Assortative Mixing in BitTorrent-like Networks

Fabrício Murai
COPPE/Computer and System Engineering Dept.
Federal University of Rio de Janeiro - Brazil
Email: fabricio@land.ufrj.br

Daniel Ratton Figueiredo
COPPE/Computer and System Engineering Dept.
Federal University of Rio de Janeiro - Brazil
Email: daniel@land.ufrj.br

I. INTRODUCTION

The tendency of having relationships form among similar vertices has been observed in several real networks. This phenomenon, known as *assortative mixing* (AM) [1], is present not only in social networks [2] and has recently been observed in peer-to-peer (P2P) file sharing networks, where similarity is given by the access bandwidth capacity of the peers [3]. Despite the fact that this "clusterization" among peers is a driving force for system scalability, application protocols such as BitTorrent do not explicitly implement mechanisms that lead the system into this state. In this work, we propose a simple model to understand the mechanisms within the BitTorrent protocol that are responsible for driving the system into high levels of AM. Using the model, we investigate the influence that system parameters have on the emergence AM. In particular, we show that the "optimistic unchoke" mechanism of BitTorrent plays a major role in this process.

II. RELATED WORK

BitTorrent is a protocol for peer-to-peer (P2P) file-sharing applications which simultaneously downloads and uploads pieces of a file of interest to the user. A detailed description of this protocol can be found in [4]. Some models have been proposed in the literature to study BitTorrent. The fluid model in [5] was used to study its scalability and performance. A game-theoretic approach for studying the equilibrium of the connections among peers is seen in [6]. However, these models does not take into account neither the dynamic process of changing connections nor the limited knowledge of each peer has about the network. There has also been some work investigating the relationship between BitTorrent mechanisms and its performance [7]. In a recent work, it was empirically observed that the network of peers in BitTorrent exhibits assortative mixing (AM) according to peer bandwidth [3].

In [1], a robust measure for AM is proposed together with a model to capture this phenomenon in networks. Another model for AM is studied in [8]. Nevertheless, in the former model the AM levels are given externally; the latter one only applies to social networks, where AM was first studied [2]. In our model, assortative mixing emerges as a property of a simple and local dynamic process.

III. THE MODEL

We propose a simple model to capture the dynamics of the peer connections in BitTorrent. In BitTorrent, a peer is only aware of the identity of some of the peers in the system, chosen uniformly at random by the tracker when the peer joins the system. We assume a fixed set of peers, namely n peers, and that each peer knows exactly k other peers. Therefore, the relationship given by this knowledge can be well represented by a k-regular random undirected graph G_k with n vertices. A peer can only establish connections to peers it knows. As in BitTorrent, we will assume that each peer establishes exactly c different connections simultaneously. The relationship "is connected to" will be encoded as a directed graph G_c, defined over G_k, with c outgoing edges from each vertex.

The choking/unchoking algorithm in BitTorrent, which guides how connections are established and torn down, is similar to the tit-for-tat strategy, in the sense that it encourages reciprocity among peers. The key idea is to prioritize more valuable relationships, i.e. a peer is willing to serve peers from which it can gain the most. The amount of data carried between peers is a function of the bandwidth of both sending and receiving peers (among other factors). We assume that the uplink and downlink rates are the same, which value is given by a tag associated to each vertex. More generally, this tag could represent the quality of service offered by a vertex in some specific network. Thus, to a vertex i, the value of serving a vertex j that is not serving i, i.e. does not reciprocate, is zero. If j reciprocates, the value of having having a connection with j is $\min\{tag(i), tag(j)\}$, which represents the bottleneck over the path between j and i.

The model dynamics are very similar: a vertex is chosen uniformly at random, and a less valuable relationship in G_c is replaced by a more valuable one. If no such substitution exists, a replacement can still take place with probability p: in this case the least valuable relationship is replaced by another one randomly chosen among the vertices which are not being served. This is similar to the optimistic unchoking in BitTorrent.

Initially, the knowledge graph G_k is built as a k-regular random graph, with n vertices and an integer tag associated to each vertex. This graph remains fixed throughout the dynamic process. At the beginning, each vertex creates c outgoing directed edges to its neighbors in G_k, chosen uniformly at random (this defines G_c). Then, the connections are established and torn down according to the dynamics described above.

The goal of the proposed model is to enable us to investigate how BitTorrent protocol can lead to high levels of assortative mixing (AM) and the influence of system parameters. Even though each vertex has only a partial view of the system and

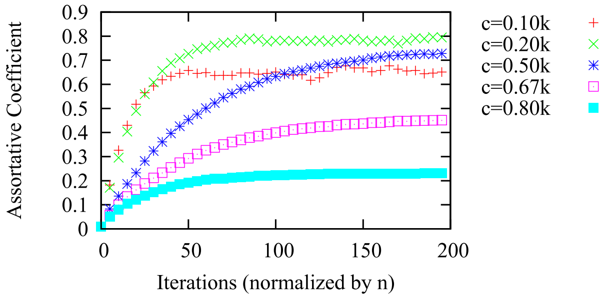

Fig. 1. r as a function of time for different c/k ratios.

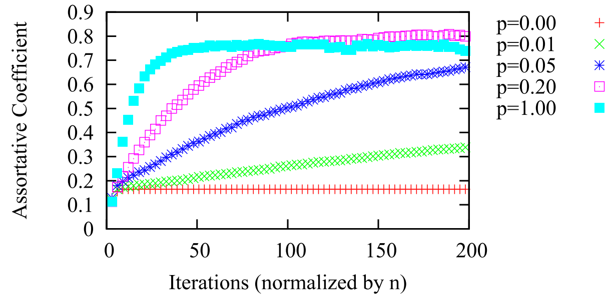

Fig. 2. r as a function of time for different values of p.

makes local greedy decisions, as we will see, the dynamic process does lead the system from complete randomness (initial state) to a network with very high levels of AM, specially when optimistic unchoking is considered.

IV. Experiments and Results

Several simulations of the model above were conducted in order to quantify the AM for different parameter choices. We have calculated the "assortative coefficient" r as defined in [1] for each step of the simulation. This metric is equal to zero when edges are placed independently and uniformly at random across the vertices, and equal to 1 when there are edges only between vertices with the same tag. In the experiments, we have set $n = 1000$, $k = 50$, $c = 10$, $p = 0.40$, half of the vertices have $tag = 1$ and the other half, $tag = 2$ (unless explicitly stated).

We empirically observed that r (which measures AM) does not depend on the number of vertices n in the network, because the knowledge and dynamics of a vertex are completely local. On the other hand, r is strongly dependent on the fraction c/k of neighbors a vertex connects to, as shown in Figure 1. As c/k gets smaller (from $c = 0.80k$ to $c = 0.20k$), higher values of r are obtained. This occurs because it is easier for a vertex to find small groups of vertices having the same tag as c/k decreases. Nevertheless, side-effects of the optimistic unchoking must be taken into account when c is too small ($c = 0.10k$). In this scenario, a vertex with the same tag is constantly being replaced by one with different tag, decreasing r. This decrease can be significant when the number of outgoing edges is small. Notice that in this case, the variability of r is greater.

The optimistic unchoking is essential to reach high levels of assortative mixing. At the beginning, since the assortative coefficient is low, when a vertex with $tag = 2$ does an optimistic unchoking, its least valuable edge is probably an edge to a vertex with $tag = 1$, which will be replaced and then r will be maintained or even increased. Thus, higher values of p make r grows faster. However, when the assortative coefficient is high, edges to vertices with the same tag will be constantly replaced, lowering the achievable r when p is high. The trade-off between faster growth and larger asymptotic values for r can be seen in Figure 2. When $p = 0.01$ the system reaches $r = 0.883$ after 4.2×10^3 iterations. If optimistic unchoking is removed ($p = 0$), the simulation stops after 14 iterations, with $r = 0.165$. Note that when $p = 0$, the simulation will stop when no peer is capable of replacing an existing edge with a more valuable one in G_c.

V. Conclusions

The proposed model for capturing the dynamics of BitTorrent's connection establishment indicates that assortative mixing (AM) can emerge in this network, despite this goal not being explicitly built into the protocol. Through simulations of our model, we observed that the convergence time and the value of AM attained asymptotically strongly depend on the probability of optimistic unchoking. In particular, we note a trade-off on the choice of p: for higher values of p, convergence occurs faster but lower levels AM are achieved.

As observed in previous works [3], the optimistic unchoke is essential for BitTorrent to exhibit good performance. Using the proposed model, we show that optimistic unchoke is also responsible for clustering peers according to their bandwidth, which consequently contributes to high system scalability.

References

[1] M. E. J. Newman, "Mixing patterns in networks," *Phys Rev E*, vol. 67, February 2003. [Online]. Available: http://arxiv.org/abs/cond-mat/0209450

[2] S. K. J. A. Catania, T. J. Coates and Fullilove, "The population-based AMEN (AIDS in Multi-Ethnic Neighborhood) study," *Am. J. Public Health*, vol. 82, pp. 284–287, 1992.

[3] A. Legout, N. Liogkas, E. Kohler, and L. Zhang, "Clustering and sharing incentives in bittorrent systems," in *Proc. of ACM SIGMETRICS' 07*. New York, NY, USA: ACM Press, 2007, pp. 301–312.

[4] B. Cohen, "Incentives build robustness in bittorrent," 2003. [Online]. Available: http://www.bittorrent.com/bittorrentecon.pdf

[5] D. Qiu and R. Srikant, "Modeling and performance analysis of bittorrent-like peer-to-peer networks," in *Proc. of SIGCOMM '04*. New York, NY, USA: ACM Press, 2004, pp. 367–378.

[6] G. Neglia, G. L. Presti, H. Zhang, and D. Towsley, "A network formation game approach to study bittorrent tit-for-tat," in *EuroFGI International Conference on Network Control and Optimization*, June 2007.

[7] A. Legout, G. Urvoy-Keller, and P. Michiardi, "Rarest first and choke algorithms are enough," in *Proc. of ACM SIGCOMM/USENIX IMC'2006*, Rio de Janeiro, Brazil, October 2006, pp. 25–27.

[8] A. Grabowski and R. A. Kosiński, "Evolution of a social network: The role of cultural diversity," *Phys Rev E.*, vol. 73, 2006.