

# **Researching Social News – Is reddit.com a mouthpiece for the ‘Hive Mind’, or a Collective Intelligence approach to Information Overload?**

By Richard Mills

## **Abstract**

Distributed moderation and filtering systems are becoming increasingly common on the web. This paper examines reddit.com, a Social News website where users vote to determine which stories will be featured in high-visibility locations. User voting activity and attention are found to be highly concentrated on the website’s front page. The consequences of this are considered, chiefly by studying reddit’s coverage of a major news story. Minority opinions are found to be slightly marginalised but not excluded from high-visibility locations. The paper concludes by discussing some positive and negative effects of the emphasis reddit places on its front page and the collective decisions about which items appear there.

## **1 Introduction**

The Web allows any connected user to publish information at very little cost, and Social Media have lowered the barriers to entry as an information producer further still. As the quantity of information available online continues to grow rapidly our capacity to navigate this vast web of data becomes increasingly important. Distributed moderation and filtering (or DM) systems appear to offer a unique, people-powered, approach to the problem of Information Overload. The principle behind these systems is that some dimension of user activity is taken as a measure of an item’s quality or relevance; this measure being used to rank and sort items, with those items which rank more highly being placed in positions of greater visibility.

The measures on which items are ranked vary by application – some websites (e.g. Social Bookmarking sites like Delicious.com, recommendation systems such as Amazon's) derive this measure from users’ actions which serve a purpose external to the ranking process (e.g. bookmarking and tagging of items, a customer's purchases). Other websites (including the ‘Social News’ sites considered here) deploy a voting system whereby their users participate directly in ranking submitted items. While the specifics of the voting and ranking procedure differ between Social News (SN) websites all of these have a ‘front page’ where the most popular posts of the moment are displayed, and which acts as a focal point for the website’s users.

### **1.1 The Social Significance of Social News**

SN websites determine what ‘the news’ is in a way which is quite dissimilar to the conventional news media. On SN websites the role of an editor is fulfilled collectively by users who vote. In the same way that an editor selects which of the items produced by their journalists to publish and where to position these; SN users vote to determine which of the submitted items will be allocated high-visibility locations (with placement on the website's front page being of particular importance). In this sense submitting users can be thought of as the journalists of SN, although in practice these users most often submit links to articles they have found rather than produced. SN, as with many other forms of Social media, employs a ‘publish-then-filter’ (Shirky, 2008) system as opposed to the ‘filter-then-publish’ approach of the mass media. On SN websites items

are publicly accessible from the moment of their submission; user voting activity determines the relative salience afforded to each item.

SN websites are also notable for the scale of the interactions they facilitate. In March 2009 68,504 users submitted 150,042 posts and cast 516,775 votes on these posts to an individual 'sub-reddit' on reddit.com. This activity produced a rolling minute-by-minute ranked list of these posts; with the top posts at any given moment being broadcast to a wide audience. Reddit applies the same voting system to the comments on posts; and a popular post can receive thousands of comments, some of these in turn receiving hundreds of votes. The weighting of comments according to voting behaviour has the effect of producing a kind of 'democratically mediated' discussion. Individuals can make their opinions known by voting on existing comments rather than adding their own, and in this way a very large number of people can participate in a discussion without it becoming so large and diffuse that it is difficult to make sense of.

An optimistic appraisal of SN would have it that this user voting activity allows the best items submitted to the website to rise to the most visible locations thereon in a somewhat organic fashion. This paper probes the workings of a SN website to scrutinise these 'organic' processes.

## **1.2 The spread of SN principles**

In addition to the novelty of a 'peer produced' (Benkler, 2006) news resource; SN websites are worthy of study because the principles and mechanisms these websites have pioneered and developed appear to be gaining traction elsewhere – most notably in the domain of politics. AmericaSpeakingOut.com is a website created by the American Republican Party which solicits political policy ideas from its users. These policy ideas are subjected to 'approve/disapprove' voting by other users, with the ideas that receive the greatest levels of approval being shown in the most prominent locations. The intention behind the website was that these most popular ideas would be put forward to Republican leaders and form the basis of policy; but the website appears to have failed in this regard (Benen, 2010; Milbank, 2010).

'Your Freedom' ([www.yourfreedom.hmg.gov.uk](http://www.yourfreedom.hmg.gov.uk)) was a website created by the current British coalition government. Your Freedom also solicited political policy ideas from the general public and allowed users to rate each other's ideas; in this case users were intended to highlight pieces of legislation which should be repealed or submit ideas for the enhancement of civil liberty.

Such applications raise a number of important questions about DM. Do these systems allow 'the best' submitted items to rise to the most prominent locations? The growth of SN websites suggests that their voting systems do allow their community of users to place items that they find interesting/appealing on their front pages; but are these items the best that are submitted, or merely a random selection of those that are good? Ongoing research aims to address this question.

Another question which has particular importance in the domain of politics relates to the treatment of minority opinions. The 'rich get richer' mechanisms which these websites are built on would suggest that for any controversial issue opinions which are opposed to that of the majority will potentially be excluded from prominent locations. Suppose 60% of the user population held some opinion on a controversial issue, if these users down-voted opposing

opinions whenever they encountered them they could presumably filter out these opposing opinions and prevent them from reaching high-visibility locations.

In fact this phenomenon is often discussed (as a facet of the 'hive mind') among the users of Social News. Digg.com, a large SN website, has had problems with this kind of behaviour in the past (Halliday, 2010; Oleoleolson, 2010) - although in this case it was a small but well-organised group of users, the 'Digg Patriots' who were responsible for suppressing liberal content.

In addition to the proliferation of websites which are based on forms of DM, many popular websites (e.g. Youtube, Facebook) now incorporate buttons whereby users can register their like or dislike of a particular item, and display summary statistics for these likes and dislikes. Furthermore, a number of newspapers' websites now incorporate similar mechanisms whereby users can submit comments on articles and 'vote' to register their opinion on these comments.

This paper presents the results of analyses conducted on reddit.com; now the largest SN website. The distribution of user voting activity and attention on the website is explored, and the process whereby an item reaches reddit's front page is considered. There follow details of an ongoing study of reddit's 'coverage' of a major news story – the release of United States' diplomatic cables by the Wikileaks organisation. Of particular interest here is the range of opinions espoused by reddit users, and the treatment of minority opinions through the voting system. The paper concludes with a discussion the 'hive mind' concept on reddit.

## **2 Reddit.com**

Reddit.com was created in 2005 and has seen rapid growth in the last three years; surpassing rival SN website Digg.com in 2010, and exceeding one billion page-views per month in January 2011. Reddit utilises a bi-directional voting system where users can up-vote or down-vote items to increase or decrease their score. An item's score, coupled with a time penalty for older votes, is used to determine the prominence with which the item will be displayed. There are two types of primary units of content (posts) on reddit; link posts and text posts. Link posts are comprised of a title and a hyper-link to some URL; while text posts consist of a title and body of text entered directly on reddit by the submitting user.

### **2.1 Sub-reddits**

Posts on reddit are submitted to a specific 'sub-reddit', and all of the pages which reddit uses to display posts exist for individual sub-reddits as well as in aggregated form. Some of these sub-reddits fit quite neatly with a conventional understanding of 'news' (e.g. the 'worldnews' sub-reddit is usually populated with links to articles on the websites of well-respected news outlets). Other sub-reddits cater to forms of interaction which rely on being situated within a DM environment. For example, on 'AskReddit' text posts are submitted which ask some question of the reddit community; users answer by making comments and vote to determine what the most popular answer is.

Reddit users are free to choose which sub-reddits to subscribe to, and when a user browses aggregated pages (such as the 'front page') they will see only posts from the sub-reddits they have subscribed to. Reddit users are also free to create their own sub-reddits, and there are now thousands of these sub-reddits which deal with particular topics or types of post. The distribution

of subscribers between sub-reddits is highly skewed and roughly approximates a power law; there are a small number of sub-reddits with more than 100,000 subscribers, and there are thousands of sub-reddits with less than 10 subscribers. This federal system allows new communities to be formed easily, and when they are popular with users they can grow quickly and become an integral part of reddit. Conversely, sub-reddits which become problematic can be easily excluded from users' experience of the website.

## 2.2 Pages

Reddit uses a number of page types which select posts for display according to different criteria. The *new* page displays posts according to their submission time in reverse chronological order; when a post is submitted it will appear at the top of the *new* page for its sub-reddit, and will be gradually pushed down (and off) this page as subsequent posts are submitted. The *rising* page shows relatively fresh posts which have attracted early positive votes. The *top* page orders posts submitted within a specified time period by their absolute score. The *controversial* page displays posts which have attracted a similar level of positive and negative votes (indicating that the opinions of users are divided).

The *hot* page is the default method of displaying posts, reddit's 'front page' is the aggregated form of this *hot* page. The *hot* page is also the page displayed upon navigating to a particular sub-reddit; and for this reason it is referred to here as a sub-reddit's 'main page'.

## 2.3 Data collection

The data considered here have been collected through reddit's Application Programming Interface (API). Many Social Media websites offer such an API, with the intention being that independent programmers can write software (e.g. mobile phone applications) which interfaces with the website. Computational restrictions mean that only a small number of sub-reddits can be monitored simultaneously. The *new* page for the sub-reddit is checked at 30 minute intervals and posts which have not previously been observed are added to a 'watch list'. These posts are then tracked at 30 minute intervals, their scores and the pages they appear on are recorded, along with the text and scores of comments on the post.

This paper also makes occasional reference to procedural data provided by reddit's administrators for the purposes of research. This data takes the form of a complete record of voting activity for the month of March 2009 (3.5 million votes). This procedural data had been anonymised with respect to post and user identity.

## 3 How does reddit work?

On SN websites positive votes or ratings for an item are associated with increased visibility; and in turn this increased visibility leads to further voting activity. This 'rich get richer' dynamic produces a highly skewed distribution of votes between items of content. On reddit in the month of March 2009, for which voting data are available, 80% of all the votes which were cast accrued to just 7.4% of the submitted posts. 167,688 posts (47.5%) received no votes; whereas the post with the most votes received 5,997 votes. The power law has been fitted to individual sub-reddits according to the method of Clauset et al. (2009) and was found to offer a plausible fit ( $p < 0.05$ ) for the majority of sub-reddits – with exponents of around 1.85. This highly skewed distribution of voting activity must constitute the background to any understanding of how reddit works.

### **3.1 How is voting activity distributed between reddit's pages?**

The most obvious explanation for a highly skewed distribution of voting activity on reddit is that user attention is unevenly distributed between reddit's pages – and most likely highly concentrated on the 'front page'. This explanation was tested through the analysis of 339,108 observations of posts submitted to the 'politics', 'science' and 'AskReddit' sub-reddits in December 2010. Each of these observations was compared to the previous observation for that post to calculate the number of new votes received since the previous observation. The observed posts received an average of 5.3 votes per minute (VPM) when they appeared on the *front page*, 0.44 VPM on the *main page* for their sub-reddit, 0.14 VPM on the *new* page and 0.02 VPM when they did not appear on any of these pages. A negative binomial regression model was fitted to ascertain the robustness of this trend and the effects were found to be significant at the 0.001 level.

The results of fitting this model suggest that reddit users' voting activity (and, it can be inferred, attention) is indeed highly concentrated on the website's front page. This allows reddit's user-base to have a somewhat shared understanding of 'what's on reddit' and the types of content which reddit tends to endorse. Indeed it is quite common for popular reddit posts and comments to make reference to posts which have previously appeared on the front page. Analyses not reported here suggest that the amount of attention received by a post (as measured by the number of hits) appearing on these various pages largely follows the distribution of voting activity. A post to a large sub-reddit which appears on reddit's front page can expect to channel around 150 hits per minute to the URL it links to. Very popular posts can persist on reddit's front page for longer than ten hours, and result in hundreds of thousands of hits for the linked-to destination.

It could be said that reddit's front page is what gives the website a sense of its own identity, and it is around this identity that feelings of community spirit and involvement form. Through the voting system users can influence not only the chance of an individual post appearing on the front page, but by extension the very definition of reddit's perceived identity or purpose. This is potentially implicated in reddit's rapid growth; the opportunity to influence a large social entity's development by clicking up and down arrows from the comfort of one's computer seems quite unique to Social News.

### **3.2 How does a post reach reddit's front page?**

Every post submitted to reddit appears initially on the *new* page with a score of one, and will be pushed down this *new* page by subsequent submissions and usually off the first page (containing 25 posts by default) within 1-2 hours. Placement on reddit's other pages is dependent on a post receiving votes, so it stands to reason that a post must start receiving positive votes quickly if it is to have any chance of reaching the front page. Previous research on Digg (Szabo & Hubermann, 2008) has shown that the ultimate success of a post can be accurately predicted based on the post's performance in its first two hours.

A logistic regression model was fitted on the data collected through the API in December 2010. The binary outcome measure was whether a post reached the front page; and the explanatory variables represent the post's voting performance at its first observation, and the number of times the post was observed on the *new* page.

This model indicated that early voting (within the post's first 30 minutes) is important if a post is to reach the front page. A post which received one vote per minute immediately after submission would be around 550 times more likely to reach the front page than a post which received no votes in its first 30 minutes. The number of times which a post was observed on the 'new' page is also a significant predictor of reaching the front page. This variable acts as a crude measurement of time, suggesting that for every 30 minutes a post remains on the new page, its chances of reaching the front page increase by around 18.5%. This implies that a post submitted at relatively quiet times (in terms of submission rate) stands a better chance of reaching the front page.

#### **4 Reddit's coverage of Wikileaks' release of US diplomatic cables**

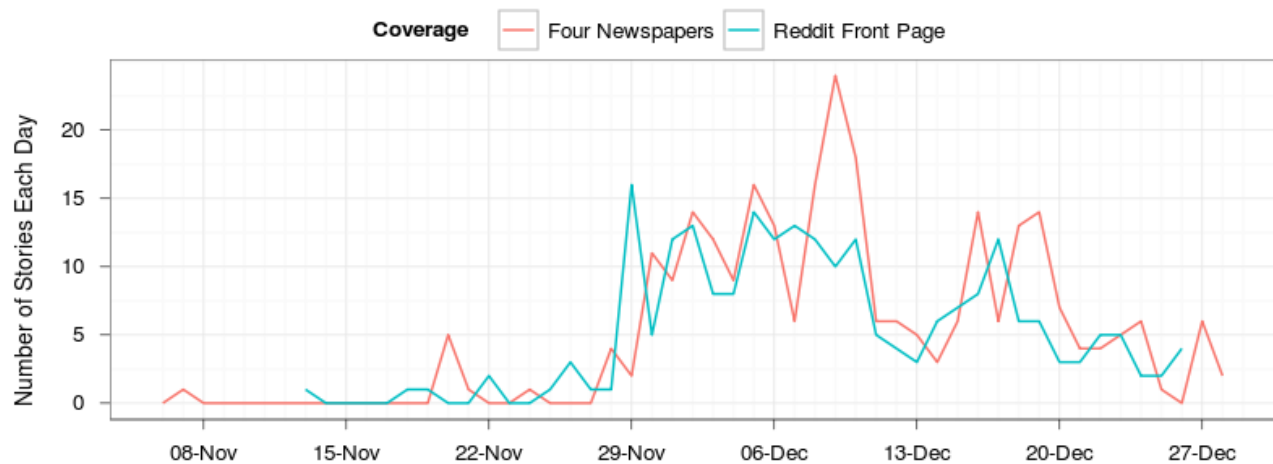
In order to consider the social impact of a website like reddit it is important to consider the qualities of the content being voted on, on as well as the functioning of the voting system. What types of content do reddit users up-vote to the front page, and what types are conspicuous by their absence?

These questions will be addressed here within the context of a case study of reddit's coverage of a major news story: the release of United States' diplomatic cables by the Wikileaks organisation in November 2010. A study period of 4<sup>th</sup> November to 27<sup>th</sup> December 2010 was selected. This time period covers some time before the release of the first batch of cables, in which charges of rape or sexual misconduct were brought against Wikileaks' spokesperson Julian Assange. This time period also incorporates several weeks following the release of the first set of diplomatic cables, when stories about the cables were being widely reported by the conventional news media.

The key words 'wikileaks', 'assange' and 'cablegate' were selected, and for a post or article to be included in the study its title or headline had to include at least one of these key words. Data to be analysed from reddit has been collected through the API. A complete record of all front page posts is available; in the study period there were 249 posts with a key word in their title which appeared on reddit's front page – these posts received a combined total of 87,416 comments. During the study period the 'politics' sub-reddit was being monitored; a complete record of every post submitted here is available, along with observations of its voting performance on reddit, its comments, and the voting performance of each of these. In total there were 3,093 posts submitted to the politics sub-reddit during the study period with at least one key word in their title. These politics posts received a combined total of 43,921 comments.

Four Newspapers were selected for the purpose of comparison; The Guardian, The New York Times, The Independent and The Times. A total of 270 articles appeared in these newspapers during the study period with at least one key word in their headline.

Figure 4-1 shows the number of articles per day of study which appeared in each of the sources being considered; The Four Newspapers being observed, the reddit front page, and the *new* page of the Politics sub-reddit. This graph reveals a strong correlation between sources in terms of the number of articles per day about the topic. This suggests that reddit users, through the voting system, have afforded this story a similar level of coverage to the newspapers being considered.

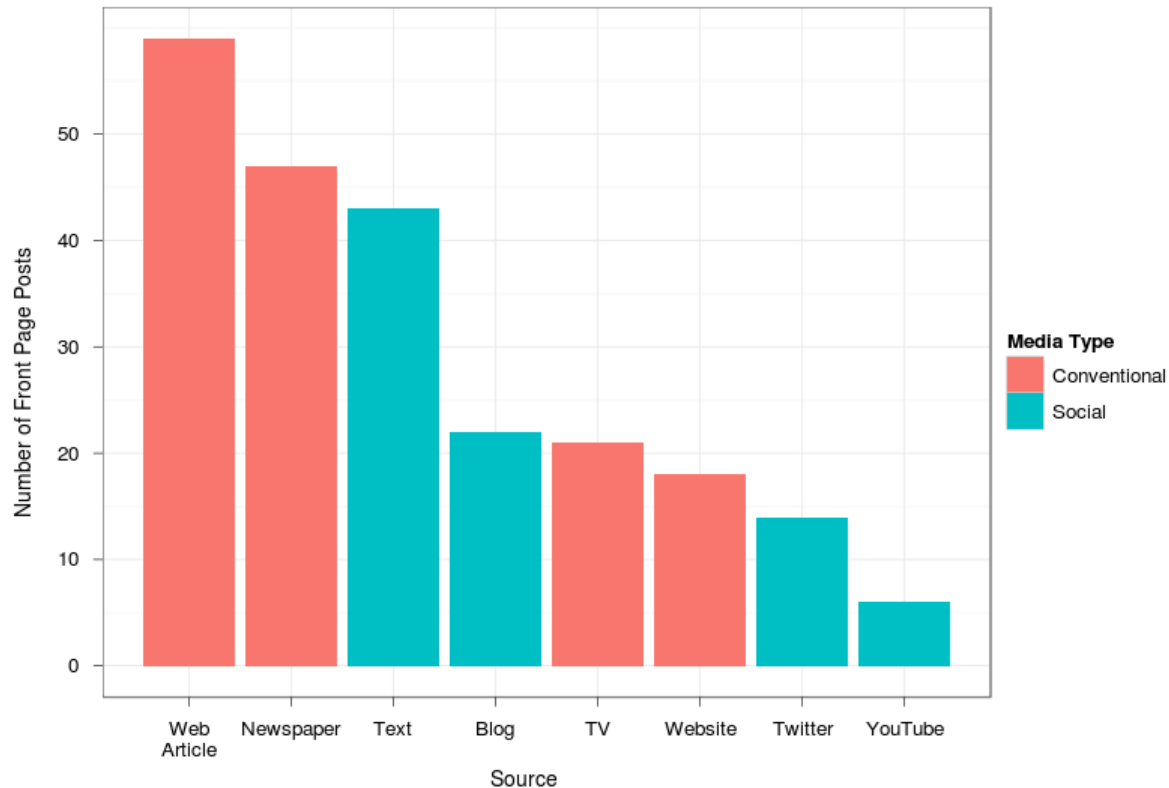


**Figure 4-1 Showing the number of articles/posts per day which appeared in the Newspapers being studied and on reddit’s front page.**

**4.1 What types of post about Wikileaks did reddit users up-vote to the front page?**

Of the 249 relevant posts which reached the front page 48 are text posts and 201 are link posts. Figure 4.1-1 shows the types of source which were linked to by front page posts. 'Old' media sources such as the websites of newspapers and television news channels account for 50 and 21 front page posts respectively. 'Web articles' (articles written by professional journalists for websites which do not have an off-line equivalent) are the most popular source being linked to by front page posts (with 61 posts linking to these sources). Blogs account for 23 front page posts, 17 link to twitter streams and 6 link to Youtube videos.

19 front page posts link to websites (i.e. the main page of another website) – many of these are links to the Wikileaks website itself. These links were submitted when Wikileaks’ domain name was revoked or it was forced to change servers; for example several are links to IP addresses where the website could be accessed when it did not have a domain name. These posts can be thought of as a counter-measure against efforts to make the Wikileaks website inaccessible; when a domain was revoked or a hosting account canceled, the Wikileaks organisation moved their website to a new location, and through social media (e.g. twitter and reddit) they could quickly spread word of this new location. Through these posts reddit users are moving beyond reporting or discussing the story (‘Wikileaks domain name has been revoked’) and towards interacting with the story (‘Wikileaks domain name has been revoked – here’s where you can find their website now’).



**Figure 4.1-1 Showing the sources linked to by relevant front page posts on reddit.**

Calls to action are also relatively commonplace on reddit. In relation to the Wikileaks story one front page post called on reddit users to vote for Julian Assange in Time magazine’s ‘Person of the Year’ award. Several of the ‘operations’ organized by the *Anonymous* group (DDoS attacks against companies which were perceived to have wronged Wikileaks, such as Visa and Amazon) were also ‘advertised’ through reddit before they had commenced. Two attempts were also made by reddit users to form groups which would work to catalogue or re-distribute the content of the leaked cables. These followed a similar trajectory to previously observed endeavours originating on reddit; first a sub-reddit was created where interested users could congregate ([reddit.com/r/wikileaksanalysis](http://reddit.com/r/wikileaksanalysis) and [reddit.com/r/leakspin](http://reddit.com/r/leakspin) respectively), and from there some of these users went on to produce websites or set up IRC channels which would further facilitate the proposed action.

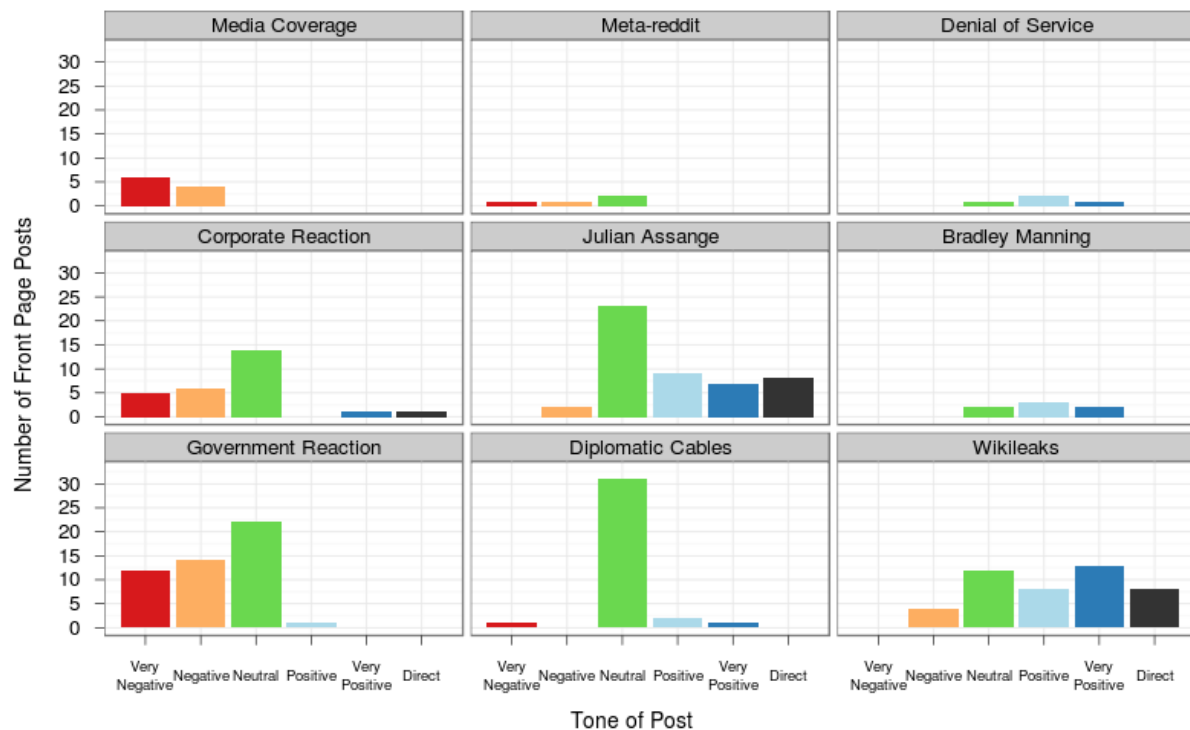
Reddit has something of a history of being involved in making the news rather than merely reporting it; a particularly striking example of this being the impetus it seems to have provided for the ‘rally to restore sanity’ (Friedman, 2010). Also see Voltier (2010 & 2011) for some further examples of events which occurred as a result of posts reaching reddit’s front page.

#### **4.2 Reddit’s bias towards Wikileaks, and the treatment of minority opinions.**

The 249 reddit posts with key words in their title were classified according to their subject and the tone they espoused towards this subject. Details of these classifications are displayed in figure 4.2-1. Many of the posts which reached the front page are neutral with respect to their subject. It does appear however that posts which concern the governments' reaction to the release



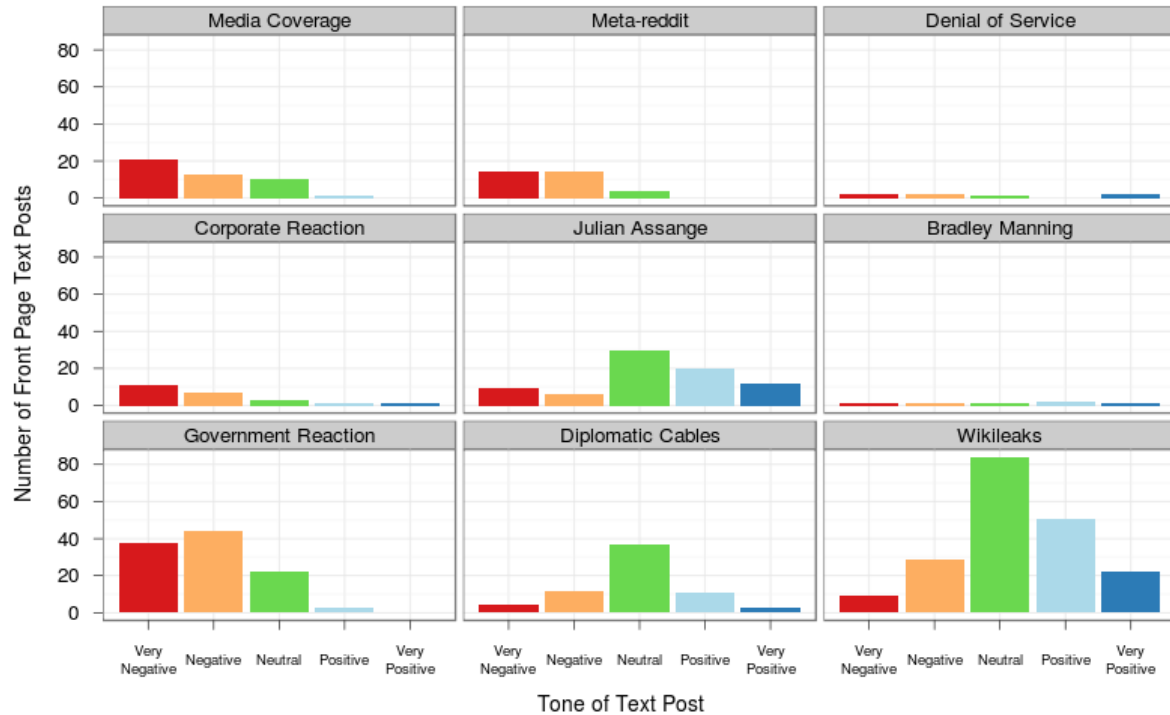
of the cables, or the reaction of large corporations (e.g. Paypal, Visa), or the coverage of the story by the mass media - are more likely to be negative than positive. Posts which appear on the front page that concern Julian Assange or the Wikileaks organisation are more likely to have a positive tone. These measures suggest that reddit's users are largely pro-Wikileaks, and critical of the government/business's reaction to the leaks. There are however a small number of posts which go against this trend, suggesting that reddit's coverage of the story is not entirely one-sided. Posts which link directly to their subject are most likely to concern Wikileaks or Julian Assange. In the case of posts concerning Wikileaks these are links to the current domain/IP where the website can be accessed, or links to announcements on Wikileaks' twitter stream.



**Figure 4.2-1 showing the subject of front page posts and the tone of the item being linked to with regard to this subject.**

This bias in the tones espoused by front page posts suggests that a majority of reddit users are pro-Wikileaks; but is this majority large or small? To take the posts which relate to the Wikileaks organisation as an example: 10% are negative, 32% are neutral, and 58% are positive – do these figures reflect the opinions of reddit users?

In the absence of reliable information about the opinions of reddit users the profile of text posts submitted to the politics sub-reddit will be taken as an alternative measure of reddit users' opinions. These text posts come directly from reddit users and therefore should provide an accurate profile of the opinions of reddit users who submit posts and are active on the politics sub-reddit. However, it should be noted that our previous research suggests that reddit users who vote and who submit are to some degree separate populations.



**Figure 4.2-2 showing the subject of politics text posts and the tone of these with regard to this subject.**

Figure 4.2-2 suggests that in relation to Wikileaks and Julian Assange there is much greater variety in the tones espoused by these text posts than on the front page. To return to the example of posts about the Wikileaks organisation; here 19% of posts are negative, 43% are neutral and 37% are positive. On the subjects of government/corporate reaction to the leaks and media coverage of the story, opinions seem to be more one-sided, with very few posts having a positive tone in relation to these subjects.

Of the text posts displayed in figure 4.2-2 only 12 reached the front page, four of these adopted a positive tone towards Wikileaks or the release of the cables and six adopted a negative outlook on the government or media’s reaction to the release of the cables. The only front-page post to go against the pro-Wikileaks anti-government/media trend was a long and well-written post about the potential damage to international diplomacy caused by the release of the cables.

Reddit’s voting system may marginalise minority opinions to some degree but in the case of Wikileaks coverage these do not seem to have been deliberately excluded from prominent locations. Reddit’s voting system does not seem to preclude the possibility of this happening; it would therefore appear that reddit’s users collectively do not wish to exclude opinions they disagree with. Reddit’s page of etiquette guidelines (‘reddiquette’) states that users should not ‘Downvote opinions just because you disagree with them’. Perhaps reddit’s users follow this guideline more than they give each other credit for – posts and comments which bemoan the reddit ‘hive mind’ are common, and often cite the exclusion of minority opinions as one of the community’s failings.

## **5 'Hive Mind' or 'Collective Consciousness'?**

Reddit's voting system contains a number of positive feedback loops. Firstly, positive votes for a post are associated with increased visibility and in turn further votes – leading to the emergence of a small number of 'superstar' posts. These 'superstar' posts are displayed on reddit's front page for the whole community to see, and serve as shining examples of the kind of content which is successful on reddit. If a user wished to submit content which the reddit community would appreciate, a potentially fruitful strategy would be to mimic the types of post which they see on the front page. It would seem that many users adopt such a strategy; it is quite common for a sub-reddit to be flooded with new submissions which emulate a post that is currently performing well in that sub-reddit.

Reddit's penchant for up-voting certain types of post is often discussed by users. The tendency for users to submit posts of a type which have previously been received well is known in the community as 'circlejerking' – and a sub-reddit exists for the purpose of highlighting and satirising these trends (the 'circlejerk' sub-reddit). This, and the perception that posts which go against certain commonly held beliefs or mores are always down-voted, are two of the main characteristics of what reddit users often refer to as the 'reddit hive mind'.

Is it possible that the same feedback loops implicated in the negatives of the 'hive mind' might also allow reddit's users to self-regulate the behaviour of their community? If reddit empowers a 'hive mind', could it also be said to facilitate the expression of a collective 'consciousness'?

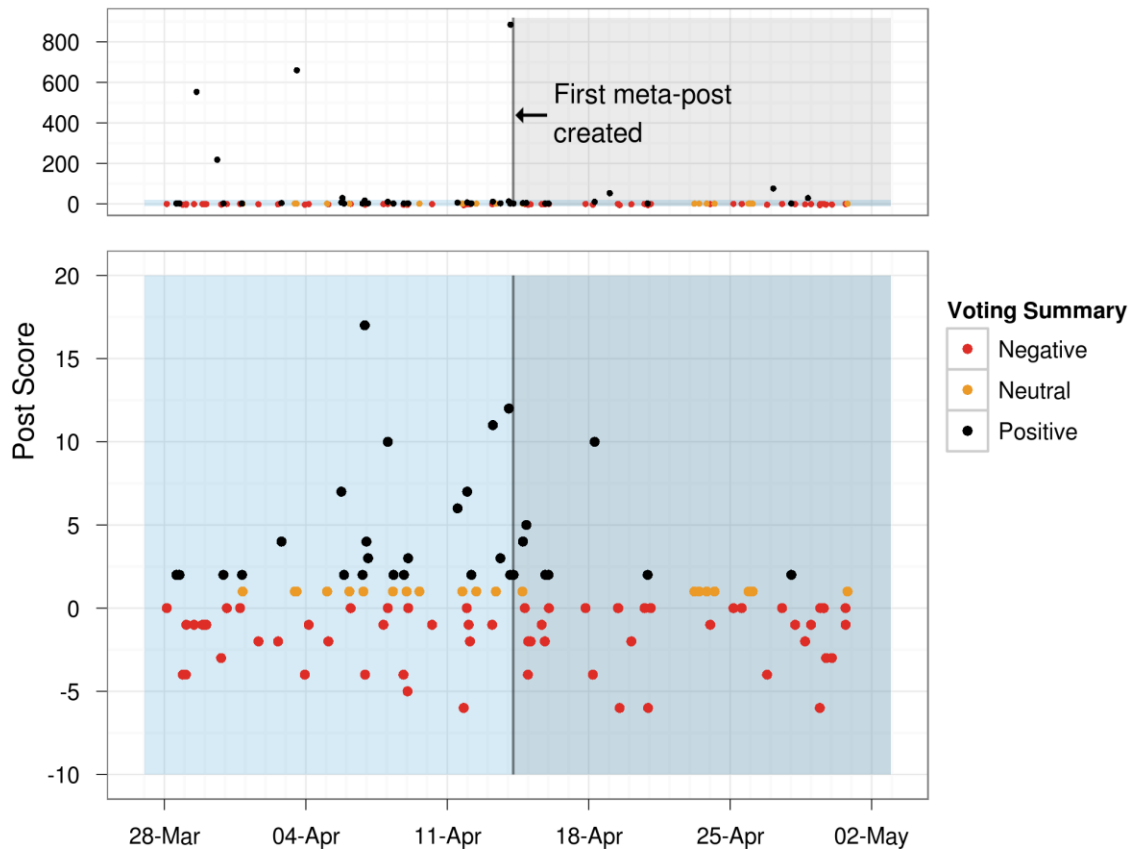
The posts in figure 4.2-2 which have been classified as 'meta-reddit' are text posts where a reddit user highlights or criticises some aspect of reddit's behaviour. In the context of the Wikileaks case study these are predominantly posts which criticise reddit's coverage of the story - the most common criticisms being that there are too many prominent posts about the Wikileaks story, or that posts about the Wikileaks story are too one-sided and out of touch with the general population. Indeed it is quite common to see a number of posts on reddit's front page about a particular topic or following a particular trend, alongside a post criticising the current popularity of said topic/trend.

When such a post appears on reddit's front page it has effectively been endorsed by the community (through the voting system) and broadcast to all of the community's members who visit reddit while the post is visible on the front page. Do these successful 'meta' posts influence the collective behaviour of the website's users? If they do, then this would constitute evidence that the voting system allows the community to self-regulate at a high level. If they do not, this suggests that the hive mind suffers from cognitive dissonance or poor discipline/memory.

### **5.1 Don't link to Fox News!**

Posts which criticise mass media news outlets often appear on reddit's front page. Fox News is regularly the subject of such criticism and posts will often link to the Fox News website for the purpose of highlighting a particular flaw or faux pas. On 14th April 2011 two 'meta' posts concerning Fox News ([redd.it/gpwzp](http://redd.it/gpwzp) and [redd.it/gpy9s](http://redd.it/gpy9s)) were submitted to the politics sub-reddit and were up-voted to the top of this sub-reddit and the default front page. Both of these posts were submitted in response to a post which was then at the top of the politics sub-reddit ([redd.it/gplqq](http://redd.it/gplqq)); and which linked directly to the Fox News website for the purpose of criticising

its content. The message of these two `meta' posts was similar; that reddit users should stop linking directly to Fox News because incoming traffic from reddit was actually helping Fox by increasing their advertising revenue. These posts received scores of 1,500 and 800 respectively, indicating that there were many reddit users who agreed with the point they made. These posts also appeared at the top of the politics sub-reddit and on the default front page for a combined total of 19 hours, meaning that many users of reddit (who subscribe to the politics sub-reddit) will have seen them during this time.



**Figure 5.1 showing the scores achieved by politics posts which linked to Fox News according to their day of submission. The bottom pane shows posts with a score of between -10 and 20, the top pane shows the full range.**

The nature of this particular suggestion is such that it is possible to investigate whether it has influenced the behaviour of reddit users – a decline in the number of posts which link to a Fox News domain ('Fox posts'), or a more negative voting response to such posts, would suggest that reddit's users have taken this community-endorsed suggestion onboard. Of these two possibilities a more negative voting response seems the more likely outcome. An individual user could single-handedly counter-balance any decline in the submission rate of Fox News links – whereas it is through the voting system that the collective will of the users is expressed.

The politics sub-reddit was being monitored for 27 days before the meta posts were submitted and 16 days afterwards. During this time 159 posts which linked to a Fox News domain were submitted; an average of 4 Fox posts were submitted per day before the meta posts, dropping slightly to 3.2 per day afterwards.

Figure 5.1 shows the scores achieved by all of the posts submitted during the observation period. In the 27 days before the meta posts there were four posts which linked to Fox News and achieved a score of greater than 200; the highest score achieved by a Fox post after the meta posts was 76. The bottom pane of Figure 5.1 shows the scores of Fox posts over a limited range (excluding posts with a score of greater than 20) – points have been coloured to reflect whether they received a positive, neutral or negative voting response. After the submission of the meta posts the proportion of Fox posts which received a negative voting response appears to have increased. A binary logistic regression model was fitted to verify the robustness of this trend; it revealed that Fox posts submitted after the meta posts were 61% less likely to receive a positive score than those submitted before the meta posts ( $p < 0.05$ ).

## 6 Conclusions

Most of the posts submitted to reddit are seen by very few people, while a small minority can reach a large audience. Reddit's front page acts as a hub for the website's community of users; the posts which appear here go on to form the basis of discussions between users, and can be thought of as defining the website's identity (in the same manner that a newspaper can have an identity based on the articles it publishes).

The front page showcases posts which have been endorsed by reddit's users, and these posts may be taken as example by users who wish to submit content that the community will appreciate – thus one popular front page post may inspire a deluge of copycat submissions. This is one aspect of what reddit's users refer to as the 'hive mind', the other being the marginalisation of posts or comments which go against certain norms or values which are held by a majority of reddit users.

Analysis of reddit's coverage of the Wikileaks story suggests that the marginalisation of minority opinions may not be as severe as reddit's users often speculate. The website was found to have a strong 'pro-Wikileaks' bias, but many of the posts which reached the front page were neutral in tone, and several posts criticising Wikileaks also managed to reach the front page. The Wikileaks case study also revealed a similarity in the level of coverage on reddit as compared to the mainstream news media; but some interesting differences in the nature of this coverage. Reddit's users showed a penchant for becoming actively involved in the story in addition to reporting it. This desire to participate manifested itself as posts which by their very presence helped the Wikileaks organisation (e.g. by helping people find their website at a new location), or which called on reddit users to engage in some form of collective action related to the story.

Attention is a valuable resource (Hubermann, 2004) - and there appears to be a growing awareness among reddit users that placement on reddit's front page is exactly such a valuable resource. As reddit has grown the nature of its interaction with the mainstream media has shifted; in its formative years this relationship was a one-way street but more recently mainstream news outlets have begun to report on things which have happened on or through reddit (e.g. Townsend,

2010; Rotham, 2010). This interplay between ‘old’ and ‘new’ media also makes reddit a particularly interesting venue for the study of ‘Convergence Culture’ (Jenkins, 2006).

Front page posts can act as a springboard for campaigns of collective action (redd.it/d7ntl, redd.it/hy24b) – or give a boost to some good-natured endeavour that has no marketing budget (redd.it/humy3 and redd.it/elpko). The ‘Humble Indie Bundle’ (redd.it/elpko) is a good example of this; several ‘independently’ developed computer games were packaged together and offered on a ‘pay what you want’ basis, with proceeds being divided between the developers and charities. There have now been three of these bundles, each earning around \$1 million, and reddit has been the primary source of referrals for each by ‘a decent margin’ (private communication).

Some of reddit's users also seem to actively police the website, there have been many popular posts in which a user has brought potential abuse of reddit's voting system to light (e.g. redd.it/gly8, redd.it/gjt26). While reddit's voting system and front page focus may give rise to a 'hive mind' it also seems to encourage pro-social behaviour among some users; the website has a history of supporting good causes and acknowledging acts of altruism, perhaps the prospect of being acknowledged on the front page inspires some users to engage in these types of behaviour.

The front page feedback loop may also allow reddit to govern its own development in an interesting fashion. Reddit users often make suggestions for improving the website; either through some modification of its software (e.g. redd.it/gnorc) or a change in the collective behaviour of its users (e.g. redd.it/fz3z1). When such suggestions appear in prominent locations they do so with the endorsement of users through the voting system; and they are quickly accompanied by a comments page containing hundreds or thousands of comments – the scores of which offer a good insight into the community's feelings on the matter.

The Fox News example in section 5.1 suggests that reddit's users may respond to such suggestions. If this is the case it is quite remarkable; a member of this large group (the politics sub-reddit has 570,000 subscribers, although some proportion of these are likely inactive) with no special standing raised a suggestion and within 24 hours this had been discussed and appears to have been 'adopted' (to the extent that it has effected users' behaviour). All of this has been facilitated by the website's software infrastructure, no individuals with special status (moderators or administrators) and no special oversight was required.

Social News websites like reddit are certainly an interesting development in the communication and decision-making of large groups. Given the youth of these websites and the lack of a thorough understanding of how they operate, there is potentially great promise in their continued development. Some important questions going forward are as follows: Do DM systems make reliable decisions? How refined are these decisions, do SN websites feature 'the best' submitted items on their front page or a random sample of those that are good? How many users or what level of activity can DM systems accommodate?

The rich data these applications collect and store (and often provide public access to) in their day-to-day operation means that there is little impediment to addressing these questions - developing a much deeper understanding of how they operate from the perspective of both the individual and the group. With this more developed understanding it should be possible to reason

in a more informed manner about the various tasks to which these systems might be fruitfully applied.

## References

- Benen, S. (2010), Checking in with ‘America Speaking Out’, online at [http://www.washingtonmonthly.com/archives/individual/2010\\_06/024361.php](http://www.washingtonmonthly.com/archives/individual/2010_06/024361.php) accessed 14.06.2011
- Benkler, Y. (2006), *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, Yale University Press.
- Clauset, A., Shalizi, C. R. and Newman, M. E. J. (2009), Power-law distributions in empirical data, *Siam Review*, 51(4), 661-703.
- Friedman, M. (2010), Reddit Campaign for Colbert Rally Breaks Donation Record, online at <http://newsfeed.time.com/2010/09/14/reddit-campaign-for-colbert-rally-breaks-charity-records/> accessed 14.06.2011
- Halliday, J. (2010), Digg investigates claims of conservative ‘censorship’, online at <http://www.guardian.co.uk/technology/2010/aug/06/digg-investigates-claims-conservative-censorship> accessed 14.06.2011
- Hubermann, B. A., Loch, C. H. and Oncular, A. (2004), Crowdsourcing, Attention and Productivity, *Social Psychology Quarterly*, 67(1), 103.
- Jenkins, H. (2006), *Convergence Culture: Where Old and News Media Collide*, NYU Press.
- Milbank, D. (2010), Republicans’ new Web site not exactly what they hoped it would be, online at <http://www.washingtonpost.com/wp-dyn/content/article/2010/05/25/AR2010052504396.html> accessed 14.06.2011
- Oleoleolson (2010), Massive Censorship of Digg uncovered, online at <http://blogs.alternet.org/oleoleolson/author/oleoleolson/> accessed 14.06.2011
- Rotham, W. (2010), ‘Cigar guy’ is Internet antidote to Sad Keanu, online at [http://www.msnbc.msn.com/id/39522903/ns/technology\\_and\\_science-tech\\_and\\_gadgets/](http://www.msnbc.msn.com/id/39522903/ns/technology_and_science-tech_and_gadgets/) accessed 14.06.2011
- Shirky, C. (2008), *Here Comes Everybody: The Power of Organizing without Organizations*, Penguin Press.
- Szabo, G. and Hubermann, B. A. (2008), Predicting the popularity of online content, *Communications of the ACM*

Townsend, A. (2010), Drink Up NYC: Meet the Tiny Crustaceans in Your Water, online at <http://newsfeed.time.com/2010/08/31/drink-up-nyc-meet-the-tiny-crustaceans-not-kosher-in-your-tap-water/?hpt=C2?> accessed 14.06.2011

Voltier Creative (2010), Reddit's Astonishing Altruism, online at <http://voltiercreative.com/2010/11/reddits-astonishin-altruism/> accessed 14.06.2011

Voltier Creative (2011), Reddit's Astonishing Altruism: Part Two, online at <http://voltiercreative.com/2011/04/reddits-astonishing-altruism2/> accessed 14.06.2011