Data, knowledge, and common sense reasoning

» **Ali Ghodsi**
KTH/Royal Institute of Technology / aligh@kth.se

Bernardo Huberman
Fellow y Vice President, Core Innovation en CableLabs, Stanford University / bernardo.huberman@hp.com

Fang Wu
HP Labs / fang.wu@hp.com

Abstract

We explore the use of pari-mutuel markets in a peer-to-peer setting to generate a wide diversity of content offerings while responding adaptively to customer demand. Files are served and paid for through a parimutuel market similar to that used for betting in horse races and in lotteries. Our simulations are based on rational agents, which all act according to a set of simple rules. The results show that a favorite-longshot bias occurs, where agents tend to bet on longshots rather than favorites when following simple expected utility maximization. Furthermore, we have confirmed that the long-tail does sustain even when the agents only have a limited view of all files to pick from. If the limited view consists of random subsets of all files, the long tail is enhanced. If the limited view consists of the top most popular items, the long tail slightly decreases. We have also explored the effect of bounded rationality. Our results show that the system is robust in presence of a large fraction of providers that have bounded rationality. If the providers with bounded rationality pick random items, the long tail is enhanced. Conversely, if the providers with bounded rationality only pick their favorite, the long tail slightly decreases.

I. INTRODUCTION

The distribution of popularity is often characterized by a long-tail, or heavy-tail distribution, such as Zipf, where very few items are immensely popular, while most are not. With the advent of large scale resource sharing systems, a central question is how to give resource providers incentives to provide, not only popular items, but also less popular items. Huberman and Wu [3] proposed a general incentive mechanism for peer-to-peer systems, which gives providers incentives to provide this long tail.

A. Motivation

Previous work has shown that an equilibrium does exist, where resource providers are content with their choices. A question is how such equilibriums can be reached. Moreover, what is the shape of such an equilibrium given that the providers have a certain initial belief about the items. In pari-mutuel horsebetting, a long-tail effect has been observed, and it has even been observed that there is a so called favorite-longshot bias, where players are biased towards longshots. Many researchers have suggested that this has to do with the psychology of bettors, which tend to be risk-loving, or that this occurs especially when betting on horses.

B. Contribution

We show by means of simulation, the equilibria that arise when the providers are simple compute agents. We show that the equilibrium will have a long-tail shape when the players are rational. Furthermore, the favoritelongshot bias occurs even though the players are compute agents that behave according to a simple utility expectation maximizing rule. Hence, the bias occurs without any psychological effects. We then explore the effect of bounded rationality, where providers either randomly bet on files, or are biased towards certain files. Similarly, we explore the effect of providers, which have limited ability in providing files. Our results confirm that all cases lead to a long-tail, with different levels of skewness.

C. Previous Work

We briefly describe the mechanism proposed by Huberman et al. [3]. Each resource provider decides which files to provide, and accordingly allocates bandwidth to each file. Users download files from multiple providers in parallel, much like in popular file-sharing applications, such as BitTorrent [1]. The price of each item is fixed to a constant amount c. Whenever a user downloads a file, it pays c units of money to the providers, giving each provider of the file a fraction of c that is proportional to fraction of the file it provided. If the total amount of bandwidth in the system is normalized to 1, and c =1, then the parallel to a pari-mutuel market, such as the one for horse betting, becomes obvious. Hence, providers bet bandwidth and each download determines a winning file, which results in the downloader proportionally distributing the total bet, 1, to the providers of the winning file.

D. Applications

We envision using the pari-mutuel market in a commercial content distribution network. In such a setting, a central server gathers statistics about files and regularly posts them to the set of content providers. These statistics include which files are the most popular currently and the current availability of the files, as provided by the providers. The providers take the popularity of the files as an input to form a subjective probability distribution over the set of files. Similarly, the current availability of the files serves as the "bets" of the system. Each provider then tries to maximize its utility by choosing which files to make available. When a user downloads a file, it downloads it in parallel from multiple providers, and pays a fixed amount of money proportionally to the providers it downloaded from.

II. SIMULATION RESULTS

This section first introduces the simulation model and thereafter presents the results from the conducted simulation experiments.

A. Model of a Pari-mutuel Market

In this section we briefly describe our simulation model, which is influenced by Eisenberg and Gale [2].

There are m resource providers, and n files or items. Each resource provider has a probability distribution over the items, describing its subjective belief of the probability that a given file will be downloaded. The subjective probability that provider ihas about file j is given by pij. Each provider i has bi amount of bandwidth, such that the total bandwidth in the system is 1, i.e. Pm bi =1.i=1 Player i will reserve bij amount of bandwidth for item j. The total amount of bandwidth in the system reserved on item j is oj, where oj =Pm bij.i=1

The pari-mutuel market gives incentives for providers to not bet on the item that they subjectively think will win, but rather compare it with the current odds such that they bet on items that have the highest ratio between their subjective opinion and the bets on the item. Hence, each provider i will bet its bandwidth bi on the item j that maximizes pij , which we refer to as the provider's ojutility for item j. We also assume that providers reserve all their bandwidth for the first file which maximizes the utility, ignoring other files with the same utility. If provider i has pij =0 and oj =0 for some item j, we assume that the utility is also zero. On the other hand, if oj =0 and pij > 0 for some provider i and some item j, then provider i will reserve its bandwidth on the item k, which has ok =0 and for which its subjective probability is highest, i.e. item k such that pik =max{pij|(8j)oj =0}.

B. Results

As previously mentioned, each provider needs to know everyone else's bet before it can calculate her own utility, and consequently her bet. This creates a circular situation where every provider requires everyone else's bets before betting. We solve this by letting the providers initially bet on a random item.

1) Simultaneous Betting:

A first attempt at simulating the pari-mutuel process is the following simple simultaneous algorithm:

Algorithm 1 Simultaneous Betting

1: Every provider i bets bi on a random item

2: For every item j, update oj =Pm biji=1

3: Every provider change its bets bij to maximize utility

4: Goto 2

However, the simultaneous algorithm need not converge to any equilibrium as seen by the following simple example. Assume there are two items and three 1providers, each with bi =. Assume that the providers 3 111initially randomly pick b11 = 3 , b22 = 3 , b31 = , and 3 2 =1otherwise bij =0. The bets will now be o1 =,o2 .33 Assume pij =0.5for all i and j. The second iteration of the loop, every provider will bet on the second item 3since it gives the highest utility:. The third iteration, 2 we have that o1 =0,o2 =1, which makes all players bet on the first item. Hence, fourth iteration results in o1 =1,o2 =0, and fifth back to o1 =0,o2 =1, which makes the bets oscillate indefinitely.

2) Sequential Betting:

If we instead let each provider bet and update the bets/odds oi between each bet, we get the following sequential algorithm:

Algorithm 2 Sequential Betting

1: Provider i =1bets bi on a random item

2: For every item j, update oj =Pm biji=1

3: i =i+1modm

4: Provider i change its bets bij to maximize utility

5: Goto 2

In our simulations, the above algorithm quickly converges to an equlibrium. At this point, we have no formal proof that this fixed point always exists, and consider the problem out of scope of this report.

As described in the introduction of this report, our goal is to apply pari-mutuel markets in settings which enhance the long tail. But we first start with each provider having a random PDF to see what kind of equilibrium the providers reach. Figure 1 shows the reached equilibrium if there are 500 providers and 50 items. We plot the average of all PDFs with the equilibrium to see how the two differ. Figure 2 shows the same results, buts the 50 items have been ranked according to the popularity of their bets.

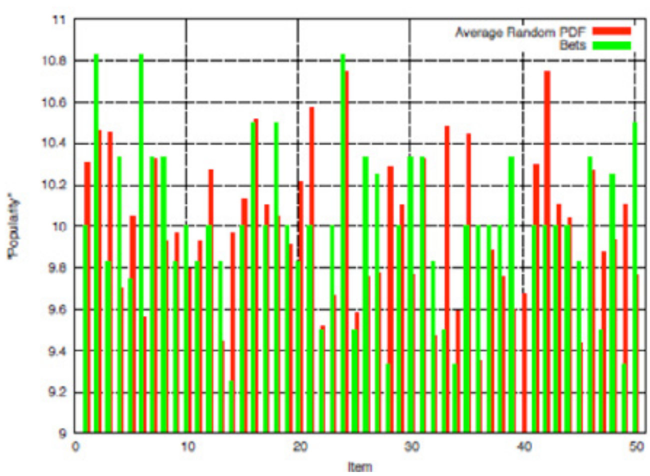

Fig. 1. The equilibrium reached with 50 items and 500 providers with random PDFs. The graph also shows the average of all PDFs.

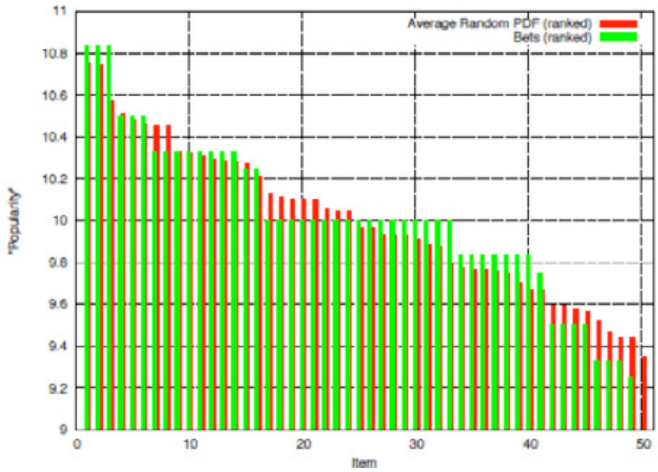

Fig. 2. Same graph as Figure 1, but the items have been ranked according to their bet popularity.

a) Zipf Distribution of Subjective Probabilities:

Our main goal is to apply the pari-mutuel market to a content distribution network. Hence, it is unlikely that the subjective probabilities of the providers is random. Popularity of items is more likely to follow a Zipf distribution, as observed in many different contexts [4]. Hence, we assume that the providers have subjective probabilities that are Zipf distributed. Hence, for N items, the probability of the i:th most popular item is: 1/is p(i)= PN 1/js j=1 The exponent s determines the slope of the line represented by taking the logarithm of both sides of the function. To model the varying risk-averseness of the providers, we choose s randomly between 1.1 to 2.2 with uniform distribution.

b) Perturbing the Subjective Probabilities:

The above described Zipf distributions will vary between different providers due to the parameter s. Nevertheless, the total order or ranking induced from the subjective probability of each provider will be identical, i.e. every provider will have the same item as the one with the highest probability, etc. To introduce perturbations, we run Algorithm 3. The algorithm assumes a subjective ranking function f : Nn+1 ! Nn+1, which ranks the items such that f(1) is the index of the highest ranked item, f(2) the index of the second highest ranked item etc. The algorithm repeatedly chooses a random number r with uniform distribution on [1,n ° 1] and swaps the subjective probability for item f(r) and f(r + 1). This is repeated Æn times.

All our simulations use Æ =0.3.

Algorithm 3 Perturbations

1: for i =1 to m do

2: for j =1 to Æn do

3: r =1+ rand(n ° 1) . r 2 [1,n ° 1]

4: x = f(r) . Index of r'th ranked item

5: y = f(r + 1)

6: swap(pix,piy)

7: end for

8: end for

Figure 3 shows the same experiment as conducted earlier with 50 items and 500 providers, except the providers now have a subjective probability which is Zipf and perturbated with Æ =0.3. Figure 4 shows the same graph ranked, which reveals that the subjective probabilities tend to favor the "long-tail" compared to the average of the probability distributions. This has been observed in real pari-mutuel markets, and is referred to as the favorite-longshot bias [5]. The bias is more clear if plotted in a log-log scale as seen by Figure 5.

c) Uniform Subjective Probability with Average Zipf Probability:

The above experiments surprisingly show that the consensus will be Zipf given that the providers have Zipf distributions that give rise to an average Zipf distribution, even though the subjective probabilities are not identical. A justified question is whether the consensus will always be similar to the average distribution. To investigate this we set up an ex-periment where the providers have a uniform distribution for a small subset of the items, and pij =0 for the rest of the items. However, the probabilities are chosen such that the average of the subjective distributions is Zipf. Hence, each provider has a uniform subjective distribution, while the average distribution is Zipf.

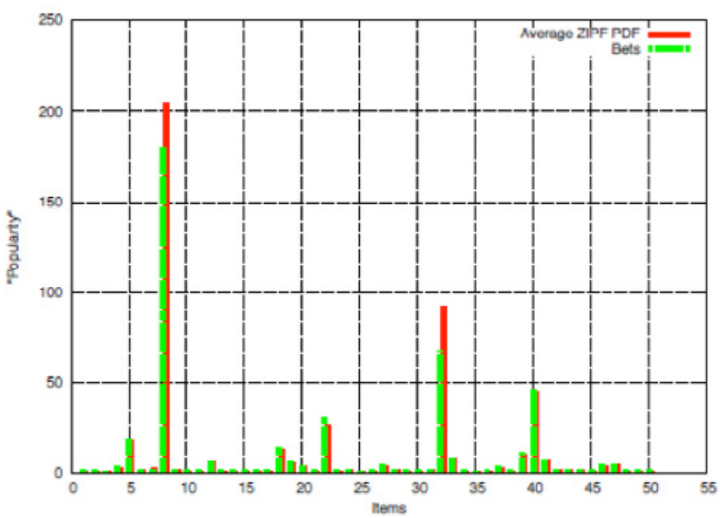

Fig. 3. The equilibrium reached with 50 items and 500 providers
with random Zipf PDFs. The graph also shows the average of all PDFs.

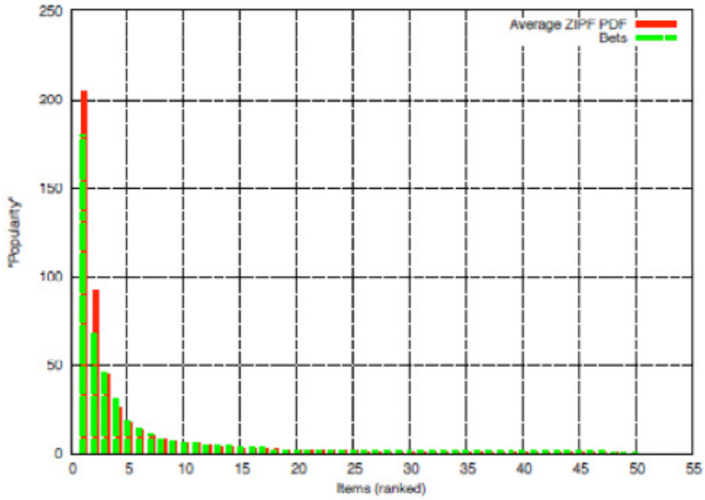

Fig. 4. Same graph as Figure 3, but the items have been ranked according to their bet popularity.

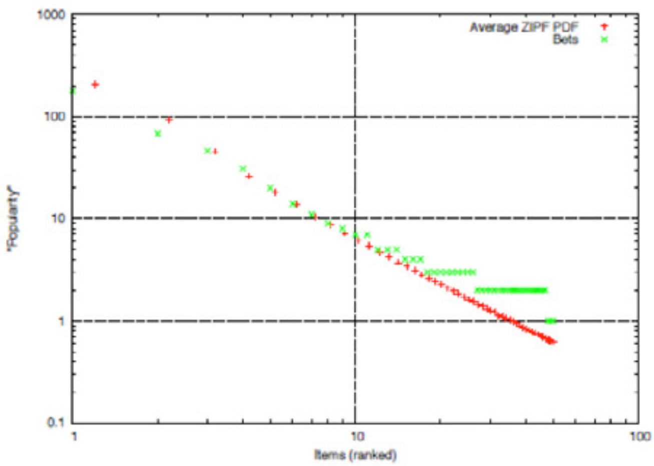

Fig. 5. A log-log scale graph of the graph shown by Figure 4.

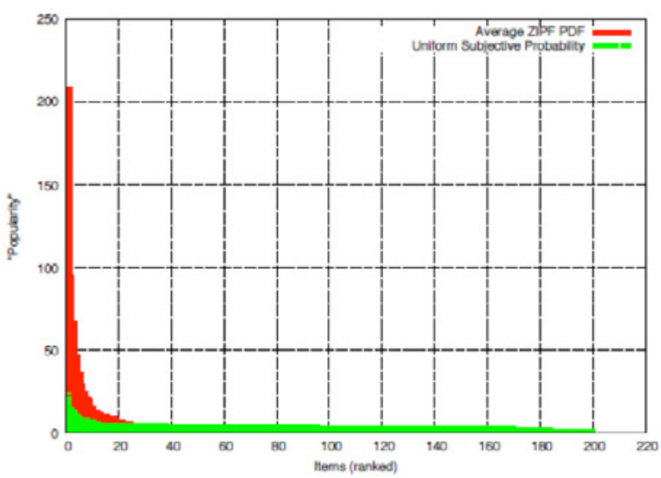

Fig. 6. Uniform Subjective Probability with Average Zipf. The graph shows the equilibrium reached with 200 items and 1000 providers with uniform subjective probabilities. The graph shows that the average of all subjective probabilities is Zipf.

The subjective probabilities were constructed from the same Zipf distribution. Figure 6 shows an experiment with 200 items and 1000 providers. Every provider's subjective probability was set to zero. Thereafter, each provider randomly picked 10 items using a global Zipf distribution, increasing each picked item's subjective probability with 1/10. Hence, each provider ended up with a subjective probability distribution that was close to uniformly distributed. As shown by Figure 6, the consensus is almost uniformly distributed, even though the average PDF is Zipf distributed. The slight deviation from the uniform distribution in the consensus is due to some items being picked several times. Hence, since the subjective probabilities are not exactly uniformly distributed, nor is the consensus.

d) Bounded Rationality — Random Bias:

To make the above model more realistic, we explore different types of bounded rationality. We first assume that a number of the providers pick random elements. Figure 7 shows the same graph as in Figure 4, except 30% of the providers have bounded rationality and bet on random items. The graph shows that this enhances the long tail of the graph.

e) Bounded Rationality — Favorite Bias:

We now explore a different kind of bounded rationality. We assume that a number of the providers have a bias towards their favorites and will pick those items regardless of the bets of the other providers. Figure 8 shows the same graph as in Figure 4, except 30% of the providers have bounded rationality and bet on their favorites. The graph shows that this does not affect the graph considerably. The rational providers will make up for the ones with bounded rationality by "filling" in where the odds are good. Nevertheless, the long tail does decrease slightly.

f) Limited View — Random Items:

It might be infeasible to assume that all providers are able to choose any file. For example, this could be because they only know of a subset of files, or because they do not know the popularity distribution of the bets of all files. We model this by letting each provider only see a fraction of the files, from which it chooses the one which maximizes the provider's utility. The files in each provider's view will be a random subset of all the files.

Figure 10 shows the same graph as in Figure 4, except each provider can only bet on 0.2 fraction of all files. The graph again shows that this will enhance the long tail of the graph. We think that this could be exploited in a system. For example, assume a central server is to provide statistics to the providers, which they use to decide which files to provide. The central server can then choose to only provide statistics for a random subset of all files.

g) Limited View — Top Items:

We now explore the effects of choosing a view, which just consists of the most popular items. Figure 9 shows the same graph as in Figure 4, except each provider can only bet on 0.2 fraction of all files. The graph again shows that this does not affect the bets considerably. The rational providers will make up for the ones with bounded rationality.

Revista Abierta de Informática Aplicada/vol 2 n1 (2018): 23-32
ISSN 2591-5320

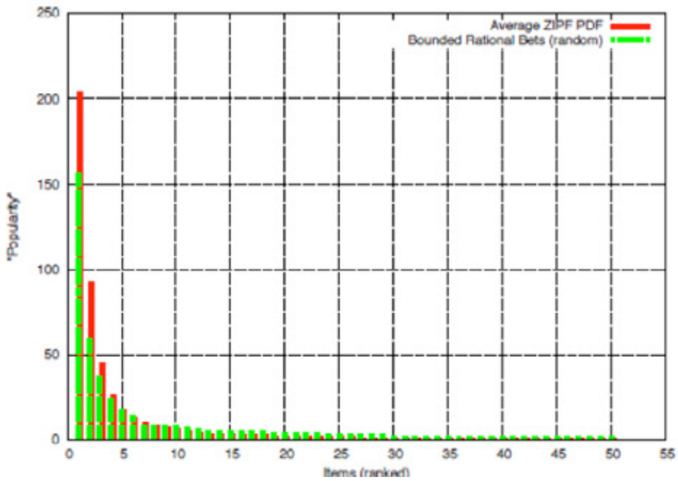

Fig. 7. Bounded Rationality. 30% of the providers have bounded rationality and bet on a random item. The graph shows the equilibrium reached with 50 items and 500 providers with Zipf PDFs. The graph also shows the average of all PDFs. Items have been ranked according to their bet popularity.

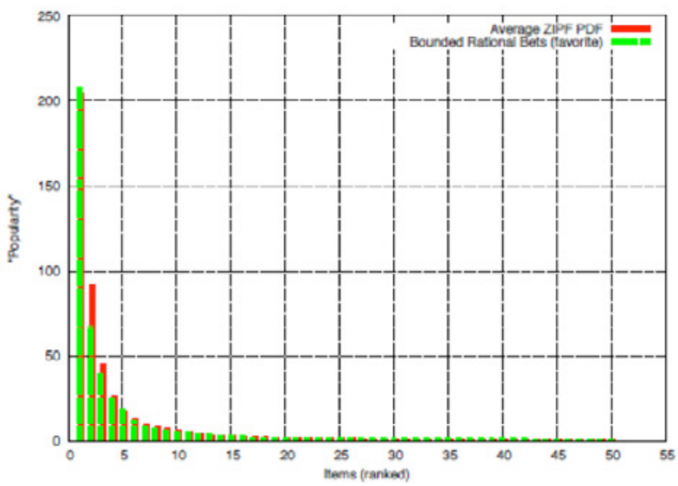

Fig. 8. Bounded Rationality. 30% of the providers have bounded rationality and bet on their favorite item. The graph shows the equilibrium reached with 50 items and 500 providers with Zipf PDFs. The graph also shows the average of all PDFs. Items have been ranked according to their bet popularity.

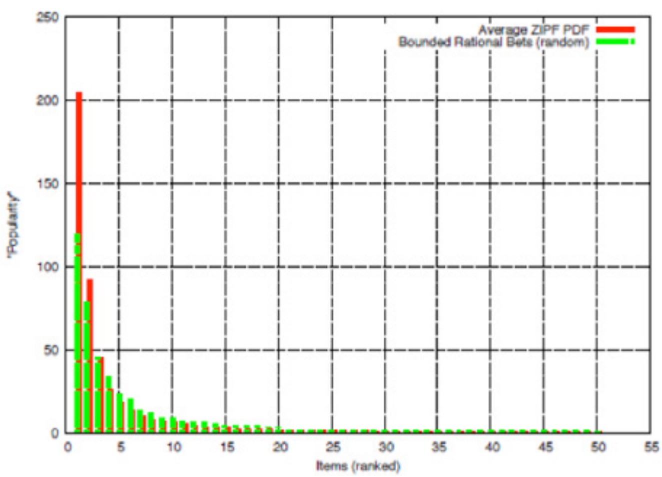

Fig. 9. Limited View -Random. The providers can only choose from 30% of the files. The view is picked randomly. of the providers have a random limited view of all files. The graph shows the equilibrium reached with 50 items and 500 providers with Zipf PDFs. The graph also shows the average of all PDFs.

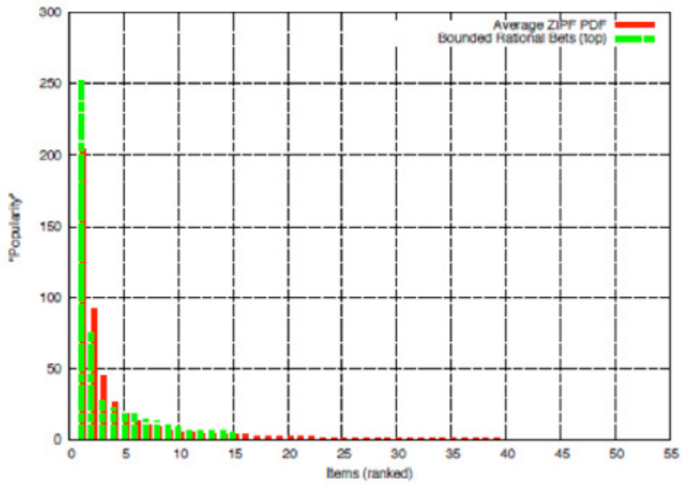

Fig. 10. Limited View -Top. The providers can only choose from 30% of the most popular files. The graph shows the equilibrium reached with 50 items and 500 providers with Zipf PDFs.

III. CONCLUSION

We have explored the use of pari-mutuel markets in a P2P setting. Our simulations are based on rational agents, which all act according to a set of simple rules. The results confirm that there is a favorite-longshot bias even when the agents' subjective probabilities have non-identical Zipf distributions that give rise to an average Zipf distribution. When the agents have uniform subject probabilities with a Zipf average, the consensus will be lie between the average Zipf distribution and the uniform distribution. Furthermore, we have confirmed that the long-tail does sustain even when the agents only have a limited view of all files to pick from. If the limited view consists of random subsets of all files, the long tail is enhanced. If the limited view consists of the top most popular items, the long tail slightly decreases. We have also explored the effect of bounded rationality. Our re-sults show that the system is robust in presence of a large fraction of providers that have bounded rationality. If the providers with bounded rationality pick random items, the long tail is enhanced. Conversely, if the providers with bounded rationality only pick their favorite, the long tail slightly decreases.

REFERENCES

» [1] Cohen, B. (2003). "Incentives Build Robustness in BitTorrent". In *First Workshop on Economics of Peer-to-Peer Systems*, Berkeley, CA, USA, June.

» [2] Eisenberg, E. and Gale, D. (1958). "Consensus of Subjective Probabilities: The Parimutuel Method". *Annals of Mathematical Statistics*.

» [3] Huberman, B. A. and Wu, F. (2006). "Bootstrapping the Long Tail in Peer to Peer Systems". In *Proceedings of the First Workshop on Economics of Networked Systems* (NetEcon'06), Ann Harbor, Michigan, USA, June.

» [4] Sripanidkulchai, K. (2001). *The popularity of gnutella queries and its implications on scalability*, http://www.cs.cmu.edu/ªekunwadee/research/p2p/gnutella.html.

» [5] Thaler, R. H. and Ziemba, W. T. (1988). "Parimutuel betting markets: racetracks and lotteries". In *Journal of Economic Perspectives*, 2(2):161–174.