Do critical users on online social Q&A communities have broader interests in topics compared to others?

Jingyuan Zhou
MACSS, Division of Social Science, University of Chicago
juliazhou@uchicago.edu

Background

Influence has been an important topic for people to understand how and why some innovations or ideas are adopted by larger population faster than others. Many empirical studies using Twitter data have shown the unreliability of determining influential users solely based on their global network measurements. Scholars have found that using different global network statistics gives different results and the combination of prevailing culture could exert more influence than those who have many followers.

Existing literature based on social Q&A websites have mainly focused on collaboration behavior and answer evaluation. In the discussion of roles, previous researches have defined specific roles including questioner, discussion people and technical editor.

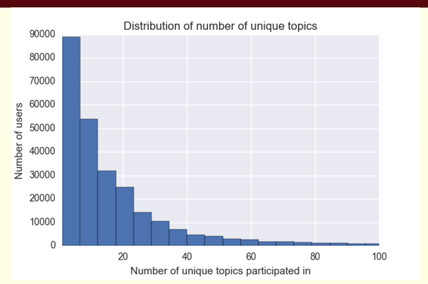
Specific Aims

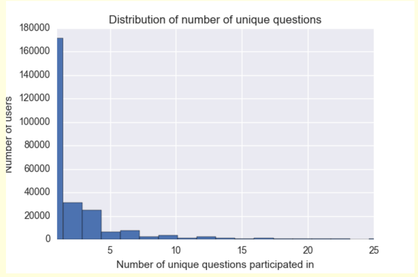
In this research, I define users' roles based on their influence within their own communities under the network structure provided by the following follower functionality. Thus, by traversing activity history of users, I plan to understand if there is a difference between critical users and the others under this definition of user roles. This approach is both new on defining influential users in social network and analyzing social Q&A sites.

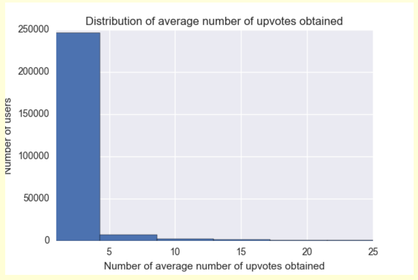
Data

I obtained a collection of data containing 314400 questions and activity history of 261376 users by scraping zhihu.com. Within this data collection, for each question, there is a list of topics that it's tagged with, answers, answerers and number of upvotes of the answers; for each user, there is a list of followers and a list of followings. This dataset is processed so that for each user, I have the total number of unique topics it has participated in, number of questions that it has answered, and the average number of upvotes it gets for each answer.

Distributions







Methodology

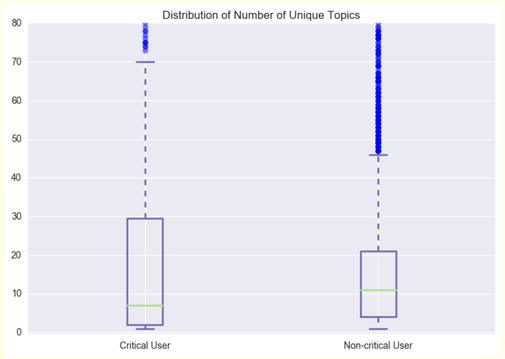
Based on the social network imposed by following, follower lists, I obtained a network of users.

- Community detection through Fastgreedy algorithm
- Within each community, top 10% nodes with highest betweenness are determined as critical users
- Split users into two groups: critical users and non-critical users
- Perform hypothesis testing between the two groups on:
- - number of questions
- - number of unique topics
- - average number of upvotes received per answer

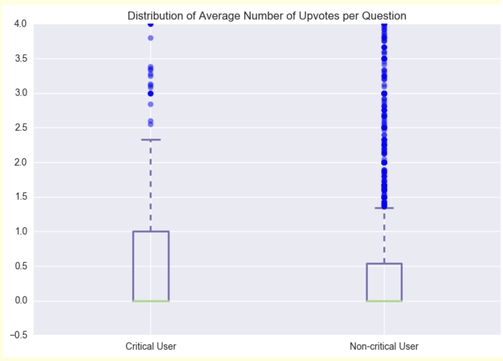
Results

- The network is split into 2 communities.

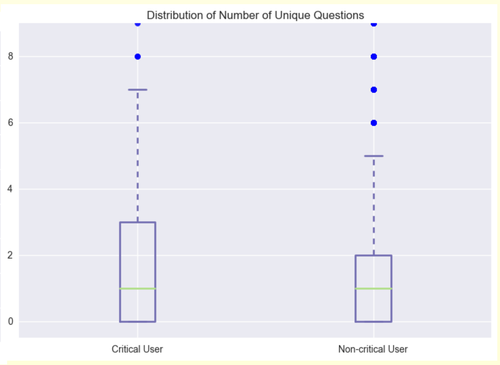
	Community 1	Community 2
Number of Vertices	139184	122192
25% quantile of betweenness	14301	14059.4
median of betweenness	40423.5	34131.5
75% quantile of betweenness	353112.5	74158.5



T – test:
Statistic = 4.223
P-value = 2.85e-05



T – test:
Statistic = 0.113
P-value = 0.910



T – test:
Statistic = 3.204
P-value = 0.001

Conclusion

- Critical users participate in less number of topics than the others, and this finding is statistically significant.
- Critical users participate in higher number of topics than the others, and this finding is statistically significant..
- Disitinction of average number of upvotes reived for each answer between critical users and the others are not statistically significant.

Limitations

- Communities of the network are dependent on fastgreedy algorithm, which sufffer from a resolution limit, i.e. communities below a given size threshoold will always be merges with neighboring communities. Since our determination of critical users depends on these communities, this could affect of result.

References

- H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? pages 591–600. ACM, 2010.
- M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummad. Measuring user influence on twitter: The million follower fallacy. In 4th Int'l AAAI Conference on Weblogs and Social Media, Washington, DC, 2010.
- E. Bakshy, B. Karrer, and A. Adamic, Lada. Social influence and the diffusion of user-created content. In 10th ACM Conference on Electronic Commerce, Stanford, California, 2009.