no loss of generality in concentrating in Gibbs distributions.

In this regard, our previously defined energy function (14) is a perfect fit as the Gibbs energy. We recall its definition:

$$\mathcal{E}(X) = \frac{1}{2}\sum_{i,j}(\mu_i x_{ij} - \mu_j x_{ji})^2.$$

Here, a potential $V_C(x)$ is assigned only to the cliques of connected peers i and j (i.e. $a_{ij} > 0$) and the potential is given by the reciprocity attained in the current configuration, i.e. $V_{(i,j)} = (\mu_i x_{ij} - \mu_j x_{ji})^2$.

This choice of energy yields the following invariant distribution

$$\pi^T(X) = \frac{\exp\left(-\frac{1}{2T}\sum\limits_{i,j\in S}(\mu_i x_{ij} - \mu_j x_{ji})^2\right)}{\sum\limits_{X'\in\Lambda^S}\exp\left(-\frac{1}{2T}\sum\limits_{i,j\in S}(\mu_i x'_{ij} - \mu_j x'_{ji})^2\right)}.$$

As we stated before, letting the temperature $T \to 0^+$ the probability concentrates on states with higher reciprocity, as desired.

Note that other energy choices are possible to favor reciprocity. In particular, one could use the Kullback-Leibler divergence between the offered rates μ and the received rates r, but this energy cannot be decomposed in sums of terms that only depend on the connections of peers that form cliques.

A major advantage of having a Gibbs measure as target invariant distribution is that we have a natural Markov chain which converges to this invariant distribution. This is often called a Gibbs sampler [16]. Another commonly used name is "Glauber dynamics", and has been used with success in the problem of resource allocation in wireless networks [24].

We remark at this point that in recent work by [17, 18], it was proposed to use this kind of approach for a P2P network utility maximization problem, and it was argued that this "reverse engineered" BitTorrent. In this regard, we make the following remarks:

- The energy function used in the Gibbs approach of [17, 18] is defined in terms of a network utility, aimed more at performance than at fairness. This would have impact in a situation where the rate of upload of peer i is not equivalent for all peers j, due to other network bottlenecks.
- The dynamics proposed in these references implies *choking* one of the current peers and replacing by a new one; the peer most likely to be choked is the one with lowest current rate *to* it in the *upload* sense. Such a rule is in fact consistent with the algorithm for *seeders* in the BitTorrent protocol (peers who already own the file). It is different, however, to a reciprocity scheme based on *download* rates received *from* other peers, as in the tit-for-tat mechanism used by *leechers*. The latter is the focus of our work, and so our Gibbs proposal will be complementary to these references.

4.1. Random sweep Gibbs sampler

We now define the continuous time Markov chain which has stationary distribution π^T and only involves neighbor interactions. The only transitions that are admissible are between configurations X and X' that only differ in one row, that is, in the outgoing connections (unchokes) of one peer. Given X, denote by $\Lambda_i^S(X) = \{X'' \in \Lambda^S : x''_{kj} = x_{kj}, \forall k \neq i, \forall j\}$, that is, all the possible configurations that can be reached from X changing only row i. For any $X' \in \Lambda_i^S(X)$, define the transition rate

$$q^T_{X,X'} = \tau \cdot p^T_{X,X'}, \text{ where} \tag{18}$$
$$p^T_{X,X'} = \frac{\exp\left(-\frac{1}{T}\sum\limits_{j\in S}(\mu_i x'_{ij} - \mu_j x_{ji})^2\right)}{\sum\limits_{X''\in\Lambda_i^S(X)}\exp\left(-\frac{1}{T}\sum\limits_{j\in S}(\mu_i x''_{ij} - \mu_j x_{ji})^2\right)},$$

and $\tau > 0$ is a parameter.

The main property of the chosen transition rates is that

$$\pi^T(X)q^T_{X,X'} = \pi^T(X')q^T_{X',X},$$

where we note that $\Lambda_i^S(X') = \Lambda_i^S(X)$ for every $X' \in \Lambda_i^S(X)$. The above *detailed balance equations* imply that the Markov chain defined by (18) is reversible [25] and has invariant distribution π^T as required.

Additionally, note that by construction we have

$$q_i^T := \sum_{X'\in\Lambda_i^S(X)} q^T_{X,X'} = \tau \sum_{X'\in\Lambda_i^S(X)} p^T_{X,X'} = \tau.$$

Therefore, the rate at which each site i transitions is common to all sites. This kind of Markov chain is called a random sweep Gibbs sampler. Peers stay at each configuration an exponential amount of time, of parameter τ, after which they choose a new configuration $X' \in \Lambda_i^S(X)$ with probability $p^T_{X,X'}$.

Remark 6. *When the temperature T goes to zero, the transitions of peer i are dominated by configurations that minimize the* local *energy $\mathcal{E}_i$; as we saw in the previous section, these correspond to a tit-for-tat rule unchoking peers from whom it is currently downloading the fastest, similar to BitTorrent. The difference between this algorithm and BitTorrent lies in the manner that we introduce its randomness. Instead of having always an optimistic connection that blindly explores other peering options, this algorithm chooses all of its connections using the same distribution. If at some point we reach a state with local energy close to zero (e.g. when the peer is exchanging with other peers with the same upload capacity), the probability of choosing a different peer is very small, making the current configuration stable. This is the key to obtaining an allocation as close as possible to proportional fairness, while retaining the capability of random search.*

4.2. Systematic sweep Gibbs sampler

An alternative to the random sweep Gibbs sampler is the *systematic sweep Gibbs sampler*, in which each site is updated in a particular deterministic order, multiplying the transition probabilities of each row as the sequence goes along. It is most convenient here to define a discrete-time Markov chain that tracks the configuration state after each full sweep, with transition probabilities $p^{sys}_{X,X'}$ now involving changes in all matrix rows, with the following form:

$$p^{sys}_{X,X'} = \prod_{i=1}^{N} p^{i}_{X,X'} = \prod_{i=1}^{N} \frac{\exp\left(-\frac{1}{T}\mathcal{E}_i(X,X')\right)}{Z_i^T},$$

where

$$\mathcal{E}_i(X,X') = \sum_{j=1}^{i-1}(\mu_i x'_{ij} - \mu_j x'_{ji})^2 + \sum_{j=i}^{N}(\mu_i x'_{ij} - \mu_j x_{ji})^2,$$

and Z_i^T are appropriate normalizing constants. $\mathcal{E}_i(X,X')$ reflects the local energy of the i-th intermediate configuration when transitioning between X and X'.

The Markov chain defined before has a finite state space and is irreducible and aperiodic, thus it eventually converges to its invariant distribution, which can be shown to be equal to π^T. This sampler would correspond to the case where each peer updates its connections after a fixed amount of time, which is the version that we chose to implement for the simulations (Section 6).

5. Mixing times of the Gibbs samplers

The previously defined Gibbs samplers effectively yield the desired approximate allocation through their invariant distribution. However, there is a delay between the start of the process and the time that it reaches its stationary regime, which could deviate the final allocation from the desired one. For a discrete time Markov chain with finite state space Λ^S, transition matrix P and unique invariant distribution π, this delay is measured by the *mixing time* defined as

$$T_{mix}(\epsilon) = \max_{X_0 \in \Lambda^S} \min\{n : d_{TV}(X_0 P^n, \pi) \leq \epsilon\}$$

where $d_{TV}(\cdot,\cdot)$ is the distance in total variation between two probability measures.

5.1. Random sweep Gibbs sampler

As all transitions occur at the same rate independently of the state, we will analyze the mixing time of the embedded chain of the random sweep Gibbs sampler whose transition matrix P_{ran} is defined by

$$p^{ran}_{X,X'} = \frac{1}{N} \frac{\exp\left(-\frac{1}{T}\sum_{j\in S}(\mu_i x'_{ij} - \mu_j x_{ji})^2\right)}{\sum_{X'' \in \Lambda_i^S(X)} \exp\left(-\frac{1}{T}\sum_{j\in S}(\mu_i x''_{ij} - \mu_j x_{ji})^2\right)}$$

Note that the embedded chain has the same invariant distribution π_T because the transition rates are uniform.

Proposition 9. *Let η be the initial distribution and P_{ran} be the transition matrix of the embedded discrete-time Markov chain. Then*

$$d_{TV}\left(\eta (P_{ran})^{Nn}, \pi_T\right) \leq d_{TV}(\eta, \pi_T)\, \delta(P_{ran})^n$$

where

$$\delta(P_{ran}) \leq \left[1 - \frac{N!}{N^N}\exp\left(-\frac{2N\mu_{max}^2}{TN_0}\right)\prod_{i=1}^{N}\frac{\binom{d_i}{N_0}}{\binom{d_{max}}{N_0}}\right].$$

where d_i is the degree of node i and d_{max} is the maximum degree between all nodes. Furthermore

$$T_{mix}(\epsilon) \leq \frac{N\log(\epsilon)}{\log\left(1 - \frac{N!}{N^N}\exp\left(-\frac{2N\mu_{max}^2}{TN_0}\right)\prod_{i=1}^{N}\frac{\binom{d_i}{N_0}}{\binom{d_{max}}{N_0}}\right)} + 1.$$

The proof is given in Appendix C.

Remark 7. *The preceding bound for mixing time is valid for all $T > 0$, in particular for low values of temperature it would indicate approximate convergence to the set of optimal energy configurations. Note however that, consistently with the combinatoric complexity of the underlying problem, the bound scales poorly (exponentially) with the network size N.*

If instead we are willing to work with a high enough T, a tighter bound on mixing time can be obtained, which grows polynomially (as $N\log(N)$) in the size of the network. This means that we have a fast mixing algorithm, which is great in practice.

Proposition 10. *If $T > \frac{2\mu_{max}^2}{N_0 \log\left(\frac{2N_0}{2N_0-1}\right)}$, then*

$$T_{mix}(\epsilon) \leq \frac{\log(\epsilon N^{-1})}{\log\left[1 - \frac{1}{N} + \frac{2N_0}{N}\left(1 - \exp\left(-\frac{2\mu_{max}^2}{TN_0}\right)\right)\right]} + 1.$$

The proof uses the path coupling technique of [26], and is given in Appendix D.

5.2. Systematic sweep Gibbs sampler

For the naturally discrete-time Systematic sweep sampler, we bound the speed of convergence in the following result.

Proposition 11. *Let η be the initial distribution and P_{sys} be the transition matrix of the discrete-time Markov chain. Then*

$$d_{TV}\left(\eta P_{sys}^n, \pi_T\right) \leq d_{TV}\left(\eta, \pi_T\right)\delta(P_{sys})^n$$

where

$$\delta(P_{sys}) \leq \left[1 - \exp\left(-\sum_{i=1}^{N} \frac{2\mu_i\mu_{\max}}{TN_0}\right)\right].$$

Furthermore

$$T_{mix}(\epsilon) \leq \frac{\log(\epsilon)}{\log\left(1 - \exp\left(-\sum\limits_{i=1}^{N} \frac{2\mu_i\mu_{max}}{TN_0}\right)\right)} + 1.$$

The proof is given in Appendix E.

Remark 8. *Here again we have a bound that holds for all $T > 0$, with a similar drawback as the one in Proposition 9; namely, it grows exponentially in N.*

In both cases there appears to be a tradeoff between the speed of convergence and the fairness of the allocation. Indeed in the simulations section we will encounter such a tradeoff. We refer the reader to [24] for a discussion of these issues on a similar problem.

6. Implementation and simulations

We now evaluate the devised systematic sweep Gibbs algorithm as a means to achieve reciprocity and fairness. In order to perform comparisons, we also implemented idealized versions of the BitTorrent unchoking mechanism, the ideal proportional reciprocity algorithm discussed in Section 2.2, the PropShare unchoking algorithm of [6] and the Markov approximation approach devised in [17].

Let us begin by briefly recalling the different algorithms. The standard BitTorrent unchoking mechanism maintains for each peer $N_0 = 4$ outgoing connections. Three of these connections are used to reciprocate other peers, and the remaining connection is an *optimistic* unchoke, designed to explore new peers. The latter is kept for several iterations in order to allow time for the optimistically unchoked peer to reciprocate. This algorithm has low overhead and enables peers to find appropriate partners [2], but it has two main disadvantages: the unchokes are based only on the ranking of better contributors, and not in the bandwidth they provided, which has incentives problems [6]. It also constantly searches for new peers, allocating a substantial proportion of the uplink bandwidth to this end, and possibly drifting away from good configurations.

The proportional reciprocity algorithm (9), on the other hand, focuses on reciprocating only, by allocating proportional shares to each connected peer. To this end, it is the best one can do and achieves a fast convergence time. A pure proportional response however, has two main drawbacks from a practical perspective: it requires to keep a large amount of connections with several peers, as well as controlling exactly the amount of bandwidth allocated to each unchoked peer, which may be difficult to implement in practice. More importantly, it can get stuck in bad configurations if the initial connectivity of peers is sparse.

The PropShare algorithm is based on the reciprocity iteration, and was devised to correct this last problem, among other improvements. This algorithm allocates proportionally to the received contributions 80% of the uplink bandwidth of a given peer. It uses the remaining 20% to explore new peers through optimistic unchokes, much like BitTorrent. This exploration mechanism enables the algorithm to increase the number of connected peers. While this may achieve a higher level of fairness, the bandwidth committed to the optimistic search can make the algorithm drift away from good configurations. This algorithm still suffers from the burden of maintaining many connections and controlling the amount of bandwidth given to each. In our simula-

tions we implemented an idealized version of PropShare based on these features, not taking into account these problems, and thus we expect our results to provide an upper bound on real life PropShare performance.

The Markov approximation algorithm of [17] has points in common with the one proposed in this paper. However, unchoke transitions in that case emphasize maximizing throughput, by discovering the best neighbors to *upload* to. Reciprocity is not taken into account, so it suffers on the fairness side, as we will see.

As for our Gibbs algorithm, connections are updated in order following the transition probabilities given Section 4.2. The temperature parameter T in these expressions has units of bandwidth squared; for normalization purposes we express all bandwidth parameters in Mbps.

To evaluate the algorithms, we simulated the following scenario: we constructed a swarm of $N = 160$ peers, where each one has an average of 40 connections with other peers, which is the typical value of neighbors in BitTorrent implementations. This is achieved by constructing a random realization of an Erdös-Renyi graph with edge probability $p = 0.25$. Participating peers divide into two classes: half of the peers have an uplink bandwidth of 1Mbps whereas the rest contribute 256Kbps =1/4 Mbps to the system. All algorithms start from the same initial unchoke condition with $N_0 = 4$ outgoing connections per peer. Updates are made every $10s$ as in BitTorrent.

As a measure of the achieved reciprocity and fairness, we evaluate two metrics: the Gibbs energy $\mathcal{E}(X)$ from (14) defined in Section 3, which is intended to be minimized by the Gibbs algorithm, and also the Kullback-Leibler divergence $D(\mu\|r)$ between the offered bandwidths μ_i and the rates received by each peer r_i. The latter serves as an objective indicator of how much each algorithm adheres to our fairness criterion, recalling from Section 2 that in the optimal allocation $D(\mu\|r^*) = 0$.

In order to correctly assess the performance of each algorithm, we take several random initial conditions (common to all), simulate each of the algorithms and plot the average results for each metric.

In Figure 1, we plot the evolution of the Gibbs energy $\mathcal{E}(x)$ for the different algorithms in log scale. The Markov approximation algorithm from [17] remains with a high energy, which is not surprising since it does not pursue reciprocity. The (ideal) proportional reciprocity iteration (labeled in the Figure as "Sinkhorn") would be theoretically the best, but when facing random initial conditions with sparse connectivity the algorithm cannot fully achieve reciprocity, and thus it does not reach minimum energy. The BitTorrent algorithm is assigning too many optimistic unchokes, and this reflects on the energy achieved. PropShare is the best alternative, at the expense of having a greater number of simultaneous connections for each peer, and controlling bandwidth on each one of them. Our Gibbs algorithm (here with $T = 0.2$) finds the lowest energy level using at any given time only $N_0 = 4$ open connections which share the uplink equally.

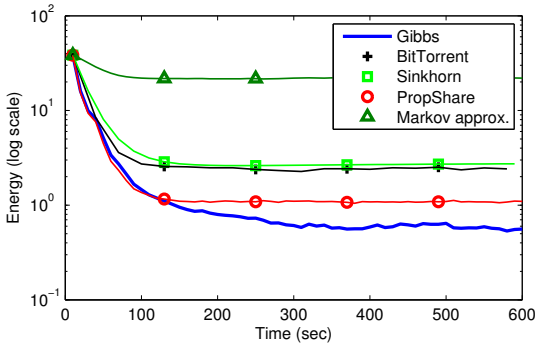

Figure 1: Gibbs energy evolution for the different algorithms.

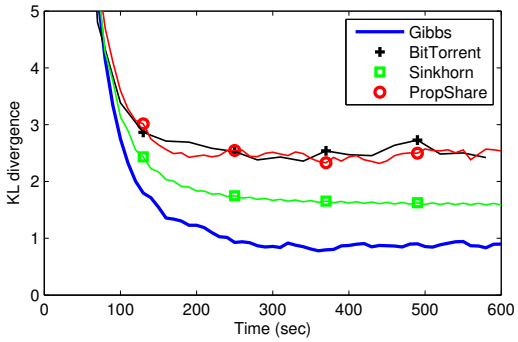

Figure 2: Evolution of the Kullback-Leibler divergence between uplink and received rate.

To evaluate fairness, in Figure 2 we plot the aforementioned KL divergence for each algorithm, leaving out in this case the Markov algorithm from [17]. Note that also for this metric the best algorithms are PropShare and the proposed Gibbs sampler, the latter achieving better results.

To explore the effect of the temperature parameter T on the Gibbs algorithm, we simulated the system for several values of T in the range 0.1 to 10. In Figure 3 we plot the steady-state energy and KL divergence as a function of T. As we can see, the Gibbs algorithm outperforms the alternatives for low enough T. The price to pay for operating at lower temperatures is convergence

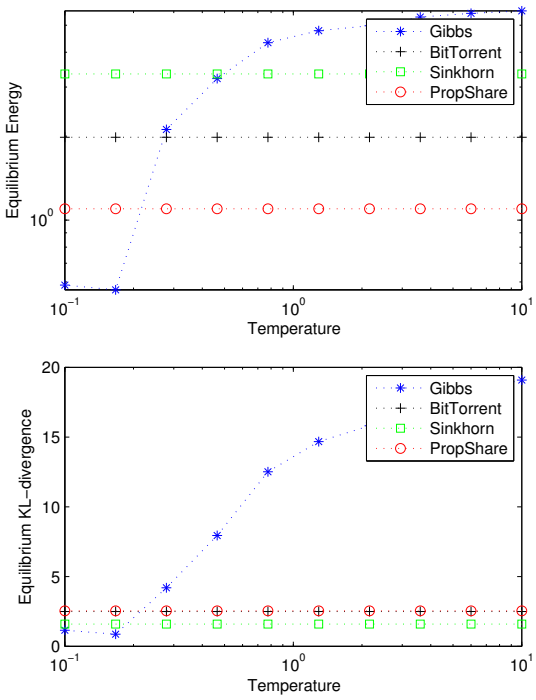

Figure 3: Attained Gibbs energy and KL divergence for different values of T.

time, as shown in Figure 4[3]. For low temperatures our algorithm takes longer to stabilize (in the order of minutes). Note however, referring back to Figure 2, that in this case even before fully converging our algorithm is already achieving an improved fairness with respect to the competitors.

Remark 9. *The convergence-time issue is relevant if one thinks of the exchange mechanism operating in a dynamic P2P swarm with arrival and departure of peers. In this regard, we are separating two time-scales: the faster, microscopic dynamics of piece exchanges, and the slower dynamics of peer populations. In the present paper we have considered the fast time scale, taking the population of peers to be fixed, and developed an algorithm to achieve fairness. Other papers by ourselves [27, 28] and others (e.g., [9]) address the slower population dynamics, assuming that fairness is imposed instantaneously. While in reality time-scale separation is not perfect, the decomposition will be approximately valid provided that fairness is imposed quickly with respect to peer inter-arrival times. Our analysis of convergence time gives credence to this idea: starting from a random place, fairness is imposed relatively quickly, in comparable times to other algorithms. If the initial condition were to deviate from fairness only in one newly arrived peer, much shorter convergence times are expected.*

[3]Precisely, we take convergence time to mean the iteration for which which our energy metric decreases by less than 1%.

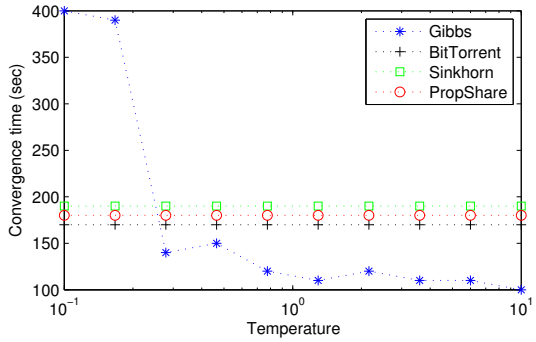

Figure 4: Convergence time as a function of temperature.

7. Conclusion

In this paper we analyzed P2P networks under heterogeneity in access bandwidth, with the aim of achieving a proportionally fair bandwidth allocation by means of decentralized peer interactions. A proportional reciprocity scheme from previous work was analyzed in some detail, using tools of convex optimization. Incorporating practical design constraints led to an approximate, randomized scheme for recursive neighbor selection, built from a Gibbs sampler dynamics based on a natural energy function. The proposal was implemented in simulation and compared against alternatives, showing good results. Natural follow-ups to the present work would be to construct a packet-level implementation, and investigating the performance of our Gibbs sampler protocol in the case of swarm of peers that varies in time.

Appendix A. Proof of Proposition 1

Proof. The *necessity* of the condition is straightforward. If a feasible Z exists,

$$\sum_{j\in J}\mu_j = \sum_{j\in J} r_j = \sum_i \sum_{j\in J} z_{ij}$$
$$= \sum_{i\notin I}\sum_{j\in J} z_{ij} \le \sum_{i\notin I}\sum_j z_{ij} = \sum_{i\notin I}\mu_i.$$

The intermediate step (from one row to the next) uses the fact that $z_{ij} = 0 \ \ \forall i \in I, j \in J$, a zero minor inherited from the structure A.

Let us now turn to the converse implication, established through a contradiction argument. Suppose that the allocation $r(Z) = \mu$ is *infeasible* within the structure A, and let Z^* be a solution to Problem 1, with p^* the optimal Lagrange multipliers. Invoking $r_j^* = p_j^* \mu_j$ from Remark 3, infeasibility implies not all prices are unity. Recalling further that peers only exchange information with others of inverse price, the set of optimal prices must satisfy

$$p_{min} = \min_i p_i^* < 1, \qquad p_{max} = \max_i p_i^* = \frac{1}{p_{min}} > 1.$$

Define the partition of indices in disjoint sets

$$I_{min} = \{i : p_i^* = p_{min}\}; \quad I_{int} = \{p_{min} < p_i^* < p_{max}\};$$
$$I_{max} = \{i : p_i^* = p_{max}\}.$$

If we choose a row and column order consistent with the index partition $(I_{min}, I_{int}, I_{max})$, Z^* must have structure

$$Z^* = \left[\begin{array}{c|c|c} 0 & 0 & X \\ \hline 0 & X & 0 \\ \hline X & 0 & 0 \end{array}\right], \tag{A.1}$$

where nonzero entries are confined to blocks marked by X. Adding the entries in the bottom-left block we have

$$\sum_{i \in I_{max}} \mu_i = \sum_{j \in I_{min}, i \in I_{max}} z_{ij}^* = \sum_{j \in I_{min}} r_j^*.$$

Now recalling that $r_j^* = p_j^* \mu_j$ we conclude that

$$\sum_{i \in I_{max}} \mu_i = p_{min} \sum_{j \in I_{min}} \mu_j. \tag{A.2}$$

We now claim that the row/column sets $(I_{min}, I_{min} \cup I_{int})$ must define a zero minor of A. In fact for any $j \in I_{min} \cup I_{int}$ we have, invoking (6):

$$\frac{1}{\min\{p_i^* : a_{ij} = 1\}} = \pi_j^* = p_j^* < \frac{1}{p_{min}},$$

therefore no $i \in I_{min}$ can be connected to j. Using symmetry of A, $(I_{min} \cup I_{int}, I_{min})$ is also a zero minor. In other words the top-left blocks of the structure (A.1) must be hard zeros imposed by A. Now applying the hypothesis (condition (8)) to $(I, J) = (I_{min} \cup I_{int}, I_{min})$ we have

$$\sum_{j \in I_{min}} \mu_j \le \sum_{i \notin I} \mu_i = \sum_{i \in I_{max}} \mu_i.$$

But then (A.2) implies

$$\sum_{j \in I_{min}} \mu_j \le p_{min} \sum_{j \in I_{min}} \mu_j,$$

a contradiction since $p_{min} < 1$. □

Appendix B. Proof of Theorem 4

Consider any optimal allocation Z^*, solution of Problem 1. For a feasible Z define the function

$$V(Z) := D(Z^* \| Z) = \sum_{i,j} z_{ij}^* \log\left(\frac{z_{ij}^*}{z_{ij}}\right), \tag{B.1}$$

which is the K-L divergence between the *matrices*. Note that $\sum_{i,j} z_{ij}^* = \sum_{i,j} z_{ij}$ so this metric is meaningful.

In order for $V(Z)$ to be finite, it must be that $z_{ij} > 0$ whenever $z_{ij}^* > 0$; this is automatically guaranteed for any Z that assigns positive rate to all allowable connections ($a_{ij} = 1$), which is an assumption of the Theorem for the initial condition $Z(0)$ of the proportional reciprocity algorithm; it is easily seen that this condition is preserved during the iteration.

Our first Lemma establishes a monotonicity condition of V after *two* steps of proportional reciprocity.

Lemma 12. *For V defined in* (B.1)*, the mapping $\mathcal{R}$ in* (10) *satisfies*

$$V(\mathcal{R}^2[Z]) \le V(Z), \tag{B.2}$$

with equality only if Z is an optimal allocation.

Proof. Let $Z^+ = \mathcal{R}[Z^*]$, recall by Proposition 2 that it is also an optimal allocation. Also denote $\hat{Z}^+ = \mathcal{R}[Z]$, $\hat{Z}^{++} = \mathcal{R}^2[Z]$, $r = r(Z)$ and $\hat{r}^+ = r(\hat{Z}^+)$. Recalling the definition the reciprocity iteration (10), we have

$$\hat{z}_{ij}^{++} = \frac{\mu_i}{\hat{r}_i^+} \cdot \frac{\mu_j}{r_j} z_{ij}. \tag{B.3}$$

Now introduce some additional factors from the optimal allocations Z^*, Z^+ to write the equivalent expression

$$\begin{aligned} \frac{\hat{z}_{ij}^{++}}{z_{ij}} &= \frac{\mu_i}{r_i^+} \cdot \frac{r_i^+}{\hat{r}_i^+} \cdot \frac{\mu_j}{r_j^*} \cdot \frac{r_j^*}{r_j} \\ &= \frac{r_i^+}{\hat{r}_i^+} \cdot \frac{r_j^*}{r_j} \quad \text{for } z_{ij}^* > 0. \end{aligned}$$

In the second step we have invoked the facts from Proposition 2, that $z_{ij}^* > 0$ implies $\frac{\mu_j}{r_j^*} = p_i$, $\frac{\mu_i}{r_i^+} = \frac{1}{p_i}$.

Taking a log and substituting in (B.1) we arrive at:

$$\begin{aligned} V(Z) - V(\hat{Z}^{++}) &= \sum_{i,j} z_{ij}^* \log\left(\frac{r_i^+}{\hat{r}_i^+}\right) + \sum_{i,j} z_{ij}^* \log\left(\frac{r_j^*}{r_j}\right) \\ &= \sum_i \mu_i \log\left(\frac{r_i^+}{\hat{r}_i^+}\right) + \sum_j r_j^* \log\left(\frac{r_j^*}{r_j}\right). \end{aligned} \tag{B.4}$$

The first term above is non-negative because Z^+ is a solution to Problem 1, and the second is the K-L divergence $D(r^*||r) \geq 0$. This establishes (B.2). Furthermore, from the second term we see that equality holds only if $r = r^*$, i.e. Z is an optimal allocation. □

We now tackle the proof of the Theorem.

Proof. Denote by $Z_k = Z(2k)$ the sequence of even iterations of the initial condition $Z(0)$. The previous Lemma implies that $V_k = V(Z_k)$ is decreasing, non-negative sequence and hence has a limit $V_\infty \geq 0$.

Now since Z_k is a bounded matrix sequence due to (1), consider a convergent subsequence $Z_{k_l} \to \tilde{Z}$. By continuity we must have $V(\tilde{Z}) = D(Z^*||\tilde{Z}) = V_\infty$. Now also $\mathcal{R}^2(Z_{k_l}) = Z_{k_l+1}$ must satisfy $V(Z_{k_l+1}) \to V_\infty$, therefore $V(\mathcal{R}^2(\tilde{Z})) = V(\tilde{Z})$. Invoking the second statement of the previous Lemma, $\tilde{Z}$ must be an optimal allocation.

So $\{Z_k\}$ can only accumulate in the set of optimal points; a slight refinement of the argument implies Z_k must actually converge to an optimal point. To see this, choose a function V as in (B.1) but where Z^* is actually $\tilde{Z}$, subsequential limit of Z_k. In that case clearly $V_\infty = 0$. Now $D(\tilde{Z}||Z_k) \to 0$ implies that $Z_k \to \tilde{Z}$, optimal point as claimed.

An analogous argument applies to the odd subsequence $Z(2k+1)$.

□

Appendix C. Proof of Proposition 9

First, we prove a useful lemma.

Lemma 13. *We have the following lower bound for the transition probabilities*

$$p^T_{X,X'} = \frac{\exp\left(-\frac{1}{T}\mathcal{E}_i(X')\right)}{\sum\limits_{X''\in\Lambda_i^S(X)} \exp\left(-\frac{1}{T}\mathcal{E}_i(X'')\right)} \geq \frac{\exp\left(-\frac{2\mu_i\mu_{max}}{TN_0}\right)}{\#\Lambda_i^S(X)}$$

which is uniform in X, as $\#\Lambda_i^S(X)$ does not actually depend on X.

Proof. Let

$$m_i(X) = \min_{X'\in\Lambda_i^S(X)} \{\mathcal{E}_i(X')\}$$

and

$$M_i(X) = \max_{X'\in\Lambda_i^S(X)} \{\mathcal{E}_i(X')\}$$

be the minimum and maximum local energy at node i that can be achieved from X. Then we have that

$$p^T_{X,X'} = \frac{\exp\left(-\frac{1}{T}\left[\mathcal{E}_i(X') - m_i(X)\right]\right)}{\sum\limits_{X''\in\Lambda_i^S(X)} \exp\left(-\frac{1}{T}\left[\mathcal{E}_i(X'') - m_i(X)\right]\right)}$$

We can bound from above each term of the sum in the denominator by 1 and thus the sum by the number of terms that is $\#\Lambda_i^S(X)$. Furthermore, we can bound from below the numerator by

$$\exp\left(-\frac{1}{T}\left[M_i(X) - m_i(X)\right]\right)$$

To find an appropriate bound, we need to find the maximum possible difference between $M_i(X)$ and $m_i(X)$. The maximum local energy in node i is achieved when the peer chooses to upload to a set of peers from which it is not downloading content. In that case

$$M_i(X) = \frac{1}{N_0^2}\left(\sum_{k=1}^{N_0}\mu_i^2 + \sum_{k=1}^{N^{in}}\mu_{j_k}^2\right).$$

On the other hand, Proposition 7 tells us that the minimum local energy is achieved when the peer chooses the fastest peers that are uploading to him. As a result, the minimum energy is

$$m_i(X) = \frac{1}{N_0^2}\left(\sum_{k=1}^{N_0}(\mu_i - \mu_{j_k})^2 + \sum_{k=N_0+1}^{N^{in}}\mu_{j_k}^2\right)$$

if $N^{in} \geq N_0$ and

$$m_i(X) = \frac{1}{N_0^2}\left(\sum_{k=1}^{N^{in}}(\mu_i - \mu_{j_k})^2 + \sum_{k=N^{in}+1}^{N_0}\mu_i^2\right)$$

otherwise. In any case, the minimum energy can be expressed as

$$m_i(X) = \frac{1}{N_0^2}\left(\sum_{k=1}^{N_0}\mu_i^2 + \sum_{k=1}^{N^{in}}\mu_{j_k}^2 - \sum_{k=1}^{\min\{N_0,N^{in}\}} 2\mu_i\mu_{j_k}\right).$$

and thus the difference is

$$M_i(X) - m_i(X) = \frac{1}{N_0^2}\sum_{k=1}^{\min\{N_0,N^{in}\}} 2\mu_i\mu_{j_k} \leq \frac{2\mu_i\mu_{max}}{N_0}.$$

Putting all together we have that

$$p^T_{X,X'} \geq \frac{\exp\left(-\frac{2\mu_i\mu_{max}}{TN_0}\right)}{\#\Lambda_i^S(X)}$$

□

Now we are ready to prove the proposition.

Proof. Consider the Markov chain with transition matrix $P' = P^N_{ran}$, which is actually the embedded chain for

the random sweep Gibbs sampler every N-th transition. This chain can now jump from one state to any other in exactly one step, while the original could not. Using Theorem 7.2 of chapter 6 in [16], we have

$$d_{TV}\left(\eta P'^{n}, \pi_T\right) \leq d_{TV}\left(\eta, \pi_T\right) \delta(P')^n$$

where $\delta(P')$ is the Dobrushin's ergodic coefficient

$$\begin{aligned}\delta(P') &= \frac{1}{2} \max_{X,X'} \left\{ \sum_{X''} |p'_{X,X''} - p'_{X',X''}| \right\} \\ &= 1 - \min_{X,X'} \left\{ \sum_{X''} \min\left\{ p'_{X,X''}, p'_{X',X''} \right\} \right\}\end{aligned}$$

The transition probabilities $p'_{X,X''}$ are sums of products of N factors that come from P_{ran}. In the worst case, when X and X'' differ in all rows, there is only $N!$ terms in the sum (one for each permutation of the order of transitions). As a result, if we get a lower bound for the elements of P_{ran}, we can obtain one for the elements of P'.

Using Lemma 13, we obtain the following uniform bound

$$p^{ran}_{X,X''} \geq \frac{1}{N} \frac{\exp\left(-\frac{2\mu_i\mu_{max}}{TN_0}\right)}{\#\Lambda_i^S(X)}$$

If $X = X_0, X_1, \ldots, X_{N-1}, X_N = X''$ is a path of configurations from X to X'' such that X_{i-1} and X_i can only differ in row $\sigma(i)$ for some permutation σ (but may be equal), then the minimum element in the matrix P' is lower bounded as follows

$$\begin{aligned}\min_{X,X''\in\Lambda^S}\{p'_{X,X''}\} &\geq N! \min_{X_0,\ldots,X_N\in\Lambda^S} \left\{ \prod_{i=1}^{N} p^{ran}_{X_{i-1},X_i} \right\} \\ &\geq \frac{N!}{N^N} \left[\min_i \frac{\exp\left(-\frac{2\mu_i\mu_{max}}{TN_0}\right)}{\#\Lambda_i^S(X)} \right]^N \\ &\geq \frac{N!}{N^N} \frac{\exp\left(-\frac{2N\mu^2_{max}}{TN_0}\right)}{\binom{d_{max}}{N_0}^N},\end{aligned}$$

where d_{max} is the maximum degree in the connectivity graph. With this we obtain the following expression

$$\begin{aligned}\delta(P') &\leq 1 - \frac{N!}{N^N} \frac{\exp\left(-\frac{2N\mu^2_{max}}{TN_0}\right)}{\binom{d_{max}}{N_0}^N} \#\Lambda^S \\ &\leq 1 - \frac{N!}{N^N} \exp\left(-\frac{2N\mu^2_{max}}{TN_0}\right) \prod_{i=1}^{N} \frac{\binom{d_i}{N_0}}{\binom{d_{max}}{N_0}}\end{aligned}$$

For the mixing time we need the total variation to be less than or equal to ϵ in the worst case regarding the initial condition

$$\begin{aligned}d_{TV}\left(\delta_{X_0}P'^{n}, \pi_T\right) &\leq d_{TV}\left(\delta_{X_0}, \pi_T\right) \delta(P')^n \\ &\leq \left[1 - \frac{N!}{N^N} \exp\left(-\frac{2N\mu^2_{max}}{TN_0}\right) \prod_{i=1}^{N} \frac{\binom{d_i}{N_0}}{\binom{d_{max}}{N_0}} \right]^n \\ &\leq \epsilon\end{aligned}$$

Taking logarithms, rearranging terms and multiplying by N concludes the proof. □

Appendix D. Proof of Proposition 10

This proposition heavily relies on a theorem from [26] which we now state.

Let C and V be finite sets with $\#V = N$. Consider a discrete-time Markov chain with state space $\Omega = C^V$ with the following transition structure: We first pick $i \in V$ from a fixed distribution J over V. Then we pick $c \in C$ according to a distribution $\kappa_{X,i}$ over C, dependent only on the current state X and i. We make the transition towards the state X' which only differs from X in i, such that $X'_i = c$ (which we denote $X_{i\to c}$). Also, assume that the chain is irreducible and aperiodic thus having a unique invariant distribution.

Theorem 14. *If*

$$\begin{aligned}\beta = \max_{X,Y\in\Omega, i\in V} \Bigg\{ &1 - J(i) + \sum_{j\in V} J(j) d_{TV}(\kappa_{X,j}, \kappa_{Y,j}) : \\ &X, Y \text{ only differ in } i,\ X \neq Y \Bigg\} < 1\end{aligned}$$

then

$$T_{mix}(\epsilon) \leq \left\lceil \frac{\log(\epsilon N^{-1})}{\log(\beta)} \right\rceil$$

Although this theorem requires a state space which is a product of identical finite sets, the same result still holds for Markov chains with state spaces which are product of different finite sets. The proof is essentially the same with minor modifications. Now we can prove the proposition.

Proof. In order to use the previous theorem, we only need to get an upper bound for β in our particular chain.

$$\beta = 1 - \frac{1}{N} + \max_{X\in\Lambda^S, Y\in\Lambda_i^S(X), i} \left\{ \frac{1}{N} \sum_{j=1}^{N} d_{TV}\left(p^T_{X,X_{j\to\cdot}}, p^T_{Y,Y_{j\to\cdot}}\right) \right\}$$

Note that because X and Y only differ on the connections of one peer (i), there can only be $2N_0$ nonzero d_{TV}, due to the fact that the peers that do not receive a connection from i in X nor in Y will be unaffected by the connections of i. Furthermore, $d_{TV}\left(p^T_{X,X_{i\to\cdot}}, p^T_{Y,Y_{i\to\cdot}}\right) = 0$ always as the distribution of the new connections does not depend on current connections.

Now, for $j \neq i$ we have that

$$\begin{aligned} d_{TV}\left(p^T_{X,X_{j\to\cdot}}, p^T_{Y,Y_{j\to\cdot}}\right) &= \frac{1}{2}\sum_{X'_j\in\Lambda_j}\left|p^T_{X,X_{j\to X'_j}} - p^T_{Y,Y_{j\to X'_j}}\right| \\ &= 1 - \sum_{X'_j\in\Lambda_j}\min\left\{p^T_{X,X_{j\to X'_j}}, p^T_{Y,Y_{j\to X'_j}}\right\} \\ &\leq 1 - \#\Lambda_j \min_{X'_j\in\Lambda_j}\min\left\{p^T_{X,X_{j\to X'_j}}, p^T_{Y,Y_{j\to X'_j}}\right\} \\ &\leq 1 - \exp\left(-\frac{2\mu^2_{max}}{TN_0}\right) \end{aligned}$$

where the last inequality comes from Lemma 13. This translates into an upper bound for β

$$\beta \leq 1 - \frac{1}{N} + \frac{1}{N}2N_0\left(1 - \exp\left(-\frac{2\mu^2_{max}}{TN_0}\right)\right) < 1$$

where the last inequality comes from the hypothesis on T. Theorem 14 concludes the proof.

□

Appendix E. Proof of Proposition 11

Proof. Using Theorem 7.2 of chapter 6 in [16], we have

$$d_{TV}\left(\eta P^n_{sys}, \pi_T\right) \leq d_{TV}(\eta, \pi_T)\delta(P_{sys})^n$$

where $\delta(P_{sys})$ is the Dobrushin's ergodic coefficient

$$\begin{aligned} \delta(P_{sys}) &= \frac{1}{2}\max_{X,X'}\left\{\sum_{X''}|p^{sys}_{X,X''} - p^{sys}_{X',X''}|\right\} \\ &= 1 - \min_{X,X'}\left\{\sum_{X''}\min\left\{p^{sys}_{X,X''}, p^{sys}_{X',X''}\right\}\right\} \end{aligned}$$

Now, lets define $\Lambda^S_i(X, X')$ as the set of all configurations $X'' \in \Lambda^S$ such that the row j of X'', denoted X''_j satisfies $X''_j = X'_j$ if $j < i$ and $X''_j = X_j$ if $j > i$. That is, $\Lambda^S_i(X, X')$ is the set of all possible configurations that can be reached in step i in the partial transition between X and X'. Using Lemma 13, we obtain the following uniform bound

$$p^i_{X,X'} \geq \frac{\exp\left(-\frac{2\mu_i\mu_{max}}{TN_0}\right)}{\#\Lambda^S_i(X, X')}$$

Then, the minimum element in the matrix P_{sys} is

$$\begin{aligned} \min_{X,X'\in\Lambda^S}\{p^{sys}_{X,X'}\} &= \min_{X,X'\in\Lambda^S}\left\{\prod_{i=1}^N p^i_{X,X'}\right\} \\ &\geq \prod_{i=1}^N \frac{\exp\left(-\frac{2\mu_i\mu_{max}}{TN_0}\right)}{\#\Lambda^S_i(X, X')} \\ &\geq \frac{\exp\left(-\sum_{i=1}^N \frac{2\mu_i\mu_{max}}{TN_0}\right)}{\prod_{i=1}^N \#\Lambda^S_i(X, X')}, \end{aligned}$$

and with this we obtain the following expression

$$\begin{aligned} \delta(P_{sys}) &\leq 1 - \frac{\exp\left(-\sum_{i=1}^N \frac{2\mu_i\mu_{max}}{TN_0}\right)}{\prod_{i=1}^N \#\Lambda^S_i(X, X')}\prod_{i=1}^N \#\Lambda^S_i(X, X') \\ &\leq 1 - \exp\left(-\sum_{i=1}^N \frac{2\mu_i\mu_{max}}{TN_0}\right) \end{aligned}$$

For the mixing time we need the total variation to be less than or equal to ϵ in the worst case regarding the initial condition

$$\begin{aligned} d_{TV}\left(\delta_{X_0}P^n_{sys}, \pi_T\right) &\leq d_{TV}(\delta_{X_0}, \pi_T)\delta(P_{sys})^n \\ &\leq \left[1 - \exp\left(-\sum_i \frac{2\mu_i\mu_{\max}}{TN_0}\right)\right]^n \\ &\leq \epsilon \end{aligned}$$

Taking logarithms and rearranging terms concludes the proof. □

Acknowledgements

The authors were partially supported by ANII-Uruguay scholarship POS_NAC_2012_1_9088, and AFOSR-US under Grant FA9550-12-1-0398.

References

[1] B. Cohen, Incentives build robustness in BitTorrent, in: Proc. of 1st Workshop on the Economics of Peer-2-Peer Systems (2003), pp. 1–5.

[2] B. Fan, J. Lui, D. Chiu, The Design Trade-offs of BitTorrent-like File Sharing Protocols, IEEE/ACM Transactions on Networking 17 (2009) 365–376.

[3] R. Kumar, K. W. Ross, Peer-assisted file distribution: The minimum distribution time, in: Proc. of the 1st IEEE Workshop on Web Systems and Technologies (2006), pp. 1–11.

[4] J. Mundinger, R. Weber, G. Weiss, Optimal scheduling of peer-to-peer file dissemination, Journal of Scheduling 11(2) (2008) 105–120.

[5] G. M. Ezovski, A. Tang, L. H. Andrew, Minimizing average finish time in P2P networks, in: Proc. of IEEE Infocom 2009, pp. 594–602.

[6] D. Levin, K. LaCurts, N. Spring, B. Bhattacharjee, BitTorrent is an Auction: Analyzing and Improving BitTorrent's Incentives, in: Proc. of the ACM SIGCOMM 2008, pp. 243–254.

[7] F. Kelly, A. Maulloo, D. Tan, Rate control in communication networks: shadow prices, proportional fairness and stability, Journal of the Operational Research Society 39 (1998) 237–252.

[8] S. Boyd, L. Vanderberghe, Convex Optimization, Cambridge University Press, Cambridge, UK, 2004.

[9] X. Yang, G. de Veciana, Performance of peer-to-peer networks: Service capacity and role of resource sharing policies, Performance evaluation 63 (2006) 175–194.

[10] F. Wu, L. Zhang, Proportional response dynamics leads to market equilibrium, in: Proc. of the 39th ACM Symposium on Theory of Computing (2007), pp. 354–363.

[11] R. Sinkhorn, A relationship between arbitrary positive matrices and doubly stochastic matrices, Annals of Mathematical statistics 35 (1964) 876–876.

[12] R. Sinkhorn, Diagonal equivalence to matrices with prescribed row and column sums II, Proceedings of the American Mathematical Society 45 (1974) 195–198.

[13] H. Balakrishnan, I. Hwang, C. Tomlin, Polynomial Approximation Algorithm for Belief Matrix Maintenance in Identity Management, in: Proc. of the 43rd IEEE Conference on Decision and Control (2004), pp. 4874–4879.

[14] L. Zhang, Proportional response dynamics in the fisher market, Theoretical Computer Science 412 (2011) 2691–2698.

[15] E. Eisenberg, Aggregation of utility functions, Management Science 7 (1961) 337–350.

[16] P. Brémaud, Markov Chains: Gibbs Fields, Monte Carlo simulation and queues, Springer, New York, NY, 1999.

[17] Z. Shao, H. Zhang, M. Chen, K. Ramachandran, Reverse-Engineering BitTorrent: A Markov Approximation Perspective, in: Proc. of the 31st IEEE International Conference on Computer Communications (2012), pp. 2996–3000.

[18] H. Zhang, Z. Shao, M. Chen, K. Ramachandran, Optimal Neighbor selection in BitTorrent-like Peer-to-Peer Networks, in: Proc. of the ACM SIGMETRICS 2011, pp. 141–142.

[19] M. Zubeldía, A. Ferragut, F. Paganini, Proportional fairness in heterogeneous peer-to-peer networks through reciprocity and Gibbs sampling, in: Proc. of the 51st Allerton conference on Communication, Control and Computing (2013), pp. 123–130.

[20] H. Yaiche, R. Mazumdar, C. Rosenberg, A Game Theoretic Framework for Bandwidth Allocation and Pricing in Broadband Networks, IEEE/ACM Transactions on Networking 8 (2000) 667–678.

[21] A. Muthoo, Bargaining Theory with Applications, Cambridge University Press, Cambridge, UK, 1999.

[22] A. Legout, N. Liogkas, E. Kohler, L. Zhang, Clustering and sharing incentives in BitTorrent systems, in: ACM SIGMETRICS Performance Evaluation Review (2007), volume 35, pp. 301–312.

[23] P. Erdős, T. Gallai, Graphs with prescribed degrees of vertices (Hungarian), Matematikai Lapok 11 (1960) 264–274.

[24] L. Jiang, J. Walrand, A Distributed CSMA Algorithm for Throughput and Utility Maximization in Wireless Networks, IEEE/ACM Transactions on Networking 18 (2010) 960–972.

[25] F. Kelly, Reversibility and stochastic networks, Wiley, 1979.

[26] B. Bubley, M. Dyer, Path coupling: A technique for providing rapid mixing in Markov chains, in: 38th Annual Symposium on Foundations of Computer Science (1997), pp. 223–231.

[27] A. Ferragut, F. Kozynski, F. Paganini, Dynamics of content propagation in BitTorrent-like P2P file exchange systems, in: Proc. of 50th IEEE Conf. on Decision and Control (2011), pp. 3116–3121.

[28] F. Paganini, A. Ferragut, PDE models for population and residual work applied to peer-to-peer networks, in: Proc. of 46th Annual Conference on Information Sciences and Systems (2012), pp. 1–6.