

Social Networks and Social Information Filtering on Digg

[Extended Abstract] *

Kristina Lerman[†]
University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina del Rey, California 90292
lerman@isi.edu

ABSTRACT

The new social media sites — blogs, wikis, Flickr and Digg, among others — underscore the transformation of the Web to a participatory medium in which users are actively creating, evaluating and distributing information. Digg is a social news aggregator which allows users to submit links to, vote on and discuss news stories. Each day Digg selects a handful of stories to feature on its front page. Rather than rely on the opinion of a few editors, Digg aggregates opinions of thousands of its users to decide which stories to promote to the front page. Digg users can designate other users as “friends” and easily track friends’ activities: what new stories they submitted, commented on or read. The friends interface acts as a *social filtering* system, recommending to user stories his or her friends liked or found interesting. By tracking the votes received by newly submitted stories over time, we showed that social filtering is an effective information filtering approach. Specifically, we showed that (a) users tend to like stories submitted by friends and (b) users tend to like stories their friends read and liked. Social filtering is a promising new technology that can be used to personalize and tailor information to individual users: for example, through personal front pages.

Keywords

Social Network analysis; collaborative filtering; social filtering

1. INTRODUCTION

Many Web sites that provide information (or sell products or services) use collaborative filtering technology to suggest relevant documents (or products and services) to its users. Collaborative filtering-based recommendation systems try to find users with similar interests by comparing their opinions about products. They will then suggest new products

*A full version of this paper is available at arxiv.org/abs/cs.HC/0612046

[†]This research is based on work supported in part by the National Science Foundation under Award Nos. IIS-0535182 and IIS-0413321. We are grateful to Dipsy Kapoor for helping with data analysis, and to Fetch Technologies (<http://www.fetch.com>) for providing wrapper building and execution tools.

that were liked by other users with similar opinions. Recommender systems based on *social filtering*, on the other hand, suggest new products or documents simply based on whether the user’s designated friends found these products or documents interesting. Researchers in the past have recognized that social networks present in the user base of the recommender system can be induced from the explicit and implicit declarations of user interest, and that these social networks can in turn be used to make new recommendations.

The new social media sites, such as the social news aggregator Digg,¹ allow users to explicitly build social networks by designating others as friends. Tracking activities of friends is common feature in many social media sites and is one of the major draws attracting users to these sites. It offers a new paradigm for interacting with information — social filtering. Rather than actively searching for new interesting content, or subscribing to a set of predefined topics, users can now put other people to task of finding and filtering information for them. We show that social networks are being used on Digg for social filtering. Specifically, we show that Digg users tend to be interested in the news stories their friends find interesting. Although social filtering, as practiced by Digg, has recently come under fire for being susceptible to “gaming,” we believe it to be a promising technology that will lead to new generation of personalization and recommendation algorithms.

2. STRUCTURE OF DIGG

Digg’s functionality is very simple: Users submit links to stories they find online, and other users vote on these stories. When a story gets enough positive votes, or diggs, it is promoted to the front page. The front page is what users see on the Digg home page, while the newly submitted stories are less visible, being “hidden” in the Upcoming stories pages. Digg also allows users to designate other users as friends and makes it easy to track friends’ activities. A section of Digg’s home page summarizes the number of stories the friends have submitted, commented on or liked recently.

Each day Digg selects a handful of stories to feature on its heaving trafficked front page. Although the exact formula for how a story is promoted to the front page is kept secret, so as to prevent users from “gaming the system,” it appears to take into account the number of diggs a story gets and

¹<http://digg.com/technology>

the rate at which it gets them. The mechanism by which the stories are promoted, therefore, does not depend on the decision of one or few editors, but emerges from the activities of many users.

In order to study the role of social networks in filtering, we tracked both new and front page stories in the technology category. We collected data in May 2006 by scraping Digg site with the help of Web wrappers, created with tools provided by Fetch Technologies. We extracted 195 front stories. For each story, we extracted the submitter’s name, story title, time submitted, number of diggs the story received and the list of the first 216 users who dugg the story (15,742 unique users total). We also collected information about the top 1020 ranked users. For each user, we extracted the list of friends and reverse friends or “people who have befriended this user.”

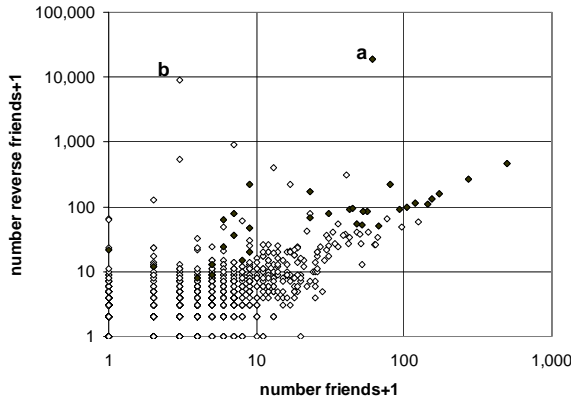


Figure 1: Scatter plot of the number of friends vs reverse friends for the top 1020 Digg users. Two of the biggest celebrities, kevinrose and diggnation, are marked a and b

3. SOCIAL FILTERING ON DIGG

To show that Digg users take advantage of the Friends interface to filter the tremendous number of new submissions, we analyze two sub-claims: (a) *users digg stories their friends submit*, and (b) *users digg stories their friends digg*.

Note that the “friend” relationship is not symmetric: if user A designates user B as a friend, user A can keep track of user B’s activities, but not vice versa. This makes A the *reverse friend* of B. Figure 1 shows the scatter plot of the number of friends vs reverse friends of the top 1020 Digg users as of May 2006. Black symbols correspond to the top 33 users. For the most part, users appear to take advantage of Digg’s social networking feature, with the top users having bigger social networks.

Users digg stories their friends submit. We compare the list of users who dugg the story, or any portion of it, with the list of reverse friends of the submitter. Submitter’s name is the first on the list. Figure 2 shows the number of diggers who are also among the reverse friends of the submitter. Dashed line shows the size of the social network (number of reverse friends) of the submitter. More than half of the stories (102) were submitted by users with one or more re-

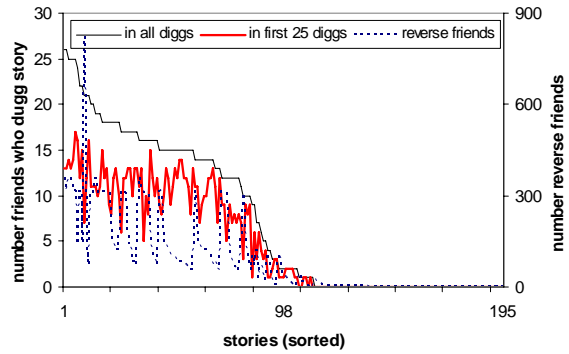


Figure 2: Number of diggers who are also among the reverse friends of the user who submitted the story

verse friends, and the rest by unknown users. We use simple combinatorics to compute the probability that k of the submitter’s friends could have dugg the story purely by chance. The probability that after picking $n = 215$ users randomly from a pool of $N = 15,742$ you end up with k that came from a group of size K is $P(k, n) = \binom{n}{k} (p)^k (1-p)^{n-k}$, where $p = K/N$. Using this formula, the probability (averaged over stories digg by at least one friend) that the observed numbers of friends dugg the story by chance is $P = 0.005$, making it highly unlikely.

Moreover, users digg stories submitted by their friends very quickly. The heavy solid line in Figure 1(b) shows the number of reverse friends who were among the first 25 diggers. The probability that these numbers could have been observed by chance is even less — $P = 0.003$. We conclude that users digg stories their friends submit. A consequence of this conclusion is that users with active and large social networks are more successful in getting their stories promoted to the front page. We believe that this explains the success of top users.

Users digg stories their friends digg. Do social networks also help users discover interesting stories that were submitted by unknown users? In other words, do users digg stories their friends like?

We looked at the 25 diggs that came after the first m diggs to see how many came from friends of the first m diggers. Of the stories posted by “unknown” users, ten were dugg by submitter’s reverse friends ($p = 0.005$). After five more diggs ($m = 6$), 75 became visible to others through the friends interface, and of these 23 ($p = 0.028$) were dugg by friends. After 15 users dugg the story, 94 are now visible and 37 ($p = 0.060$) are dugg by friends. After 25 diggs, all 96 stories were visible, and almost half of these were dugg by friends ($p = 0.077$). The probabilities that these many friends could have dugg the story by chance are above the 0.05 significance level for after 25 diggs, possibly reflecting the story’s increased visibility on the front page. Although the effect is not quite as dramatic as one in the previous section, we believe that the data shows that users do use the friends interface to find new interesting stories.