

Rigg

Gaming Within Digg

Srikanth Narayan, Isaac Salier-Hellendag, Xiaomeng Zhong
UC Berkeley School of Information
i247 – Information Visualization and Presentation, Spring 2008

Project Goals

Digg is a social content aggregation website. In the company's own words, they're trying to create "a place where people can collectively determine the value of content" and ultimately, to change "the way people consume information online". The concept of social content aggregation is quite a novel one, introduced by Digg itself a few years ago. Unsurprisingly, the site has grown on to become immensely popular since then.

In a social content aggregation website such as Digg, users submit news articles, webpages, videos, or anything that they find interesting. These submissions are reviewed by other users of the site, and if they find it interesting, they "digg" it, which is a form of a positive vote. Stories that receive sufficient votes bubble up to the frontpage, a part of the website that is accessed by its millions of users. This process is continuous and never-ending, providing a steady churn of stories that surface to the frontpage. In essence, it relies on the community using the website to collectively create this public good, may it be submitting stories, recognizing worthy stories or developing a conversation around these stories.

As a consequence of its natural appeal, Digg has grown on to become a very busy place. After the expansion of the website to include categories beyond 'Technology', the growth has been exponential. Currently, it has been speculated that Digg has around 2-3 million registered users [<http://www.jgc.org/blog/2008/01/how-many-users-does-digg-have.html>]. Most of these users visit the website to check frontpage stories, which means that a frontpage story would direct hundreds of thousands of hits to the website hosting that story. Naturally, getting a story on the frontpage has become extremely competitive, with more than 10,000 stories submitted by users everyday, and only two percent (roughly 200) of these stories making it to the frontpage.

The advantages of getting a story on the frontpage are manifold for a website operator. Firstly, it could prove quite lucrative, providing a generous spike in the ad revenue generated. Secondly, it provides free exposure, and highly effective marketing, with its reach to millions of users. In fact, startup companies and bloggers routinely target Digg to amass their first wave of users.

With these advantages laid out, it is hardly a surprise that people attempt to game submissions so that they reach the frontpage. By providing enough votes during the 'Upcoming' phase of a story, one could provide the necessary momentum to propel the story on to the frontpage. Since the early days of Digg, it has been widely suspected that people reach out to their social network to garner votes for their submissions. A person who submitted a new story could notify friends of this submission using outside media such as email, and get the necessary votes. With the rise of Digg, these attempts became much more pointed. Enterprising members of the community

started services such as User/Submitter[<http://www.usersubmitter.com/>], SpikeTheVote[<http://www.spikethevote.com/>], Subvert and Profit[<http://www.subvertandprofit.com/>] and many other services that promised frontpage glory in exchange for money. We derived our initial hypothesis from this, that there are latent social networks that work to push stories on the frontpage, and that one could detect these latent social networks from an analysis of behavior across stories from a particular submitter.

Another side of this issue is with the Top 100 users of Digg. Time and again, it has been verified that these users exert an unnatural control over the content in Digg. The Top 100 users are collectively responsible for 50-60% of the stories that make it to the frontpage, and the Top 20 are responsible for around 30% of the stories. While these users usually do not have any motives in promoting a particular website, they intend to promote stories they submit for popularity and respect among other Digg users.

Recently, the milieu of Digg was given a major overhaul by the operators of the website. This included social features such as sending 'shouts' to friends in one's network, and many others. After these features were introduced, people who try to game the system concluded that the best way to game Digg would be to grow one's network within the website, and reach out to this network to exchange votes for stories submitted. This practice is rampant within Digg today, as evidenced by profile pages of top users, which is often filled with a flurry of requests to vote on stories submitted. In deciding the focus of our project, we decided that this would be more interesting to visualize than our initial hypothesis, as it is a much more prevalent problem.

After setting the focus of our project, we had decide on the goals of our visualization. We decided that the visualization should be able to bring out patterns in gaming across categories, within the upcoming and the frontpage sections, and across stories. We wanted to provide an unified view that communicated how stories accumulate votes in their Upcoming phase, how they are received on the frontpage, and what the history of each one of these stories is. The viewer should be able to discern stories with high/low percentage of Friend Diggs easily, and should be able to query for additional details about any of these stories.

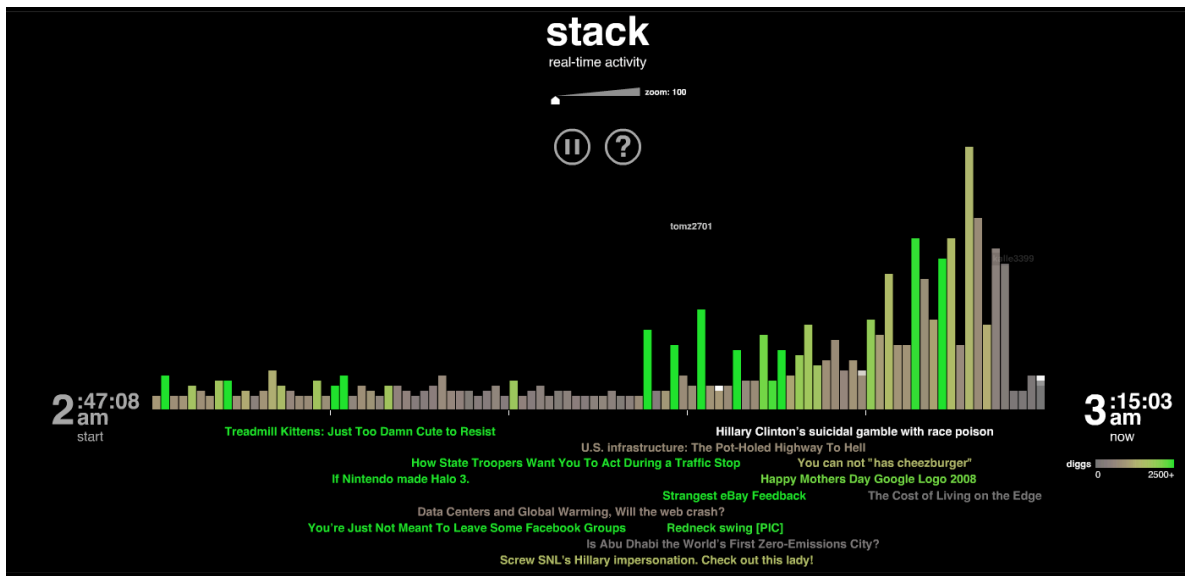
Related Work

Digg Labs

Digg Labs is a collection of visualizations of real-time Digg data, created for the purposes of providing a "broader (and deeper) view of Digg." While our initial intent in this project was not to display or animate Digg activity in real-time as these do, Digg Labs visualizations are exceptional in their style and ability to pique interest in the viewer. This is, in part, because they depict Digg as a living, active community. Static visualizations of Digg activity could provide the deeper understanding we seek, but our experience with Digg Labs encouraged us to embrace the dynamic nature of Digg as our ideas and iterations evolved.

One of the visualizations, "Stack", served as inspiration for our visualization of the upcoming queue. Each story is represented by a block, colored to reflect digg volume and labeled with a story title. Incoming diggs are symbolized by smaller blocks that fall onto stories from above. By

implementing a similar queue of stacks with diggs animated upward, we aim to show how incoming diggs help push upcoming stories "upward" onto the front page. The inverted "Stack" is a key component of the bubbling metaphor we employ in our visualization.



Digg Labs' "Stack" - each bar represents a story, incoming diggs fall from above

Social Blade

Our project is not the first to compile data related to the phenomenon of collusion on Digg. Social Blade is a website that collects Digg data for display in tabular, sortable format. It specifically focuses on several factors that we find most compelling in our own study of Digg: total diggs, friend diggs, time required for a story to reach the front page, and other factors of user popularity. The website does not provide further analysis of the data, so we carried out our own analysis in Tableau as we developed our ideas and hypotheses.

#	Story	Diggs	F Diggs	%FD	Buries*	Comments	Wait Time	DPH	User	Profile Views	Submits	Recent FPAs*	SubCategory	Category	Popular Time	Delay
108	Audi R8 Test Drive (Iron Man Car)	82	88	1.07	1	17	00:05:17:13	16	kevinrose	801314	359	9	Autos	Lifestyle	May 10 2008 18:22:07	47
141	Australia: Biological Weapons To Control Cane Toad Invasion	41	39	0.95	0	0	00:14:12:36	3	vick3ii	9565	389	27	Pets & Animals	Offbeat	May 10 2008 10:27:03	43
62	iC Mad PrideK™ Fights a Stigma	60	56	0.93	1	2	00:21:50:08	3	petsheep	22294	1260	31	Health	Lifestyle	May 11 2008 08:19:54	61
60	Sexy orchids do more than embarrass wasps	57	52	0.91	1	0	00:15:35:30	4	bamafun	21432	976	41	General Sciences	Science	May 11 2008 08:33:07	0
47	Bendito Machine	54	49	0.91	0	2	01:08:16:35	2	reflex768	11570	493	23	Comics & Animation	Entertainment	May 11 2008 10:16:25	102
148	U.S. Panel Wants King Statue Reworked	71	64	0.9	0	1	00:23:37:03	3	TalSach	13871	671	58	Arts & Culture	Lifestyle	May 10 2008 10:00:30	0
154	10 Everyday Financial Challenges (& How to Overcome Them)	103	92	0.89	0	1	00:17:23:53	6	skored	45356	1903	25	Business & Finance	World & Business	May 10 2008	34

Social Blade – Tabular collection of Digg data, including friend diggs

Visualization

Initial Concept

When we initially began discussing our ideas about how to produce a deeper level of analysis of Digg interaction, we focused on the potential for latent networks to form among users of the service. We hypothesized that users would digg each other's stories and implicitly form groups based on shared interests or other similarities.

In order to analyze these networks, our initial visualization concept consisted of a network structure, with nodes to represent individual users. Nodes would be connected by directional edges that would indicate diggs by each user to other users' stories. By combining nodes and edges into a network graph, we hoped to identify close-knit groups of users that consistently digg each other's stories. We would then aim to pick out influential users, opinion leaders, and areas of interest most likely to develop these latent communities.

Further research into the Digg community, however, led us down a different path. We found that while there may be latent activity in the system as we had hypothesized, many (if not most) of the diggs traded between users are the result of explicit collusion.

Shifting Focus

As described in the introduction to this paper, there are dozens of regulation-skirting methods Digg members use to push their stories to the front page. With help from the Social Blade dataset, along with Tableau and ManyEyes, we found that these methods do play a significant role, consistently promoting certain users and stories within the system. We then decided to shift our focus to address the impact of this "gaming" by the Digg community, and in doing so, we adjusted our visualization concept as well.

Rather than break data into a user network graph, we decided to represent stories individually. In this way, we would be able to track the proportions of diggs by friends for each story, then follow certain stories throughout their stay on the upcoming queue and front page. In addition, while we did not originally intend to use real-time data for our visualization, we changed our plans when changing our focus. We felt that producing the ability to view Digg collusion in real-time would be an excellent way to analyze its impact, help inform the Digg community with salient data, and follow in the footsteps of the well established Digg Labs real-time visualizations.

Final Visualization Details

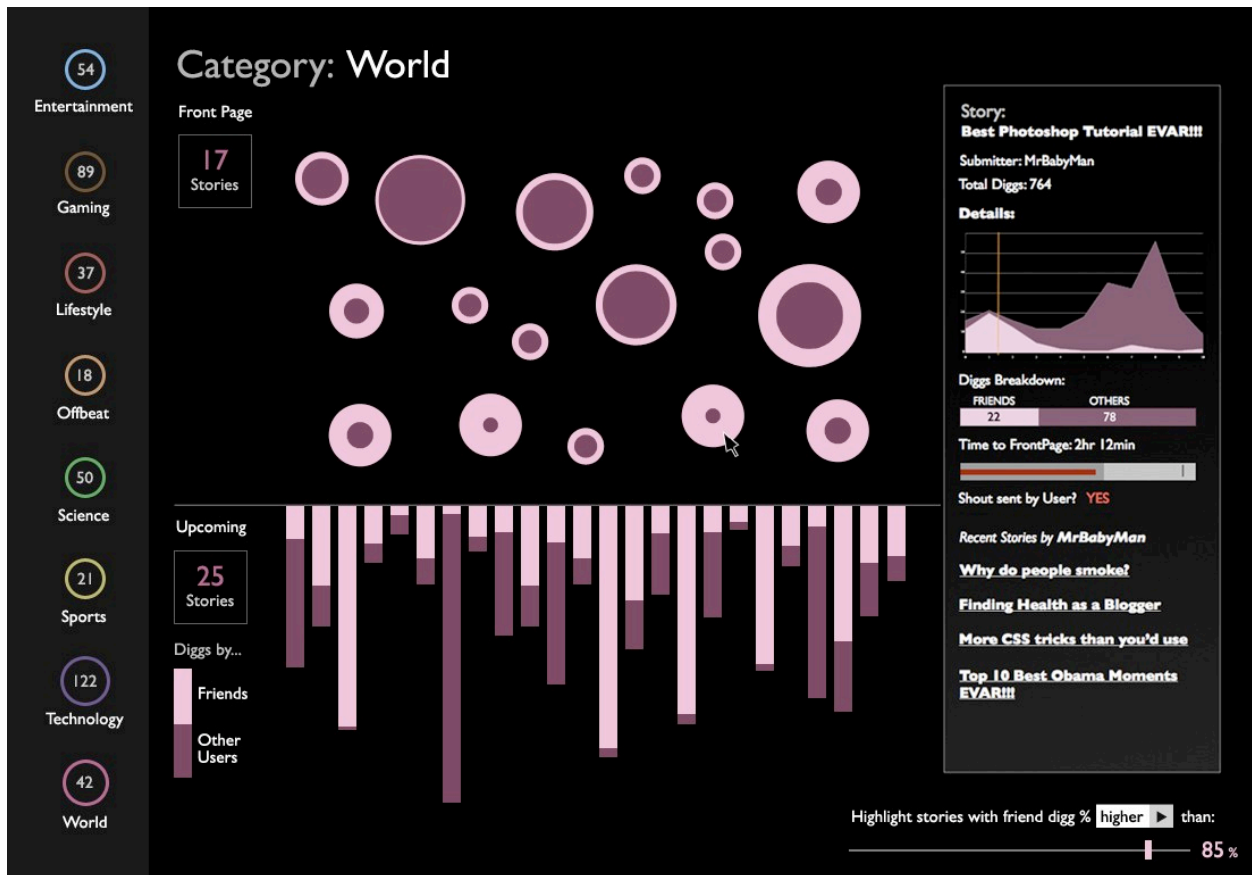
As mentioned before, our goal with the visualization was to provide a unified view of real-time activity on Digg, and aid the viewer in understanding patterns of gaming within Digg.

Because Digg is an extremely multivariate environment, it was initially hard to create a visualization that would explicate the goals discussed above. Refining to the set of variables that would aid us most in understanding gaming patterns in Digg was an imperative first step. The

variables we recognized as most important were:

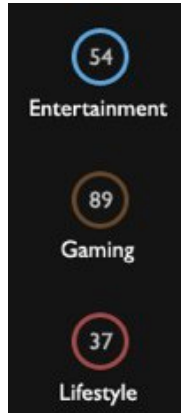
- Total number of Diggs for a particular story.
- Upcoming or Frontpage status of a story.
- Percentage of Friend/Non-Friend Diggs for each story.
- Rate of Diggs per hour for the story, both in the Upcoming and Frontpage sections.
- If the story made it to the Frontpage, total time spent in the Upcoming queue.
- Submitter username.
- Category to which the story was submitted.

The Visualization



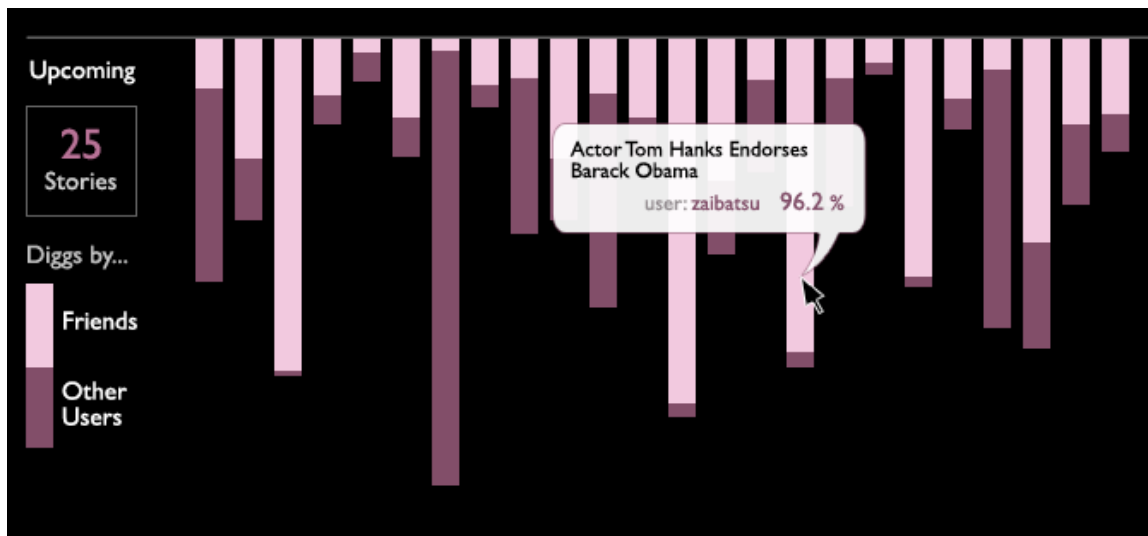
Shown in the above picture is a mockup of our visualization. The interface is broken down to four distinct sections, representing the categories within Digg, the Upcoming section, the Frontpage section and individual story details. Detailed explanation of each one of these sections and the visual elements contained within follows:

Categories



On the left, all the categories to which stories belong are shown. Each one is assigned a unique color, and the number within each circle represents the total number of frontpage stories for that category at any given point. Clicking on any of these circles brings up a detailed analysis for that category, visualizing the upcoming and frontpage sections for that category.

Upcoming



The Upcoming section is visualized using an inverted stacked bar graph. As mentioned before, this was inspired by the 'Stack' visualization done by Stamen for Digg Labs. Like the Stack, diggs fly in and stack on to the bars of the visualization. While the Stack visualization is mainly aimed towards discerning stories with high growth rates, the bars in our visualization correspond to the total number of votes that the story has gotten since submission. These bars are further broken down to represent Friend diggs and Non-Friend Diggs. A viewer of the visualization would be able to easily notice stories with large number of diggs and high percentages of Friend/Non-Friend diggs.

The Upcoming queue is usually a noisy environment, with hundreds to thousands of stories contending for the frontpage at any given point. A large number of these stories are mediocre and uninteresting, and most of these stories never reach a two-digit vote count. In visualizing the Upcoming queue, it would be important to weed out these uninteresting stories, and capture stories that are interesting in terms of high friend/non-friend digg percentage, growth rate or digg

count.

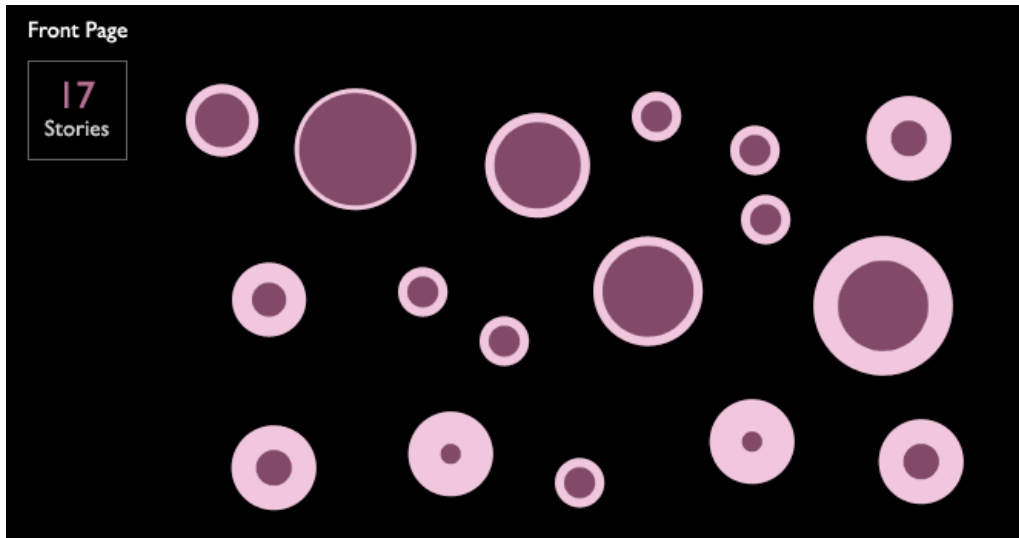
Frontpage

When stories in the Upcoming queue cross a certain threshold in Digg, they are promoted to the frontpage. There are hundreds to thousands of stories contending for the frontpage at any point, but only a few tens make it to the frontpage. Usually, stories that are promoted to the frontpage gain an average of 1000-2000 diggs and persist on the frontpage for around half a day. If the story is of particular interest to the community, it could receive tens of thousands of votes and persist longer than usual. If the story is not recognized as frontpage-worthy, it could be 'buried' and pushed off the frontpage quickly.

In our visualization, we indicate the transition from the Upcoming to the Frontpage using a bubbling metaphor. Stories that are promoted from the Upcoming queue bubble to the frontpage section of the visualization. In this section, each story is represented by a circle. The size of the circle corresponds to the total number of diggs for the story, making it easy to discern the most popular frontpage stories.

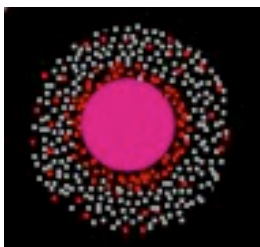
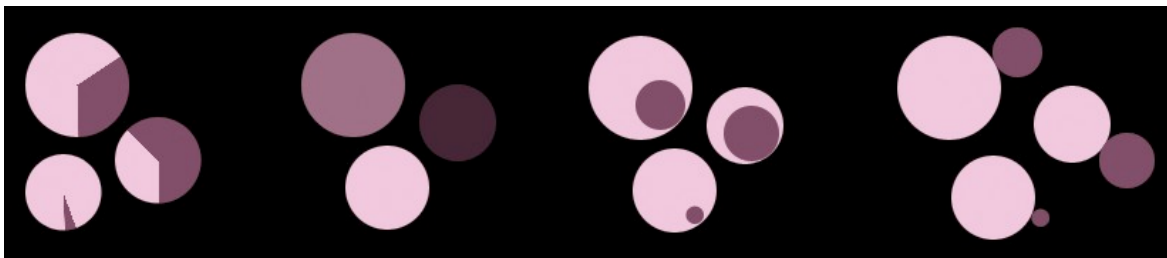
Though it could be contended that the frontpage visualization might be hard to interpret, employing a circle as the visual element to represent stories provides many advantages:

- As mentioned above, the number of diggs for a story could range from a few hundred to tens of thousands. If these raw numbers were to be represented using an element such as the bar graph, where the quantitative value is directly proportional to the length, outliers would suppress the average frontpage story. When using a circle, the quantitative value is proportional to the square of the radius. Therefore, these values undergo a square-root transform before being plotted, and can be used to represent a larger range.
- Carrying forward our bubbling metaphor, we could use the two-dimensional space of the frontpage section to make stories rise and fall within this section. The animation could help discern stories with high growth rates, or stories that are 'buried' quickly by Digg mobs.
- Again, because of the extreme amounts of activity on the frontpage, it offers us a slew of variables to visualize. Visualizing the frontpage activity was definitely the most debated part of our project, and we had to come up with tens of mockups before deciding on a particular visualization.

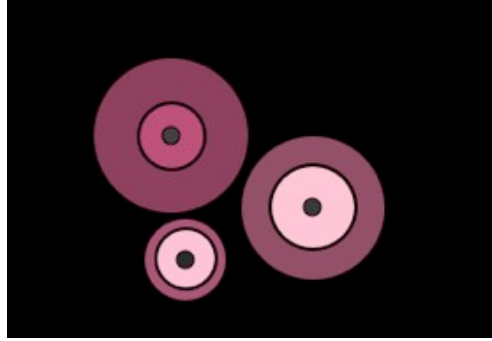


In the mockup shown above, each circle represents the percentage of Friend/Non-Friend digs for the particular story. For genuinely interesting stories, gamed or not, the percentage of Non-Friend digs would gradually increase to cover (almost) the entire circle, whereas uninteresting gamed stories wouldn't experience such a change.

After several users complained during the user testing that the percentages might be hard to discern with the visualization shown above, we experimented with pie chart, placing one circle at the periphery of another, or placing them next to each other.

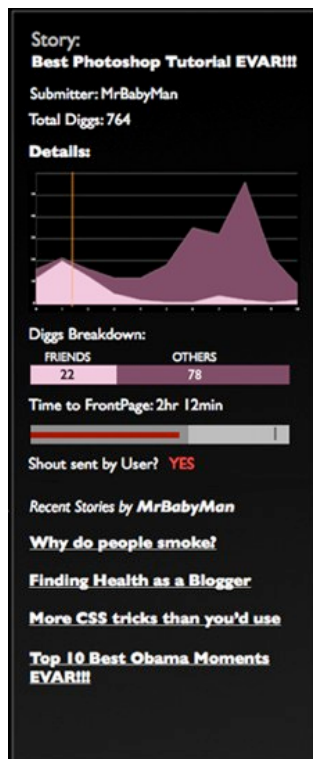


One particular problem with the above visualizations was that it did not communicate the history of a story well. In fact, during our user testing, users were surprised to learn that there wasn't any provision to tell if a particular story was gamed or not from looking at the visualization at a certain point in time. An early attempt to correct this used dots to represent individual diggs. These dots were differentiated by color to represent Friend/Non-Friend digg. One could potentially see dots of certain color clustering around stories, and greater number of dots as the story became more popular.



Another attempt to incorporate the history of a story into the visualization is shown above. In this visualization, the circle in the center represents a story, and is the same size and color for every story. Around this circle is a ring that represents the Upcoming phase of a story. The thickness of this ring is proportional to the number of digs it received in the Upcoming queue, and the color corresponds to the percentage of Friend Digs. A Frontpage ring grows around this ring during the lifetime of the story, again encoding the same variables as the Upcoming ring. Using this visualization, one could make small multiple comparisons of the statuses of these stories, quickly discerning the relative popularity, and the probability that the story was gamed. The use of the small dummy circle in the center makes it easy to compare relative thickness of the two rings of a story. If the Upcoming phase were to be represented using a whole circle, it would have proved harder to do this relative comparison. While this comparison might not be beneficial for stories that were received well on the Frontpage, it is definitely helpful with stories that are struggling on the Frontpage.

Extended Details Panel



Clicking on a story in the Upcoming or the Frontpage sections brings up an Extended Details panel for that story. This panel provides a rigorous analysis of the variables around a particular story, and aids the viewer in completing what was observed in the Upcoming or the Frontpage sections.

- Within the Extended Details panel, relevant metadata pertaining to the selected story such as story title, username, and total digs are displayed.
- In addition, the panel includes a layered graph to show rate of digs for the story over time, further broken down to indicate distribution of friends and non-friends digs. A conspicuous visual cue, shown as an orange line in the graph, denotes the point at which the particular story transitioned from Upcoming to the Frontpage section. This graph could potentially show interesting patterns with respect to how stories are received on the Frontpage. For example: An uninteresting gamed story might fall off steeply once it hits the frontpage, whereas an interesting albeit gamed story

would sustain to receive non-friend votes.

- Other visualizations contained within this panel include a simple ratio chart, and a bullet graph that denotes the single variable, Time to Frontpage, in comparison with the average time.
- The panel also denotes whether "shouts" were sent by the user after submission, and provides a list of other stories submitted by this user, offer quick links for comparison to reveal trends.

User Evaluations

After designing the initial visualization mockup, we interviewed four users to get feedback and suggestions. Overall, our users liked the visualization. They found it to be informative. They felt that breaking the visualization into different views and panels allowed the application to successfully display large amounts of information simultaneously in one screen. They appreciated that the design aims to highlight potential gaming activities in the upcoming queue, since those actions would often be otherwise hard to spot during normal use of the site. In addition, the Extended Details Panel received high marks from all the users we interviewed. They liked that the small panel highlights the relevant information on a more granular scale, especially the graph comparing friend and non-friend diggs before and after the story reached the front page. Finally, our users enjoyed the aesthetics of the visualization, and found the rising bubble metaphor to be interesting.

We also received many constructive suggestions from the users during the evaluation. Most of the comments are directed to the upcoming and the front page views.

Upcoming Queue

Inverted Bar Graph

Two users mentioned that the inverted bar graph could require some time to get used to. They felt that it might be counter-intuitive to see bars grow toward the bottom of the page despite potentially being closer to jumping to the front page.

Front Page

Circles

One of our biggest challenges in designing the visualization was to try to elegantly combine several variables within the circles on the front page section. Three of the users we interviewed expressed difficulties in accurately comparing the percentages of friends and non-friends diggs within a circle. They felt that concentric circles imply an overlapping of the two circles. Of course, this was not our intention -- the inner circle reflects non-friend diggs, and the outer ring indicates friend diggs. It is also difficult to determine the area of the outer ring in a glance, which makes it very hard to discern the ratio of friends to non-friends. In addition, since the circles are intended to represent current distribution of diggs, two users felt that the circles would lose the

friend/non-friend data collected during the story's growth in the upcoming queue. Indeed, we anticipate diggs coming from non-friends much more than from friends once a story reaches the front page. This might mean that the circles would slowly converge to look very similar over time, and the distinction between gamed and non-gamed stories could disappear. Finally, two users found the movement of circles could be confusing and easy to ignore without explicit understanding that vertical movement conveys any information. As one of our users noted, vertical movement for front page story circles might just end up looking "like a Lava Lamp," with no clear understanding of what the movement entails.

Tracking stories

Each user felt it would be important to track stories across the front page and the upcoming views, which we did not initially support in our visualization mockup. They felt that when upon noticing a potentially gamed story, they would want to be able to see how it progresses if it were to reach the front page. Without any form of story tracking, it is easy to lose stories when they cross the threshold to the front page.

Filtering data

Since the visualization embodies large amount of information, all of our users felt it would be challenging to focus on the stories they cared about. They would prefer to have a means to adjust each view to make some more prominent than others and avoid distraction from less relevant data.

Bar-to-bubble transformation

One user noted that the bubble metaphor could be considered inconsistent. This user felt that the change from bar to circle could be difficult to grasp, partly because of the inverted bar graph issue, and partly because it could make it harder for users to track stories.

Labeling gamed stories

Three out of the four users wanted the visualization to set an established threshold and explicitly mark gamed stories. They felt the application's objective is to highlight gaming activities in Digg's system, so they would like to see that as clearly as possible. This could also help the visualization appear less busy, and bring the most relevant stories to the forefront.

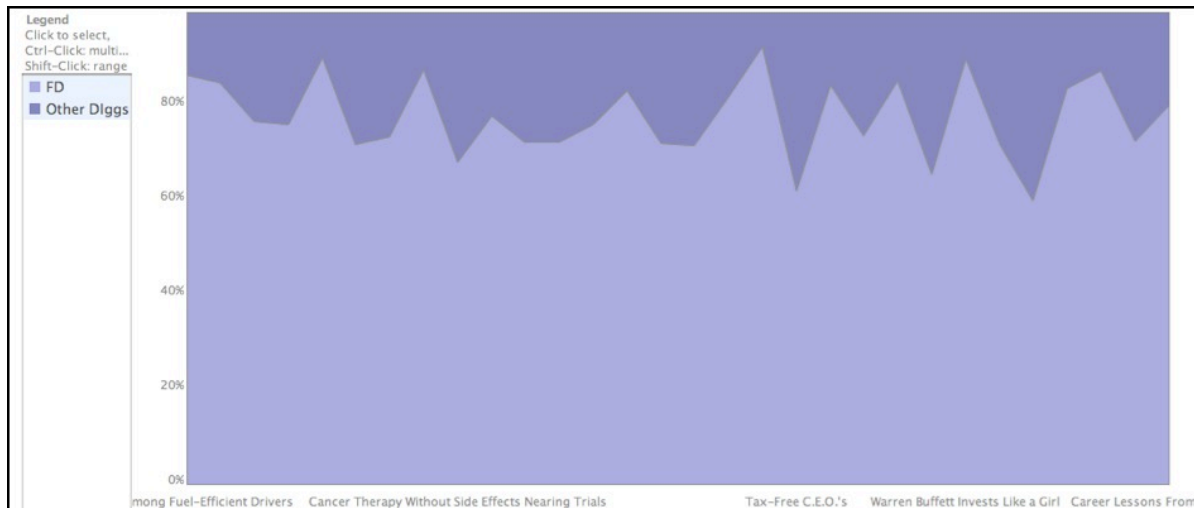
Dataset

One dataset that we initially found helpful with the Hypothesis Validation was the Social Blade dataset. For the rest of the development of the application, we envision using the Digg API[] by Stamen to query real-time data from Digg. The Digg API provides a REST interface that can be used to query (almost) all actions on Digg, and receive it in an easily processable JSON or XML.

Tools Used

During the initial evaluation of our hypothesis, we used a data set provided by Social Blade (csv format) within ManyEyes and Tableau to get an overall sense of how often friend diggs had a significant impact on story movement. We also looked at the specific users who most often use this strategy, and whether their influence often pushes their stories to the front. In short, we

wanted to get a static view of real data we would later develop into our real-time visualization.



ManyEyes visualization - % friend/non-friend diggs by story

To create our visualization mockup, we first created rough paper prototypes to flesh out our ideas on how best to illustrate the variables we wanted to address. We then developed our prototype predominantly in Photoshop, with some design elements created in Omnigraffle.

To build a fully functional, animated visualization with real-time data, we chose to work with ActionScript 3 in Adobe Flex Builder, using the Flare toolkit.

Steps to Achieve Goals

Previous Work and Preliminary Digg Analysis

An essential part of our research was to gain a better understanding of the milieu of Digg. Since none of us were regular users of Digg, we spent considerable time trying to understand user behavior in this system. After carrying out an in-depth analysis of the system, we were better able to comprehend the system holistically and formulate our current hypothesis.

We looked at previous visualizations created by Digg Labs to get an overview on current Digg analysis. "Stack" was particularly interesting, since it was a nice way to visualize real-time digging activities on stories. We later incorporated the idea as part of the upcoming queue design. We also gained some inspiration from Digg Labs' "Swarm" and the We Feel Fine application, most notably the way in which the animated circles indicated real-time responses and the intensity of the user activities.

In addition, we gathered an upcoming-queue dataset which highlights the percentage of friends and non-friends diggs. We plotted it in Tableau/ManyEyes to get initial overview of the data. This helped us uncover some salient information, and helped provide an initial validation of our hypotheses. For instance, we found some of the top users have an average of 80%-90% friend diggs for their stories.

Design

Brain-storming

Focusing on the refined set of hypotheses, we had multiple meetings where we brain-stormed visualization goals and design ideas. We identified all the factors we wished to include for the visualization, and drew up several paper prototypes.

Prototyping initial mockups

Collating ideas from our paper prototypes, we used Photoshop and made the initial mockups. We focused on one category and showed views for both the upcoming queue and front page stories. In addition, we added a vertical navigation bar to serve as an access point to other categories, and the Extended Details Panel, which gives complete overviews of the selected stories.

User evaluation

We interviewed four users to get feedback on the initial mockup. Two of them were ISchool students, and the other two were working professionals.

Iteration of design

Utilizing the feedback from user evaluations, we improved the initial mockup. We added filtering and tracking features to the current views, and based on user suggestions, we developed several new ideas for designing the front page story circles.

Coding

This is the part we are hoping to accomplish during the summer. None of us have previous experience in programming in ActionScript. However, we have done some quick prototyping using it and had some simple animation working. We also hope to refine and broaden the design as we progress in coding.

Testing

Once we have a working application finished, we hope to conduct a larger user testing session to observe its effectiveness. The results will lead us to improve and finalize the application.

Future Work

To reiterate what was mentioned before, the first task that awaits us is to program the application and test our hypotheses with real-time data. In addition to that, there might be several other avenues that might be worth exploring. One possibility is to look at the data we have from other perspectives, so that we can get a deeper understanding of how users are trying to game the system.

- A particular behavior we observed during our initial research on Digg is that several users visit their 'Friend's Activity' page, and vote on stories without reading them and assessing its interestingness. Such a user would essentially produce a large number of diggs in a short duration, and it might be interesting to create a visualization that detects this particular occurrence by looking at the feed of diggs.

- To get an understanding of how top users exert an unnatural control in the system, it might be helpful to visualize daily activity of these users and compare it with average users. Interesting variables to visualize would be number of visits to the website, timing of submissions, timing of shout activity and digg reciprocation.
- Another interesting path to explore would be to visualize changes that particular functionality can bring about to the system as a whole. ‘Shout’ & some other social features were added in Sep 2007, influence that such changes in functionality could have on the user behavior would provide much insight into the design of these systems.
- Further refining our existing visualization to present additional information regarding gaming across categories, trends over time or with respect to specific users.

Links to Documents and Demos

Midpoint Presentation can be found at - <http://people.ischool.berkeley.edu/~srikanth/rigg/RiggMidpoint.pdf>

Final Presentation can be found at - <http://people.ischool.berkeley.edu/~srikanth/rigg/RiggFinal.pdf>

Posters used for the Poster Exhibition can be found at - <http://people.ischool.berkeley.edu/~srikanth/rigg/PosterFinal.pdf>

Thumbnail

Thumbnail for the course webpages can be found at - <http://people.ischool.berkeley.edu/~srikanth/rigg/Thumb.png>