

Accelerating Peer-to-Peer File Sharing with Social Relations: Potentials and Challenges

Haiyang Wang
Simon Fraser University
British Columbia, Canada
Email: hwal17@cs.sfu.ca

Feng Wang
Simon Fraser University
British Columbia, Canada
Email: fwal@cs.sfu.ca

Jiangchuan Liu
Simon Fraser University
British Columbia, Canada
Email: jcliu@cs.sfu.ca

Abstract—Peer-to-peer file sharing systems, most notably BitTorrent (BT), have achieved tremendous success among Internet users. Recent studies suggest that long-term relationships among BT peers could be explored for peer cooperation, so as to achieve better sharing efficiency. However, whether such long-term relationships exist remain unknown. From an 80-day trace of 100,000 real world swarms, we find that less than 5% peers can meet each other again throughout the whole period, which largely invalidates the fundamental assumption of these peer cooperation protocols.

Yet the recent emergence of online social network applications sheds new light on this problem. In particular, a number of BT swarms are now triggered by Twitter, reflecting a new trend for initializing sharing among communities. In this paper, we for the first time examine the challenges and potentials of accelerating peer-to-peer file sharing with Twitter social networks. We show that the peers in such swarms have stronger temporal locality, thus offering great opportunity for improving their degree of sharing. We further demonstrate a practical cooperation protocol that utilizes the social relations. Our PlanetLab experiments indicate that the incorporation of social relations remarkably accelerates the downloading time.

I. INTRODUCTION

Peer-to-peer file sharing systems, in particular, BitTorrent (BT), have achieved tremendous success among Internet users. To ensure that the system grows organically, the existing BT protocol relies on a Tit-for-Tat mechanism to penalize free-riders [1]. This incentive mechanism deals well with certain peers' selfish behaviors, and has indeed become a key factor toward the prevailing popularity of BitTorrent. Unfortunately, it also hinders decent peers or peers of close relations from more efficient collaboration. Recent studies suggest that long-term relationships among certain BT peers could be explored to achieve better sharing efficiency [2][3]. However, whether such long-term relationships exist and how could they be effectively identified remain unknown. We have collected trace-data from more than 100,000 real world swarms spanning over 80 days. We find that peers' online patterns in conventional BT swarms are highly diverse: less than 5% peers can meet each other again in our entire measurement duration¹. As such, the peers hardly have a chance to help each other, implying the cooperation protocols that blindly assume the existence of long-term relations may not work well.

¹Note that the peers are not necessarily online at the same time even they are downloading identical contents in the same swarms.

The recent emergence of online social network applications, for example, Facebook [4] and Twitter [5], sheds new light into this problem. Such applications have been quickly changing the Internet users' behaviors as well as the content sharing landscape. In particular, we have noticed that a number of BT swarms are now triggered by Twitter, reflecting a new trend for initializing sharing among communities. In our 80-day trace, we found that there are 2,106 Twitter-triggered swarms among the 100,000 real world swarms, and its percentage steadily grows in our more recent data (as we finished this paper, it reached above 7%).

In this paper, we for the first time examine the challenges and potentials of accelerating peer-to-peer file sharing with social networks, particularly *Twitter-trigger BT swarms*, whose downloads are initialized/shared in Twitter communities. We show that, in these swarms, the peers' online periods are much better overlapped. In particular, more than 35% peers can meet each other again, thus being able to perform data exchange. A closer look into individual peers suggests that a number of peers indeed exhibit very similar online patterns. This temporal locality partly reflects their common social interests, and offers a great opportunity for improving their degree of sharing. We further demonstrate a practical cooperation protocol that utilizes the social relations. Preliminary PlanetLab experiments indicate that the incorporation of social relations remarkably accelerates the downloading time for BT peers.

The rest of this paper is organized as follows: In Section II, we present the related works. Based on the measurement of real world Twitter swarms in Section III, we examine three key issues on accelerating peer-to-peer file sharing with social relations in Sections IV, V and VI, respectively. Finally, the paper is concluded in Section VII.

II. RELATED WORKS

There have been numerous studies on the implementation, analysis, and optimization on peer-to-peer file sharing, particularly on BitTorrent [6]. To deal with certain peers' selfish behaviors, BitTorrent introduces the Tit-for-Tat incentive mechanism, which largely prevents a peer from free riding [7]. The effectiveness of Tit-for-Tat has been evaluated through both theoretical analysis and practical experiments [8] [9]. Fan et al. [10] further proposed strategies for assigning rates to connections, which ensures fairness if universally adopted.

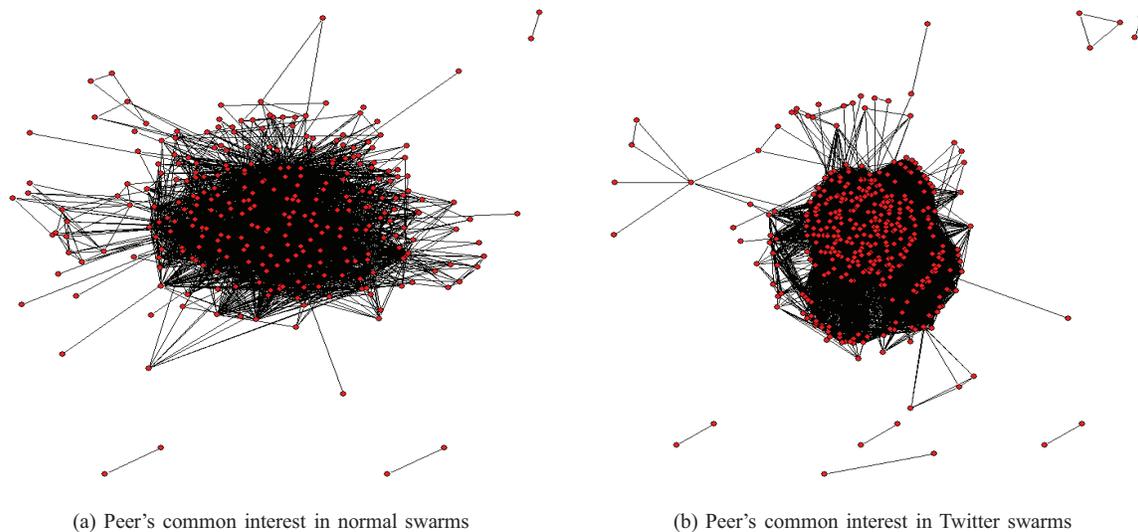


Fig. 1: Sampled graph visualization of: (a) Q_{normal} ; (b) $Q_{twitter}$

Neely et al. [11] explored the utility optimization for dynamic networks with Tit-for-Tat. Unfortunately, recent studies have also identified potential problem in this incentive mechanism, e.g., data pollution [12] and weak robustness [13].

More importantly, it is known that Tit-for-Tat hinders decent peers or peers of close relationship from more efficient collaboration. To address this problem, private torrents that do not solely rely on Tit-for-Tat have been introduced for closed communities [3] [14]. For public torrents, long-term relationships among peers have been explored to improve the content availability [15] and the sharing efficiency [16]. However, whether such long-term relationships exist and how they could be effectively identified and then properly utilized remains unknown, which will be addressed in this paper. We also extend the earlier works on community-based BitTorrent [2] [17] by explicitly incorporating the Twitter communities, a real-world social network that have gained great popularity among BitTorrent users.

III. BACKGROUND AND MEASUREMENT SCHEME

A. BitTorrent and Social Applications

Despite its name, peer-to-peer file sharing is often a solitary pursuit, where the peers swap bits of contents, but each of them remains anonymous to one another. Yet, an increasing number of users as well as the BitTorrent itself is trying to make downloading more socialized by incorporating social relationships [18]. Twitter, one of the most popular social applications on the Internet, has therefore attracted significant attention from BT users and developers². A new feature in the latest version of the uTorrent [19] client called “Torrent Tweets” allows users to talk about a given download from the

²It is worth noting that Twitter itself also highly depends on the BitTorrent to manage its thousands of data servers. The BitTorrent-powered system in Twitter’s new setup has made the Twitter server deployment 75 times faster than before [5].

application and see what everyone else is saying on Twitter. These social functionalities, have already started changing the way of Internet torrents sharing, similar to the increasing video sharing on Facebook [4].

We have found that more than 10,000 groups on Twitter site are built for torrent sharing. It is well known that Twitter emphasizes the up-to-date sharing of instant information among friends. Once a *tweet*(Twitter user) updates a message/torrent link, his/her followers will be able to see it at the same time through updating notifications to their PCs or smartphones. We believe that this feature will potentially change the sharing behavior of peers and is thus worth investigation.

B. Measurement Scheme

To this end, we first collected over 100,000 torrents from *www.btmon.com*, one of the most popular BT torrent sites. We further crawled the Twitter pages from a cluster of servers in our campus to check whether these torrents are also shared among Twitter communities. We found that about 2% (2,106 out of 100,000) of torrents in our dataset are shared on Twitter by Feb 2010, and this ratio has steadily increased to 7% when we finished this paper. For ease of exposition, we call these swarms *Twitter swarms* and others *Normal swarms*.

To learn the peers’ online behaviors in these swarms, we passively monitored the BitTorrent traffic from the out-going switch of a local ISP from Oct 2009 to Jan 2010 (for over 80 days). We generated a tracker list based on the collected torrent files (resulting in 683 active trackers in total). According to this tracker list, we obtained the updating message between these trackers and the peers that are located in this ISP. Based on the torrent information in these messages, we found 334 torrents (10,120 peers) belonging to our Twitter swarm dataset and 2,271 torrents (33,240 peers) belonging to our normal swarm dataset. In other words, the peers in this ISP have participated in 2,605 torrents (out of 100,000) over the 80-day duration.

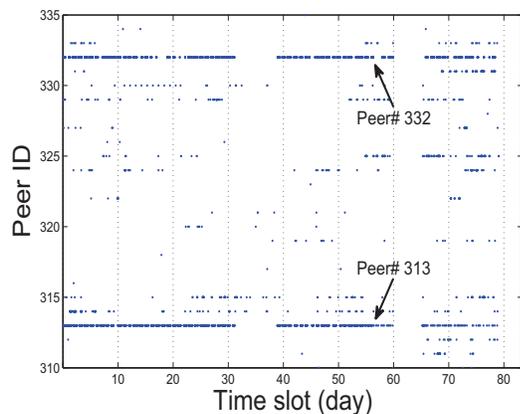


Fig. 2: An illustration to show the existence of highly overlapped peers showing that the peers in Twitter swarms are more tightly related with their common interests.

Considering that many peers may not belong to the ISP that we have measured, we also actively probed the peer information from PlanetLab nodes [20] to obtain more detailed peer information in the swarms. We ran a modified BT client on over 250 PlanetLab nodes, which actively joined the torrents and recorded the observed peer information, as in [21]. As such, we successfully detected the IP addresses of over 95% peers for most of the swarms³. We use this result to infer the common interests among BitTorrent peers.

IV. COMMON INTERESTS AMONG PEERS

We start from examining the peers of commonly interested files. We model the peer relationship across different swarms in a $n \times n$ matrix, Q , where n is the number of peers. Each component of Q , $Q^{i,j}$, is a binary value which indicates whether peers i and j share at least one common torrent (1=yes, 0=no). We use Q_{normal} and $Q_{twitter}$ to record the peer relationships in normal swarms and in Twitter swarms, respectively.

A sample graph visualization of Q_{normal} and $Q_{twitter}$ is presented in Figure 1a and Figure 1b (with 400 sampled peers), where the distance between two nodes corresponds to the number of torrents that the peers shared in common; in other words, peers will be closer to each other if they have downloaded more torrents in common. Here we only plot the peers with a degree being greater than 1, i.e., having relationship with others. We can see that $Q_{twitter}$ is denser than Q_{normal} (with 374 and 291 peers respectively). Intuitively, this implies that more peers in Twitter swarms share clear interests with others and have downloaded at least one torrent in common.

A closer look shows that both graphs are not random, but rather having certain community behaviors. This is quantified by evaluating their *clustering coefficient*⁴. The clustering

³This ratio is calculated by comparing the number of detected peers against the total number of peers advertised by the tracker of each torrent.

⁴The clustering coefficient of node i is the fraction of all possible edges between neighbors of i that are present, while the clustering coefficient of a graph is the average of the coefficient across all nodes [22].

coefficient of $Q_{twitter}$ is over 0.25, whereas the clustering coefficient of Q_{normal} is 0.2. Both of them are noticeably higher as compared to a random graph (nearly 0), and the Twitter swarms exhibit greater communitized behaviors that could be explored.

V. WHETHER PEERS CAN MEET EACH OTHER AGAIN?

Unfortunately, simply having common interests is not enough to enable efficient sharing among these peers. A more critical question is whether these peers have similar/overlapping online patterns. Otherwise, the peers will have no chance to help their friends at all. Figure 2 presents an illustration of two peers with similar online patterns. These two peers join the BitTorrent networks regularly every day and their online time slots are highly overlapped following a clear 7-day pattern. A closer look of Figure 3 showed that peer#313 is not only overlapped with peer#332 but also very likely to meet other peers in its cluster. However, it is not clear that whether such overlapped patterns are pervasive in BitTorrent networks.

We first check the number of encounters between the peer pairs, i.e., how many times a peer's online duration is overlapped with another peer's. From Figure 4, we can see that the peers in normal BT swarms are not likely to meet others again; in particular, less than 5% peers can meet others more than once in the BT networks over 80 days. On the other hand, we observe that peers' online patterns are much better overlapped in the Twitter-triggered swarms. In particular, the ratio is increased from 5% to 35%, indicating that more peers are eligible to provide constant helps to others (we call these peers *socially active* peers as discussed in section I). Note that a study from Piatek et al. [23] shows that the peers can hardly have direct data exchange with others again (be assigned as neighbors again). Our study is, however, focusing on peers' online patterns and seeking for the potential to build direct data exchange among social friends.

Given this higher encounter ratio, we further investigate the total length of the overlapped time slots in Twitter swarms. As shown in Figure 5, we can see that most (around 60%) peers overlapped with others for more than 15 hours in our measurement. This is relatively a long time that could be utilized for effectively exchanging data. An intuitive explanation of Figure 4 and Figure 5 is that Twitter emphasizes the up-to-date sharing of instant information among friends. Once a tweet updates a message/torrent, his/her followers will be able to see this message at the same time (through updating notifications) and then start to download. Therefore, the peers are very likely to share common interests and to have very similar online patterns. Since the Twitter communities consist of largely trusted friends, a better sharing incentive can naturally be expected.

VI. CAN SOCIAL NETWORKS ACCELERATE CONTENT SHARING?

In this section, we will discuss the performance gain of the peer incorporation based on social relationship. A possible

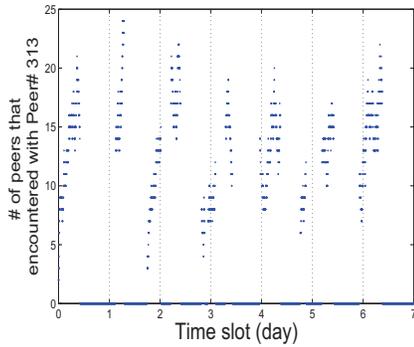


Fig. 3: # of peers that encountered with peer# 313

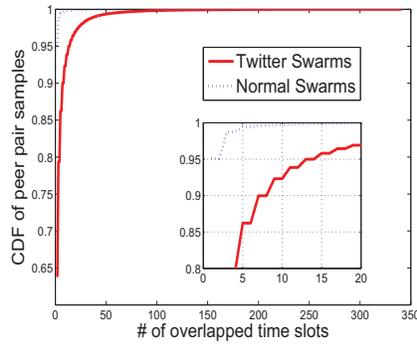


Fig. 4: # of peer encounters

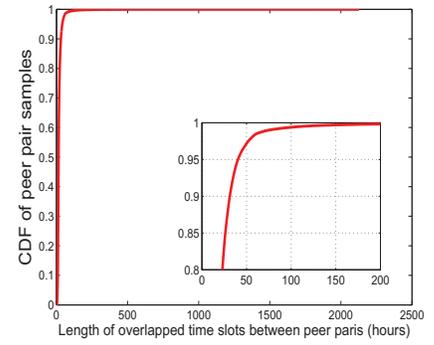


Fig. 5: Length of overlap time slots between peer pair samples (in Twitter swarms)

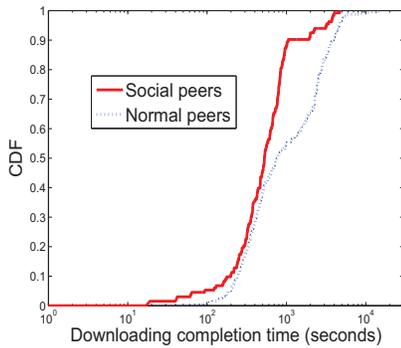


Fig. 6: Downloading completion time

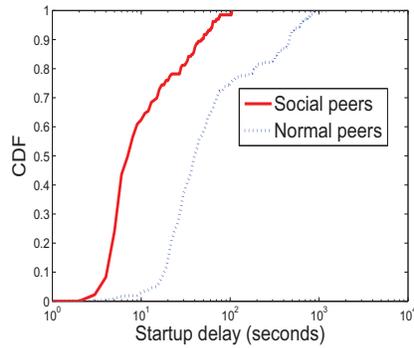


Fig. 7: Startup delay

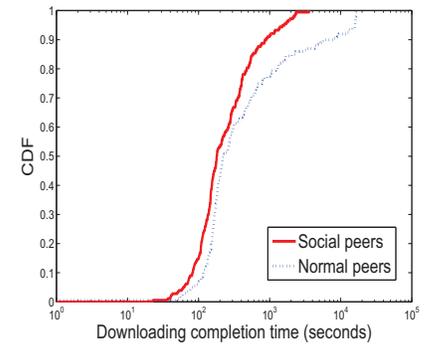


Fig. 8: Downloading completion time (larger swarm)

social network based protocol is proposed and evaluated through preliminary Planet-lab experiment based on our trace.

A. Collaboration Among BT Peers: A Simple Solution

We assume that peers' social relationships can be obtained by the trackers, and the trackers will select the majority, but not all, of the peers' social friends to build the peers' neighbor lists (with a maximum of 8 social friends out of 10 neighbors in our design).

The standard choking algorithm is designed by only changing who's choked once every 10 seconds. This is processed by unchoking the 4 peers which it has the best downloading rates from and are interested. If a leecher has completed the downloading (became a seeder) it will use its uploading rate rather than its downloading rate to decide whom to unchoke (note that the optimistic unchoking is not discussed in here).

It is worth noting that for any leecher who wants to fetch data from other leechers, the key requirement is that this leecher should be interested by others. This design guarantees the instant rewards for every bit that the leechers uploaded (except for optimistic unchoking cases), which is considered robust to peers' possible selfish behaviors. However, it also hinders decent peers or peers of close relations from more efficient cooperations; for example, the friend peers in social networks. Therefore, we make a very simple modification to leechers' choking algorithm. In particular, the leechers will use

the uploading rate to choke their social friends (as a seeder in the standard BT protocol). In our design, the leechers will unchoke the 3 peers among its social friends, with which it has the highest uploading rates (in the past 10 seconds), and 1 other peer that has the highest uploading rate.

B. Evaluation

To evaluate the benefit of social network based content delivery, we carry out Planet-lab experiments with a modified version of BT client. Our goal is to investigate the possible gain in an extreme case where all the peers are social friends.

In the first experiment, we investigate the sharing efficacy in two BT swarms S_{social} and S_{normal} (both with 350 peers). S_{social} consists of social friends, and S_{normal} consists of normal peers. The standard BT protocol is applied to the clients in S_{normal} , whereas our modified choking algorithm is applied to the clients in S_{social} . The content size is 900MB with the piece size of 1024KB. We use a local server in our campus network to run both the seeder and tracker functions, and the seeder's maximum uploading capacity is set to 10M. There are 350 peers arriving over a very short period of 2 minutes. Note that the peers in S_{normal} will leave the swarm as soon as they finished the downloading. On the other hand, the social peers will continue to contribute their uploading if their friends are still downloading the content (this is further discussed in section VIII).

Figure 6 shows the download completion time of swarms S_{social} and S_{normal} . It is easy to see that the social-relation-based enhancement significantly improves the peers' download completion time. In S_{social} , 70% peer will finish their downloading within 800 seconds, and the maximum downloading completion time is 4,000 seconds. In S_{normal} , only 40% peer can finish their download within 4,000 seconds and the maximum downloading completion time reaches over 20,000 seconds. Figure 7 further shows that the startup delay (the delay till receiving the first data piece) is also greatly improved. In our new protocol, most peers (90%) in S_{social} receive their first piece within 1 minute. Yet only 60% peers in S_{normal} can achieve this speed with the conventional optimistic uncorking. We believe that it is because the peers' average downloading rate is greatly improved with the social-relation-based enhancement.

We have also examined their performance in larger swarms, and a typical result for a 550-peer and 4-seeder system is shown in Figure 6. It is easy to see that the peers in both swarms benefit from the increasing number of seeders. Since the seeders are not selfish and act like "common friends" to all the peers, when we keep increasing the number of seeders in the swarm, the downloading performance of S_{social} and S_{normal} will become closer. However our experiments show that, the peers' downloading completion time in S_{social} remains much faster than S_{normal} .

VII. CONCLUSIONS

In this paper, we for the first time examined the challenges and potentials of accelerating peer-to-peer file sharing with Twitter social networks. Our trace analysis showed that the BitTorrent system has enough potential to apply social network based enhancements. The PlanetLab experiments further indicated that the incorporation of social relations remarkably accelerates the downloading time. Given the growing trend of spreading torrents through social networks, we believe that there is a great opportunity to improve the data distribution efficiency in peer-to-peer file sharing systems, which is worth further exploration.

ACKNOWLEDGEMENT

This research was supported by a Canadian NSERC Discovery Grant, a Discovery Accelerator Supplements Award,

an NSERC Engage Grant, a MITACS Project Grant, and a China NSFC Major Program of International Cooperation Grant (61120106008).

REFERENCES

- [1] T. Locher, P. Moor, S. Schmid, and R. Wattenhofer, "Free Riding in BitTorrent is Cheap," in *Proc. ACM HOTNETS*, 2006.
- [2] N. Andrade, M. Mowbray, A. Lima, G. Wagner, and M. Ripeanu, "Influences on Cooperation in BitTorrent Communities," in *Proc. ACM P2PECON*, 2005.
- [3] P. Dhungel, D. Wu, Z. Liu, , and K. Ross, "BitTorrent Darknets," in *Proc. IEEE INFOCOM*, 2010.
- [4] Facebook. [Online]. Available: <http://www.facebook.com/>
- [5] Twitter. [Online]. Available: <http://twitter.com/>
- [6] D. Qiu and R. Srikant, "Modeling and Performance Analysis of BitTorrent-Like Peer-to-Peer Networks," in *Proc. ACM SIGCOMM*, 2004.
- [7] B. Cohen, "Incentives Build Robustness in BitTorrent," in *Workshop on Economics of Peer-to-peer Systems 2003*.
- [8] R. Axelrod, "The Evolution of Cooperation," in *Basic Books*, 1985.
- [9] A. Bharambe, C. Herley, and V. Padmanabhan, "Analyzing and Improving a BitTorrent Networks Performance Mechanisms," in *Proc. IEEE INFOCOM*, 2006.
- [10] B. Fan, D.-M. Chiu, and J. Lui, "BitTorrent-like File Sharing Protocol Design," in *Proc. ICNP*, 2006.
- [11] M. J. Neely and L. Golubchik, "Utility Optimization for Dynamic Peer-to-Peer Networks with Tit-For-Tat Constraints," in *Proc. IEEE INFOCOM*, 2011.
- [12] N. Liogkas, R. Nelson, E. Kohler, and L. Zhang, "Exploiting BitTorrent for fun (but not profit)," in *Proc. USENIX IPTPS*, 2006.
- [13] M. Piatek, T. Isdal, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "Do Incentives Build Robustness in BitTorrent?" in *Proc. USENIX NSDI*, 2007.
- [14] M. Meulpolder, L. D'Acunto, M. Capota, M. W. and J.A. Pouwelse, D. Epema, and H. Sips, "Public and Private BitTorrent Communities: A measurement Study," in *Proc. USENIX IPTPS*, 2010.
- [15] L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, and X. Zhang, "Measurements, Analysis, and Modeling of BitTorrent-like Systems," in *Proc. ACM/USENIX IMC*, 2005.
- [16] M. Piatek, T. Isdal, A. Krishnamurthy, and T. Anderson, "One hop Reputations for Peer to Peer File Sharing Workloads," in *Proc. USENIX NSDI*, 2008.
- [17] D. Choffnes, J. Duch, D. Malmgren, R. Guerm, F. Bustamante, and L. A. N. Amaral, "Strange Bedfellows: Community Identification in BitTorrent," in *Proc. USENIX IPTPS*, 2010.
- [18] BitTorrent. [Online]. Available: <http://www.bittorrent.com/>
- [19] uTorrent. [Online]. Available: <http://www.utorrent.com/>
- [20] Planetlab. [Online]. Available: <http://www.planet-lab.org/>
- [21] J. Liu, H. Wang, and K. Xu, "Understanding Peer Distribution in Global Internet," *IEEE Network Magazine*, 2010.
- [22] D. Watts and S. Strogatz, "Collective Dynamics of Small-world Networks," *Nature*, 393(6684):409, 1998.
- [23] M. Piatek, T. Isdal, A. Krishnamurthy, and T. Anderson, "One hop Reputations for Peer to Peer File Sharing Workloads," in *Proc. USENIX NSDI*, 2008.