TWITTER: PART 1 REPORT

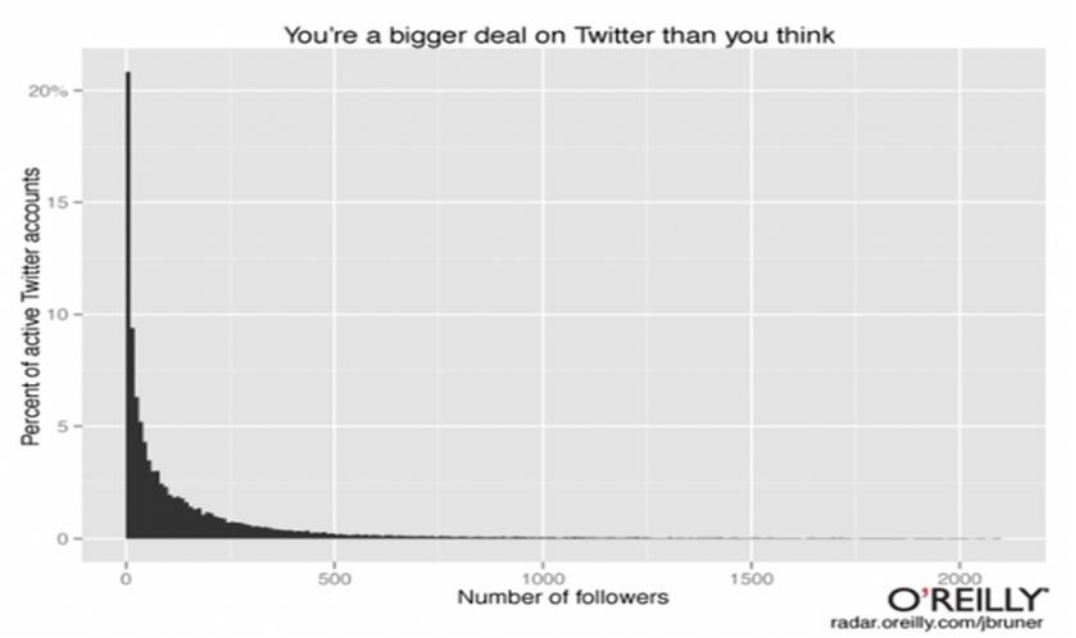
How the user behavior is simulated?

The user behavior is controlled by the following parameters:

1. Next Tweet Interval: Interval after which the user will tweet (default 1 sec)
2. Followers list: The number of followers a user has

We used two methods to generate the followers list for each user:

1. The first one is based on the twitter data of 400,000 users



Percentile of active Twitter accounts	Number of follwers
10	3
20	9
30	19
40	36
50	**61**
60	98
70	154
80	246
90	458
95	819
96	978
97	1,211
98	1,675
99	2,991
99.9	24,964

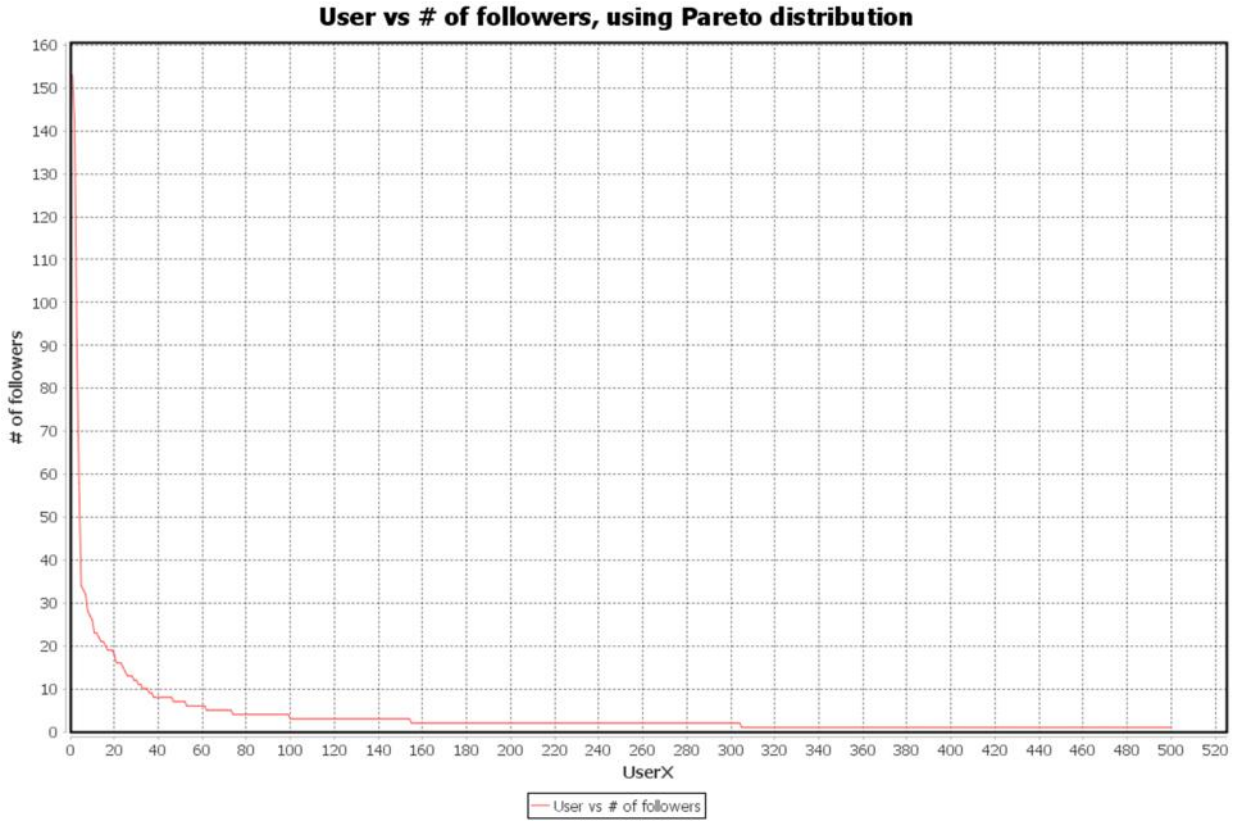
The number of followers is linearly scaled for the number of users in the system which maintaining a lower bound on followers for a user falling a certain percentile.

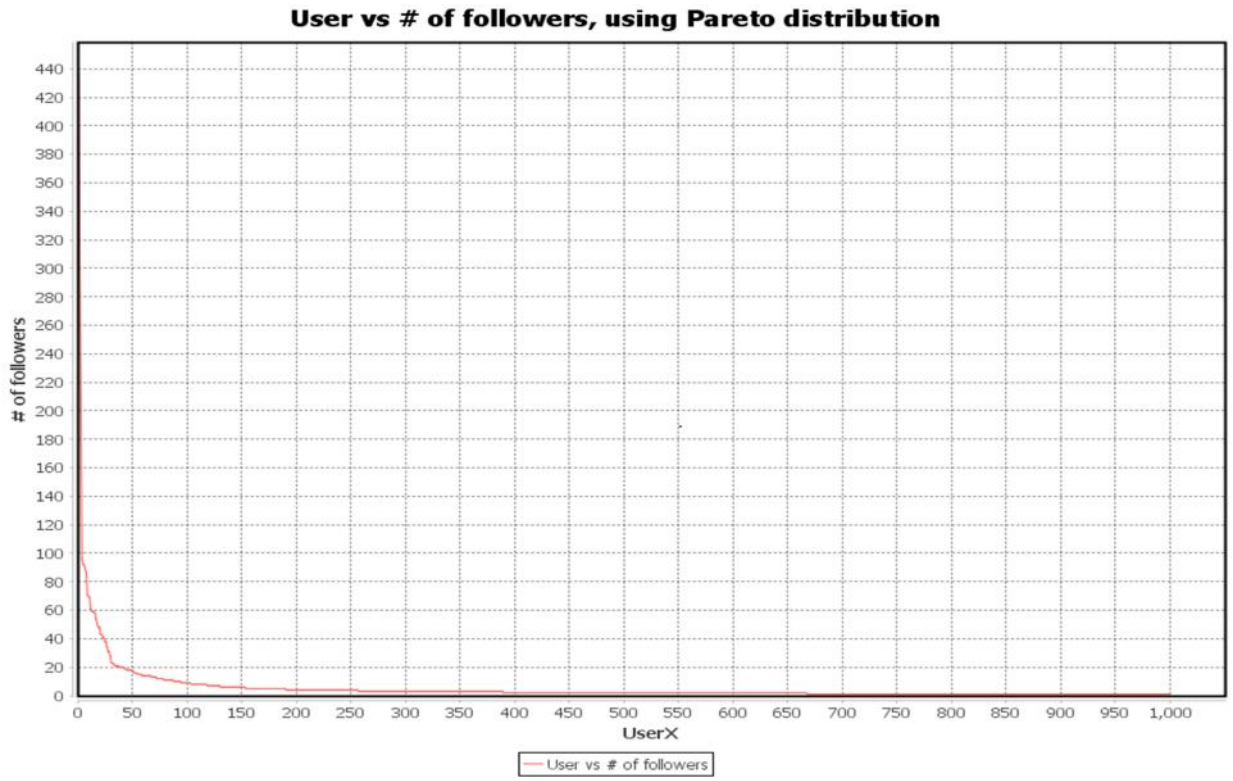
2. For the second method we used Pareto principle, also known as the 80-20 rule, by using the following probability density function of Pareto distribution

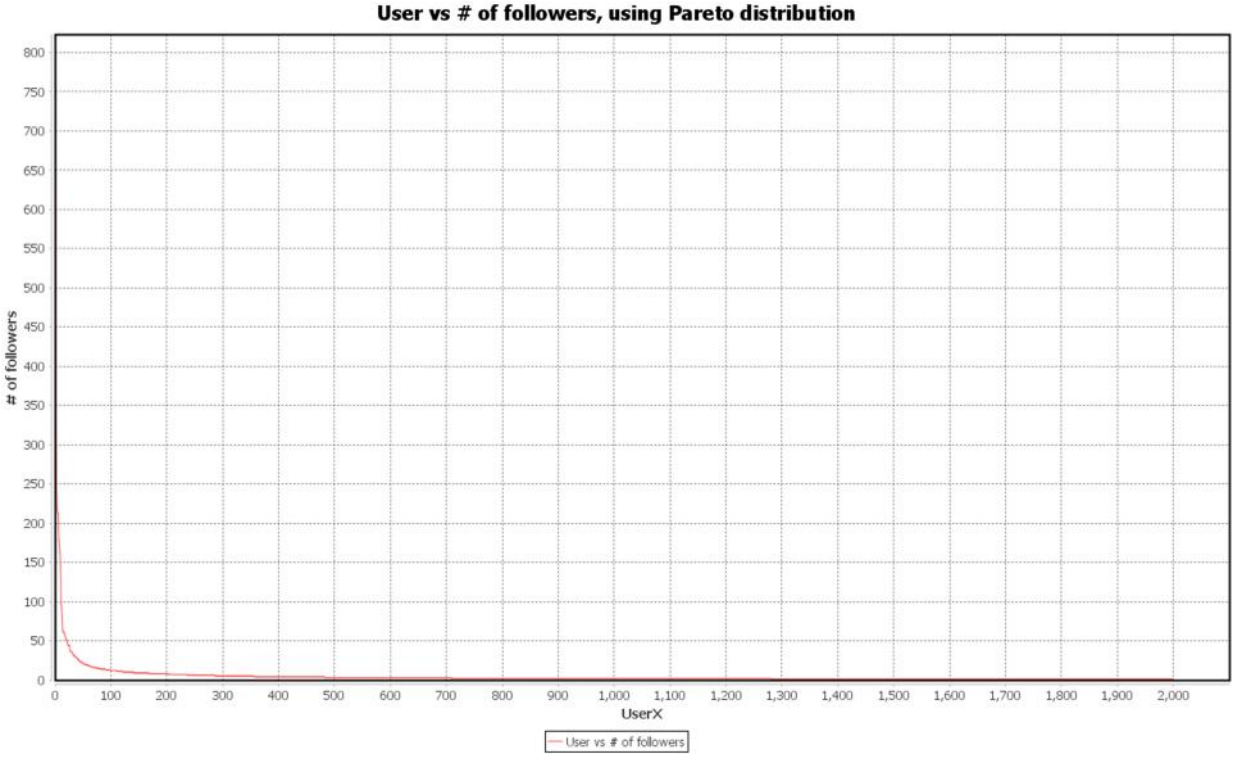
$$f_X(x) = \begin{cases} \frac{\alpha x_{\mathrm{m}}^{\alpha}}{x^{\alpha+1}} & x \geq x_{\mathrm{m}}, \\ 0 & x < x_{\mathrm{m}}. \end{cases}$$

Where $\alpha = \log_4(5)$, and x_m is dependent on the number of users in the system.

Following are the Users vs No. of followers graph generated using 2nd method:







From the graphs, we can see that Pareto distribution almost correctly simulated the followers distribution of users.

Results :

a) Simulation on laptop:

On a single dual core machine, which runs both the server and the client, we were able to simulate up to 20,000 users. The number of tweets received by the server varied from 16000 – 20,000, whereas the number of tweets sent by the server varied from 16000 – 123800.

b) Simulation using three(8 cores) CISE lab machines

No. of users	Peak Achieved	Average	Followers
10000	77000	75000	48000
20000	308562	110000	212060
50000	643000	150000	586262
100000	761043	180000	706262