An Analysis of BG's Implementation of the Zipfian Distribution

Jason Yap, Shahram Ghandeharizadeh, Sumita Barahmand

Database Laboratory Technical Report 2013-02

Computer Science Department, USC

Los Angeles, California 90089-0781

{jyap,shahram,barahman}@usc.edu

November 4, 2013

A Introduction

This research note reports on the behavior of BG's implementation of the Zipfian distribution [2] with different exponents that control its degree of skew. We focus on how BG generates requests while preserving the intended distribution. We conclude with an analysis of the degree of skew with the different exponents for the Zipfian distribution.

BG [1] is a benchmark that emulates a member of a social networking site performing an interactive action such as viewing another member's profile, inviting a member to be friends, confirming friendship requests and others. One may utilize a Zipfian distribution to emulate the frequency at which a member issues a social networking action. Assuming M is the number of members, the probability of member i issuing an action is: $p_i(M, \theta) = \frac{\frac{1}{i^{(1-\theta)}}}{\sum_{m=1}^{M}(\frac{1}{m^{(1-\theta)}})}$ where the exponent θ characterizes the Zipfian distribution. The value of θ is greater than zero and less than one. While a low θ value causes one member to issue actions more frequently than the others, a θ value closer to one results in a uniform distribution where one member is as likely as another to issue requests. This is depicted in Figure 1

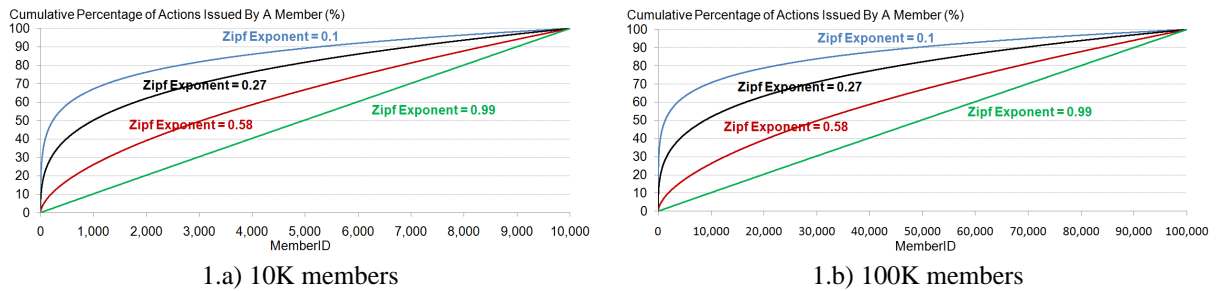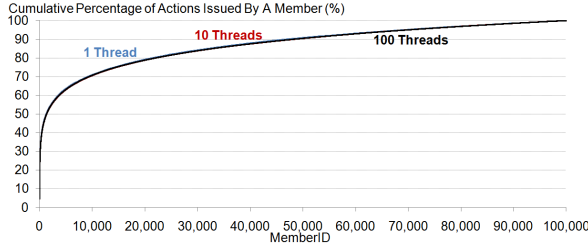
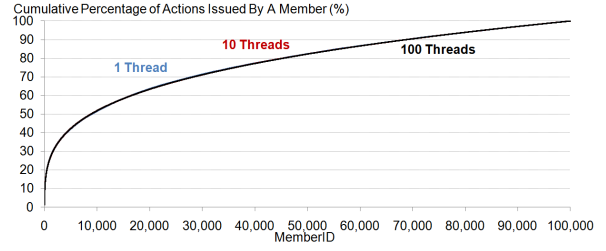1.a) 10K members   1.b) 100K members

Figure 1: Cumulative percentage of actions issued by members of a social graph as a function of the exponent used with the Zipfian distribution.

2.a) Zipf Exponent of 0.1

2.b) Zipf Exponent of 0.27

Figure 2: Cumulative percentage of actions issued as a function of the number of members with three different settings for the number of threads selecting unique members.

Zipfian Exponent	20% of members M issuing requests		40% of members M issuing requests	
	M=10K	M=100K	M=10K	M=100K
0.1	75%	78%	85%	87%
0.27	62%	63%	76%	77%
0.58	40%	39%	58%	59%
0.99	20%	20%	40%	40%

Table 1: Cumulative percentage of BG actions issued by 20% and 40% of members with different exponents for the Zipfian distribution.

that shows the cumulative percentage of actions generated by different members of a social graph consisting of 10K and 100K members. In these experiments, each action has a fixed service time of 1 millisecond. The distribution of requests was measured by post-processing the logs of read requests and generating a histogram of requests made by each MemberID. BG was configured to run with different values for the Zipfian exponent, the size of the social graph and the number of simultaneously executing threads.

In Figure 1, the linear line is generated by the exponent 0.99. It corresponds to a uniform selection of members where one member is as likely as another to generate requests. As the value of the exponent is reduced to zero, the Zipfian distribution becomes skewed and selects some members to issue requests more frequently than others. The knee of the curve is steep when the exponent equals 0.1 such that more than 75% of actions are issued by 20% of the members, see Table 1. (Note that 20% of members generate 20% of actions with the exponent 0.99.)

BG emulates a closed model with multiple concurrent threads issuing requests. Each thread emulates a member of a social networking site. At times, one may want the concurrent threads to emulate unique members at any given instance in time. This means with a load of 100 members (threads), each thread must emulate a unique member at any instance in time. (See [1] for details of how BG does this.) A key question is whether the uniqueness criterion impacts the distribution produced by a given Zipfian exponent. Figure 2 shows the cumulative percentage of actions issued by different members for the exponents 0.1 and 0.27. These results are for 1, 10, and 100 simultaneous threads and show the distribution is not sensitive to how BG guarantees the uniqueness of simultaneous members issuing actions.

Finally, the distribution is the same for different number of members in a social graph. Table 1 shows the percentage of actions issued by 20% and 40% of members with the different Zipfian exponents with a social graph consisting of

10K and 100K members. As expected, these percentages are almost identical for the two social graphs as both the number of members and the raw number of members corresponding to the percentage of members (e.g., 20%) scale. For example, with 20% of members, the total number of members contributing the cumulative percentage is 2K with 10K members and 20K with 100K members.

References

[1] S. Barahmand and S. Ghandeharizadeh. BG: A Benchmark to Evaluate Interactive Social Networking Actions. *CIDR*, January 2013.

[2] S. Barahmand and S. Ghandeharizadeh. D-Zipfian: A Decentralized Implementation of Zipfian. In *ACM SIGMOD DBTest Workshop*, 2013.