Reddit.com - A census of subreddits

Richard A. Mills
The Psychometrics Centre
Psychology Department
Cambridge University
rm747@cam.ac.uk

ABSTRACT

The Social News site reddit.com is composed of thousands of independent user-created subreddits where people use the site's submission and voting features in a variety of ways. This paper offers a brief overview of different types of subreddit and how user activity is distributed between these.

1. INTRODUCTION

On reddit users submit posts and comments, and vote on these to determine the prominence with which they will be displayed. Each post must be submitted to a particular subreddit, users choose which subreddits they subscribe to and on aggregated pages (like the 'Front page') they are presented with posts from these subreddits. Any user can create a new subreddit, of which they become the first moderator—giving them the power to define the subreddit's purpose through a description and moderate content which deviates from this. Research about reddit tends to concern a particular subreddit or small selection of subreddits [2, 3], or to consider the site as a whole without reporting differences between subreddits [1]—although [4] is a notable exception.

2. SUBREDDIT CENSUS

On 27th February 2014, 246,542 subreddits had been created[1]. It is very easy for a user to create a new subreddit and many of these are created but seldom used—user activity is concentrated on a small number of subreddits (Figure 1). 241,398 subreddits with less than 1,000 posts each accounted for just 1% of voting activity and 2% of commenting activity on the site. Posts to these subreddits received a median 2 votes and 0.17 comments—these subreddits were active for a median 1.3 weeks.

---

[1]Based on data obtained from redditanalytics.com (all reddit posts from 2008 onwards) in February 2014. A data-set containing all subreddits with the variables utilised in this research is available at: www.researchmills.com/subreddits.csv
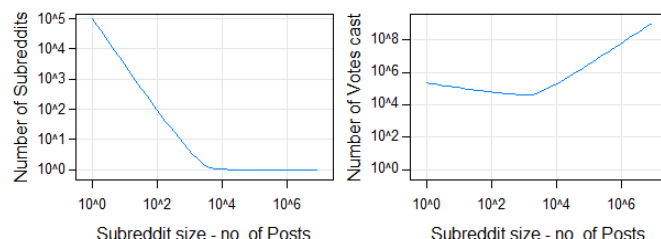
Figure 1: Number of subreddits by total number of post submissions (left), and the manner in which voting activity is distributed between these (right)
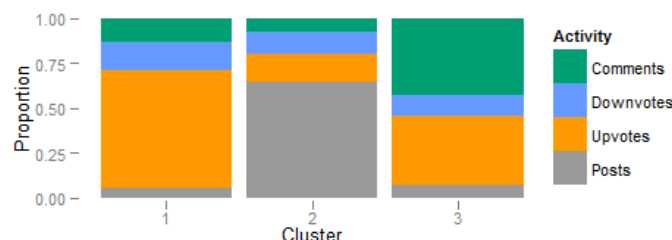

Figure 2: Centroids for 3 k-means clusters based on activity proportions for the 5,144 subreddits with more than 1,000 posts

To explore the nature of interaction on the remaining 5,144 subreddits with more than 1,000 posts each, these were clustered using k-means on proportion of user activity (posting, voting, commenting). Centroid proportions for the 3-cluster solution fitted from 25 random starting points are shown in Figure 2. Voting dominates the first cluster (2,667 cases), the second (384 cases) is characterised by a high proportion of post submissions (65%), the third (2,203 cases) by a high commenting rate (43%). Most of the large default subreddits[2] fall into the first cluster, subreddits in the second cluster can be likened to an RSS feed (some, like /r/risingthreads[3] are automatically populated by a bot and have disabled the voting buttons), whereas those in the third cluster are more similar in use to discussion boards. Some redditors have categorised a range of subreddits[4], and these categories have been used to model the proportion of commenting activity for 992 subreddits. Taking 'Entertainment' as the reference category, subreddits devoted to discussion and particular regions have significantly higher commenting proportions, those devoted to images/gifs, sciences and arts have lower commenting proportions.

---

[2]New users are automatically subscribed to the 'default' subreddits, and users who are not signed into an account see posts from these subreddits on the site's Front page

[3]/r/ is short-hand for a subreddit name, reddit.com/r/risingthreads is the main page for that subreddit

[4]reddit.com/r/ModeratorDuck/wiki/subreddit_classification/

Table 1: Some large subreddits - percentage of posts ignored, post count, vote and comment means

Subreddit	% Ignored Posts	Posts	Votes per post	Comments per post
starcraft	4.8%	294k	84.9	17.2
teenagers	6%	121k	33.8	27.4
AdviceAnimals	7.5%	2686k	231	5.5
AskReddit	10.5%	3614k	32.5	29.6
funny	14.8%	4422k	379.1	7.2
science	36.1%	429k	86.4	9.1
worldnews	42.2%	1122k	75.4	9
technology	50.6%	941k	59.7	6

In February 2014 there were 35 subreddits with over 3 million comments each, together accounting for 51% of all comments, 42% of post submissions and 74% of voting activity on the site. Posts submitted to large subreddits have the potential to reach a very large audience if they outscore the posts they appear alongside. Conversely, many posts slip down the 'New Queue' of these subreddits without ever being voted or commented upon. There is considerable variation between these large subreddits in the percentage of posts which receive no votes or comments—and no clear relationship between a subreddit's activity levels and the percentage of posts which are 'ignored'.

A previous study [3] analysed observations of reddit pages for default subreddits recorded at 30-minute intervals over a one month period in 2012, fitting negative binomial regression models to estimate the level of voting and commenting activity associated with appearance on the New, Rising and 'Hot' pages for each subreddit and for the site's default Front page. The /r/funny subreddit showed the expected stepping-up of voting rates as posts progressed through the New (0.3 votes per minute), Rising(4.1 vpm) and Hot (12.2 vpm) pages towards the Front page (75.7 vpm). Voting rates for /r/worldnews were however negligible on the New (0.01 vpm) and Rising (0.04 vpm) pages, low on the Hot page (0.5 vpm) but still quite high for posts which appeared on the site's Front page (20.2 vpm). This suggests that /r/worldnews subscribers want to be presented with news, and will vote on posts which they see on their Front page, but are less likely to browse and vote on the latest submissions. The New page is the only page which a post is guaranteed to appear on, where it will be pushed down the order by subsequent submissions—the low voting rate for /r/worldnews on this page relative to the number of incoming submissions accounts for the high percentage of posts which are ignored.

Figure 3 illustrates the outcome of these voting rates in relation posts which appeared on the subreddit's Hot page. For /r/funny, voting on the New and Rising pages filtered out the poorest submissions, the final observed scores of posts which appeared on the Hot page have close to a normal distribution. For /r/worldnews this filtering process is not functioning, many of the posts observed on the Hot page finished with small or negative scores. Although the voting rate on the /r/worldnews Hot page was not very high, the best-scoring posts were displayed on the Front page, and once they appeared on the Front page the high voting rate associated with that territory made it difficult for other /r/worldnews posts to displace them. Comparing these two subreddits suggests that /r/funny was making more reliable
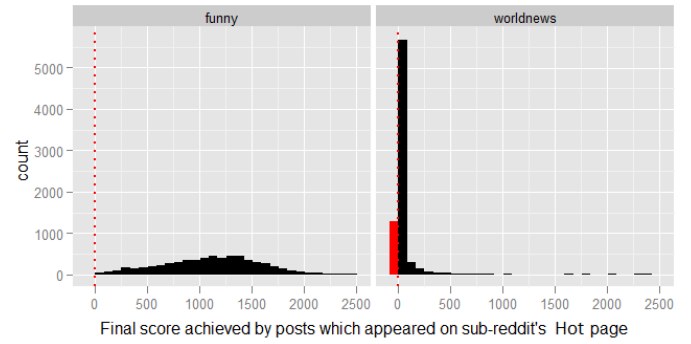

Figure 3: Final observed scores for posts which appeared on the subreddit's Hot page (ranks 1-100)

decisions about which posts to promote to the Front page than /r/worldnews—but still a large number of posts were being quickly filtered out on the New page based on the votes of relatively few users. These results support a previous finding that under-provision of votes on the /r/pics New page resulted in many posts failing when first submitted but later upvoted to the Front page upon subsequent submission(s)[2]—and suggest that the situation is much worse for subreddits like /r/worldnews.

3. CONCLUSIONS

Since reddit began allowing users to create their own subreddits in 2008, over 240,000 have been created. Although most of these new subreddits falter and are soon abandoned, a significant minority are embraced by subscribers and form sustained communities. This paper has considered the population of subreddits in only the broadest terms, it has not attempted to describe the variety of subjects and uses to which reddit's post and comment voting systems have been applied, the inter-subreddit dynamics (like subreddit invasions and voting brigades) which have been unleashed by 'meta' subreddits and cross-posting, or the ways in which reddits' users have tried to mitigate the worst of these effects. As the number of active subreddits continues to grow, each with its own subject and driven by its own independent community—reddit's offer to users shifts away from the once unifying, once unitary, 'Front page of the Internet', and towards a highly customisable 'Daily Me'.

4. REFERENCES

[1] Y. Benkler, H. Roberts, R. Faris, A. Solow-Niederman, and B. Etling. Social mobilization and the networked public sphere - mapping the sopa-pipa debate. *Berkman Center Research Publication*, (16), 2013.

[2] E. Gilbert. Widespread underprovision on reddit. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 803–808. ACM, 2013.

[3] R. A. Mills. A computational study of how and why reddit.com was an effective platform in the campaign against sopa. In *Proceedings of the 2015 conference on Human Computer Interaction*. Springer, in press.

[4] P. Singer, F. Flöck, C. Meinhart, E. Zeitfogel, and M. Strohmaier. Evolution of reddit: from the front page of the internet to a self-referential community? In *Proceedings of WWW'14*, pages 517–522, 2014.