# On the Topologies Formed by Selfish Peers

Thomas Moscibroda, Stefan Schmid, Roger Wattenhofer
{moscitho,schmiste,wattenhofer}@tik.ee.ethz.ch
Computer Engineering and Networks Laboratory, ETH Zurich, 8092 Zurich, Switzerland

## TIK Report 252

**Abstract**

Current peer-to-peer (P2P) systems often suffer from a large fraction of freeriders not contributing any resources to the network. Various mechanisms have been designed to overcome this problem. However, the selfish behavior of peers has aspects which go beyond resource sharing. This paper studies the effects on the topology of a P2P network if peers selfishly select the peers to connect to. In our model, a peer exploits locality properties in order to minimize the latency (or response times) of its lookup operations. At the same time, the peer aims at not having to maintain links to too many other peers in the system. By giving tight bounds on the price of anarchy, we show that the resulting topologies can be much worse than if peers collaborated. Moreover, the network may never stabilize, even in the absence of churn. Finally, we establish the complexity of Nash equilibria in our game theoretic model of P2P networks. Specifically, we prove that it is NP-hard to decide whether our game has a Nash equilibrium and can stabilize.

# 1 Introduction

The power of peer-to-peer (P2P) computing arises from the collaboration of its numerous constituent parts, the peers. If all the participating peers contribute some of their resources—for instance bandwidth, memory, or CPU cycles—, highly scalable decentralized systems can be built which significantly outperform existing server based solutions. Unfortunately, in reality, many peers are selfish and strive for maximizing their own utility by benefiting from the system without contributing much themselves. Hence the performance—and thus its success in practice!—of a P2P system crucially depends on its capability of dealing with *selfishness*. A well-known mechanism designed to cope with this freeriding problem is the *tit-for-tat policy* which is for instance employed by the file-distribution tool *BitTorrent*.

However, selfish behavior in peer-to-peer networks has numerous important implications even beyond the peer's unwillingness to contribute bandwidth or memory. For example, in unstructured P2P systems—the predominant P2P architectures in today's Internet—, a peer can select to which and to how many other peers in the network it wants to connect. With a clever choice of neighbors, a peer can attempt to optimize its lookup performance by minimizing the latencies—or more precisely, the *stretch*—to the other peers in the network. Achieving good stretches by itself is of course simple: A peer can establish links to a large number of other peers in the system. Because the memory and maintenance overhead of such a neighbor set is large, however, egoistic peers try to exploit locality as much as possible, while avoiding to store too many neighbors. It is this fundamental trade-off between the need to have small latencies and the desire to reduce maintenance overhead that governs the decisions of selfish peers.

This paper investigates the impact of selfish neighbor selection on the quality of the resulting network topologies. An appropriate tool to study such selfish behavior is *game theory*. In particular, this paper studies the *Price of Anarchy* of P2P overlay creation, which is the ratio between an optimal solution obtained by perfectly collaborating participants compared to a solution generated by peers that act in an egoistic manner, optimizing their individual benefit. The importance of studying the

Price of Anarchy in peer-to-peer systems stems from the fact that it quantifies the possible degradation caused by selfishness. Specifically, a low Price of Anarchy indicates that a system does not require an incentive-mechanism (such as *tit-for-tat*), because selfishness does not overly bog down the overall system performance. If the Price of Anarchy is high, however, specific cooperation incentives (whose goals are to reduce the Price of Anarchy) need to be enforced in order to ensure that the system can perform efficiently. Hence, in peer-to-peer systems the Price of Anarchy is a measure that helps explaining the necessity (or non-necessity) of cooperation mechanisms.

The contribution of this paper is threefold. First, we show that the topologies of selfish, unstructured P2P systems can be much worse than in a scenario in which peers collaborate. More precisely, we show that the Price of Anarchy is $\Theta(\min(\alpha, n))$, where $\alpha$ is a parameter that captures the tradeoff between lookup performance (low stretches) and the cost of neighbor maintenance, and $n$ is the number of peers in the system, respectively. Thereby, the upper bound $O(\min(\alpha, n))$ holds for peers located in *arbitrary* metric spaces, including the popular *growth-bounded* and *doubling metrics*. On the other hand, intriguingly, this bound is tight even in such a simple metric space as the 1-dimensional *Euclidean space*. As a second contribution, we prove that the topology of a static peer-to-peer system consisting of selfish peers may never converge to a stable state. That is, links may continuously change even in environments without *churn* (causing the network to be inherently instable). Finally, we consider the complexity of Nash equilibria. We show that deciding whether there exists a pure Nash equilibrium in a given network is NP-complete. Consequently, it is infeasible in practice to determine if a P2P network of selfish peers can stabilize.

The remainder of the paper is organized as follows. In Section 2, we present our model and introduce the game theoretic approach. Related work is reviewed in Section 3. In the subsequent Section 4 we give tight bounds on the Price of Anarchy of P2P topologies. Then, in Section 5, we show that a system consisting of selfish peers may never stabilize. Section 6 investigates the complexity of Nash equilibria before Section 7 concludes the paper.

# 2 Model

We model the peers of a P2P network as points in a *metric space* $\mathcal{M} = (V, d)$, where $d : V \times V \rightarrow [0, \infty)$ is the *distance function* which describes the underlying latencies between all pairs of peers.

In this paper, the effects of selfish peer behavior is studied from a *game-theoretic* perspective. We consider a set of $n$ peers $V = \{\pi_0, \pi_1, \ldots, \pi_{n-1}\}$. A peer can choose to which subset of other peers it wants to store pointers (IP addresses). Formally, the *strategy space* of a peer $\pi_i$ is given by $S_i = 2^{V \setminus \{\pi_i\}}$, and we will refer to the actually chosen links as $\pi_i$'s *strategy* $s_i \in S_i$. We say that $\pi_i$ *maintains or establishes a link* to $\pi_j$ if $\pi_j \in s_i$. The combination of all peers' strategies, i.e., $s = (s_0, ..., s_{n-1}) \in S_0 \times \cdots \times S_{n-1}$, yields a (directed) graph $G[s] = (V, \cup_{i=0}^{n-1}(\{\pi_i\} \times s_i))$, which describes the resulting P2P topology.

Selfish peers exploit *locality* in order to maximize their lookup performance. Concretely, a peer aims at minimizing the *stretch* to all other peers. The stretch between two peers $\pi$ and $\pi'$ is defined as the shortest distance between $\pi$ and $\pi'$ using the links of the resulting P2P topology $G$ divided by the direct distance, i.e., for a topology $G$, $stretch_G(\pi, \pi') = d_G(\pi, \pi')/d(\pi, \pi')$. Clearly, it is desirable for a peer to have low stretch to other peers in order to keep its latency small. By establishing a link to all peers in the system, a peer reaches every peer with minimal stretch 1, and the potential lookup performance is optimal. However, storing and especially maintaining a large number of links is expensive. Therefore, the individual cost $c_i(s)$ incurred at a peer $\pi$ is composed not only of the stretches to all other peers, but also of its *degree*, i.e., the number of its neighbors:

$$c_i(s) = \alpha \cdot |s_i| + \sum_{i \neq j} stretch_{G[s]}(\pi_i, \pi_j).$$

Note that this cost function captures the classic P2P trade-off between the need to minimize latencies and the desire to store and maintain only few links, as it has been addressed by many existing systems, for example Pastry [18]. Thereby, the relative importance of degree costs versus stretch costs is expressed by the parameter $\alpha$.

The objective of a selfish peer is to minimize its individual cost. In order to evaluate the topologies constructed by selfish peers—and compare them with the topologies achieved by collaborating peers—, we use the notion of a *Nash equilibrium*. A P2P topology constitutes a Nash equilibrium if no peer can reduce its individual cost by changing its set of neighbors given that the connections of all other peers remain the same. More formally, a (pure) Nash equilibrium is a combination of strategies $s$ such that, for each peer $\pi_i$, and for all alternative strategies $s'$ which differ only in the $i^{th}$ component (different neighbor sets for peer $\pi_i$), $c_i(s) \leq c_i(s')$. This means that in a Nash equilibrium, no peer has an incentive to change its current set of neighbors, that is, Nash equilibria are *stable*.

While peers try to minimize their individual cost, the system designer is interested in a good overall quality of the P2P network. The *social cost* is the sum of all peers' individual costs, i.e.,

$$C(G) = \sum_i c_i = \alpha|E| + \sum_{i \neq j} stretch_G(\pi_i, \pi_j).$$

The lower this social cost, the better is the system's performance.

Determining the parameter $\alpha$ in real unstructured peer-to-peer networks is an interesting field for study. As mentioned, $\alpha$ measures the relative importance of low stretches compared to the peers' degrees, and thus depends on the system or application: For example, in systems with many lookups where good response times are crucial, $\alpha$ is smaller than in distributed archival storage systems consisting mainly of large files. In the sequel, we denote the link and stretch costs by

$$C_E(G) = \alpha|E| \quad \text{and} \quad C_S(G) = \sum_{i \neq j} stretch_G(\pi_i, \pi_j).$$

Typically, a given distribution of peers in a metric space can result in different Nash equilibria, depending on the order in which peers change their links. To gain an understanding of the impact of selfishness on the social cost, we are particularly interested in the social cost of the *worst* possible Nash equilibrium. That is, we study topologies in which no selfish peer has an incentive to change its neighbors, but in which all peers together could be much better off if they collaborated. More precisely and using the terminology of game theory, we are interested in the *Price of Anarchy*, the ratio between the social cost of the worst Nash equilibrium and the social cost of the optimal topology.

# 3 Related Work

The lack of cooperation in traditional P2P file-sharing systems has been well-documented over the last years [3, 20], and research on the causes and possible counter-measures is very active, e.g., [5] and [13]. Most of the current literature focuses on the issue of free resource consumption, *freeriding*. In contrast, the impact of other aspects of selfishness has received much less attention. In fact, to the best of our knowledge, this is the first paper to take a step towards studying the consequences of selfish neighbor selection on the topologies of P2P networks.

Our game-theoretic model of network creation has been inspired by the paper by Fabrikant et al. [10] which studies the Internet's architecture as built by economic agents, e.g., by Internet providers or *autonomous systems*. Recent subsequent work on network creation in various settings includes [4, 6, 8, 9]. In contrast to all these works, our model takes into account many of the intrinsic properties of P2P systems. For instance, it captures the important *locality properties* of P2P systems, i.e., the desire to reduce the latencies (expressed as the stretch) experienced when performing look-up operations. Furthermore, the fact that a peer can decide to which other peers it wishes to store pointers and thus maintain links yields a scenario with *directed* links.

Building structured systems that explicitly exploit locality properties has been a flourishing research area in networking and P2P computing (e.g. [1, 18, 19]). In early literature on distributed hash tables (DHT), the major measure of system quality has been the number of hops required for look-up operations. While this hop-distance is certainly of importance, it has been argued that the delay of communication (i.e., the stretch between pairs of peers) is a more relevant quality measure. Based on results achieved in [17], systems such as [1, 2, 18, 21] guarantee a provably bounded stretch with a limited number of links per peer. All of these systems are structured and peers are supposed to participate in a carefully predefined topology. Our paper complements this line of research by analyzing topologies as they are created by *selfish peers*, which are interested only in optimizing their *individual* trade-off between locality and maintenance overhead.

# 4 Price of Anarchy

The Price of Anarchy is a measure to bound the degradation of a globally optimal solution caused by selfish individuals. In this section, we show that the topologies created by selfish peers deteriorate more (compared to collaborative networks) as the cost of maintaining links becomes more important (larger $\alpha$). Concretely, in Section 4.1 we prove that for *arbitrary* metric spaces—thus, including the important and well-studied *growth-bounded* [14] and *doubling* (e.g. [7]) metrics—, the Price of Anarchy never exceeds $O(\min(\alpha, n))$. We then show in Section 4.2 that this bound is tight even in the "simplest" metric space, the 1-dimensional Euclidean space, where there exist Nash equilibria with a Price of Anarchy of $\Omega(\min(\alpha, n))$.

## 4.1 Upper Bound

Assume the most general setting where $n$ peers are arbitrarily located in a given metric space $\mathcal{M}$, and consider a peer $\pi$ which has to find a suitable neighbor set. Clearly, the *maximal* stretch from $\pi$ to any other peer $\pi'$ in the system is at most $\alpha + 1$: If $stretch(\pi, \pi') > \alpha + 1$, $\pi$ could establish a direct link to $\pi'$, reducing the stretch from more than $\alpha + 1$ to 1, while incurring a link cost of $\alpha$. Therefore, in any Nash equilibrium, no stretch exceeds $\alpha + 1$. Because there are at most $n(n-1)$ directed links (from each peer to all remaining peers), the social cost of a Nash equilibrium is $O(\alpha n^2 + \alpha n^2)$. In the social optimum on the other hand, all stretches are at least 1 and there must be at least $n - 1$ links in order to keep the topology connected. This lower bounds the social cost by $\Omega(\alpha n + n^2)$ and yields the following

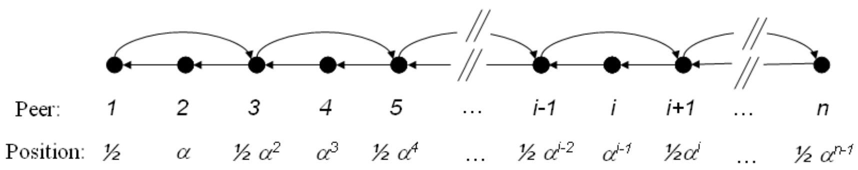


Figure 1: Example topology $G$ where the Price of Anarchy is $\Theta(\min(\alpha, n))$ for $3.4 \leq \alpha$. The peers are arranged on a 1-dimensional Euclidean line, with exponentially increasing distances. Even peers are only connected to the nearest peer on the left, while odd peers additionally have a link to the second nearest peer on their right. Observe that every peer has stretch 1 to all peers on the left.

result.

**Theorem 4.1.** *For any metric space $\mathcal{M}$, the Price of Anarchy is $O(\min(\alpha, n))$.*

Theorem 4.1 implies that if the relative importance of the peers' stretch is large, the Price of Anarchy is small. That is, for small $\alpha$, the selfish peers have an incentive to establish links to many other peers, while also the optimal network is highly connected.

## 4.2 Lower Bound

We now show that there are P2P networks in which the Price of Anarchy is as bad as $\Omega(\min(\alpha, n))$, which implies that the upper bound of Section 4.1 is asymptotically tight. Intriguingly, the Price of Anarchy can deteriorate to $\Theta(\min(\alpha, n))$ even if the underlying latency metric describes a simple 1-dimensional Euclidean space.

Consider the topology $G$ in Figure 4.2 in which peers are located on a line, and the distance (latency) between two consecutive peers increases exponentially towards the right. Concretely, peer $i$ is located at position $\alpha^{i-1}/2$ if $i$ is odd, and at position $\alpha^{i-1}$ if $i$ is even. The peers of $G$ maintain links as follows: All peers have a link to their nearest neighbor on the left. Odd peers additionally have a link to the second nearest peer on their right. After proving that $G$ constitutes a Nash equilibrium, we derive the lower bound on the Price of Anarchy by computing the social cost of this topology.

**Lemma 4.2.** *The topology $G$ shown in Figure 4.2 forms a Nash equilibrium for $\alpha \geq 3.4$.*

*Proof.* We distinguish between even and odd peers. For both cases, we show that no peer has an incentive to deviate from its strategy.

***Case even peers:*** Every even peer $i$ needs to link to at least one peer on its left, otherwise $i$ cannot reach the peers $j < i$. A connection to peer $i-1$ is optimal, as the stretch to all peers $j < i$ becomes 1. Observe that every alternative link to the left would imply a larger stretch to at least one peer on the left without reducing the stretch to peers on the right. Furthermore, $i$ cannot reduce the distance to any—neither left nor right—peer by adding further links to the left. Hence, it only remains to show that $i$ cannot benefit from adding more links to the right.

By adding a link to the right, peer $i$ shortens the distance to *all* peers on the right. However, we show that the cost reduction per peer decreases as a geometric series, and any such link to the right would strictly increase $i$'s costs. We consider two cases: $i$ linking to an odd peer on the right, and to an even peer on the right.

*Link to an odd peer:* Consider the benefit of $i$ adding a link to its odd neighbor $i+1$. For an odd peer $j > i$, we define the *benefit* $B_{i,j}$ as the stretch cost reduction caused by the addition of the link $(i, i+1)$. We have, for $i \geq 2$,

$$\begin{aligned}
B_{i,j} &= stretch_{old}(i,j) - stretch_{new}(i,j) \\
&= \frac{d(i,i-1) + d(i-1,j)}{d(i,j)} - \frac{d(i,j)}{d(i,j)} \\
&= \frac{\alpha^{i-1} - \frac{1}{2}\alpha^{i-2} + \frac{1}{2}\alpha^{j-1} - \frac{1}{2}\alpha^{i-2}}{\frac{1}{2}\alpha^{j-1} - \alpha^{i-1}} - 1 \\
&= \frac{2\alpha^{i-1} - \alpha^{i-2}}{\frac{1}{2}\alpha^{j-1} - \alpha^{i-1}} = \frac{2 - \frac{1}{\alpha}}{\frac{1}{2}\alpha^{j-i} - 1}
\end{aligned}$$

Similarly, the savings $B_{i,j}$ for an even peer $j > i$ and $i \geq 2$ amount to $B_{i,j} = stretch_{old}(i,j) - stretch_{new}(i,j) = (d(i,i-1) + d(i-1,j+1) + d(j+1,j))/(d(i,j)) - (d(i,j+1) + d(j+1,j))/(d(i,j)) = (\alpha^{i-1} - \alpha^{i-2} + \alpha^{j} - \alpha^{j-1})/(\alpha^{j-1} - \alpha^{i-1}) - (\alpha^{j} - \alpha^{i-1} - \alpha^{j-1})/(\alpha^{j-1} - \alpha^{i-1}) = (2\alpha^{i-1} - \alpha^{i-2})/(\alpha^{j-1} - \alpha^{i-1}) = (2 - \frac{1}{\alpha})/(\alpha^{j-i} - 1)$ Hence, for all $\alpha \geq 3.4$, the total savings $B_i$ for peer $i$ are less than

$$
\begin{aligned}
B_i &= \sum_{\text{odd } j > i} B_{i,j} + \sum_{\text{even } j > i} B_{i,j} \\
&= \sum_{\delta=1}^{\infty} \frac{2-\frac{1}{\alpha}}{\frac{1}{2}\alpha^{2\delta-1}-1} + \sum_{\delta=1}^{\infty} \frac{2-\frac{1}{\alpha}}{\alpha^{2\delta}-1} \\
&\underset{(\alpha\geq 3)}{\leq} \sum_{\delta=1}^{\infty} \frac{2-\frac{1}{\alpha}}{\frac{1}{2}\alpha^{2\delta-2}} + \sum_{\delta=1}^{\infty} \frac{2-\frac{1}{\alpha}}{\alpha^{2\delta-1}} \\
&= \left(2-\frac{1}{\alpha}\right) \sum_{\delta=1}^{\infty} \left(\frac{1}{\frac{1}{2}\alpha^{2\delta-2}} + \frac{1}{\alpha^{2\delta-1}}\right) \\
&= \left(2-\frac{1}{\alpha}\right) \left(\frac{2\alpha^2}{\alpha^2-1} + \frac{\alpha}{\alpha^2-1}\right) \\
&= \frac{4\alpha^2-1}{\alpha^2-1} \underset{(\alpha\geq 3.4)}{<} \alpha+1
\end{aligned}
$$

Therefore, the construction of link $(i, i+1)$ would be of no avail (benefit smaller than cost). The benefit of alternative or additional links to odd neighbors on the right is even smaller.

*Link to an even peer:* A link to an even peer $j > i$ entails a stretch 1 to the corresponding peer instead of $stretch_{old}(i,j) = (\alpha^j - \alpha^{j-1} + \alpha^{i-1} - \alpha^{i-2})/(\alpha^{j-1} - \alpha^{i-1}) < \alpha + 1$ for $\alpha > 2$. However, the stretch from $i$ to all other peers remains unchanged, since the path $i \rightsquigarrow (i-1) \rightsquigarrow (i+1)$ is shorter than $i \rightsquigarrow (i+2) \rightsquigarrow (i+1)$: $\alpha^{i-1} - \frac{1}{2}\alpha^{i-2} + \frac{1}{2}\alpha^i - \frac{1}{2}\alpha^{i-2} < \alpha^{i+1} - \alpha^{i-1} + \alpha^{i+1} - \frac{1}{2}\alpha^i$ for $\alpha > 1$. Therefore, an even peer $i$ has no incentive to build links to any even peer on its right.

***Case odd peers:*** An odd peer $i$ needs to link to peer $i-1$, otherwise there is no connection to $i-1$ and the stretch from $i$ to $i-1$ is infinite. Moreover, if the link $(i, i-1)$ is established, $stretch(i,j) = 1$ for all $j < 1$. Therefore, peer $i$ does not profit from building additional or alternative links to the left.

It remains to study links to the right. In order to reach all peers with a finite stretch, peer $i$ needs a link to some peer $j \geq i+2$. In the following, we first show that peer $i$ can always benefit from a link $(i, i+2)$, independently of additional links to the right. Secondly, we prove that if $i$ has a link $(i, i+2)$, it has no incentive to add further links.

Assume peer $i$ has no direct link to peer $i+2$. Then, $stretch(i, i+2) \geq (2\alpha^{i+2} - \frac{1}{2}\alpha^{i-1} - \frac{1}{2}\alpha^{i+1})/(\frac{1}{2}\alpha^{i+1} - \frac{1}{2}\alpha^{i-1}) > \alpha + 1$. Hence, no matter which links it already has, peer $i$ can benefit by additionally pointing to peer $i+2$. On the other hand, if $i$ maintains the link $(i, i+2)$, any other links to the right only reduce $i$'s gain. For *odd* peers, this is obvious, since the corresponding stretches are already optimal. A link $(i, j)$ to some *even* peer $j > i$ only improves the stretch to peer $j$ itself, but not to other peers. The stretch to peer $j$ becomes 1 instead of $stretch_{old}(i, j) = (\frac{1}{2}\alpha^{j+1} - \frac{1}{2}\alpha^{i-1} + \frac{1}{2}\alpha^{j+1} - \alpha^{j})/(\alpha^{j} - \frac{1}{2}\alpha^{i-1}) = (\alpha^{j+1} - \alpha^{j} - \frac{1}{2}\alpha^{i-1})/(\alpha^{j} - \frac{1}{2}\alpha^{i-1}) < \alpha + 1$ for $\alpha > 0$. Thus, also this link would increase $i$'s costs. □

Having verified that the topology of Figure 4.2 is a Nash equilibrium, we compute its social cost.

**Lemma 4.3.** *The social cost $C(G)$ of the topology $G$ shown in Figure 4.2 is $C(G) \in \Theta(\alpha n^2)$.*

*Proof.* The topology $G$ has $n-1$ links pointing to the left and $\lfloor n/2 \rfloor$ links pointing to the right. Hence, the total link costs are

$$C_E(G) = \alpha \left[(n-1) + \lfloor n/2 \rfloor\right] \in \Theta(\alpha n).$$

It remains to compute the costs of the stretches.

The stretch from an odd peer $i$ to an even peer $j > i$ is $stretch(i, j) = (\alpha^{j} - \alpha^{j-1} - \frac{1}{2}\alpha^{i-1})/(\alpha^{j-1} - \frac{1}{2}\alpha^{i-1}) > (\frac{1}{2}\alpha^{j} - \frac{1}{2}\alpha^{i-1})/(\alpha^{j-1} - \frac{1}{2}\alpha^{i-1}) > \frac{1}{2}\alpha$ for $\alpha > 2$. Thus, the sum of the stretches of an odd peer $i$ is

$$\begin{aligned} C_S(i) &= \sum_{j<i} stretch(i, j) + \sum_{j>i} stretch(i, j) \\ &> (i-1) + \frac{1}{2}\alpha \left\lfloor \frac{n-i-1}{2} \right\rfloor + \left\lfloor \frac{n-i}{2} \right\rfloor. \end{aligned}$$

The stretch between two even peers $i$ and $j$ is $stretch(i, j) = (\alpha^{j} - \alpha^{j-1} + \alpha^{i-1} - \alpha^{i-2})/(\alpha^{j-1} - \alpha^{i-1}) > (\frac{1}{2}\alpha^{j} - \frac{1}{2}\alpha^{i-1})/(\alpha^{j-1} - \alpha^{i-1}) > \frac{1}{2}\alpha$ for $j > i$ and all $\alpha > 2$. Thus, the stretch costs are at least

$$C_S(i) > (i-1) + \frac{1}{2}\alpha \left\lfloor \frac{n-i-1}{2} \right\rfloor - 1 + \left\lfloor \frac{n-i-1}{2} \right\rfloor.$$

Adding up the stretches of odd and even peers yields a lower bound on the total stretch costs:

$$\begin{aligned} C_S(G) &= \sum_{i\ even} C_S(i) + \sum_{i\ odd} C_S(i) \\ &> \frac{n(n-2)}{2} + \alpha\frac{(n-3)(n-2)-n}{8} \\ &+ \frac{(n-1)(n-2)}{4} \in \Omega(\alpha n^2). \end{aligned}$$

Thus, in combination with Theorem 4.1, it follows that $C_S(G) \in \Theta(\alpha n^2)$. The proof is concluded by combining link and stretch costs, $C(G) = C_E(G) + C_S(G) \in \Theta(\alpha n^2)$. □

**Theorem 4.4.** *The Price of Anarchy of the peer topology G shown in Figure 4.2 is* $\Theta(\min(\alpha, n))$.

*Proof.* The upper bound follows directly from the result obtained in Theorem 4.1. As for the lower bound, if $\alpha < 3.4$, the theorem holds because $\Theta(\min\{\alpha, n\}) \in O(1)$ in this case. By Lemma 4.2, the topology $G$ constitutes a Nash equilibrium for $\alpha \geq 3.4$. Moreover, by Lemma 4.3, the social cost of $G$ are in order of $\Theta(\alpha n^2)$. In the following, we prove that the optimal social cost is upper bounded by $O(n^2 + \alpha n)$ from which the claim of the theorem follows by dividing the two expressions.

Consider again the peer distribution shown in Figure 4.2, and assume that there are no links. If every peer connects to the nearest peer to its left and to the nearest peer to its right, there are $2(n-1)$ links, and all stretches are 1. Thus, the social cost of this resulting topology $\widetilde{G}$ is $C(\widetilde{G}) = \alpha \cdot 2(n-1) + n(n-1) \in O(n^2 + \alpha n)$. The optimal social cost is at most the social cost of $\widetilde{G}$. □

# 5 Existence of Nash Equilibria

In this section, we show that a system of selfish peers may never converge to a stable state, even in the absence of churn, mobility, or other sources of dynamics. Interestingly, this result even holds if we assume latencies to form simple metric spaces, such as a 2-dimensional Euclidean space. Specifically, there may not exist a pure Nash equilibrium for certain P2P networks in our "locality game".

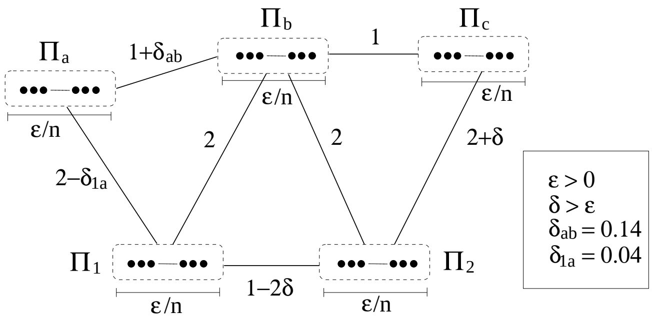


Figure 2: Instance $I_k$ has no pure Nash equilibrium when $\alpha = 0.6k$, where $k = n/5$. The number of peers in each cluster is $k$.

**Theorem 5.1.** *Regardless of the magnitude of $\alpha$, there are metric spaces $\mathcal{M}$, for which there exists no pure Nash equilibrium, i.e. certain P2P networks cannot converge to a stable state. This is the case even if $\mathcal{M}$ is a 2-dimensional Euclidean space.*

Instead of presenting the formal proof (which will be implicit in the proof of Theorem 6.1), we attempt to highlight the main idea only. Assume that the parameter $\alpha$ is a multiple of 0.6, i.e., $\alpha_k = 0.6k$ for an arbitrary integer $k > 0$. Given a specific $k$, the 2-dimensional Euclidean instance $I_k$ of Figure 2 has no pure Nash equilibrium. Specifically, $I_k$ constitutes a situation in which there are peers $\pi_1 \in \Pi_1$ and $\pi_2 \in \Pi_2$ that continue to deviate to a better strategy ad infinitum, i.e., the system cannot converge.

The $n$ peers of instance $I_k$ are grouped into five clusters $\Pi_1$, $\Pi_2$, $\Pi_a$, $\Pi_b$, and $\Pi_c$, each containing $k = n/5$ peers. Within a cluster, peers are located equidistantly in a line, and each cluster's diameter is $\epsilon/n$, where $\epsilon > 0$ is an arbitrarily small constant. The *inter-cluster distance* $d(\Pi_i, \Pi_j)$ between $\Pi_i$ and $\Pi_j$ is the minimal distance between any two peers in the two clusters. Distances not explicitly defined in Figure 2 follow implicitly from the constraints imposed by the underlying Euclidean plane.

The proof unfolds in a series of lemmas that characterize the structure of the resulting topology $G[s]$ if the strategies $s$ form a Nash equilibrium in $I_k$. First, it can be shown that in $G[s]$, two peers in the same cluster are always connected by a path that does not leave the cluster. Secondly, it can be shown that there exists exactly one link in both directions between clusters $\Pi_a$ and $\Pi_b$, $\Pi_b$ and $\Pi_c$, as well as between $\Pi_1$ and $\Pi_2$. A third structural characteristic of any Nash equilibrium is that for every $i$ and $j$, there is *at most one* directed link from a cluster $\Pi_i$ to peers in a cluster $\Pi_j$.

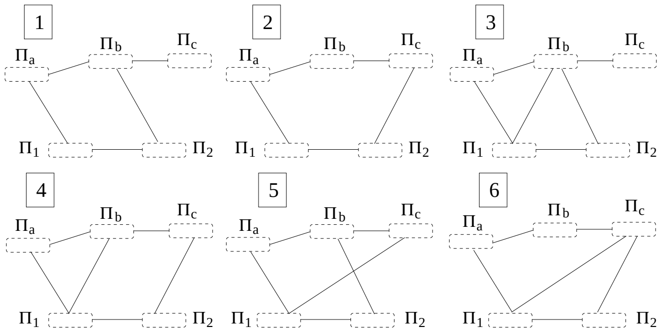


Figure 3: Candidates for a Nash equilibrium.

To preserve connectivity, some peers in $\Pi_1$ and $\Pi_2$ must have links to top-peers. Based on the aforementioned observations, the set of possible strategies can further be narrowed down as follows.

i) Neither peers in $\Pi_1$ nor $\Pi_2$ select three links to top-peers.

ii) There exists a peer $\pi_1 \in \Pi_1$ that establishes a link to $\Pi_a$.

iii) There is exactly one link from cluster $\Pi_2$ to either cluster $\Pi_b$ or $\Pi_c$, but there is no link to $\Pi_a$.

Correctness of all three properties is proven by verifying that there exists some peer $\pi_1 \in \Pi_1$ or $\pi_2 \in \Pi_2$ that has an incentive to change its strategy in case the property is not satisfied. If, for instance, there are two peers $\pi_2, \pi_2' \in \Pi_2$ that simultaneously maintain links to $\Pi_b$ and $\Pi_c$, (thus violating case iii)), $\pi_2'$ can lower its costs by dropping its link to $\Pi_c$. This holds because the sum of the stretches $\sum_{\pi_c \in \Pi_c} stretch(\pi_2', \pi_c)$ entailed by the indirection $\pi_2' \rightsquigarrow \pi_2 \rightsquigarrow \Pi_b \rightsquigarrow \Pi_c$ does not justify the additional cost $\alpha$.

It can be shown that only the six structures depicted in Figure 3 remain valid candidates for Nash topologies. In each scenario, however, at least one peer benefits from deviating from its current strategy.

***Case 1:*** In this case, a peer $\pi_1 \in \Pi_1$ can reduce its cost by adding a link to a peer in $\Pi_b$.

***Case 2:*** If the only outgoing link from $\Pi_1$ to a top-cluster is to cluster $\Pi_a$, the peer $\pi_2 \in \Pi_2$ maintaining the link to $\Pi_c$ can be shown to profit from switching its link from $\Pi_c$ to $\Pi_b$.

***Case 3:*** The availability of the link from $\Pi_1$ to $\Pi_b$ changes the optimal choice of the above mentioned peer $\pi_2 \in \Pi_2$. Unlike in the previous case, $\pi_2$ now prefers linking to $\Pi_c$ instead of $\Pi_b$.

***Case 4:*** Due to the existence of a link from a peer $\pi_2 \in \Pi_2$ to $\Pi_c$, the peer $\pi_1 \in \Pi_1$ with the link to $\Pi_b$ has an incentive to drop this link and instead use the detours via $\Pi_2$ and $\Pi_a$ to connect to $\Pi_c$ and $\Pi_b$, respectively.

***Case 5:*** In this case, the peer $\pi_1 \in \Pi_1$ having the link to $\Pi_c$ reduces its cost by replacing this link with a link to a peer in $\Pi_b$.

***Case 6:*** Finally, this case is similar to Case 4: $\pi_1 \in \Pi_1$ with the link to $\Pi_b$ has an incentive to remove its link to $\Pi_c$

These cases highlight how the system is ultimately trapped in an infinite loop of strategy changes, without ever converging to a stable situation. There is always at least one peer which can reduce its cost by changing its strategy. For instance, the following sequence of topology changes could repeat forever (cf. Figure 3): $1 \leadsto 3 \leadsto 4 \leadsto 2 \leadsto 1 \leadsto 3 \ldots$ In other words, selfish peers will not achieve a stable network topology.

# 6 Complexity of Nash Equilibria

The question is whether for a given P2P network, it can be determined if it will eventually converge to a stable state or not. In the following, we show that it is NP-hard to decide whether there exists a pure Nash equilibrium. This result establishes the *complexity of stability* in unstructured P2P networks, showing that in general, it is computationally infeasible to determine whether a peer-to-peer network consisting of selfish peers can stabilize or not.

Ever since Papadimitriou's influential survey on game theoretic aspects of the Internet [15], the complexity of Nash equilibria has become one of the most active and fruitful areas in recent theoretical computer science research. Numerous profound results on important families of games have been presented, e.g. in [11, 16]. While these works investigate the complexity of *finding* a Nash equilibrium,

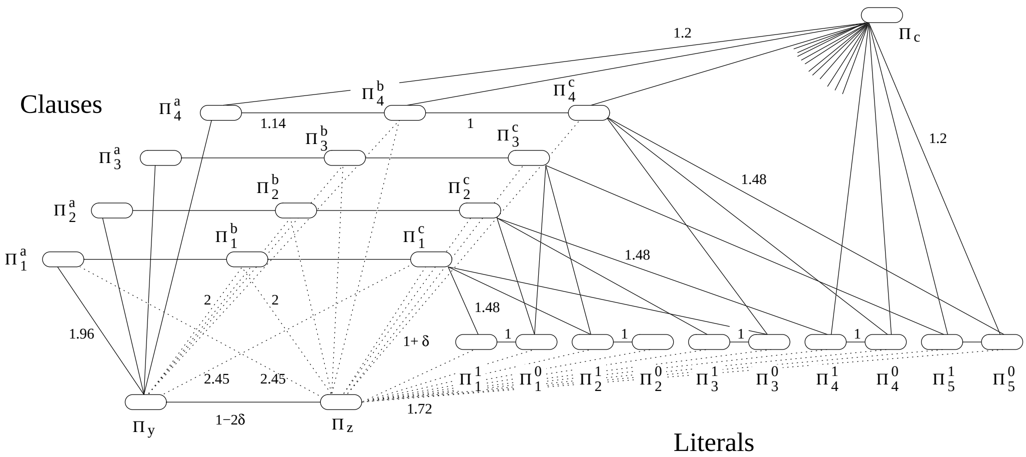


Figure 4: The graph $G_I$ for instance $I = (x_1 \vee x_2 \vee \overline{x_3}) \wedge (\overline{x_1} \vee x_3 \vee x_4) \wedge (\overline{x_1} \vee x_2 \vee x_5) \wedge (\overline{x_3} \vee \overline{x_4} \vee \overline{x_5})$. Each cluster contains $k$ peers with pairwise distance of $\epsilon$. $\delta$ is an arbitrarily constant such that $\delta > \epsilon > 0$.

we show that it is NP-hard to *decide whether there exists* a pure Nash equilibrium in a given P2P network.

**Theorem 6.1.** *Regardless of the magnitude of $\alpha$, determining whether a given P2P network represented by a metric space $\mathcal{M}$ has a pure Nash equilibrium (and can therefore stabilize) is NP-hard.*

The proof being rather technical, we first describe its main intuition. Due to lack of space, most parts of the proof are omitted from this extended abstract. The proof is based on a reduction from an NP-complete form of 3-SAT in which each variable appears in at most 3 clauses [12]. For any $\alpha$ a multiple of 0.6, i.e., $\alpha_k = 0.6k$ for an integer $k > 0$, we give a polynomial time construction of a metric space $\mathcal{M}_I^k$ from an instance $I$ of 3-SAT, such that the following holds: There exists a pure Nash equilibrium in $\mathcal{M}_I^k$ if and only if $I$ is satisfiable.

The reduction is illustrated in Figure 4, each rectangular box representing a cluster of $k$ peers. Assume that the 3-SAT instance is given in standard CNF normal form. For each clause $C_j$, we employ a gadget of three *clause-clusters* $\Pi_j^a$, $\Pi_j^b$, and $\Pi_j^c$. For every variable $x_i$, the two *literal-clusters* $\Pi_i^0$ and $\Pi_i^1$ represent the negative and positive literal of the variable, respectively. Finally, the construction's peer set is completed with three special clusters $\Pi_c$, $\Pi_y$, and $\Pi_z$. The pairwise distances between two peers in $\mathcal{M}_I^k$ are determined by the undirected weighted graph $G_I^k$ shown in Figure 4 (a formal definition appears in Section 6.1). Two nodes within the same cluster have a distance of $\epsilon$, for some arbitrarily small $\epsilon < (k(2n+3m+3))^{-2}$, where $m$ and $n$ denote the number of clauses and variables in $I$, respectively. An edge of $G_I^k$ describes the *cluster-distance* between two clusters: The mutual distance between *every pair* of two peers $\pi_i \in \Pi_i$ and $\pi_j \in \Pi_j$ in neighboring clusters $\Pi_i$ and $\Pi_j$ with cluster-distance $X$ is $d(\pi_i, \pi_j) = X$. All other distances are determined by the length of the shortest path between the peers in $G_I^k$, that is, $\mathcal{M}_I^k$ corresponds to the *shortest path metric* induced by $G_I^k$. Note that while $\mathcal{M}_I^k$ cannot be embedded in the Euclidean space, it still forms a valid metric space, i.e., it fulfils symmetry and triangle inequality.

Consider an arbitrary clause $C_j$. Its clause-clusters $\Pi_j^a$, $\Pi_j^b$, and $\Pi_j^c$ in combination with the two special clusters $\Pi_y$ and $\Pi_z$ form an instance similar to $I_k$ as used in the discussion of Theorem 5.1 (cf. Figure 2). Hence, intuitively, when considering such a clause-gadget by itself, it does not have a pure Nash equilibrium. In order to make a clause-gadget stable, however, literal clusters may be used. For this purpose, the cluster-distance between each pair of corresponding literals is 1 and peers in $\Pi_z$ have a distance of 1.72 to all literal-peers. Furthermore, the distance between a clause-cluster $\Pi_j^c$ and a literal-cluster depends on whether the corresponding literal appears in the clause. Specifically, if the positive literal $x_i$ appears in clause $C_j$, $x_i \in C_j$, the distance between $\Pi_i^1$ and $\Pi_j^c$ is small, i.e., only 1.48. Similarly, if $\overline{x_i} \in C_j$, then $d(\Pi_i^1, \Pi_j^c) = 1.48$. And finally, if neither literal is in $C_j$, then there exists no short connection between the clusters, and the shortest distance between peers in these clusters is via $\Pi_c$.

The proof comprises two ingredients. First, we prove that if the underlying SAT instance $I$ is *not satisfiable*, then there exists no Nash equilibrium. Towards this end, we show that in any Nash equilibrium two "neighboring" clusters (clusters connected by a short link in $G_I^k$, such as two clause-clusters in the same clause, a literal-cluster $\Pi_i^1$ to a clause-cluster $\Pi_j^c$ if $x_i \in C_j$, or $\Pi_c$ to all clause-clusters and literal-clusters,...) always establish links in both directions between them. Between such close-by clusters, there are always exactly two links, one in each direction. Furthermore, for every variable $x_i$, there is exactly one peer $\pi_z \in \Pi_z$ that establishes a link to exactly either $\Pi_i^1$ or $\Pi_i^0$ (but not both!), while no other peer in $\Pi_z$ links to these clusters.

From these lemmas, it then follows that because $I$ is not satisfiable, there must exist a clause $C_{j^*}$ for which the path from $\pi_z \in \Pi_z$ to peers in $\Pi^c_{j^*}$ via any literal-peer has length at least $d(\Pi_z, \Pi^{\mu}_i) + d(\Pi^{\mu}_i, \Pi^{1-\mu}_i) + d(\Pi^{1-\mu}_i, \Pi^c_{j^*}) = 4.2$, for $\mu \in \{0, 1\}$. This path being long, it follows that it is worthwhile for $\pi_z$ to build an additional link directly to some peer in $\Pi^c_{j^*}$ or even in $\Pi^b_{j^*}$ instead. Based on these observations, we show that the subset of $\mathcal{M}^k_I$ induced by peers in $\Pi_y$, $\Pi_z$, and the clause-peers of $C_{j^*}$ behaves similarly as in instance $I_k$ of Figure 2. That is, peers in $\Pi_y$ and $\Pi_z$ continue to change their respective strategies forever, thus preventing the system from stabilizing.

On the other hand, if the SAT instance $I$ has a *satisfying assignment* $A_I$, we explicitly construct a set of pure strategies that constitute a Nash equilibrium. In this strategy vector, one peer in $\Pi_z$ builds a direct link to a peer in $\Pi^1_i$ if $x_i$ is set to true in $A_I$ and to a peer in $\Pi^0_i$ otherwise. Since $A_I$ is a satisfying assignment, there must exist a path from $\Pi_z$ via a single literal-cluster (i.e., without the additional detour of going from one literal-cluster to the other) to peers in every cluster $\Pi^c_j$. This path can be shown to have length at most $k\epsilon + d(\Pi_z, \Pi^{\mu}_i) + k\epsilon + d(\Pi^{\mu}_i, \Pi^c_j) + k\epsilon = 3.2 + 3k\epsilon$ from $\Pi_z$ via a literal-cluster to peers in every cluster $\Pi^c_j$. It follows that in any satisfied clause $C_j$, the achievable reduction in stretch costs at a peer in $\Pi_z$ when connecting directly to clusters $\Pi^b_j$ or $\Pi^c_j$ is significantly smaller than in an unsatisfied clause. Specifically, it can be shown that peers in $\Pi_y$ and $\Pi_z$ are in a *stable* situation if one peer $\pi_y \in \Pi_y$ connects to $\Pi^a_j$ and $\Pi^b_j$ of every clause $C_j$, and no peer in $\Pi_z$ directly builds a link to any clause-peer. Since $A_I$ is a satisfying assignment, peers in $\Pi_y$ and $\Pi_z$ are stable relative to *all* clauses in the SAT instance.

Furthermore, we also prove that in our strategy vector, no other peer in the network (i.e., peers in $\Pi_c$, $\Pi^a_j$, $\Pi^b_j$, $\Pi^c_j$, $\Pi^1_i$, or $\Pi^0_i$) has an incentive to deviate from its strategy. For this final ingredient of the proof, the existence of cluster $\Pi_c$ is essential, because it ensures that all helper peers are mutually connected by optimal paths.

All in all, the P2P network induced by the metric space $\mathcal{M}^k_I$ has a pure Nash equilibrium if and only if the underlying SAT instance $I$ is satisfiable. Hence, determining whether a given P2P network can stabilize is NP-hard. Section 6.1 defines the construction of $G_I^k$ (and consequently $\mathcal{M}_I^k$) from the 3-SAT instance $I$. In Sections 6.2 and 6.3, we show that there exists a Nash equilibrium in $\mathcal{M}_I^k$ if and only if $I$ is satisfiable. Theorem 6.1 then follows from Lemmas 6.11 and 6.13, as well as the NP-hardness of 3-SAT.

## 6.1 The Construction of $\mathcal{M}_\mathbf{I}^\mathbf{k}$

Let $I$ be an instance of 3-SAT expressed in conjunctive normal form (CNF), in which each clause contains 3 literals. Without loss of generality, we can assume that each variable in $I$ appears in at most 3 clauses [12]. Furthermore, we can restrict our attention to those instances of 3-SAT in which every variable appears in most 2 positive and 2 negative literals, because otherwise, the variable appears as a positive or negative literal only, which renders assigning a feasible value to this variable trivial. The set of clauses and variables of $I$ is denoted by $\mathcal{C}$ and $\mathcal{X}$, respectively. Further, we write $m = |\mathcal{C}|$ and $n = |\mathcal{X}|$. Given $I$, we construct an undirected weighted graph $G_I^k = (V_I, E_I)$ in which each node represents a peer of the underlying network. Nodes are grouped into clusters of $k$ peers and each cluster is illustrated as a rectangular box in Figure 4. Within each cluster, the pairwise distance between two peers is $\epsilon < (k(2n+3m+3))^{-2}$, and the distance between two peers in neighboring clusters is described by the *cluster-distance* $d(\Pi_i, \Pi_j)$ illustrated in Figure 4. The P2P network is then characterized by $\mathcal{M}_I^k$, which is induced by the *shortest path metric* of $G_I^k$, i.e., the distance between two peers corresponds to the length of the shortest path in $G_I^k$.

In more detail, $G_I^k$ is defined as follows. The node-set $V_I$ consists of three clusters of peers per clause $C_j \in \mathcal{C}$, denoted as *clause-clusters* $\Pi_j^a$, $\Pi_j^b$, and $\Pi_j^c$. Also, we add a pair of *literal-clusters* $\Pi_i^0$ and $\Pi_i^1$ for each of the $n$ variables, with $\Pi_i^0$ representing the positive literal $x_i$, and $\Pi_i^1$ representing the negative literal $\overline{x_i}$. The set of clause-peers and literal-peers is denoted by $C_P$ and $L_P$, respectively. Finally, there are three additional special clusters $\Pi_c$, $\Pi_x$, and $\Pi_y$. Call the union of $\Pi_c$ and all clusters in $C_P$ and $L_P$ *top-layer clusters*. Peers in top-layer clusters are *top-layer peers*. The total number of peers $N$ in the network $\mathcal{M}_I^k$ is therefore $N = k(2n + 3m + 3)$. Notice that $N \cdot \epsilon$ is smaller than

$(k(2n+3m+3))^{-1}$.

The pairwise distances between the peers in different clusters—as illustrated in Figure 4—are formally defined as follows. Let $\delta$ be an arbitrarily small constant with $\delta > 10k\epsilon$, and $\mu \in \{0,1\}$. For all $\pi_c \in \Pi_c$ and $\pi_w \in C_P \cup L_P$, it holds that $d(\pi_c, \pi_w) = 1.2$. For every $C_j \in \mathcal{C}$, the following distances apply.

$$
\begin{aligned}
\forall \pi_y \in \Pi_y, \forall \pi_j^a \in \Pi_j^a: &\qquad d(\pi_y, \pi_j^a) = 1.96 \\
\forall \pi_y \in \Pi_y, \forall \pi_j^b \in \Pi_j^b: &\qquad d(\pi_y, \pi_j^b) = 2 \\
\forall \pi_y \in \Pi_y, \forall \pi_j^c \in \Pi_j^c: &\qquad d(\pi_y, \pi_j^c) = 2.45 \\
\forall \pi_z \in \Pi_z, \forall \pi_j^a \in \Pi_j^a: &\qquad d(\pi_z, \pi_j^a) = 2.45 \\
\forall \pi_z \in \Pi_z, \forall \pi_j^b \in \Pi_j^b: &\qquad d(\pi_z, \pi_j^b) = 2 \\
\forall \pi_z \in \Pi_z, \forall \pi_j^c \in \Pi_j^c: &\qquad d(\pi_z, \pi_j^c) = 2+\delta \\
\forall \pi_j^a \in \Pi_j^a, \forall \pi_j^b \in \Pi_j^b: &\qquad d(\pi_j^a, \pi_j^b) = 1.14 \\
\forall \pi_j^b \in \Pi_j^b, \forall \pi_j^c \in \Pi_j^c: &\qquad d(\pi_j^b, \pi_j^c) = 1.
\end{aligned}
$$

For every variable $x_i \in \mathcal{X}$, it holds that

$$
\begin{aligned}
\forall \pi_i^0 \in \Pi_i^0, \forall \pi_i^1 \in \Pi_i^1: &\qquad d(\pi_i^0, \pi_i^1) = 1 \\
\forall \pi_z \in \Pi_z, \forall \pi_i^\mu \in \Pi_i^\mu: &\qquad d(\pi_z, \pi_i^0) = d(\pi_z, \pi_i^1) = 1.72.
\end{aligned}
$$

Furthermore, it holds for all $C_j \in \mathcal{C}, x_i \in C_j$ that

$$
\forall \pi_i^1 \in \Pi_i^1, \forall \pi_j^c \in \Pi_j^c: \quad d(\pi_i^1, \pi_j^c) = 1.48
$$

and for all $C_j \in \mathcal{C}, \overline{x_i} \in C_j$ that

$$\forall \pi_i^0 \in \Pi_i^0, \forall \pi_j^c \in \Pi_j^c : \quad d(\pi_i^0, \pi_j^c) = 1.48.$$

Finally, the distance between any two peers $\pi_y \in \Pi_y$ and $\pi_z \in \Pi_z$ is $d(\pi_y, \pi_z) = 1 - 2\delta$. All distances not explicitly defined follow from the shortest path metric induced by the above definitions.

Intuitively, the idea of the construction is the following. Each clause $C_j \in \mathcal{C}$ is represented by a gadget consisting of the two clusters $\Pi_y$, $\Pi_z$, as well as the clause-clusters $\Pi_j^a$, $\Pi_j^b$, and $\Pi_j^c$. By itself, each such gadget is reminiscent of the construction shown in Figure 2. Specifically, this implies that the sub-network induced by each such clause-gadget does not have a pure Nash equilibrium when considered independently from the rest of the network.

In order to render a clause-gadget stable, literal-peers can be used. In particular, it can be shown that for $\mu \in \{0, 1\}$, the peers in every literal-cluster $\Pi_i^\mu$ construct links to those (at most two) clause-clusters $\Pi_j^c$ in whose clause the literal occurs, that is, if $x_i^\mu \in C_j$. Based on this and other structural properties of Nash equilibria in $\mathcal{M}_I^k$, it can further be shown that in a Nash equilibrium, there is exactly one link from cluster $\Pi_z$ to each variable $x_i \in \mathcal{X}$, i.e., one peer in $\Pi_z$ connects to a peer in either $\Pi_i^0$ or $\Pi_i^1$ for all $x_i \in \mathcal{X}$.

Consider a clause $C_j$. If there is a peer $\pi_z \in \Pi_z$ that connects to at least one literal-cluster that is directly connected to $\Pi_j^c$, the length of the path from $\pi_z$ to peers in $\Pi_j^c$ via this literal-cluster is at most $k\epsilon + d(\Pi_z, \Pi_i^\mu) + k\epsilon + d(\Pi_i^\mu, \Pi_j^c) + k\epsilon = 3.2 + 3k\epsilon$. In this case, the detour from $\pi_z$ to $\Pi_j^c$ via some "satisfying" literal-cluster $\Pi_i^\mu$—while being suboptimal compared to the direct connection—is relatively small. Specifically, it is small enough to ensure that no peer in $\Pi_z$ has an incentive to construct an additional direct link to $\Pi_j^b$ or $\Pi_j^c$. Once peers in $\Pi_z$ have no further need to establish direct links to a clause-peer of $C_j$, the best possible strategy of peers in $\Pi_y$ becomes fixed, too. In other words, this satisfying literal helps in *stabilizing* the clause-gadget.

Conversely, if there is a clause $C_j$ for which no peer in $\Pi_z$ connects to a satisfying literal-cluster, there exists no efficient detour. Specifically, the length of the path from $\pi_z \in \Pi_z$ to $\pi_j^c \in \Pi_j^c$ via a literal-cluster is at least 4.2, including the distance between the positive and negative literal-cluster of the variable. The increased length of the detour renders the resulting stretch from $\Pi_z$ to $\Pi_j^c$ too large, and it becomes worthwhile for $\pi_z \in \Pi_z$ to construct direct links to $\Pi_j^c$, and even to $\Pi_j^b$. That is, in a sense, the network induced by the unsatisfied clause $C_j$ becomes independent of the remainder of the network and therefore does not stabilize.

Finally, the special cluster $\Pi_c$ ensures that the shortest path in $G_I^k$ (and hence the distance in $\mathcal{M}_I^k$) between two top-layer peers is small. In fact, it can be shown that there are links in both directions from every top-layer cluster to $\Pi_c$. This implies that all top-layer clusters are connected to one another almost optimally in every Nash equilibrium, thus facilitating the proof that such an equilibrium exists in case $I$ is satisfiable. We end the section with a series of lemmas that capture structural properties of $\mathcal{M}_I^k$.

**Lemma 6.2.** *Consider two peers $\pi_g$ and $\pi_g'$ in an arbitrary cluster $\Pi_g$. In a Nash equilibrium, there exists a path from $\pi_g$ to $\pi_g'$ of length at most $k\epsilon$.*

*Proof.* Because the distance between $\pi_g$ and $\pi_g'$ is $\epsilon$, it is easy to see that the shortest path between these two nodes must be located entirely in $\Pi_g$. Because the distance between each pair of peers in a cluster is $\epsilon$ and there are $k$ peers in the cluster, the claim follows. □

**Lemma 6.3.** *Consider two arbitrary clusters $\Pi_g$ and $\Pi_h$. In a Nash equilibrium, there is at most one peer $\pi_g \in \Pi_g$ that has a link to a peer in $\Pi_h$.*

*Proof.* Assume for contradiction that there are two nodes $\pi_g$ and $\pi_g'$ that maintain links to peers in $\Pi_h$. Then, $\pi_g'$ can reduce its cost by dropping its link. Doing so, the stretches to each peer in the network can increase by at most $2k\epsilon$. By the definition of $\epsilon$, it holds that $2Nk\epsilon < \alpha$ and hence, dropping the link is worthwhile. □

Based on these two lemmas, we can go on to prove more elaborate properties.

**Lemma 6.4.** *Let $\Pi_g$ and $\Pi_h$ be two clusters with cluster distance at most $d(\Pi_g, \Pi_h) \leq 1.48$. In any Nash equilibrium, there is exactly one peer $\pi_g \in \Pi_g$ that has a link to a peer in $\Pi_h$.*

*Proof.* By Lemma 6.3, there cannot be more than one peer in $\Pi_g$ that has a link to $\Pi_h$. It therefore remains to show that at least one link exists. We divide the proof in two parts and begin by showing that the claim holds for all pairs of clusters with cluster distance $d(\Pi_g, \Pi_h) \leq 1.2$. In a second step, we prove the claim for pairs of clusters with cluster distance $d(\Pi_g, \Pi_h) = 1.48$, which suffices because there are no cluster distances between $1.2$ and $1.48$ in $G_I^k$.

Consider any two clusters in the network $\mathcal{M}_I^k$ with cluster distance at most $1.2$. It follows from the construction of $G_I^k$ that the shortest path between peers in these clusters via a third cluster has length at least 2.2 (e.g., from $\Pi_i^0$ via $\Pi_i^1$ to $\Pi_c$). In other words, if there is no direct connection between the two clusters, $\pi_g$ has a stretch of at least $2.2/1.2$ to each peer in $\Pi_h$. Because $\frac{2.2k}{1.2} > \alpha + k(1 + 2k\epsilon)$, it is beneficial for $\pi_g$ to establish a direct link to the other cluster.

For the second part of the proof, consider pairs of clusters with cluster distance $d(\Pi_g, \Pi_h) = 1.48$. Specifically, we need to show the existence of a link in each direction between clusters $\Pi_j^c$ and $\Pi_i^1$, if $x_i \in C_j$, or between $\Pi_j^c$ and $\Pi_i^0$, if $\overline{x_i} \in C_j$. The shortest indirect connection between two such clusters has length at least 2.4 (via cluster $\Pi_c$) and hence, the cumulated stretch to all peers in the respective cluster without a direct link is $\frac{2.4k}{1.48} > \alpha + k(1 + 2k\epsilon)$. That is, peers in both clusters decrease their cost by paying for this direct link. □

Lemma 6.4 implies that within a clause, neighboring clause-clusters (i.e., $\Pi_j^a \leftrightarrow \Pi_j^b$ and $\Pi_j^b \leftrightarrow \Pi_j^c$, respectively) are connected in both directions in any Nash equilibrium. The same holds for corresponding literal-cluster $\Pi_i^1$ and $\Pi_i^0$, as well as for a literal-cluster $\Pi_i^1$ (or $\Pi_i^0$) and a $\Pi_j^c$ if $x_i \in C_j$ (or $\overline{x_i} \in C_j$). Also, there are links in both directions from any top-layer (clause or literal) cluster to $\Pi_c$ and vice versa. All in all, this implies that in a Nash equilibrium, every pair of top-layer peers is connected almost optimally, i.e., with stretch of less than $1+2k\epsilon$. The value $\epsilon$ being smaller than $(k(m+n+3))^{-2}$, this stretch is virtually as good as 1. Finally, there are also links between $\Pi_y$ and $\Pi_z$ in any Nash equilibrium. In the sequel of the proof, we use the fact that these "short" links are available in any Nash equilibrium without particular mention.

**Lemma 6.5.** *In a Nash equilibrium, there is exactly one peer $\pi_y \in \Pi_y$ that has a link to a peer in $\Pi_j^a$, for all $C_j \in \mathcal{C}$, and vice versa.*

*Proof.* Consider a specific $\Pi_j^a$. If there exists no direct link from $\Pi_y$ to $\Pi_j^a$, the stretch of a peer $\pi_y \in \Pi_y$ to each peer in $\Pi_j^a$ is at least $\frac{3.14}{1.96}$. Because for small enough $\epsilon$, we have $\frac{3.14k}{1.96} > \alpha + k(1 + 2k\epsilon)$, it is always worthwhile for some $\pi_y$ to build an additional link to $\Pi_j^a$. Clearly, the argument also holds for the opposite direction. □

**Lemma 6.6.** *Assume that there is a link between $\Pi_z$ and at least one literal-cluster of every variable $x_i \in \mathcal{X}$ and that there is a link between $\Pi_y$ and $\Pi_j^a$, for all $C_j \in \mathcal{C}$. Assume further that there are links in both directions between clusters with cluster distance at most 1.48. Finally, assume that all peers are connected within their cluster with a path of length at most $k\epsilon$. It holds for all $j$ that the shortest path from a peer $\pi_y \in \Pi_y$ to a peer in $V \setminus (\Pi_j^a \cup \Pi_j^b \cup \Pi_j^c)$ is not via $\Pi_j^a$, $\Pi_j^b$, or $\Pi_j^c$, even when directly connecting to such a cluster. The same holds for $\pi_z \in \Pi_z$.*

*Proof.* Recall that by assumption there exists a link from $\Pi_y$ to $\Pi_j^a$ (for every $C_j \in \mathcal{C}$) and $\Pi_z$. Hence, connecting to $\Pi_j^b$ or $\Pi_j^c$ clearly cannot reduce the stretch to peers in $\Pi_z$, $\Pi_c$, and any $\Pi_{j'}^a$, $j \neq j'$. Furthermore, the distance in the topology to any clause-peer in $\Pi_{j'}^b$ and $\Pi_{j'}^c$ via $\Pi_{j'}^a$ is at most $3.1 + 3k\epsilon$ and $4.1 + 4k\epsilon$, respectively, which is strictly smaller than $2 + 2 \cdot 1.2 = 4.4$, which is the shortest achievable distance via $\Pi_j^b$ or $\Pi_j^c$. Finally, the path from $\pi_y \in \Pi_y$ to any literal-peer in $\Pi_i^\mu$ has length at most $3.72 - 2\delta + 3k\epsilon$. This is because there exists a link between $\Pi_y$ and $\Pi_z$, and between $\Pi_i^0$ and $\Pi_i^1$, and because there is a link from $\Pi_z$ to either $\Pi_i^0$ or $\Pi_i^1$. On the other hand, the path from $\pi_y \in \Pi_y$ to a literal-peer via $\Pi_j^b$ or $\Pi_j^c$ has length at least $2.45 + 1.48 = 3.93$. Similar arguments show that the same holds for $\pi_z \in \Pi_z$. □

## 6.2 Satisfiable Instances

In this section, we show that if $I$ has a satisfying assignment $A_I$, then there exists a Nash equilibrium in $\mathcal{M}_I^k$. For this purpose, we explicitly construct a set of strategies $s$, which we prove to constitute a Nash equilibrium. As for notation, we define $A_I(x_i)$ to be the assignment of $x_i$ in $A_I$, i.e.,

$$A_I(x_i) := \begin{cases} 1\,,\ x_i \text{ is set to 1 in } A_I \\ 0\,,\ x_i \text{ is set to 0 in } A_I. \end{cases} \tag{1}$$

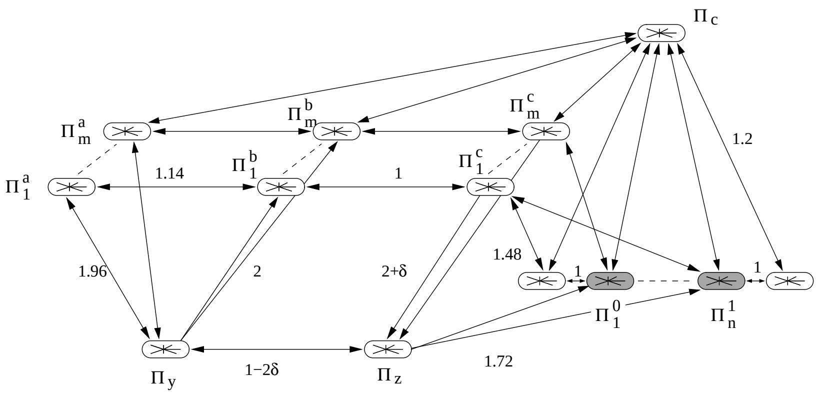


Figure 5: An example instance $G_I^k$ with the topology resulting from strategy $s$. Within each cluster, the peers are connected as a star. Directed arrows between clusters indicate inter-cluster links between cluster-leaders. Cluster-leader $\hat{\pi}_z$ connects to those leaders of literal-peers that appear in the satisfying assignment $A_I$. In the example, $A_I$ sets $x_1 = 0$ and $x_n = 1$.

Furthermore, we define in every cluster $\Pi_g$ a single *leader peer*, which we denote by $\hat{\pi}_g$. The role of this leader-peer is to construct all inter-cluster links going from this cluster to peers located in other clusters. The strategy of the remaining *non-leader peers* $\check{\pi}_g \in \Pi_g \setminus \{\hat{\pi}_g\}$ is to connect to the unique leader peer within their cluster. Formally, the strategy $s_g$ for a non-leader peer $\check{\pi}_g \in \Pi_g \setminus \{\hat{\pi}_g\}$ is $s_g := \{\hat{\pi}_g\}$. For each leader-peer, we define the set of strategies $s$ as follows:

$$
\begin{aligned}
s_y &:= \Pi_y \cup \{\hat{\pi}_z\} \cup \bigcup_{C_j \in \mathcal{C}} \{\hat{\pi}_j^a, \hat{\pi}_j^b\} \\
s_z &:= \Pi_z \cup \{\hat{\pi}_y\} \cup \bigcup_{x_i \in \mathcal{X}} \{\hat{\pi}_i^{A_I(x_i)}\} \\
s_c &:= \Pi_c \cup \bigcup_{x_i \in \mathcal{X}} \{\hat{\pi}_i^0 \cup \hat{\pi}_i^1\} \cup \bigcup_{C_j \in \mathcal{C}} \{\hat{\pi}_j^a \cup \hat{\pi}_j^b \cup \hat{\pi}_j^c\} \\
s_j^a &:= \Pi_j^a \cup \{\hat{\pi}_c, \pi_y, \hat{\pi}_j^b\} && , \forall C_j \in \mathcal{C} \\
s_j^b &:= \Pi_j^b \cup \{\hat{\pi}_c, \hat{\pi}_j^a, \hat{\pi}_j^c\} && , \forall C_j \in \mathcal{C} \\
s_j^c &:= \Pi_j^c \cup \{\hat{\pi}_c, \hat{\pi}_z, \hat{\pi}_j^b\} \cup \bigcup_{x_i^\mu \in C_j} \{\hat{\pi}_i^\mu\} && , \forall C_j \in \mathcal{C} \\
s_i^\mu &:= \Pi_i^\mu \cup \{\hat{\pi}_c, \hat{\pi}_z, \hat{\pi}_i^{1-\mu}\} \cup \bigcup_{x_i^\mu \in C_j} \{\hat{\pi}_j^c\} && , \forall x_i \in \mathcal{X}
\end{aligned}
$$

Strategy $s$ is illustrated in Figure 5. Our goal is to show that $s$ constitutes a Nash equilibrium for

$A_I$. The topology resulting from strategy $s$ contains all "short" links, i.e., links between cluster leaders of clusters that have a distance of at most $1.48$ (cf. Lemma 6.4). Additionally, peer $\hat{\pi}_y$ builds links to clause-cluster leaders $\hat{\pi}_j^a$ and $\hat{\pi}_j^b$ for all $C_j \in \mathcal{C}$. On the other hand, leaders $\hat{\pi}_j^a$ and $\hat{\pi}_j^c$ have a link to $\hat{\pi}_y$ and $\hat{\pi}_z$, respectively. Most importantly, however, for every variable $x_i \in \mathcal{X}$, leader-peer $\hat{\pi}_z$ maintains a link to the literal-peers $\hat{\pi}_i^{A_I(x_i)}$ that are used in the satisfying assignment. Note that because in $s$, peer $\hat{\pi}_z$ has exactly one connection to a literal-peer of every variable, we can apply Lemma 6.6. That is, no peer in clusters $\Pi_y$ and $\Pi_z$ can reduce its stretch to any peer $V \setminus (\Pi_j^a \cup \Pi_j^b \cup \Pi_j^c)$ by connecting to one of the clause-peers of clause $C_j$. Finally, note that non-leaders are directly connected to their cluster leader, and cluster leaders maintain direct links to each peer in their cluster.

The next three lemmas prove that no peer has an incentive to single-handedly deviate from strategy $s$. In the proofs, we use the notation $\Delta_i(\psi)$ to denote the change in cost when peer $\pi_i$ changes its strategy according to action $\psi$, $\psi$ being clear from the context. Specifically, if $\Delta_i(\psi) \geq 0$, peer $\pi_i$ has no incentive to perform action $\psi$ because doing so would increase its cost.

We begin with a lemma that shows that no peer can unilaterally benefit from changing its links within its own cluster.

**Lemma 6.7.** *In $s$, no peer in an arbitrary cluster $\Pi_g$ has an incentive to change its strategy within the cluster, i.e., to add, replace, or remove links to peers in $\Pi_g$.*

*Proof.* The cluster leader $\hat{\pi}_g$ cannot remove any link because the topology would become disconnected without it. Next, consider a non-leader $\check{\pi}_g$. If $\check{\pi}_g$ removes its link to the cluster-leader, it disconnects itself from the network. Adding one or more new link to a non-leader costs $\alpha$ per link, while the resulting stretch reduction per link is $\frac{2\epsilon}{\epsilon} - 1 = 1$ only. Finally, replacing the link to the leader with a link to another non-leader strictly increases the stretch to all but one peer in the network and therefore cannot be beneficial. □

Based on Lemma 6.7, we can consider the topology within each cluster in $s$ to be fixed. It remains to show that no peer has an incentive to add, remove, or replace its inter-cluster links. As shown next, peers in $\Pi_y$ cannot unilaterally reduce their costs in $s$.

**Lemma 6.8.** *No peer in $\Pi_y$ has an incentive to change its strategy, given that all other peers follow strategy $s$.*

*Proof.* By Lemma 6.7, no peer $\pi_y \in \Pi_y$ has an incentive to change its intra-cluster links. Furthermore, $\hat{\pi}_y$ does not benefit from switching its link from a leader peer to a non-leader peer, because this would only decrease the stretch to that particular peer, while increasing the stretch to all other peers (at least) in this cluster. It follows from Lemmas 6.5 and 6.4 that $\hat{\pi}_y$ must keep its links to $\hat{\pi}_j^a$ and $\hat{\pi}_z$, and hence, we must only consider that leader peers connect to leader peers.

We now show that no peer in $\Pi_y$ can reduce this cost by deviating from its strategy in any other way.

***Case 1:*** Some $\check{\pi}_y$ or $\hat{\pi}_y$ adds one of more additional links:

In the topology resulting from $s$, every peer in $\Pi_y$ is connected with stretch at most $1 + 2\epsilon$ with all peers except from peers in $\Pi_j^c$ (for all $C_j \in \mathcal{C}$) and peers in those literal-clusters to which $\hat{\pi}_z$ does not have a direct connection. With any additional link, a peer in $\Pi_y$ can reduce its stretch to peers in exactly one of these clusters only. Hence, for every additional link would increase the peer's cost: $\Delta_y(+) \geq -\frac{k(4.72+\epsilon)}{3.72} + \alpha + k > 0$.

Observe that because non-leader peers $\check{\pi}_y \in \Pi_y$ do not have inter-cluster links, Case 1 in combination with Lemma 6.7 implies that no $\check{\pi}_y$ can benefit from changing its strategy.

***Case 2:*** $\hat{\pi}_y$ changes its link from $\hat{\pi}_j^b$ to $\hat{\pi}_j^c$:

While the stretch to peers in $\Pi_j^c$ is reduced, the stretch to peers in $\Pi_j^b$ increases. The relative cost difference is $\Delta_y(\hat{\pi}_j^b \rightarrow \hat{\pi}_j^c) \geq -\frac{k(3+\epsilon)}{2.45} + \frac{(1.96+1.14)k}{2} > 0$.

***Case 3:*** $\hat{\pi}_y$ removes its link from $\hat{\pi}_j^b$:

By removing such a link, $\hat{\pi}_y$ can save the link's cost $\alpha$. On the other hand, the stretch to both $\Pi_j^b$ and $\Pi_j^c$ increase. Specifically, the shortest connection to peers in these clusters is now via $\hat{\pi}_j^a$ and $\hat{\pi}_j^b$, i.e., $\Delta_y(-\hat{\pi}_j^b) \geq -\alpha - k(1+\epsilon) - \frac{k(3+\epsilon)}{2.45} + \frac{(1.96+1.14)k}{2} + \frac{(1.96+1.14+1)k}{2.45} > 0$.

The only other thing that could potentially lead to an advantage for $\hat{\pi}_y$ is to replace a link $\hat{\pi}_j^b$ by some leader peer in $\Pi_i^\mu$ to which $\hat{\pi}_z$ is not connected, formally $\mu \neq A_I(x_i)$. Doing so clearly increases the stretch to peers in $\Pi_j^b$ and $\Pi_j^c$, but like in Case 3, the shortest connection between $\hat{\pi}_y$ to peers in $\Pi_j^c$ is via $\hat{\pi}_j^a$ and $\hat{\pi}_j^b$. In particular, this path has length at most $4.1 + \epsilon$, whereas the shortest path via a literal-cluster has length at least $1 - 2\delta + 1.72 + 1.48 = 4.2 - 2\delta$, which is larger. Hence, replacing one or more links to $\hat{\pi}_j^b$ by links to literal-peers reduces to Cases 1 and 3, respectively, and therefore cannot be worthwhile. Finally, no combination of the above cases can reduce the cost of any peer in $\Pi_y$ either. □

**Lemma 6.9.** *No peer in $\Pi_z$ has an incentive to change its strategy, given that all other peers follow strategy $s$.*

*Proof.* Again, we discuss the various cases and show that none of them is beneficial for a peer in $\Pi_z$. Recall that by Lemma 6.6, connecting to any clause-peer cannot improve the stretch to any other peer outside this clause. Furthermore, because $A_I$ is a satisfying assignment, the topology of $s$ contains a path of length at most $\epsilon + d(\Pi_z, \Pi_i^\mu) + d(\Pi_i^\mu, \Pi_j^c) + \epsilon = 3.2 + 2\epsilon$ between peers in $\Pi_z$ and peers in $\Pi_j^c$, for *every* clause $C_j \in \mathcal{C}$. Consequently, connecting to a so far unconnected literal-peer cannot decrease the stretch to any clause-peer $\pi_j \in C_P$ in the system.

It follows from Lemma 6.7 that no peer $\pi_z \in \Pi_z$ has an incentive to change its intra-cluster links. Also, as shown in the proof of Lemma 6.8, no peer can benefit from connecting to a non-leader peer in the network, because this bears strictly higher costs than connecting to the corresponding leader peer of the same cluster. Hence, we only need to verify the cases in which peers in $\Pi_z$ connect to leader peers.

In the following, we discuss the various cases how peers in $\Pi_z$ could improve their situation and derive that none of them is actually beneficial.

***Case 1:*** Some peer in $\Pi_z$ adds an additional link to $\hat{\pi}_j^b$:

The reduction of the stretches to peers in $\Pi_j^b$ and $\Pi_j^c$ resulting from the additional link does not outweigh the link's cost. Specifically, we have $\Delta_z(+\hat{\pi}_j^b) \geq -\frac{k(3-2\delta+2\epsilon)}{2} + k - \frac{k(3.2+2\epsilon)}{2+\delta} + \frac{3k}{2+\delta} + \alpha \geq k(4\delta + 2\delta^2) > 0$. Notice that in the second term, the stretch to each of the $k$ peers in $\Pi_j^b$ is at least 1, and in the third term, the distance $3.2 + 2\epsilon$ holds because $A_I$ is a satisfying assignment.

***Case 2:*** Some peer in $\Pi_z$ adds an additional link to $\hat{\pi}_j^c$:

Again, the stretches to $\Pi_j^b$ and $\Pi_j^c$ are not reduced enough to render the additional link worthwhile. In fact, the stretch to peers in $\Pi_j^b$ is not reduced by the addition of this link, nor is the stretch to any other peer in the network except from peers in $\Pi_j^c$ (Lemma 6.6). It follows that $\Delta_z(+\hat{\pi}_j^c) \geq -\frac{k(3.2+2\epsilon)}{2+\delta} + k + \alpha = k(1.6\delta - 2\epsilon) > 0$.

***Case 3:*** Some peer in $\Pi_z$ adds an additional link to $\hat{\pi}_j^a$:

Clearly, this option is even worse than Cases 1 and 2.

***Case 4:*** Some peer in $\Pi_z$ adds an additional link to $\hat{\pi}_i^\mu$:

Adding a link to a literal-cluster that is not used in $A_I$ reduces the stretch to peers in this cluster only, because there is already a short connection from $\Pi_z$ to every $\Pi_j^c$ through the literal-clusters $\Pi_i^{A_I(x_i)}$. Hence, $\Delta_z(+\hat{\pi}_i^\mu) \geq -\frac{k(2.72+2\epsilon)}{1.72} + k + \alpha > 0$.

Observe that because non-leader peers $\check{\pi}_z \in \Pi_z$ do not have inter-cluster links, Cases 1 to 4 in combination with Lemma 6.7 implies that no $\check{\pi}_z$ can benefit from changing its strategy.

***Case 5:*** $\hat{\pi}_z$ replaces some $\hat{\pi}_i^{A_I(x_i)}$ by $\hat{\pi}_i^{1-A_I(x_i)}$:

Again, the new link to a previously unconnected literal-cluster cannot decrease the stretch to any clause-peer, because $A_I$ is a satisfying assignment and $\hat{\pi}_z$ already had a path of length $3.2$ to every $\hat{\pi}_j^c$ via some $\hat{\pi}_i^{A_I(x_i)}$. Furthermore, by a symmetry argument, the stretch cost gained by adding the link to $\hat{\pi}_i^{1-A_I(x_i)}$ is lost by removing the link to $\hat{\pi}_i^{A_I(x_i)}$. Thus, $\Delta_y(\hat{\pi}_i^{A_I(x_i)} \rightarrow \hat{\pi}_i^{1-A_I(x_i)}) \geq 0$.

***Case 6:*** $\hat{\pi}_z$ removes or replaces some $\hat{\pi}_i^{A_I(x_i)}$:

If $\hat{\pi}_z$ does not have a connection to any literal-cluster of a variable $x_i$, the resulting stretch to each peer in these two clusters is at least $\frac{3+\delta+1.48}{1.72}$. Because $\frac{k(4.48+\delta)}{1.72} > k(1+2\epsilon) + \alpha$, it follows that $\hat{\pi}_z$ must maintain at least one link to such a peer.

Any other possible strategy deviation can either be reduced to one of the above five cases or to Lemma 6.4. □

Having shown that peers in $\Pi_y$ and $\Pi_z$ have no incentive to deviate from $s$, it remains to show that no other peer can improve its situation either.

**Lemma 6.10.** *No top-layer peer can benefit from changing its strategy, given that all other peers follow* $s$.

*Proof.* First, by Lemma 6.7, it holds that no peer can improve its situation by adding, replacing, or removing a link within its cluster. Also, no peer can benefit from connecting to a non-leader, as opposed to the leader peer in the same cluster. Both claims can be proven with exactly the same argument as in the proof of Lemma 6.8.

It is important to observe that in $s$, all top-layer peers are almost optimally connected with each other, either via the central cluster $\Pi_c$ or because their respective clusters are neighbors in the graph. More specifically, the stretch between any pair of top-layer peers in $s$ is at most $1 + 2\epsilon$ (via the own cluster leader, $\hat{\pi}_c$, and the other cluster leader). Besides removing the final $2\epsilon$ from these small stretches, adding additional links can only help in reducing the stretches to peers in $\Pi_y$ and $\Pi_z$. By Lemma 6.4, no link between cluster leaders whose clusters have a distance of less than $1.48$ can be removed from $s$. Hence, the possible strategy deviations by other nodes is actually limited.

***Peers in $\Pi^{\mathrm{a}}_{\mathrm{j}}$:*** A peer $\hat{\pi}^a_j$'s link to $\hat{\pi}_y$ cannot be removed by Lemma 6.5. For every peer $\pi^a_j \in \Pi^a_j$, it further holds that building an additional link to $\hat{\pi}_z$ is too costly, $\Delta^a_j(+\hat{\pi}_z) \geq -\frac{k(2.96-2\delta+2\epsilon)}{2.45} + k + \alpha > 2kN\epsilon$. Hence, even if this additional link could reduce all other less than $N$ stretches to top-level peers by the remaining $2\epsilon$, the cost of an additional link would still be too high.

***Peers in $\Pi^{\mathrm{b}}_{\mathrm{j}}$:*** Peer $\hat{\pi}^b_j$ does not have a link longer than $1.48$ in $s$ and hence, cannot remove any of them. We show that neither building a link to $\hat{\pi}_y$ nor to $\hat{\pi}_z$ decreases the cost of any peer in $\Pi^b_j$. In the first case, we have $\Delta^b_j(+\hat{\pi}_y) \geq -\frac{k(1.96+1.14+2\epsilon)}{2} + k - \frac{k(3+\delta+2\epsilon)}{2} + \frac{k(3-2\delta)}{2} + \alpha > 2kN\epsilon$. As for the second case, $\Delta^b_j(+\hat{\pi}_z) \geq -\frac{k(1.96+1.14+2\epsilon)}{2} + \frac{k(3-2\delta)}{2} - \frac{k(3+\delta+2\epsilon)}{2} + k + \alpha > 2kN\epsilon$. Clearly, building both links is even less worthwhile.

***Peers in $\Pi^{\mathrm{c}}_{\mathrm{j}}$:*** The potential strategy deviations that could decrease peer $\hat{\pi}^c_j$'s costs are to add a link to $\hat{\pi}_y$, to remove its link from $\hat{\pi}_z$, or to replace the link to $\hat{\pi}_z$ by a link to $\hat{\pi}_y$. However, none of these alterations are beneficial for $\hat{\pi}^c_j$ (or for any non-leader peer in $\Pi^c_j$ in the case of link addition). First, it holds that $\Delta^c_j(+\hat{\pi}_y) \geq -\frac{k(3-\delta+2\epsilon)}{2.45} + k + \alpha > 2kN\epsilon$ and $\Delta^c_j(-\hat{\pi}_z) \geq -\alpha - k(1+2\epsilon) - \frac{k(3-\delta+2\epsilon)}{2.45} + \frac{3.2k}{2+\delta} + \frac{4.1k}{2.45} > 2kN\epsilon$. Also, switching the link from $\hat{\pi}_z$ to $\hat{\pi}_y$ is not helpful, $\Delta^c_j(\hat{\pi}_z \rightarrow \hat{\pi}_y) \geq \frac{3.2k}{2+\delta} - \frac{k(3-\delta+2\epsilon)}{2.45} > 2kN\epsilon$.

***Peers in $\Pi^{\mu}_{\mathrm{i}}$:*** Each leader of a literal-cluster maintains a link to $\hat{\pi}_z$, and we show that they (as well as any non-leader peer in these clusters) do not have an incentive to change that strategy. It is clear that neither adding a link to $\hat{\pi}_y$ nor switching from $\hat{\pi}_z$ to $\hat{\pi}_y$ can be beneficial. In the first case, the stretch is reduced by at most $2\epsilon$ by the additional link, which does not render the link cost $\alpha$ worthwhile. In the second case, the stretch is strictly increased. If $\hat{\pi}^\mu_i$ removes its link to $\hat{\pi}_z$ and connects via its neighboring literal-cluster, the stretches to both $\Pi_y$ and $\Pi_z$ increase. Particularly, we have $\Delta^\mu_i(-\hat{\pi}_z) \geq -\alpha - k(1 + 2\epsilon) + \frac{2.72k}{1.72} + \frac{k(3.72-2\delta)}{2.72-2\delta} > 2kN\epsilon$.

***Peers in*** $\mathbf{\Pi_c}$***:*** Finally, peers in $\Pi_c$ are connected with stretch at most $2\epsilon$ to all peers in the network. To top-clusters, the connection is via links shorter than $1.48$. As for the remaining two clusters, it is connected to $\hat{\pi}_z$ via one of the literal-clusters and to $\hat{\pi}_y$ via some $\hat{\pi}_j^a$. By the definition of $\epsilon$ and $\alpha$, it is clear that no peer in $\Pi_c$ can improve its strategy. □

By combining Lemmas 6.8, 6.9, and 6.10, we know that no peer in the network has an incentive to change its strategy. Hence, $s$ constitutes a pure Nash equilibrium.

**Lemma 6.11.** *If $I$ is satisfiable, there exists a pure Nash equilibrium in $\mathcal{M}_I^k$.*

## 6.3 Non-satisfiable Instances

It remains to prove the other direction, that is, there exists no pure Nash equilibrium in the network if the underlying 3-SAT instance $I$ has no satisfying assignment. We proceed by defining structural properties that any Nash equilibrium must fulfil, and show that the intersection of all these properties is empty. Besides the basic properties derived in Section 6.1, an important characteristic of any Nash equilibrium is the fact that exactly one peer in $\Pi_z$ connects to *exactly one* literal-peer (either in $\Pi_i^0$ or $\Pi_i^1$) for every variable $x_i \in \mathcal{X}$.

**Lemma 6.12.** *In any Nash equilibrium, exactly one peer in $\Pi_z$ connects to either a peer $\pi_i^1 \in \Pi_i^1$ or $\pi_i^0 \in \Pi_i^0$, for every $x_i \in \mathcal{X}$.*

*Proof.* We have already shown in Lemma 6.9 (Case 6) that there must be a peer $\pi_z \in \Pi_z$ that has at least one link to a literal-peer of every variable. Furthermore, we know by Lemma 6.3 that no other peer in $\Pi_z$ connects to the same cluster as $\pi_z$. Hence, we only need to show that in a Nash equilibrium no two peers in $\Pi_z$ connect to both literal-clusters of the same variable.

Assume for the sake of contradiction that peers $\pi_z$ and $\pi_z'$ (potentially $\pi_z = \pi_z'$) maintain links to both $\Pi_i^0$ and $\Pi_i^1$ for some $x_i \in \mathcal{X}$. In this case, it would be worthwhile for one of the two peers to remove its link and replace it with a link to some peer in $\Pi_j^c$ (if this link is not there, already), such that $x_i \in C_j$. By the definition of our special 3-SAT instance and the construction of $G_I$, we know that of the two literal-clusters, one, say $\Pi_i^\mu$, has clause-cluster $\Pi_j^c$ at distance $1.48$, and the other literal-cluster, say

$\Pi_i^{1-\mu}$, has two such close-by clause-clusters. Let $\pi'_z$ be the peer that connects to cluster $\Pi_i^\mu$ (otherwise, replace $\pi_z$ for $\pi'_z$ for the remainder of the proof).

Assume for the first case that the length of the shortest path from $\pi'_z$ to this $\Pi_j^c$ without the link via $\Pi_i^\mu$ is 3.2 or longer. In this case, the change in $\pi'_z$'s costs when switching from its link to literal-cluster $\Pi_i^\mu$ that has only a single close-by clause-cluster $\Pi_j^c$ directly to a peer in $\Pi_j^c$ is $\Delta_z(\pi_i^\mu \to \pi_j^c) \leq +\frac{k(2.72+2k\epsilon)}{1.72} - \frac{3.2k}{2+\delta} + \frac{k(2+\delta+2k\epsilon)}{2+\delta} < 0$. If the length of the path from $\pi_z$ to $\Pi_j^c$ is strictly shorter than 3.2, then the link to $\Pi_i^\mu$ can simply be dropped, resulting in a gain of $\Delta_z(-\pi_i^\mu) \leq -\alpha - k + \frac{k(1.72+2k\epsilon)}{1.72} + \frac{k(2.72+2k\epsilon)}{1.72} < 0$. Hence, $\pi'_z$ is always better off not connecting to a literal-cluster if $\pi_z$ already connects to a literal-cluster. ¿From this, the claim follows. □

Lemma 6.12 is an important ingredient for the remainder of the proof, because it gives us a one-to-one correspondence between the connections of $\Pi_z$ to literal-clusters, and an assignment of variables in the 3-SAT instance $I$. Also, note that when combining Lemma 6.12 with Lemma 6.6, it follows that in a Nash equilibrium, peers in $\Pi_y$ and $\Pi_z$ cannot reduce their stretch to any peer in $V \setminus \{\Pi_j^a \cup \Pi_j^b \cup \Pi_j^c\}$ by connecting to one of the clause-peers of clause $C_j$.

**Lemma 6.13.** *If $I$ is non-satisfiable, there exists no pure Nash equilibrium in $\mathcal{M}_I^k$.*

*Proof.* By Lemma 6.12, exactly one peer in $\Pi_z$ connects to either the positive or negative literal-cluster of every variable $x_i$. Because there exists no satisfying assignment, it follows that regardless of how $\Pi_z$ is connected to the literal-clusters, there must exist at least one clause $C_{j^*}$ that is "not satisfied". In the resulting topology, this means that the path from a peer in $\Pi_z$ to a clause-peer in $\Pi_{j^*}^c$ of this unsatisfied clause via any literal-cluster must be of length at least $d(\Pi_z, \Pi_i^\mu) + d(\Pi_i^\mu, \Pi_i^{1-\mu}) + d(\Pi_i^{1-\mu}, \Pi_{j^*}^c) = 4.2$. Particularly, every such path must include the additional distance of 1 between $x_i^1$ and $x_i^0$. In the sequel, we consider this *unsatisfied clause* $C_{j^*}$ in more detail.

First, we show that in a Nash equilibrium, no peer $\pi_y \in \Pi_y$ establishes a link to $\Pi_{j^*}^c$. We distinguish two cases. In the first case, if some peer in $\Pi_y$ already has a link to $\Pi_{j^*}^b$, then the cost reduction for $\pi_y$ when omitting its link to $\Pi_{j^*}^c$ is $\Delta_y(-\pi_{j^*}^c) \leq -\alpha - k + \frac{k(3+2k\epsilon)}{2.45} < 0$. In the other case, the cost reduction when switching the link from $\Pi_{j^*}^c$ to a peer in $\Pi_{j^*}^b$ is at least $\Delta_y(\pi_{j^*}^c \to \pi_{j^*}^b) \leq -\frac{k(3-2\delta)}{2} + \frac{k(3+2k\epsilon)}{2.45} < 0$. That is, in either case it is beneficial for $\pi_y$ not to connect directly to $\Pi_{j^*}^c$.

For the next step, we establish that in any Nash equilibrium, exactly one peer $\pi_z \in \Pi_z$ connects to either a peer in $\Pi^b_{j^*}$ or in $\Pi^c_{j^*}$. To see this, assume that no peer in $\Pi_z$ establishes any links to peers in the two clusters. In this case (because there is no link from $\Pi_y$ to $\Pi^c_{j^*}$, and because $C_{j^*}$ is not satisfied), the sum of the stretches to peers in $\Pi^c_{j^*}$ is at least $\frac{k(4-2\delta)}{2+\delta} > k(1+2k\epsilon)+\alpha$. That is, $\pi_z \in \Pi_z$ can reduce its cost by connecting to $\pi^c_{j^*}$.

It remains to show that no peer in $\Pi_z$ connects to $\Pi^a_{j^*}$, and particularly, that no two peers in $\Pi_z$ simultaneously connect to both $\Pi^b_{j^*}$ or $\Pi^c_{j^*}$. Because there is at least one link from $\Pi_z$ to either $\Pi^b_{j^*}$ or $\Pi^c_{j^*}$, it follows that a link to $\Pi^a_{j^*}$ can only reduce the stretch to peers in this particular cluster. However, the incurred cost exceeds the savings due to the reduced stretch, i.e., $\Delta_z(+\pi^a_{j^*}) = -\frac{k(2.96-2\delta+2k\epsilon)}{2.45}+\alpha+k > 0$. For the last case, assume that two peers $\pi_z$ and $\pi'_z$ (potentially the same) connect to both $\Pi^b_{j^*}$ and $\Pi^c_{j^*}$, respectively. Then, $\pi'_z$ has an incentive to drop its link to $\Pi^c_{j^*}$: $\Delta_z(-\pi^c_{j^*}) = \frac{k(3+2k\epsilon)}{2+\delta} - k - \alpha < 0$. Hence, in any Nash equilibrium, there is exactly one link from $\Pi_z$ to either $\Pi^b_{j^*}$ or $\Pi^c_{j^*}$, but not to both.

Studying the above rules, it can be observed that there remain only four possible sets of strategies for peers in $\Pi_y$ and $\Pi_z$ that could potentially result in a pure Nash equilibrium. The four cases can be distinguished by whether or not a peer in $\Pi_y$ directly connects to $\Pi^b_{j^*}$, and by whether a peer in $\Pi_z$ connects to $\Pi^b_{j^*}$ or $\Pi^c_{j^*}$.

***Case 1:*** Some peer $\pi_z \in \Pi_z$ connects to $\pi^b_{j^*}$: In this case, some peer $\pi_y \in \Pi_y$ has an incentive to add a link to a peer in $\Pi^b_{j^*}$, because this significantly reduces its stretches to peers in $\Pi^b_{j^*}$ and $\Pi^c_{j^*}$. Specifically, $\pi_y$ could reduce its cost by at least $\Delta_y(+\pi^b_{j^*}) \leq -\frac{k(3-2\delta)}{2} - \frac{k(4-2\delta)}{2.45} + \alpha + k(1+2k\epsilon) + \frac{k(3+2k\epsilon)}{2.45} < 0$.

***Case 2:*** Peers $\pi_z \in \Pi_z$ and $\pi_y \in \Pi_y$ connect to $\Pi^b_{j^*}$: In this case, the peer $\pi_z$ can profit from switching its link to a peer in $\Pi^c_{j^*}$. Specifically, $\Delta_z(\pi^b_{j^*} \rightarrow \pi^c_{j^*}) \leq -\frac{3k}{2+\delta} + \frac{k(3-2\delta+2k\epsilon)}{2} < 0$.

***Case 3:*** Some peer $\pi_z \in \Pi_z$ connects to $\Pi^c_{j^*}$: Unlike in the previous case, $\pi_z$ prefers switching its link from $\Pi^c_{j^*}$ to a peer in $\Pi^b_{j^*}$ in the absence of a link from $\Pi_y$ to $\Pi^b_{j^*}$. By doing so, it can reduce its cost by $\Delta_z(\pi^c_{j^*} \rightarrow \pi^b_{j^*}) \leq \frac{k(3+2k\epsilon)}{2+\delta} - \frac{k(3+\delta)}{2} = k(-5\delta - \delta^2 + 4k\epsilon) < 0$.

***Case 4:*** Some peer $\pi_z \in \Pi_z$ connects to $\Pi^c_{j^*}$ and some peer $\pi_y \in \Pi_y$ connects to $\Pi^b_{j^*}$: In this configuration, peer $\pi_y$ benefits from removing its link to $\Pi^b_{j^*}$. The decrease of its costs is $\Delta_y(-\pi^b_{j^*}) < -\alpha - k + \frac{k(3.1+2k\epsilon)}{2} < 0$.

Finally, since none of these four cases is a Nash equilibrium, the proof is concluded. □

# 7 Conclusion

This paper has presented and investigated a *locality game* for selfish P2P topologies. We have established a tight bound on the Price of Anarchy and have shown that it is generally a hard problem to decide whether a system will ever stabilize. In particular, our results indicate that topologies may degrade more severely when selfish peers value maintenance cost relatively higher than latency costs.

It is interesting to compare our game to the network creation game by Fabrikant et al. [10]. In their game, links are undirected, and a hop metric is considered rather than the stretch. There always exist pure Nash equilibria, which is not the case in our setting.

# References


[1] I. Abraham, A. Badola, D. Bickson, D. Malkhi, S. Maloo, and S. Ron. Practical Locality-Awareness for Large Scale Information Sharing. In *Proc.* $4^{th}$ *Intl. Workshop on Peer-to-Peer Systems (IPTPS)*, 2005.

[2] I. Abraham, D. Malkhi, and O. Dobzinski. LAND: Stretch $(1+e)$ Locality Aware Networks for DHTs. In *Proc.* $15^{th}$ *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 550–559, 2004.

[3] E. Adar and B. Huberman. Free Riding on Gnutella. *First Monday*, 5(10), 2000.

[4] S. Albers, S. Eilts, E. Even-Dar, Y. Mansour, and L. Roditty. On Nash Equilibria for a Network Creation Game. In *Proc.* $17^{th}$ *ACM Symposium on Discrete Algorithms (SODA)*, 2006.

[5] N. Andrade, M. Mowbray, A. Lima, G. Wagner, and M. Ripeanu. Influences on Cooperation in BitTorrent Communities. In *Proc. 2005 ACM SIGCOMM on Economics of Peer-to-Peer Systems*, 2005.

[6] E. Anshelevich, A. Dasgupta, J. Kleinberg, E. Tardos, T. Wexler, and T. Roughgarden. The Price of Stability for Network Design with Fair Cost Allocation. In *Proc.* $45^{th}$ *Symposium on Foundations of Computer Science (FOCS)*, pages 295–304, 2004.

[7] H. T.-H. Chan, A. Gupta, B. M. Maggs, and S. Zhou. On Hierarchical Routing in Doubling Metrics. In *Proc.* $17^{th}$ *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 762–771, 2005.

[8] J. Corbo and D. C. Parkes. The Price of Selfish Behavior in Bilateral Network Formation. In *Proc.* $24^{th}$ *ACM Symposium on Principles of Distributed Computing (PODC)*, pages 99–107, 2005.

[9] S. Eidenbenz, V. Kumar, and S. Zust. Equilibria in Topology Control Games for Ad Hoc Networks. In *Proc. DIALM-POMC*, 2003.

[10] A. Fabrikant, A. Luthra, E. Maneva, C. H. Papadimitriou, and S. Shenker. On a Network Creation Game. In *Proc.* $22^{nd}$ *ACM Symposium on Principles of Distributed Computing (PODC)*, pages 347–351, 2003.

[11] A. Fabrikant, C. H. Papadimitriou, and K. Talwar. The Complexity of Pure Nash Equilibria. In *Proc. $36^{th}$ ACM Symposium on Theory of Computing (STOC)*, pages 604–612, 2004.

[12] M. R. Garey and D. S. Johnson. *Computers and Intractability, A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, 1979.

[13] S. Jun and M. Ahamad. Incentives in BitTorrent Induce Free Riding. In *Proc. 2005 ACM SIGCOMM on Economics of Peer-to-Peer Systems*, pages 116–121, 2005.

[14] D. R. Karger and M. Ruhl. Finding Nearest Neighbors in Growth-Restricted Metrics. In *Proc.* $34^{th}$ *ACM Symposium on Theory of Computing (STOC)*, pages 741–750, 2002.

[15] C. H. Papadimitriou. Algorithms, Games, and the Internet. In *Proc.* $33^{rd}$ *ACM Symposium on Theory of Computing (STOC)*, pages 749–753, 2001.

[16] C. H. Papadimitriou. Computing Correlated Equilibria in Multi-Player Games. In *Proc. $37^{th}$ ACM Symposium on Theory of Computing (STOC)*, pages 49–56, 2005.

[17] C. Plaxton, R. Rajaraman, and A. W. Richa. Accessing Nearby Copies of Replicated Objects in a Distributed Environment. In *Proc.* $9^{th}$ *ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, pages 311–320, 1997.

[18] A. Rowstron and P. Druschel. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In *Proc. of the IFIP/ACM International Conference on Distributed Systems Platforms (Middleware)*, pages 329–350, 2001.

[19] B. Wong, A. Slivkins, and E. G. Sirer. Meridian: A Lightweight Network Location Service without Virtual Coordinates. In *Proc. SIGCOMM*, 2005.

[20] M. Yang, Z. Zhang, X. Li, and Y. Dai. An Empirical Study of Free-Riding Behavior in the Maze P2P File Sharing System. In *Proc.* $4^{th}$ *Intl. Workshop on Peer-to-Peer Systems(IPTPS)*, 2005.

[21] B. Y. Zhao, L. Huang, J. Stribling, A. D. Joseph, and J. D. Kubiatowicz. Tapestry: A Resilient Global-scale Overlay for Service Deployment. *IEEE Journal on Selected Areas in Communications*, 22(1), 2004.