Algorithmica
DOI 10.1007/s00453-010-9398-9

Topological Implications of Selfish Neighbor Selection in Unstructured Peer-to-Peer Networks

Thomas Moscibroda · Stefan Schmid · Roger Wattenhofer

Received: 11 March 2009 / Accepted: 22 February 2010

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.

Keywords Game theory · Peer-to-peer · Price of anarchy · NP-hardness · Metric spaces

Preliminary versions of this work have been published at the 5th International Workshop on Peer-to-Peer Systems (IPTPS) [27] and at the 25th ACM Annual Symposium on Principles of Distributed Computing (PODC) [28].

T. Moscibroda
Distributed Systems Research Group, Microsoft Research, Redmond, USA
e-mail: moscitho@microsoft.com

S. Schmid (✉)
Deutsche Telekom Laboratories and Technical University of Berlin, 10587 Berlin, Germany
e-mail: stefan@net.t-labs.tu-berlin.de

R. Wattenhofer
Computer Engineering and Networks Laboratory, ETH Zurich, 8092 Zurich, Switzerland
e-mail: wattenhofer@tik.ee.ethz.ch

1 Introduction

The power of peer-to-peer computing arises from the collaboration of the system's 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 outperform existing server-based architectures. However, in reality, peers may act selfishly 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.

One of the few systems that actively tries to tackle the non-cooperation challenge is *BitTorrent*—which is also one of the most popular applications on the Internet. Unfortunately, in contrast to common belief, BitTorrent can still be cheated by selfish users, e.g., with the BitThief client. Motivated by the apparent weaknesses, we are interested in the impact of selfish behavior in p2p systems. Game theory provides tools to quantify such effects.

This article investigates the impact of selfish behavior in unstructured peer-to-peer topologies. Concretely, we study the quality of the network topologies which result if peers selfishly select to which other peers they connect. One contribution will be the computation of the *Price of Anarchy* of p2p overlay creation, which is the ratio between an optimal solution 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.

We will first 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 α 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. 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**-hard. Consequently, it is infeasible in practice to determine if a p2p network of selfish peers can stabilize.

2 A P2P Network Creation Game

We model the peers of a p2p network as points in a *metric space* $\mathcal{M} = (V, d)$, where $d : V \times V \to [0, \infty)$ is the *distance function* which describes the underlying latencies between all pairs of peers. The effects of selfish peer behavior is studied from a game-theoretic perspective. We consider a set of n peers $V = \{\pi_1, \pi_2, \ldots, \pi_n\}$. A peer can choose to which subset of other peers it wants to store pointers (IP addresses). Formally, the *strategy space* of a peer π_i is given by $S_i = 2^{V \setminus \{\pi_i\}}$, and we will refer to the actually chosen links as π_i's *strategy* $s_i \in S_i$. We say that π_i *maintains or establishes a link* to π_j if $\pi_j \in s_i$. The combination of all peers' strategies, i.e., $s = (s_1, \ldots, s_n) \in S_1 \times \cdots \times S_n$, yields a (directed) graph $G[s] = (V, \bigcup_{i=1}^{n} (\{\pi_i\} \times s_i))$, 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 π and π' is defined as the shortest distance between π and π' 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 π 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 [36]. Thereby, the relative importance of degree costs versus stretch costs is expressed by the parameter α.

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 to 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 π_i, and for all alternative strategies s' which differ only in the ith component (different neighbor sets for peer π_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 α in real unstructured peer-to-peer networks is an interesting field for study. As mentioned, α 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 vital, α is smaller than in distributed archival storage systems consisting mainly of large files. In the sequel, we will 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

Papadimitriou [31] has argued that the Internet has surpassed the *von Neumann computer* as the most complex computational artifact of our time. In particular, he pointed out that the Internet has a *socio-economic* complexity whose understanding requires techniques from mathematical economics and game theory [30]. Since then, game theoretic approaches have become increasingly popular to study selfish behavior on all layers of distributed systems. Specifically, researchers have been keen to study the inherent loss of efficiency in a system caused by the participant's selfishness in networks. Consequently, the Price of Anarchy and its complexity have been investigated in various system settings, for example in *routing* [12, 35] (see also the book by Roughgarden [34]): Roughgarden has observed that the data packets which usually "rocket along the Internet at the speed of light" [6] can be significantly slowed down by selfish routers—a phenomenon which has been known also in civil engineering in the context of road planning. Besides being a useful toolkit to understand strategic interactions occurring in today's Internet, game theory itself offers exciting questions for mathematicians and computer scientists. For instance, while it is known that Nash equilibria always exist if the players' strategies are convex sets (such sets can for example be obtained by using probability distributions over strategies), in general, the complexity of finding a Nash equilibrium is believed to be one of the most important open questions on the boundary of the complexity class **P**, besides *factoring* [31]. Despite decades of effort, the computational complexity of computing a Nash equilibrium for a general-sum normal-form game remains unknown.

Adar and Huberman [3] noticed that selfish behavior is a reality also in peer-to-peer systems, and that there exists a large fraction of free riders in the file sharing network Gnutella. The problem of selfish behavior in peer-to-peer systems has been

a hot topic in p2p research ever since, e.g. [23, 38], and many mechanisms to encourage cooperation have been proposed, for example in [19, 22, 37, 39, 41]. Perhaps the simplest fairness mechanism is to directly incorporate contribution monitoring into the client software. For instance, in the file-sharing system *Kazaa*, the client records the contribution of its user. However, such a solution can simply be bypassed by implementing a different client that hard-wires the contribution level to the maximum, as it was the case with *Kazaa Lite*. Inspired by real economies, some researchers have also proposed the introduction of some form of virtual money which is used for the transactions. However, these monetary or credit based approaches have a substantial overhead in terms of communication costs and infrastructure, and are inefficient [21, 42]. Often, these systems also require market regulating mechanisms [41] to cope with inflation or deflation—a complex issue. Additionally, monetary based systems may deter users from participating [29].

BitTorrent [11] has incorporated a fairness mechanism from the beginning. Although this mechanism has similarities to the well known *tit-for-tat scheme* [7], the mechanism employed in BitTorrent distinguishes itself from the classic tit-for-tat mechanism in many respects [24]. It has also been the subject of active research recently (e.g., [8, 20, 33]). Based on PlanetLab tests, [24] has argued that BitTorrent lacks appropriate rewards and punishments and therefore peers might be tempted to freeload. The authors further propose a tit-for-tat-oriented mechanism based on the iterated prisoner's dilemma [7] in order to deter peers from freeloading. However, in their work, a peer is already considered a free rider if it contributes considerably less than other peers. BitThief [26], on the other hand, aims at attaining fast downloads strictly without uploading any data. This is often desirable, since in many countries downloading certain media content is legal whereas uploading is not.

The game-theoretic model of our locality game has been inspired by the paper by Fabrikant et al. [18] 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, 5, 10, 14–17]. 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. [2, 36, 43]). In early literature on distributed hash tables, 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 [32], systems such as [1, 2, 36, 44] 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 work 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 α). Concretely, in Sect. 4.1 we prove that for *arbitrary* metric spaces—thus, including the important and well-studied *growth-bounded* [25] and *doubling* (e.g. [9]) metrics—the Price of Anarchy never exceeds $O(\min(\alpha, n))$. We then show in Sect. 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 π which has to find a suitable neighbor set. Clearly, the *maximal* stretch from π to any other peer π' in the system is at most $\alpha+1$: if $stretch(\pi, \pi') > \alpha + 1$, π could establish a direct link to π', reducing the stretch from more than $\alpha + 1$ to 1, while incurring a link cost of α. 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 $O(\alpha n^2)$. 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 optimal social cost by $\Omega(\alpha n + n^2)$ and yields the following 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 α, 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 Sect. 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 which peers are located on a line, and the distance (latency) between two consecutive peers increases exponentially towards the right. Concretely, peer i, for i from 1 to n, is located at position $\alpha^{i-1}/2$ if i is odd, and at position α^{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 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

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

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 $j < i$. 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. □

Lemma 4.3 *The social cost $C(G)$ of the topology G 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:

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

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 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 of G, 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".

Theorem 5.1 *Regardless of the magnitude of α, 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 α 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 Fig. 1 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.

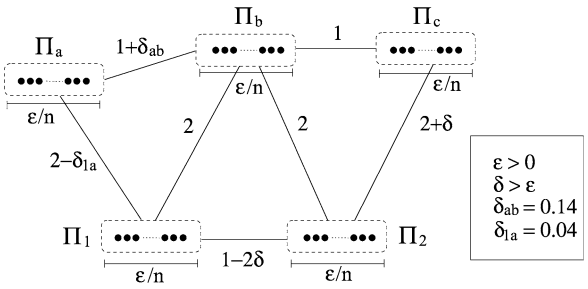

Fig. 1 Instance I_k has no pure Nash equilibrium when $\alpha = 0.6k$, where $k = n/5$. The number of peers in each cluster is k

The n peers of instance I_k are grouped into five clusters Π_1, Π_2, Π_a, Π_b, and Π_c, each containing $k = n/5$ peers. Within a cluster, peers are located equidistantly in a line, and each cluster's diameter is ϵ/n, where $\epsilon > 0$ is an arbitrarily small constant. The *inter-cluster distance* $d(\Pi_i, \Pi_j)$ between Π_i and Π_j is the minimal distance between any two peers in the two clusters. Distances not explicitly defined in Fig. 1 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 Π_a and Π_b, Π_b and Π_c, as well as between Π_1 and Π_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 Π_i to peers in a cluster Π_j.

To preserve connectivity, some peers in Π_1 and Π_2 must have links to peers in the "upper clusters" Π_a, Π_b, Π_c in Fig. 1; henceforth, we will refer to the peers in these upper clusters Π_a, Π_b, Π_c as *top-peers*. Based on the aforementioned observations, the set of possible strategies can further be narrowed down as follows.

(i) Neither peers in Π_1 nor Π_2 select three links to top-peers.
(ii) There exists a peer $\pi_1 \in \Pi_1$ that establishes a link to Π_a.
(iii) There is exactly one link from cluster Π_2 to either cluster Π_b or Π_c, but there is no link to Π_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 Π_b and Π_c (thus violating Case (iii)), π_2' can lower its costs by dropping its link to Π_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 α.

It can be shown that only the six structures depicted in Fig. 2 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 Π_b.

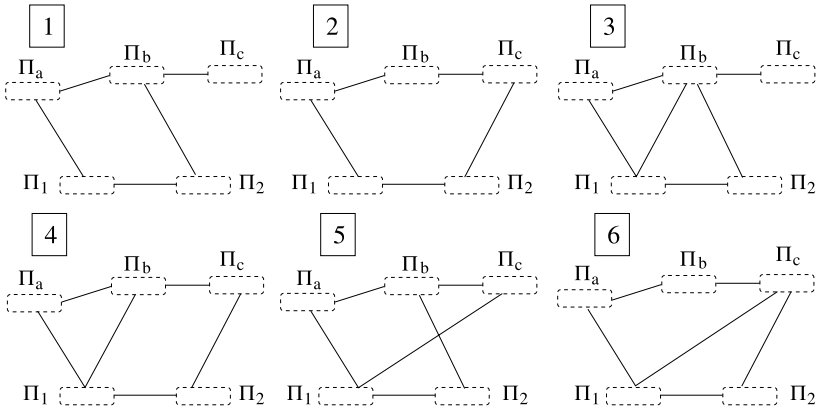

Fig. 2 Candidates for a Nash equilibrium

Case 2: If the only outgoing link from Π_1 to a top-cluster is to cluster Π_a, the peer $\pi_2 \in \Pi_2$ maintaining the link to Π_c can be shown to profit from switching its link from Π_c to Π_b.

Case 3: The availability of the link from Π_1 to Π_b changes the optimal choice of the above mentioned peer $\pi_2 \in \Pi_2$. Unlike in the previous case, π_2 now prefers linking to Π_c instead of Π_b.

Case 4: Due to the existence of a link from a peer $\pi_2 \in \Pi_2$ to Π_c, the peer $\pi_1 \in \Pi_1$ with the link to Π_b has an incentive to drop this link and instead use the detours via Π_2 and Π_a to connect to Π_c and Π_b, respectively.

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

Case 6: Finally, this case is similar to Case 4: $\pi_1 \in \Pi_1$ with the link to Π_c has an incentive to remove its link to Π_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. Fig. 2): $1 \rightsquigarrow 3 \rightsquigarrow 4 \rightsquigarrow 2 \rightsquigarrow 1 \rightsquigarrow 3 \ldots$. In other words, selfish peers will not achieve a stable network topology.

5.1 Decoupling α and n

In the construction above, the network size n and α are coupled: the larger the edge cost α, the more peers are needed per cluster in order for the construction to hold. The question is whether Nash equilibria may always exist if α is constant, whereas n is large. We now show that there are settings without an equilibrium for arbitrary ratios n/α, even for constant α.

Consider again the construction of Fig. 1 and let each cluster consist of one peer only, that is, $n = 5$, and hence $k = 1$ and $\alpha = 0.6$. Let us refer to this 5-peer network

by $\mathbb{N}_5$. From Theorem 5.1, we already know that $\mathbb{N}_5$ does not have a pure Nash equilibrium. For any $i \in 1, 2, \ldots$, we construct a network with $n = 5 \cdot i$ peers and $\alpha = 0.6$ by simply arranging i copies of $\mathbb{N}_5$ horizontally using a sufficiently large distance between two consecutive networks. Clearly, in order to prevent infinite costs, the different entities of $\mathbb{N}_5$ will be connected. Moreover, due to the distance between the $\mathbb{N}_5$ networks, the rational about the edges created within $\mathbb{N}_5$ remains independent of the peers outside, and there is only one peer in a given $\mathbb{N}_5$ connecting to another given $\mathbb{N}_5$. Therefore, the proof of Theorem 5.1 holds for the individual $\mathbb{N}_5$s which gives us, using $\alpha = 0.6$, the following claim for our construction.

Corollary 5.2 *Regardless of the magnitude of the ratio n/α, there are metric spaces $\mathcal{M}$, for which no pure Nash equilibrium exists.*

Observe that for the corresponding claim with larger αs, again, single peers can be replaced by peer clusters.

6 Complexity of Nash Equilibria

It remains to answer the question 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.

Theorem 6.1 *Regardless of the magnitude of α, 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. The proof is based on a reduction from an **NP**-complete form of the Boolean satisfiability problem SAT which is restricted to instances with 2 or 3 variables per clause and at most 3 occurrences per variable [40]. For any α 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 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 Fig. 3, each rectangular box representing a cluster of k peers. Assume that the SAT instance is given in standard conjunctive normal form (CNF). For each clause C_j, we employ a gadget of three *clause-clusters* Π_j^a, Π_j^b, and Π_j^c. For every variable x_i, the two *literal-clusters* Π_i^0 and Π_i^1 represent the negative and positive literal of the variable, respectively. Finally, the construction's peer set is completed with three special clusters Π_c, Π_y, and Π_z. The pairwise distances between two peers in $\mathcal{M}_I^k$ are determined by the graph G_I^k shown in Fig. 3 (a formal definition appears in Sect. 6.1). Two nodes within the same cluster have a distance of ϵ, for some arbitrarily small $\epsilon < (k(2n + 3m + 3))^{-2}$, where m and n denote

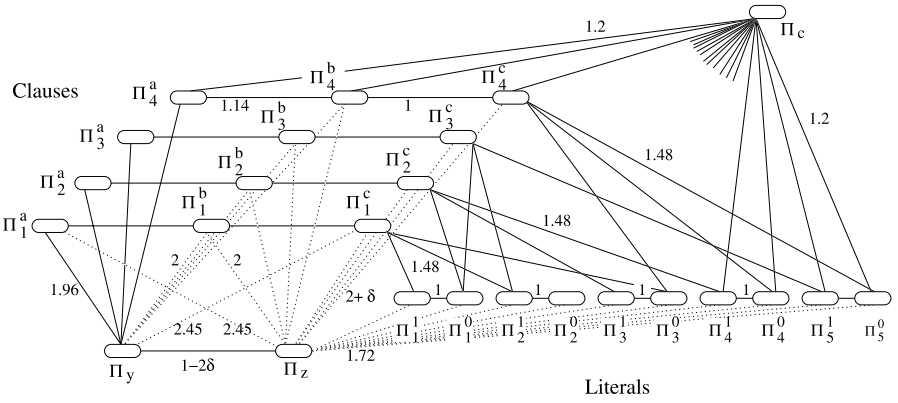

Fig. 3 The graph G_I for instance $I = (x_1 \vee \overline{x_3}) \wedge (\overline{x_1} \vee x_3 \vee x_4) \wedge (\overline{x_1} \vee x_2 \vee x_5) \wedge (\overline{x_3} \vee \overline{x_4} \vee \overline{x_5})$. Each cluster contains k peers with pairwise distance ϵ. δ is an arbitrary constant such that $\delta > \epsilon > 0$

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 Π_i and Π_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 Π_j^a, Π_j^b, and Π_j^c in combination with the two special clusters Π_y and Π_z form an instance similar to I_k as used in the discussion of Theorem 5.1 (cf. Fig. 1). 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 Π_z have a distance of 1.72 to all literal-peers. Furthermore, the distance between a clause-cluster Π_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 Π_i^1 and Π_j^c is small, i.e., only 1.48. Similarly, if $\overline{x_i} \in C_j$, then $d(\Pi_i^0, \Pi_j^c) = 1.48$. 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 Π_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 Π_i^1 to a clause-cluster Π_j^c if $x_i \in C_j$, or Π_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 Π_i^1 or Π_i^0 (but not both!), while no other peer in Π_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 a length of 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 π_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 Π_y, Π_z, and the clause-peers of C_{j^*} behaves similarly as in instance I_k of Fig. 1. That is, peers in Π_y and Π_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 Π_z builds a direct link to a peer in Π^1_i if x_i is set to true in A_I and to a peer in Π^0_i otherwise. Since A_I is a satisfying assignment, there must exist a path from Π_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 Π^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 Π_z via a literal-cluster to peers in every cluster Π^c_j. It follows that in any satisfied clause C_j, the achievable reduction in stretch costs at a peer in Π_z when connecting directly to clusters Π^b_j or Π^c_j is significantly smaller than in an unsatisfied clause. Specifically, it can be shown that peers in Π_y and Π_z are in a *stable* situation if one peer $\pi_y \in \Pi_y$ connects to Π^a_j and Π^b_j of every clause C_j, and no peer in Π_z directly builds a link to any clause-peer. Since A_I is a satisfying assignment, peers in Π_y and Π_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 Π_c, Π^a_j, Π^b_j, Π^c_j, Π^1_i, or Π^0_i) has an incentive to deviate from its strategy. For this final ingredient of the proof, the existence of cluster Π_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^k_I (and consequently $\mathcal{M}^k_I$) from the SAT instance I. In Sects. 6.2 and 6.3, we will show that there exists a Nash equilibrium in $\mathcal{M}^k_I$ if and only if I is satisfiable. Theorem 6.1 then follows from Lemmas 6.12 and 6.14, as well as the **NP**-hardness of SAT. The following theorem and proof is due to Tovey [40]:

Theorem 6.2 *Boolean satisfiability is* **NP***-hard when restricted to instances with* 2 *or* 3 *variables per clause and at most* 3 *occurrences per variable.*

Proof Consider a general 3-SAT instance. For each variable x which appears in more than three clauses perform the following procedure: suppose x appears in k clauses. Create k new variables $x_1, \ldots, x_k$ and replace the ith occurrence of x with x_i, $i = 1, \ldots, k$. Append the clause $\{x_i \vee \overline{x}_{i+1}\}$ for $i = 1, \ldots, k-1$ and the clause $\{x_k \vee \overline{x}_1\}$. In the new instance the clause $\{x_i \vee \overline{x}_{i+1}\}$ implies that if x_i is false, x_{i+1} must be false as well. The cyclic structure of the clauses therefore forces the x_i to have the same assignment, and hence the new instance is satisfiable if the original one is. As

the transformation requires polynomial time, and as 3-SAT is **NP**-hard [13], the claim follows. □

6.1 The Construction of $\mathcal{M}_I^k$

Let I be an instance of SAT expressed in conjunctive normal form (CNF), in which each clause contains 2 or 3 variables. Without loss of generality, we can assume that each variable in I appears in at most 3 clauses [40]. Furthermore, we can restrict our attention to those instances of SAT in which each variable appears both as a positive and a negative literal at least once, because otherwise, assigning a feasible value to this variable is 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 a 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 Fig. 3. 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 given by the *cluster-distance* $d(\Pi_i, \Pi_j)$ illustrated in Fig. 3. 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* Π_j^a, Π_j^b, and Π_j^c. Also, we add a pair of *literal-clusters* Π_i^0 and Π_i^1 for each of the n variables, with Π_i^1 representing the positive literal x_i, and Π_i^0 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 Π_c, Π_z, and Π_y. Call the union of Π_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 Fig. 3—are as follows. Let δ 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.

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

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

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

$$\forall \pi_i^0 \in \Pi_i^0, \forall \pi_i^1 \in \Pi_i^1 : \quad d(\pi_i^0, \pi_i^1) = 1$$

$$\forall \pi_z \in \Pi_z, \forall \pi_i^\mu \in \Pi_i^\mu : \quad d(\pi_z, \pi_i^0) = d(\pi_z, \pi_i^1) = 1.72.$$

Furthermore,

$$\forall C_j \in \mathcal{C}, x_i \in C_j \forall \pi_i^1 \in \Pi_i^1, \forall \pi_j^c \in \Pi_j^c : \quad d(\pi_i^1, \pi_j^c) = 1.48$$

$$\forall C_j \in \mathcal{C}, \overline{x_i} \in C_j \forall \pi_i^0 \in \Pi_i^0, \forall \pi_j^c \in \Pi_j^c : \quad d(\pi_i^0, \pi_j^c) = 1.48.$$

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 Π_y, Π_z, as well as the clause-clusters Π_j^a, Π_j^b, and Π_j^c. By itself, each such gadget is reminiscent of the construction shown in Fig. 1. 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 Π_i^μ construct links to those (at most two) clause-clusters Π_j^c in whose clause the literal occurs. 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 Π_z to each variable $x_i \in \mathcal{X}$, i.e., one peer in Π_z connects to a peer in either Π_i^0 or Π_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 Π_j^c, the length of the path from π_z to peers in Π_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 π_z to Π_j^c via some "satisfying" literal-cluster Π_i^μ—while being suboptimal compared to the direct connection—is relatively small. Specifically, it is small enough to ensure that no peer in Π_z has an incentive to construct an additional direct link to Π_j^b or Π_j^c. Once peers in Π_z have no further need to establish direct links to a clause-peer of C_j, the best possible strategy of peers in Π_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 Π_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 Π_z to Π_j^c too large, and it becomes worthwhile for $\pi_z \in \Pi_z$ to construct direct links to Π_j^c, and even to Π_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 Π_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 Π_c. This implies that all top-layer clusters are connected to one another almost optimally (i.e., with low stretch) 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 6.3 *Consider two peers π_g and π'_g in an arbitrary cluster Π_g. In a Nash equilibrium, there exists a path from π_g to π'_g of length at most $k\epsilon$.*

Proof Because the distance between π_g and π'_g is ϵ, it is easy to see that the shortest path between these two nodes must be located entirely in Π_g. Because the distance between each pair of peers in a cluster is ϵ and there are k peers in the cluster, the claim follows. □

Lemma 6.4 *Consider two arbitrary clusters Π_g and Π_h. In a Nash equilibrium, there is at most one peer $\pi_g \in \Pi_g$ that has a link to a peer in Π_h.*

Proof Assume for contradiction that there are two nodes π_g and π'_g that maintain links to peers in Π_h. Then, π'_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 ϵ, 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 6.5 *Let Π_g and Π_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 Π_h.*

Proof By Lemma 6.4, there cannot be more than one peer in Π_g having a link to Π_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 a length of at least 2.2 (e.g., from Π_i^0 via Π_i^1 to Π_c). In other words, if there is no direct connection between the two clusters, π_g has a stretch of at least $2.2/1.2$ to each peer in Π_h. Because $\frac{2.2k}{1.2} > \alpha + k(1 + 2k\epsilon)$, it is beneficial for π_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 Π_j^c and Π_i^1, if $x_i \in C_j$, or between Π_j^c and Π_i^0, if $\overline{x_i} \in C_j$.

The shortest indirect connection between two such clusters has a length of at least 2.4 (via cluster Π_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 6.5 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 Π_i^1 and Π_i^0, as well as for a literal-cluster Π_i^1 (or Π_i^0) and a Π_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 Π_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 ϵ being smaller than $(k(m + n + 3))^{-2}$, this stretch is virtually as good as 1. Finally, there are also links between Π_y and Π_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 6.6 *In a Nash equilibrium, there is exactly one peer $\pi_y \in \Pi_y$ that has a link to a peer in Π_j^a, for all $C_j \in \mathcal{C}$, and vice versa.*

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

Lemma 6.7 *Assume that there is a link between Π_z and at least one literal-cluster of every variable $x_i \in \mathcal{X}$ and that there is a link between Π_y and Π_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 Π_j^a, Π_j^b, or Π_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 Π_y to Π_j^a (for every $C_j \in \mathcal{C}$) and Π_z. Hence, connecting to Π_j^b or Π_j^c clearly cannot reduce the stretch to peers in Π_z, Π_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 Π_j^b or Π_j^c. Finally, the path from $\pi_y \in \Pi_y$ to any literal-peer in Π_i^μ has a length of at most $3.72 - 2\delta + 3k\epsilon$. This is because there exists a link between Π_y and Π_z, and between Π_i^0 and Π_i^1, and because there is a link from Π_z to either Π_i^0 or Π_i^1. On the other hand, the path from $\pi_y \in \Pi_y$ to a literal-peer via Π_j^b or Π_j^c has a length of 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. Let $A_I(x_i)$ be the assignment of x_i in A_I, i.e.,

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

Furthermore, we define in every cluster Π_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:

$$s_y := \Pi_y \cup \{\hat{\pi}_z\} \cup \bigcup_{C_j \in \mathcal{C}} \{\hat{\pi}_j^a, \hat{\pi}_j^b\}$$

$$s_c := \Pi_c \cup \bigcup_{x_i \in \mathcal{X}} \{\hat{\pi}_i^0 \cup \hat{\pi}_i^1\} \cup \bigcup_{C_j \in \mathcal{C}} \{\hat{\pi}_j^a \cup \hat{\pi}_j^b \cup \hat{\pi}_j^c\}$$

$$s_j^c := \Pi_j^c \cup \{\hat{\pi}_c, \hat{\pi}_z, \hat{\pi}_j^b\} \cup \bigcup_{x_i^\mu \in C_j} \{\hat{\pi}_i^\mu\}, \quad \forall C_j \in \mathcal{C}$$

$$s_i^\mu := \Pi_i^\mu \cup \{\hat{\pi}_c, \hat{\pi}_z, \hat{\pi}_i^{1-\mu}\} \cup \bigcup_{x_i^\mu \in C_j} \{\hat{\pi}_j^c\}, \quad \forall x_i \in \mathcal{X}$$

$$s_z := \Pi_z \cup \{\hat{\pi}_y\} \cup \bigcup_{x_i \in \mathcal{X}} \{\hat{\pi}_i^{A_I(x_i)}\}$$

$$s_j^a := \Pi_j^a \cup \{\hat{\pi}_c, \pi_y, \hat{\pi}_j^b\}, \quad \forall C_j \in \mathcal{C}$$

$$s_j^b := \Pi_j^b \cup \{\hat{\pi}_c, \hat{\pi}_j^a, \hat{\pi}_j^c\}, \quad \forall C_j \in \mathcal{C}$$

Strategy s is illustrated in Fig. 4. 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 6.5). 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 6.7. That is, no peer in clusters Π_y and Π_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.

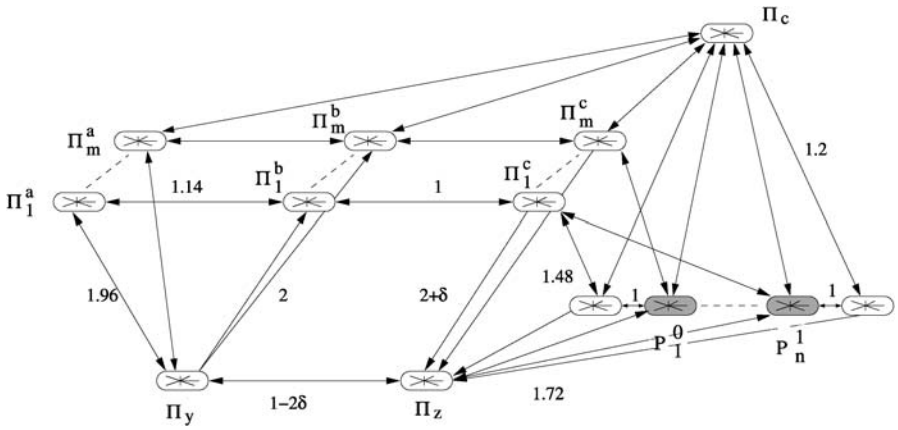

Fig. 4 An example instance G_I^k with the topology resulting from strategy s. Within each cluster, the peers are connected as a star. Directed arrows between clusters indicate inter-cluster links between cluster-leaders. Cluster-leader $\hat{\pi}_z$ connects to those leaders of literal-peers that appear in the satisfying assignment A_I. In the example, A_I sets $x_1 = 0$ and $x_n = 1$

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 π_i changes its strategy according to action ψ, ψ being clear from the context. Specifically, if $\Delta_i(\psi) \geq 0$, peer π_i has no incentive to perform action ψ 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 6.8 *In s, no peer in an arbitrary cluster Π_g has an incentive to change its strategy within the cluster, i.e., to add, replace, or remove links to peers in Π_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 α 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 6.8, we can regard the topology within each cluster in s as 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 Π_y cannot unilaterally reduce their costs in s.

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

Proof By Lemma 6.8, 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 6.6 that $\hat{\pi}_y$ must keep its links to $\hat{\pi}_j^a$ and $\hat{\pi}_z$. We now show that no peer in Π_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 or more additional links: in the topology resulting from s, every peer in Π_y is connected with stretch at most $1+2\epsilon$ with all peers except from peers in Π_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 Π_y can reduce its stretch to peers in exactly one of these clusters only. Hence, 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 6.8 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 Π_j^c is reduced, the stretch to peers in Π_j^b increases. The relative cost difference is $\Delta_y(\hat{\pi}_j^b \to \hat{\pi}_j^c) \geq -\frac{k(3+\epsilon)}{2.45} + \frac{(1.96+1.14)k}{2} > 0$.

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 α. On the other hand, the stretch to both Π_j^b and Π_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 Π_i^μ to which $\hat{\pi}_z$ is not connected, formally $\mu \neq A_I(x_i)$. Doing so clearly increases the stretch to peers in Π_j^b and Π_j^c, but like in Case 3, the shortest connection between $\hat{\pi}_y$ to peers in Π_j^c is via $\hat{\pi}_j^a$ and $\hat{\pi}_j^b$. In particular, this path has a length of at most $4.1+\epsilon$, whereas the shortest path via a literal-cluster has a length of 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 Π_y either. □

Lemma 6.10 *No peer in Π_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 Π_z. Recall that by Lemma 6.7, 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 Π_z and peers in Π_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 6.8 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 6.9, 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 Π_z connect to leader peers.

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

Case 1: Some peer in Π_z adds an additional link to $\hat{\pi}_j^b$: the reduction of the stretches to peers in Π_j^b and Π_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 Π_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 Π_z adds an additional link to $\hat{\pi}_j^c$: again, the stretches to Π_j^b and Π_j^c are not reduced enough to render the additional link worthwhile. In fact, the stretch to peers in Π_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 Π_j^c (Lemma 6.7). 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 Π_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 Π_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 Π_z to every Π_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 6.8 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, $\Delta_y(\hat{\pi}_i^{A_I(x_i)} \to \hat{\pi}_i^{1-A_I(x_i)}) \geq 0$.

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 6.5. □

Having shown that peers in Π_y and Π_z have no incentive to deviate from s, we have to prove that no other peer can improve its situation either.

Lemma 6.11 *No top-layer peer can benefit from changing its strategy, given that all other peers follow s.*

Proof First, by Lemma 6.8, 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 6.9.

It is important to observe that in s, all top-layer peers are almost optimally connected with each other, either via the central cluster Π_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 2ϵ from these small stretches, adding additional links can only help in reducing the stretches to peers in Π_y and Π_z. By Lemma 6.5, 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 Π_j^a: A peer $\hat{\pi}_j^a$'s link to $\hat{\pi}_y$ cannot be removed by Lemma 6.6. For every peer $\pi_j^a \in \Pi_j^a$, it further holds that building an additional link to $\hat{\pi}_z$ is too costly, $\Delta_j^a(+\hat{\pi}_z) \geq -\frac{k(2.96-2\delta+2\epsilon)}{2.45} + k + \alpha > 2kN\epsilon$. Hence, even if this additional link could reduce all other less than N stretches to top-level peers by the remaining 2ϵ, the cost of an additional link would still be too high.

Peers in Π_j^b: Peer $\hat{\pi}_j^b$ 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 Π_j^b. In the first case, we have $\Delta_j^b(+\hat{\pi}_y) \geq -\frac{k(1.96+1.14+2\epsilon)}{2} + k - \frac{k(3+\delta+2\epsilon)}{2} + \frac{k(3-2\delta)}{2} + \alpha > 2kN\epsilon$. As for the second case, $\Delta_j^b(+\hat{\pi}_z) \geq -\frac{k(1.96+1.14+2\epsilon)}{2} + \frac{k(3-2\delta)}{2} - \frac{k(3+\delta+2\epsilon)}{2} + k + \alpha > 2kN\epsilon$. Clearly, building both links is even less worthwhile.

Peers in Π_j^c: The potential strategy deviations that could decrease peer $\hat{\pi}_j^c$'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 $\hat{\pi}_j^c$ (or for any non-leader peer in Π_j^c in the case of link addition). First, it holds that $\Delta_j^c(+\hat{\pi}_y) \geq -\frac{k(3-\delta+2\epsilon)}{2.45} + k + \alpha > 2kN\epsilon$ and $\Delta_j^c(-\hat{\pi}_z) \geq -\alpha - k(1+2\epsilon) - \frac{k(3-\delta+2\epsilon)}{2.45} + \frac{3.2k}{2+\delta} + \frac{4.1k}{2.45} > 2kN\epsilon$. Also, switching the link from $\hat{\pi}_z$ to $\hat{\pi}_y$ is not helpful, $\Delta_j^c(\hat{\pi}_z \to \hat{\pi}_y) \geq \frac{3.2k}{2+\delta} - \frac{k(3-\delta+2\epsilon)}{2.45} > 2kN\epsilon$.

Peers in Π_i^μ: 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 2ϵ by the additional link, which does not render the link cost α worthwhile. In the second case, the stretch is strictly increased. If $\hat{\pi}_i^\mu$ removes its link to $\hat{\pi}_z$ and connects via its neighboring literal-cluster, the stretches to both Π_y and Π_z increase. Particularly, we have $\Delta_i^\mu(-\hat{\pi}_z) \geq -\alpha - k(1+2\epsilon) + \frac{2.72k}{1.72} + \frac{k(3.72-2\delta)}{2.72-2\delta} > 2kN\epsilon$.

Peers in Π_c*:* Finally, peers in Π_c are connected with stretch at most 2ϵ 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 ϵ and α, it is clear that no peer in Π_c can improve its strategy. □

By combining Lemmas 6.9, 6.10, and 6.11, we know that no peer in the network has an incentive to change its strategy. Hence, s constitutes a pure Nash equilibrium.

Lemma 6.12 *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 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 Sect. 6.1, an important characteristic of any Nash equilibrium is the fact that exactly one peer in Π_z connects to *exactly one* literal-peer (either in Π_i^0 or Π_i^1) for every variable $x_i \in \mathcal{X}$.

Lemma 6.13 *In any Nash equilibrium, exactly one peer in Π_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 6.10 (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 6.4 that no other peer in Π_z connects to the same cluster as π_z. Hence, we only need to show that in a Nash equilibrium no two peers in Π_z connect to both literal-clusters of the same variable.

Assume for the sake of contradiction that peers π_z and π_z' (potentially $\pi_z = \pi_z'$) maintain links to both Π_i^0 and Π_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 Π_j^c if this link does not already exist. By the definition of our special SAT instance and the construction of G_I, we know that of the two literal-clusters, one, say Π_i^μ, has clause-cluster Π_j^c at distance 1.48, and the other literal-cluster, say $\Pi_i^{1-\mu}$, has one or two such close-by clause-clusters. Let π_z' be the peer that connects to cluster Π_i^μ (otherwise, replace π_z with π_z' for the remainder of the proof).

Assume for the first case that the length of the shortest path from π_z' to this Π_j^c without the link via Π_i^μ is 3.2 or longer. In this case, the change in π_z''s costs when switching from its link to literal-cluster Π_i^μ that has only a single close-by clause-cluster Π_j^c directly to a peer in Π_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 π_z to Π_j^c is strictly shorter than 3.2, then the link to Π_i^μ 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, π_z' is always better off not connecting to a literal-cluster if π_z already connects to a literal-cluster. From this, the claim follows. □

Lemma 6.13 is an important ingredient for the remainder of the proof, because it gives us a one-to-one correspondence between the connections of Π_z to literal-clusters, and an assignment of variables in the SAT instance I. Also, note that when combining Lemma 6.13 with Lemma 6.7, it follows that in a Nash equilibrium, peers in Π_y and Π_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 6.14 *If I is non-satisfiable, there exists no pure Nash equilibrium in $\mathcal{M}_I^k$.*

Proof By Lemma 6.13, exactly one peer in Π_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 Π_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 Π_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 Π_y already has a link to $\Pi_{j^*}^b$, then the cost reduction for π_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 π_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 $\Pi_{j^*}^b$ or in $\Pi_{j^*}^c$. To see this, assume that no peer in Π_z establishes any links to peers in the two clusters. In this case (because there is no link from Π_y to $\Pi_{j^*}^c$, and because C_{j^*} is not satisfied), the sum of the stretches to peers in $\Pi_{j^*}^c$ 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 $\pi_{j^*}^c$.

It remains to show that no peer in Π_z connects to $\Pi_{j^*}^a$, and particularly, that no two peers in Π_z simultaneously connect to both $\Pi_{j^*}^b$ or $\Pi_{j^*}^c$. Because there is at least one link from Π_z to either $\Pi_{j^*}^b$ or $\Pi_{j^*}^c$, it follows that a link to $\Pi_{j^*}^a$ 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., $\Delta_z(+\pi_{j^*}^a) = -\frac{k(2.96-2\delta+2k\epsilon)}{2.45} + \alpha + k > 0$. For the last case, assume that two peers π_z and π_z' (potentially the same) connect to both $\Pi_{j^*}^b$ and $\Pi_{j^*}^c$, respectively. Then, π_z' has an incentive to drop its link to $\Pi_{j^*}^c$: $\Delta_z(-\pi_{j^*}^c) = \frac{k(3+2k\epsilon)}{2+\delta} - k - \alpha < 0$. Hence, in any Nash equilibrium, there is exactly one link from Π_z to either $\Pi_{j^*}^b$ or $\Pi_{j^*}^c$, but not to both.

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

Case 1: Some peer $\pi_z \in \Pi_z$ connects to π^b_{j*}: in this case, some peer $\pi_y \in \Pi_y$ has an incentive to add a link to a peer in Π^b_{j*}, because this significantly reduces its stretches to peers in Π^b_{j*} and Π^c_{j*}. Specifically, π_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 Π^b_{j*}: in this case, the peer π_z can profit from switching its link to a peer in Π^c_{j*}. Specifically, $\Delta_z(\pi^b_{j*} \to \pi^c_{j*}) \leq -\frac{3k}{2+\delta} + \frac{k(3-2\delta+2k\epsilon)}{2} < 0$.

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

Case 4: Some peer $\pi_z \in \Pi_z$ connects to Π^c_{j*} and some peer $\pi_y \in \Pi_y$ connects to Π^b_{j*}: in this configuration, peer π_y benefits from removing its link to Π^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

The analysis of our locality game reveals that the efficiency of p2p topologies can suffer if peers act selfishly. Already in a simple 5-peer network, no equilibrium may exist. Moreover, our results indicate that topologies may degrade more severely when selfish peers value maintenance cost relatively higher than latency costs. Finally, it has been shown that it is generally a hard problem to decide whether a p2p system can stabilize if peers select their neighbors in a selfish manner.

7.1 Practical Issues and Future Directions

The main contributions of this paper are rather theoretic in nature, however, we believe that our model indeed the main trade-offs that might govern selfish peers' decisions. We have gained some experience with selfish behavior by collecting data in the popular BitTorrent network (using our own client *BitThief* [26]) which provides evidence—merely by counting the number of users that actually use BitThief every day—that peers are indeed selfish. Unfortunately, the collected data cannot directly be used to derive scientific statements about topological structures, as in BitTorrent, peers learn about new neighbors by recontacting a so-called tracker. However, our measurements indicate that peers are indeed willing to accept large degrees and connect to many other peers if this improves the download speed (up to 500 TCP connections do not cause any performance loss). Thus, in a file-sharing swarm, peers tend to have relatively small αs. Gnutella would be an interesting network for a more detailed measurement study to verify our results. While we expect that some of our

insights can be confirmed, it may turn out that different forms of "rationality" may appear on the level of different client versions rather than on the level of individual users. This line of research is left for future work.

7.2 Open Theoretical Problems

Many theoretic questions are left for future research as well. For instance, we understand that the **NP**-hardness construction makes use of network topologies that are most unlikely to occur in practice. Whether unique Nash equilibria exist for the networks we observe in practice today and whether these equilibria can be computed in polynomial time remains an open question. It would also be interesting to know whether incentive mechanisms can be designed such that the resulting topologies have desirable properties, e.g., are hypercubic or pancake networks. Some of our assumptions should also be weakened; e.g., a peer may not have complete knowledge of the other peers' states. Finally, we have not investigated how optimal topologies (with respect to our social cost function) can be computed, and approximate or mixed Nash equilibria have not been considered yet either.

References

1. Abraham, I., Malkhi, D., Dobzinski, O.: LAND: Stretch $(1 + e)$ locality aware networks for DHTs. In: Proc. 15th ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 550–559 (2004)
2. Abraham, I., Badola, A., Bickson, D., Malkhi, D., Maloo, S., Ron, S.: Practical locality-awareness for large scale information sharing. In: Proc. 4th Int. Workshop on Peer-to-Peer Systems (IPTPS) (2005)
3. Adar, E., Huberman, B.A.: Free riding on Gnutella. First Monday **5**(10) (2000)
4. Albers, S., Eilts, S., Even-Dar, E., Mansour, Y., Roditty, L.: On Nash equilibria for a network creation game. In: Proc. 17th ACM Symposium on Discrete Algorithms (SODA) (2006)
5. Anshelevich, E., Dasgupta, A., Kleinberg, J., Tardos, E., Wexler, T., Roughgarden, T.: The price of stability for network design with fair cost allocation. In: Proc. 45th Symposium on Foundations of Computer Science (FOCS), pp. 295–304 (2004)
6. Austen, I.: Like a swerving commuter, a selfish router slows traffic. The New York Times (2003)
7. Axelrod, R.: The evolution of cooperation. Science **211**(4489), 1390–1396 (1981)
8. Bharambe, A.R., Herley, C., Padmanabhan, V.N.: Analyzing and improving a BitTorrent network's performance mechanisms. In: Proc. IEEE Conference on Computer Communications (INFOCOM), pp. 36–46 (2006)
9. Chan, H.T.-H., Gupta, A., Maggs, B.M., Zhou, S.: On hierarchical routing in doubling metrics. In: Proc. 17th ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 762–771 (2005)
10. Chen, H.-L., Roughgarden, T.: Network design with weighted players. In: Proc. 18th ACM Symposium on Parallel Algorithms and Architectures (SPAA), pp. 29–38 (2006)
11. Cohen, B.: Incentives build robustness in BitTorrent. In: Proc. 1st Workshop on Economics of Peer-to-Peer Systems (P2PEcon) (2003)
12. Cole, R., Dodis, Y., Roughgarden, T.: How much can taxes help selfish routing? J. Comput. Syst. Sci. **72**(3), 444–467 (2006)
13. Cook, S.A.: The complexity of theorem-proving procedures. In: Proc. 3rd ACM Symposium on Theory of Computing (STOC), pp. 151–158 (1971)
14. Corbo, J., Parkes, D.C.: The price of selfish behavior in bilateral network formation. In: Proc. 24th ACM Symposium on Principles of Distributed Computing (PODC), pp. 99–107 (2005)
15. Demaine, E.D., Hajiaghayi, M.T., Mahini, H., Zadimoghaddam, M.: On the topologies formed by selfish peers. In: Proc. 26th Annual Symposium on Principles of Distributed Computing (PODC) (2007)
16. Demaine, E.D., Hajiaghayi, M., Mahini, H., Zadimoghaddam, M.: The price of anarchy in cooperative network creation games. In: Proc. 26th International Symposium on Theoretical Aspects of Computer Science (STACS) (2009)

17. Eidenbenz, S., Kumar, V., Zust, S.: Equilibria in topology control games for ad hoc networks. In: Proc. ACM Joint Workshop on Foundations of Mobile Computing (DIALM-POMC) (2003)
18. Fabrikant, A., Luthra, A., Maneva, E., Papadimitriou, C.H., Shenker, S.: On a network creation game. In: Proc. 22nd ACM Symposium on Principles of Distributed Computing (PODC), pp. 347–351 (2003)
19. Feldman, M., Chuang, J.: Overcoming free-riding behavior in peer-to-peer systems. ACM Sigecom Exch. **6** (2005)
20. Ganesan, P., Seshadri, M.: On cooperative content distribution and the price of barter. In: Proc. 25th IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 81–90 (2005)
21. Garcia, F.D., Hoepman, J.-H.: Off-line karma: A decentralized currency for peer-to-peer and grid applications. In: Proc. 3rd Applied Cryptography and Network Security (ACNS)
22. Grolimund, D., Meisser, L., Schmid, S., Wattenhofer, R.: Havelaar: A robust and efficient reputation system for active peer-to-peer systems. In: Proc. 1st Workshop on the Economics of Networked Systems (NetEcon), June 2006
23. Hughes, D., Coulson, G., Walkerdine, J.: Free riding on Gnutella revisited: The bell tolls? IEEE Distrib. Syst. Online **6**(6) (2005)
24. Jun, S., Ahamad, M.: Incentives in BitTorrent induce free riding. In: Proc. 3rd ACM SIGCOMM Workshop on Economics of Peer-to-Peer Systems (P2PEcon) (2005)
25. Karger, D.R., Ruhl, M.: Finding nearest neighbors in growth-restricted metrics. In: Proc. 34th ACM Symposium on Theory of Computing (STOC), pp. 741–750 (2002)
26. Locher, T., Moor, P., Schmid, S., Wattenhofer, R.: Free riding in BitTorrent is cheap. In: Proc. 5th Workshop on Hot Topics in Networks (HotNets) (2006)
27. Moscibroda, T., Schmid, S., Wattenhofer, R.: On the topologies formed by selfish peers. In: Proc. 5th International Workshop on Peer-to-Peer Systems (IPTPS) (2006)
28. Moscibroda, T., Schmid, S., Wattenhofer, R.: On the topologies formed by selfish peers. In: Proc. 25th Annual Symposium on Principles of Distributed Computing (PODC) (2006)
29. Odlyzko, A.M.: The case against micropayments. In: Financial Cryptography, pp. 77–83 (2003)
30. Osborne, M., Rubinstein, A.: A Course in Game Theory. MIT Press, Cambridge (2000)
31. Papadimitriou, C.H.: Algorithms, games, and the Internet. In: Proc. 33rd ACM Symposium on Theory of Computing (STOC), pp. 749–753 (2001)
32. Plaxton, C., Rajaraman, R., Richa, A.W.: Accessing nearby copies of replicated objects in a distributed environment. In: Proc. 9th ACM Symposium on Parallel Algorithms and Architectures (SPAA), pp. 311–320 (1997)
33. Qiu, D., Srikant, R.: Modeling and performance analysis of BitTorrent-like peer-to-peer systems. In: Proc. ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (2004)
34. Roughgarden, T.: Selfish Routing and the Price of Anarchy. MIT Press, Cambridge (2005)
35. Roughgarden, T., Tardos, E.: How bad is selfish routing? J. ACM **49**(2) (2002)
36. Rowstron, A., Druschel, P.: Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems. In: Proc. IFIP/ACM Int. Conference on Distributed Systems Platforms (Middleware), pp. 329–350 (2001)
37. Sanghavi, S., Hajek, B.: A new mechanism for the free-rider problem. In: Proc. 3rd ACM SIGCOMM Workshop on Economics of Peer-to-Peer Systems (P2PEcon) (2005)
38. Shneidman, J., Parkes, D.C.: Rationality and self-interest in peer to peer networks. In: Proc. 2nd Int. Workshop on Peer-to-Peer Systems (IPTPS) (2003)
39. Tamilmani, K., Pai, V., Mohr, A.: SWIFT: A system with incentives for trading. In: Proc. 2nd Workshop on Economics of Peer-to-Peer Systems (P2PEcon) (2004)
40. Tovey, C.A.: A simplified NP-complete satisfiability problem. Discrete Appl. Math. **8**, 85–89 (1984)
41. Vishnumurthy, V., Chandrakumar, S., Sirer, E.G.: KARMA: A secure economic framework for P2P resource sharing. In: Proc. 1st Workshop on Economics of Peer-to-Peer Systems (P2PEcon) (2003)
42. Wang, W., Li, B.: Market-driven bandwidth allocation in selfish overlay networks. In: Proc. IEEE Conference on Computer Communications (INFOCOM), pp. 36–46 (2005)
43. Wong, B., Slivkins, A., Sirer, E.G.: Meridian: A lightweight network location service without virtual coordinates. In: Proc. ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (2005)
44. Zhao, B.Y., Huang, L., Stribling, J., Rhea, S.C., Joseph, A.D., Kubiatowicz, J.D.: Tapestry: A resilient global-scale overlay for service deployment. IEEE J. Sel. Areas Commun. (2003)