A Framework For Determining Influence on Reddit

Adrian Kwok (#200136359)
Vivienne Wallace (#301145687)

CMPT 741 Final Project
December 3rd, 2010

1. Introduction

Social news networks are an increasingly important method of proliferating information, opinions and ideas. As the number of sites and users grows, so too do the possibilities for research, the potential for influence by individuals on the social media landscape and the demand for businesses to utilize this influence. While there has been some research focusing mainly on social networking sites like Facebook and Twitter, there has been relatively little research done on social news sites. Since social news sites have a different purpose and structure to generic social networking sites, it is possible that the networks and interactions will be different and thus influence can be gained and exploited differently. The answer to the this question will prove valuable regardless of the outcome, as it will either point the way towards new avenues for study, or reinforce the work that has already been done. So to this end, we concentrate on identifying a number of features and metrics for building an overall influence rating of users on the social news site Reddit[20].

2. Motivation

User-driven news aggregation sites have been prevalent for many years; technologically driven sites such as Slashdot were some of the first precursors of Reddit. However, Reddit's biggest rival, Digg, saw a decline in its user base after its most powerful users started selling their services to advertisers[5], undermining the initial goal of the site – to allow all users equal opportunities in submitting news. This downward spiral only accelerated when Digg released version four of its website, tying in advertisers with news stories in a quiet and deceitful manner[28]. In November of 2010, many mainstream media outlets have started taking notice and have been declaring Digg as being dead[23][24][25][26][27].

Being Digg's foremost rival, over the years Reddit has tried to mitigate a lot of the issues that plagued Digg by keeping as many aspects of its site private – that is, they intended to prevent advertisers from influencing the site by obscuring as many of the site's details as possible. Important metrics used in developing models and driving research in Digg, such as a user's followers, likes and dislikes, ranking, etc. are completely hidden on Reddit. However, as with any social network, there always exists groups of users who are clearly more influential than others. Since the focus of the Reddit community is content, rather than personality driven – the Reddiquette [3] for the site encourages users to vote based on merit – this makes influential users on the site a significant and valuable target for advertisers and market researchers.

Furthermore, the impact of Reddit on traditional media and the real world is far flung: although Reddit may be considered a lesser known social media portal than Twitter and Facebook, there is evidence that Reddit is increasingly proving to be influential on people outside of the confines of site-based discussion. This can be seen most recently and prominently in the 215,000 (est.) person rallies in Washington D.C. [1] on October 30, 2010. Often credited as having its origins from an early Reddit campaign encouraging comedian Stephen Colbert to begin organization of the rallies [2], this rally serves to demonstrate that the community of Reddit could be a potential breeding ground for other influential movements.

Even though many aspects of Reddit are hidden from plain view, it should still be possible to infer a user's influence by analysing his behaviour and interactions with other users on the site. However, the conceptual differences that have resulted in removing many public metrics may have changed the underlying behaviour of users on Reddit. As there has been some important prior research on Digg [6][7][16][17]– for example, to determine user behaviour or to predict user success – we have attempted to apply the principles of this work to match the site specific differences on Reddit. Since Digg is rapidly dropping in popularity, further research on the platform is unrealistic, and thus the aim of this paper is to provide a stepping stone in diverting future research to the next viable social news site, Reddit – a site for which no known scholarly research has been conducted. Due to the fact that information that is explicit on Digg is only available through implicit analysis on Reddit, we believe that our work will supply fundamental metrics that can be used to adapt models and classifiers defined in prior research on Digg to a new platform, Reddit.

3. Description of Reddit

Reddit is a social news site consisting of users, known as Redditors, who create original content or submit links to interesting stories. The functionality available to Redditors mostly involves one of three categories of actions: submitting stories, commenting on stories, and voting on both stories and comments. More specifically, they are allowed to both vote positively and vote negatively. When a story is submitted by a Redditor, it is placed in the "new" section of the website for other users to vote on – stories that receive a large amount of positive votes are systematically moved from the "new" section to the front page of Reddit. Redditors are also allowed to befriend other users so that they can be notified of their friends' activities, but all friend-related information is purposely kept private to other users.

3.1 Specific Differences: Digg vs. Reddit

On Digg, most information about a user is public, while on Reddit only a user's submissions and comments are public. The specific differences are represented in tabular form in Figure 1.

Explicit Public Features	**Digg**	**Reddit**
*List of users following the user (i.e. other users interested in the user's activities)	✓	×
*List of users the user is following	✓	×
List of submissions by the users the user is following	✓	✓
List of comments by the users the user is following	✓	✓
List of submissions the user has voted up	✓	×
List of submissions the user has submitted	✓	✓
List of comments the user has made	✓	✓
How frequently the user posts	✓	×
How frequently the user posts new submissions	✓	×

Figure 1: Differences between Digg and Reddit

Even when a user is logged into their own account on Reddit, the majority of these properties about themselves are still not visible, except for the list of submissions they have voted up. Most importantly, this means that no single user on Reddit knows how many other users are following their activities – implying that building a classifier to predict followers is infeasible since there does not exist a publicly available test set.

4. Problem Statement

Katz and Lazarsfeld[21] defined "opinion leaders" as individuals who are able to exert influence over the ideas and opinions on other individuals within their social network. They presented a "two-step flow" model where these opinion leaders act as an intermediary between the media and the general public. Dodds and Watts[22] emphasize that influentials are not in fact high profile members of society, be they media, political, government or social. Instead, influentials are people who are members of a social network whose opinion carrying an unduly large amount of weight.

We propose that the influence of a user on Reddit can be quantified by focusing on three main factors: the number of high quality potential followers, the previous activities and publication success rate of a user, and the influence rating of followers who interact most often with the user.

Unfortunately, as mentioned previously, these properties have to be inferred through several different techniques, since they are completely hidden from most users. In order to determine the number of friends a user has, we analyse all of the users who have commented on the user's submissions and assign a score for each commenter conveying the likelihood that the commenter is a friend of the user; this score will be drawn from specific trends and behaviours of the comments. Furthermore, we analyse a user's previous activities on the site to determine their initial influence rating including such measures as the success and amount of discussion of previous submissions, length of time as user has been active, and user karma rating. Finally, we converge the initial influence from each user and the influence of potential followers to determine each user's final influence rating.

5. Related Work

In the past few years there has been a significant rise in research on social networks; in particular, researchers have been trying to glean useful information from popular social networking sites such as Twitter and Facebook. Examples of this include using these new forms of communication to recommend news articles[8], inferring trends on Twitter using collaborative filtering from Wikipedia[9], filtering out rubbish posts from quality posts on Yahoo Questions and Answers [10], and determining influential bloggers in a community[11].

In the their paper analysing user influence in Twitter, Cha et al. [12] focused on using the indegree and popularity of posts of each user to determine his influence on other users; their findings confirmed the results by Avnit [13], who posits that the number of followers of a user on Twitter plays a small role in the user's overall influence, since users tend to befriend other users on Twitter solely to improve their status and be polite. This is in stark contrast to our paper though, since we believe that users on Reddit tend to follow users for genuine, non aesthetic reasons.

Jamali and Rangwala[6] focused on co-participation: that is, if two users commented on the same article, they would be deemed as co-participants and their co-participation score would be incremented. The main goal of the paper was to determine the focus of active users amongst different subcategories, as well as using this information to predict submission success. Its emphasis is on the activeness of users as a predictive measure, not the users' influence amongst their peers. However, the paper focused on a combination of 16 simple and complex features for each user in their classification model for predicting user submission success – some of these features are similar to those that we use.

Instead of using a classifier, Lerman[17] describes two mathematical models to represent the dynamics of submissions and user rank on Digg. However, these models rely heavily on the notion of a user's friends on Digg, which is not available on Reddit. Fortunately, we believe that these models can be easily adapted for Reddit following our work on inferring the number of friends a user may have.

6. Determining Influence

6.1. Inferring Number of Followers

As mentioned previously, inferring the number of friends a user has is an important factor not only in determining influence, but also in predicting submission success and other features – most existing literature on social news sites is dependent on this metric.

For this section, we define the target user (i.e. the user for which we wish to determine the number of followers) as u. The relationship between u and any other user on Reddit, v, is defined as $p(u,v)$, which denotes the likelihood that v is a follower of u. Specifically, $\forall u, \forall v$: $0 \leq p(u,v) \leq 1$, where $p(u,v) = 0$ implies that v is highly unlikely to be a follower of u, and $p(u,v) = 1$ implies that v is very likely a follower of u.

We will now decompose $p(u,v)$ into three parts. These formulations are preliminary hypotheses, and will require further analysis to determine feasibility (discussed in the Challenges section).

6.1.1 Follower Likelihood via Comment Frequency and Quality

The most naive and obvious determinant of follower likelihood would be to analyse a potential follower v's comments on user u's submissions. We believe there are two situations in which a user would choose to follow the actions of another user: if they strongly agreed with the poster's positions or if they strongly disagreed with the poster's positions. For this reason, we look for posters who are frequent commentators at the root-level of a story and whose comments elicit a strong reaction from the audience, either positive or negative. We count the number of comments by a distinct user as well as the absolute value of the overall rating (determined by the net community votes) on each comment.

We analyse all of a user u's prior submissions $S = \{S_1, S_2, \dots, S_n\}$ dating back a predefined number of t days – the rationale for looking at a smaller, most-recent subset of submissions instead of analysing **all** of u's prior submissions is that most users tend to slowly accumulate followers over time, and the varying idle period between account creation and a user's first submission will skew calculations. We require further definitions shown in Figure 2.

A preliminary formulation is shown in Figure 3. In this formulation, we aggregate the ratings of a potential follower's comments on each of u's submissions, and normalise it using the value of the highest rated comment in each submission. We then sum up all of these values and calculate a weighted average across all submissions. The results of this formulation will be a probability between 0 and 1, defining the likelihood that a user v is a potential follower of u based on the criteria for this sub-definition.

$0 \leq p_{cf}(u, v) \leq 1$	Likelihood that v is a follower of u based on comment frequency
c_i	Sum of the absolute value of each rating for all comments by v on u's i^{th} submission (only first-level comments are counted)
c_i^*	Highest absolute value rated comment by any user on u's i^{th} submission
c_w	Predefined comment threshold (e.g. $c_w = 2$ implies that comments that have at least ½ the rating of the top rated comments indicate the user's comments are interesting)
c_f	Predefined comment frequency (e.g $c_f = 0.5$ implies a likely follower should comment on at least half of u's submissions)
α	Predefined dampening factor (default = 2)

Figure 2: Relevant Features to 6.1.1

$$p_{cf}(u, v) = min\left(\frac{\sum_{i=1}^{|S|} min\left(\frac{c_i}{c_i^*} c_w, 1\right)}{c_f \cdot |S|}, 1\right)^{\alpha}$$

Figure 3: Likelihood Formulation via Comment Frequency

6.1.2 Follower Likelihood via Comments on Low-ranked Submissions

Another determinant of follower likelihood is the frequency of v's comments on u's low-ranked submissions; the idea is that low rated submissions are often caused by low visibility in the "new stories" queue. If v frequently comments on stories with low public visibility there is a greater likelihood that he is discovering u's stories via other means – particularly, via his friends feed. This an abstraction of a property of Digg first proposed by Lerman[7].

As in Section 6.1.1, we analyse user u's prior submissions $S = \{S_1, S_2, \ldots, S_n\}$ dating back a predefined number of t days. Both positively and negatively voted comments by v positively affect the likelihood. We require further definitions shown in Figure 3. A preliminary formulation is shown in Figure 4 –the idea behind this formulation is that we just count the number of comments among all low-ranked submissions.

$0 \leq p_{lr}(u, v) \leq 1$	Likelihood that v is a follower of u based on low-ranked submissions
S_l	Predefined low-rank submission threshold (e.g. $S_l = 10$ implies that we only look at submissions with fewer than 10 votes)
S'	All submissions in $S = \{S_1, S_2, \ldots, S_n\}$ that have at most S_l votes
c_{lf}	Predefined comment frequency (e.g. $c_{lf} = 0.5$ implies a likely follower should comment on at least half of u's low-ranked submissions)
$c_i = \{0,1\}$	Whether v has commented on submission S_i'
β	Predefined dampening factor (default = 2)

Figure 3: Relevant Features to 6.1.2

$$p_{lr}(u,v) = min\left(\frac{\sum_{i=1}^{|S\prime|} c_i}{c_{lf} \cdot |S'|}, 1\right)^{\beta}$$

Figure 4: Likelihood Formulation via Comments on Low-ranked Submissions

6.1.3 Global Follower Likelihood via High-ranked Submissions with Low Activity

In [19], Gómez indicates that controversy often drives up views on a topic, so a popular submission with few comments would seem to be lacking in controversial content. In this situation, there must be other reasons for the popularity of this situation, and the factors involved could be multi-faceted. We hypothesize that one such cause for the unusual popularity ranking could be a user with many followers casting "knee-jerk" votes. Due to the potentially myriad of factors that could affect this popularity behaviour, the threshold and weighting variables will require considerable tuning work after the data analysis.

As in previous sections, we analyse user u's prior submissions $S = \{S_1, S_2, \dots, S_n\}$ dating back a predefined number of t days. It may be beneficial to mine the submissions such that we get snapshots of the submissions only a small period of time after it has been posted. We require further definitions shown in Figure 5. A preliminary formulation is shown in Figure 6. The reasoning behind this formulation is that we should only consider highly rated submissions with the number of comments below some predefined threshold – furthermore, the fewer the comments on a highly rated submission, the higher the weighting of the submission to the overall likelihood total of pre-existing followers.

$0 \leq p_{hl}(u,*) \leq 1$	Likelihood, in general, that u has pre-existing followers
S_h	Predefined high-rank submission threshold (e.g. $S_h = 200$ implies that we only look at submissions with more than 200 votes)
S_{num}	Predefined number of suspicious submissions threshold
$votes(S_i')$	Net up/down votes on submission S_i'
$0 \leq c_{nw} \leq 1$	Predefined comment threshold (e.g. $c_{nw} = \frac{1}{10}$ implies that we're only suspicious of submissions with a 1:10 ratio of comments to votes)
S'	All submissions in $S = \{S_1, S_2, \dots, S_n\}$ that have at least S_h votes and a comment to votes ratio below c_{nw}.
c_{ni}	Number of comments on the i^{th} submission
γ	Predefined dampening factor (default = 2)

Figure 5: Relevant Features

$$p_{hl}(u,*) = \frac{\sum_{i=1}^{|S\prime|}\left(1 - \left(\frac{c_{ni}}{votes(S_i') \cdot c_{nw}}\right)^{\gamma}\right)}{|S'|} \cdot min\left(\frac{|S'|}{S_{num}}, 1\right)$$

Figure 6: Likelihood Formulation via High-ranked Submissions with Low Activity

6.1.4 Tying it all together

We must now define a way to combine the various probabilities we have defined into an overall equation for the total probability of a user v following user u. Each of the subprobabilites will be weighted with tuning factors A, B, and C to be determined after experimentation. Since $p_{hl}(u,*)$ defines the general likelihood of prior followers of u, we distribute this probability equally amongst all potential followers (defined by $p_{cf}(u,v)$ and $p_{lr}(u,v)$). We do this in a two-step process.

1.) For all users u, v:

$$p_1(u,v) = A \cdot p_{cf}(u,v) + B \cdot p_{lr}(u,v)$$

2.) Let N be the number of nonzero entries of $p_1(u,*)$. For all users u, v:

$$p(u,v) = \begin{cases} p_1(u,v) + C \cdot \frac{p_{hl}(u,*)}{N} & \text{if } p_1(u,v) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

6.2. General Influence Metrics

There are a number of metrics that play a large part in determining the history of a user's activity by determining the quantity and quality of actions performed over a period of time. Therefore, we will be tracking a number of general features over the time period that we are examining. We will use these features to define an initial influence rating, $I_{init}(u)$ for all users u. Ideally, we would specifically formulate each of these features into $I_{init}(u)$, but this will require heavy analysis of the data (see Future Work section).

1. Poor comments and submissions should negatively affect a user's influence rating so we will use a feature that counts the number of submissions and comments falling below a threshold rating. After analysis, a tuning constant can set to indicate the negative weight these bad comments and submissions have.
2. The number of popular stories submitted by a user in an important feature. Stories considered popular will be counted by the overall rating of the story, defined as "link karma" on Reddit. This naturally leads to a similar feature counting the number of popular or successful comments, defined as "comment karma", over a certain ratings threshold.
3. The length of time a user has been registered, and the total amount of activity/over the time registered gives a picture of whether or not a user has a history of being an active participant or more of a lurker/infrequent participant as inspired by similar work by Jamali and Rangwala[6]. This will be tied closely with the amount of activity during the period of time we are mining the data. A user who has a history of a large amount of activity, but is inactive during the study period still has the potential to be an influential user.
4. As discussed by Lerman[17], the time period at which a user is active is also an important consideration; since most users of Reddit are based in the United States, users who post during the

site's most active times are in general more influential than users who post during times when most users are asleep.

5. Classification of the subreddits a user posts in most often, similar to the work done by[6] can help to classify the interests and knowledge base of a user, as well as used in determining the extent and areas where a user may have influence. This feature can be determined by counting all submissions in comments in the various subreddits and creating aggregate values.
6. A feature that is closely related to classification will be the "entropy measure" discussed in [6], which determines the amount of focus of a user. Focus is defined by the proportion of activity in a certain area. High focus indicates that the user is mostly active in only one area while low focus indicates the users spreads their attention by commenting sporadically in many different areas.
7. Trophies are a function of Reddit that requires more study. Trophies are given out by the Reddit administrators, so they may or may not act as a useful metric of determining user influence.

6.3. Calculating final influence ratings for every user on Reddit

By this point, for each user, we have an initial influence rating $I_{init}(u)$, and a probability rating $p(u,v)$ denoting the likelihood that v follows u. Ideally, we wish to determine the final influence of user u by combining the initial influence rating with some β-weighted influence rating of all potential followers. We formulate this loosely by:

$$I(u) = I_{init}(u) + \beta \cdot \sum_{\forall v \neq u} p(u,v) \cdot I(v)$$

However, we note that there is interdependence between each user's calculations of influence. We can formulate it in matrix form, and compute it in a power-iteration manner similar to Pagerank.

7. Challenges

One of the major challenges of our research has been adequately defining the problem that we were looking to solve. Initial plans were focused almost entirely on creating a model for predicting popular stories based on the content either being created or supported by so called influentials. After further thought, we realized that the key work of identifying influentials would not be trivial, especially on our chosen platform, Reddit.

A large issue with identifying influentials was how to know if they had been correctly identified. While we have used the theory that a story is popular because of the influence of an opinion leader, there is the issue that we may be falling into a circular logic trap. Opinion leaders create influential stories and stories are influential because they are created by opinion leaders. To this end, we recognize that our model for identifying influential leaders may require some kind of verification.

Verification became a large stumbling block in the progress of defining the research that we wanted to do. While we have defined some methods to infer the number of possible followers of a user, how can we actually verify these findings without any publicly available sample data sets? This is a similar issue with our influence rating calculations – it seems that there will be a need for a human oracle to classify influential users to test our theories. However, the discussion of the role of an oracle is definitely beyond the scope of this paper but may prove to be a vital topic in future research.

Further challenges arose in creating our equations to calculate influential ratings. Without a data set to study, it is difficult to tell if there will be noise and outlier data points or if hidden factors may effect our results. To account for this we have frequently included tuning variables into our equations.

8. Future Work and Conclusion

Our paper focuses on determining influence, and we defined metrics to calculate a user's influence. However, aside from intuitive conjectures, we have no idea whether our methods are successful in the real world or not. Ideally, once we have determined who the influential users on Reddit are, we'd wish to monitor their future submissions and see if they are statistically more likely to be successful than other, non-influential members. Furthermore, it would have been preferable to capture concrete data on Reddit to develop more robust and thorough mathematical models, but due to the time constraints of this project we were unable to do so.

There are some concerns about the actual use of these influence ratings; how can these values be used in a meaningful way? Its obvious nefarious use, as described previously, would be for marketers to target and monitor influential members, but it is unclear how we can use our findings conversely to prevent these activities from occurring. Furthermore, as shown in collaborative models defined by Lerman[17], predicting submission success was very effective on Digg; it would be ideal to incorporate our influence ratings to Reddit in a similar manner.

In our work on identifying influential users, naturally the subject on how a user becomes influential has frequently risen. Also, the related topic of how the influence rating of a user ebbs and flows with time and actions, including the idea that two or more influentials may at some point come into conflict with each other and how these "battles" could affect the landscape of the site. Another possible factor in the influence of a user is their stance on controversial issues. This may positively or negatively effect the number of followers a user has. These are all very interesting topics, and ideally we would like to focus on them in the future.

Similarly, the issue of a user's "global" influence or influence outside of the Reddit site is extremely interesting, but difficult to classify in a simple matter. Although we initially were interested in Reddit due to the main stream influence of some of their postings, we believe the study of this would be far outside the scope of this paper.

In our paper, we have discussed some of the differences between other social news sites and Reddit, and we proposed some methods to infer private information on Reddit. To the best of our knowledge, this is the first paper that has focused primarily on Reddit – we hope that it entices and diverts research from Digg to further research on this new platform.

9. References

1. M. Hartenstein. "Jon Stewart's 'Rally to Restore Sanity' drew 200,000, beating estimated attendance at Glenn Beck's", *available at http://tinyurl.com/2cwnocs*
2. M. Bell. "'Rally to Restore Sanity' to meet 'March to Keep Fear Alive;' Reddit users talk about starting the online campaign", available at http://tinyurl.com/35ty25h
3. http://www.reddit.com/help/faq ; http://www.reddit.com/help/reddiquette
4. D. Parrack. Is Reddit now bigger than Digg? *Blorge.com*, Sept. 2, 2010 available at http://tinyurl.com/38vthb2

5. khalid. An Interview with Digg Top User, *Invesp.com*, Sept. 3, 2008 available at http://tinyurl.com/69szq7, M. Arrington. Want on the Digg Home Page? That'll be $1200, *TechCrunch.com*, Sept 3, 2008. available at http://tinyurl.com/38ejapy
6. S. Jamali and H. Rangwala. Digging Digg: Comment Mining, Popularity Prediction, and Social Network Analysis. *Proceedings of the 2009 international Conference on Web information Systems and Mining* (November 07 - 08, 2009). WISM. IEEE Computer Society, Washington, DC, 32-38.
7. K. Lerman. Dynamics of collaborative document rating systems. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis* (San Jose, California, August 12 - 12, 2007). WebKDD/SNA-KDD '07. ACM, New York, NY, 46-55.
8. O. Phelan, K. McCarthy and B. Smyth, Using twitter to recommend real-time topical news. *Proceedings of the third ACM conference on Recommender systems*, October 23-25, 2009, New York, New York, USA
9. Q. Chen, T. Shipper and K. Latifur. Tweets mining using WIKIPEDIA and impurity cluster measurement. *2010 IEEE International Conference on Intelligence and Security Informatics*, May 23-26, 2010, Vancouver, British Columbia, Canada
10. E. Agichtein, C. Castillo, D. Donato, A. Gionis, G. Mishne, Finding high-quality content in social media. *Proceedings of the international conference on Web search and web data mining*, February 11-12, 2008, Palo Alto, California, USA
11. N. Agarwal, H. Liu, L. Tang and P. Yu. Identifying the influential bloggers in a community. *Proceedings of the international Conference on Web Search and Web Data Mining* (Palo Alto, California, USA, February 11 - 12, 2008). WSDM '08. ACM, New York, NY, 207-218.
12. M. Cha, H. Haddadi, F. Benevenuto and K. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*. May, 2010. Washington DC, USA.
13. A. Avnit. The Million Followers Fallacy, *Internet Draft, Pravada Media. Available at http://tinyurl.com/nshcjg*
14. D. M. Romero, W. Galuba, S. Asur and B. A. Huberman. Influence and Passivity in Social Media (August 4, 2010). Available at SSRN: http://ssrn.com/abstract=1653135
15. CRAN - Package egonet: http://cran.r-project.org/web/packages/egonet/index.html
16. R. Ghosh and K. Lerman. Predicting Influential Users in Online Social Networks. *The fourth SNA-KDD Workshop, Held in conjunction with The 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010)*
17. K. Lerman, Dynamics of Collaborative Document Rating Systems. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis(San Jose, California)*, 2007, New York, NY, USA, 46-55
18. S. Jamali, H. Rangwala, Defining a Coparticipation Network Using Comments on Digg. *IEEE Intelligent Systems*, pp. 36-45, July/August, 2010
19. V. Gómez, A. Kaltenbrunner, V. López, Statistical Analysis of the Social Network and Discussion Threads in Slashdot. *Proceeding of the 17th international conference on World Wide Web* (WWW '08). ACM, New York, NY, USA, 645-654
20. Reddit.com. http://www.reddit.com
21. E. Katz, P. Lazarsfeld, E. Roper, Personal Influence, the Part Played by People in the Flow of Mass Communication. Transaction Publishers, Oct. 2005. ISBN 1412805074
22. D.J. Watts, P.S. Dodds, Influentials, Networks, and Public Opinion Formation. Journal of Consumer Research, Dec. *2007*.
23. J. Errett , Digg is Dead - The downfall of Digg and what it says about online news: http://tinyurl.com/23vpc6k
24. D. Parrack, Is Digg dead or just seriously wounded? October 26, 2010. http://tinyurl.com/39nvpaw
25. P. Robles, Digg's demise: are VCs and founder cash-outs partly to blame? October 26, 2010. http://tinyurl.com/29qboed
26. M. Issac, Former CEO Jay Adelson on His Life at (and the Death of) Digg October 25, 2010. http://tinyurl.com/28f95cg
27. D. Frommer, FIVE DAYS LEFT: Enter The "Fix Digg" Contest! November 8, 2010. http://tinyurl.com/23pg588
28. R. Quigley, The Rebellion Against Digg v4. August 30, 2010. http://tinyurl.com/2fhpxqe