



Analysis of Social Voting Patterns on Digg

Kristina Lerman
Aram Galstyan
USC Information Sciences Institute
{lerman,galstyan}@isi.edu



Content, content everywhere and not a drop to read



- Explosion of user-generated content
 - 2G/day of "authored" content
 - 10-15G/day of user generated content

•How do users/consumers find relevant content?

How do producers promote their content to potential consumers?



Social networks for promoting content



Viral or word-of-mouth marketing

- Exploit social interactions between users to promote content
- But, does it really work?

Previous empirical studies have conflicting results

- Study showed popularity of albums did affect user's choice of what music to listen to [Salganik et al., 2006]
- Study showed recommendation might not lead to new purchases on Amazon [Leskovec, Adamic & Huberman, 2006]
 - Showed sensitivity to **type** and **price** of products



In this work



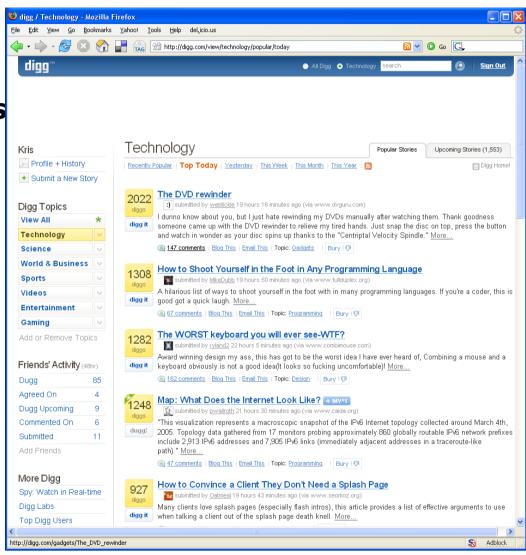
- •Do those results apply to free content?
- •How do social networks affect spread of free content?
- Empirical study on social news aggregator Digg



Social news aggregator Digg



- Users submit and moderate news stories
- Digg automatically promotes stories for the front page
- Digg allows social networking
 - Users can add other users as Friends
 - This results in a directed social network
 - Friends of user A are everyone A is watching
 - Fans of **A** are all users who are watching **A**





Lifecycle of a story



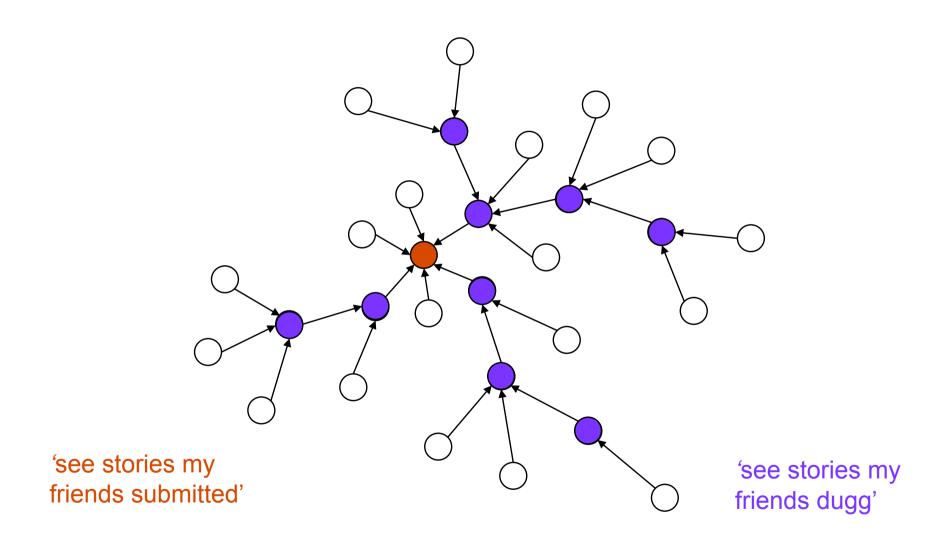
- 1. User submits a story to the Upcoming Stories queue
- 2. Other users vote on (digg) the story
- 3. When the story accumulates enough votes (diggs>50), it is promoted to the Front page
- 4. The Friends Interface lets users can see
 - 1. Stories friends submitted
 - 2. Stories friends voted on,

🐸 digg / Technology - Mozilla Firefox File Edit View Go Bookmarks Yahoo! Tools Help del_icio.us 👍 🕶 🚽 🔞 🔞 🚹 🔐 http://digg.com/view/technology/popular/today N ▼ (D) Gn [C] diaa All Digg
 Technolog Technology pcoming Stories (1,553 Popular Stories Profile + History Recently Popular | Top Today | Yesterday | This Week | This Month | This Year + Submit a N hours 16 minutes ago (via www.dvguru.com) Diga Topics o know about you, but I just hate rewinding my DVDs manually after watching them. Thank goodness View All someone came up with the DVD rewinder to relieve my tired hands. Just snap the disc on top, press the button Technology and watch in wonder as your disc spins up thanks to the "Centriptal Velocity Spindle." More... (a) 147 comments | Blog This | Email This | Topic: Gadgets | Bury | (7) Science World & Business How to Shoot Yourself in the Foot in Any Programming Language Sports submitted by MikeDubb 19 hours 50 minutes ago (via www.fullduplex.org) A hilarious list of ways to shoot yourself in the foot with in many programming languages. If you're a coder, this is Videos digg it good got a quick laugh. More... Entertainment (a) 67 comments | Blog This | Email This | Topic: Programming | Bury | ♥ Gaming The WORST keyboard you will ever see-WTF? submitted by ryland2 22 hours 5 minutes ago (via www.combimouse.com) Award winning design my ass, this has got to be the worst idea I have ever heard of, Combining a mouse and a Friends' Activity (48) digg it keyboard obviously is not a good idea(It looks so fucking uncomfortable)! More... (a) 162 comments | Blog This | Email This | Topic: Design | Bury | □ Duaa Agreed On Map: What Does the Internet Look Like? → MY=1 Dugg Upcoming submitted by pwallroth 21 hours 30 minutes ago (via www.caida.org) Commented On "This visualization represents a macroscopic snapshot of the IPv6 Internet topology collected around March 4th, 2005. Topology data gathered from 17 monitors probing approximately 860 globally routable IPv6 network prefixes Submitted include 2,913 IPv6 addresses and 7,905 IPv6 links (immediately adjacent addresses in a traceroute-like dd Friends (a) 47 comments | Blog This | Email This | Topic: Programming | Bury | (3) More Digg How to Convince a Client They Don't Need a Splash Page Spy: Watch in Real-time asubmitted by Oatmeal 19 hours 43 minutes ago (via www.seomoz.org) Didd Labs Many clients love splash pages (especially flash intros), this article provides a list of effective arguments to use digg it when talking a client out of the splash page death knell. More... Top Digg Users http://digg.com/gadgets/The_DVD_rewinder Adblock



How the Friends Interface works







Research questions



•What are the patterns of "vote diffusion" on the Digg network?

•Can these patterns in early dynamics predict story's eventual popularity?



Digg datasets



Stories

Collected by scraping Digg ... now available through the API

- ~200 stories promoted to the Front page on 6/30/2006
- ~900 newly submitted stories (not yet promoted) on 6/30/2006
- For each story
 - Submitter's id
 - Time-ordered votes the story received
 - Ids of the users who voted on the story

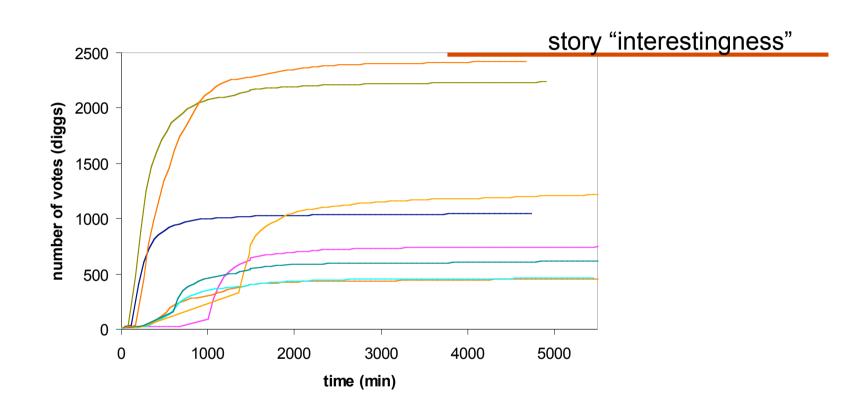
Social networks

- Friends: outgoing links A → B := B is a friend of A
- Fans: incoming links A → B := A is a fan of B
- Enables to reconstruct the diffusion process



Dynamics of votes



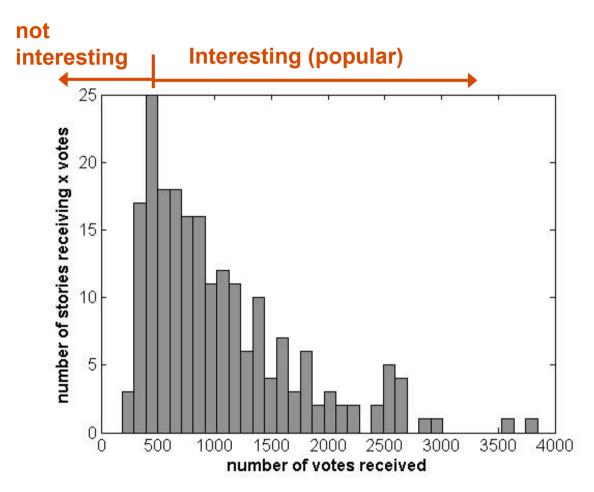


- Shape of the curves (votes vs time) is qualitatively similar
- Large spread in the final number of votes
 Implicitly defines the "interestingness", or popularity, of a story

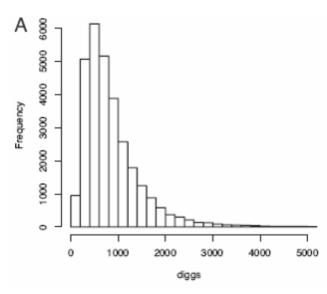


Distribution of votes





~200 front page stories submitted in June 29-30, 2006



Wu & Huberman, 2007

~30,000 front page stories submitted in 2006



Dynamics of voting on Digg



- Two main mechanisms for voting
 - Voting is influenced by intrinsic attributes of a story
 - E.g., some stories are more interesting and have more popular appeal than others
 - Voting is also impacted by social interactions (e.g, through the Friends Interface)
 - Diffusive spread on a network
- We can not measure "interestingness", but we can analyze the patterns of "social voting"
- Can we use those patterns to predict the eventual popularity of a story?



Patterns of network spread



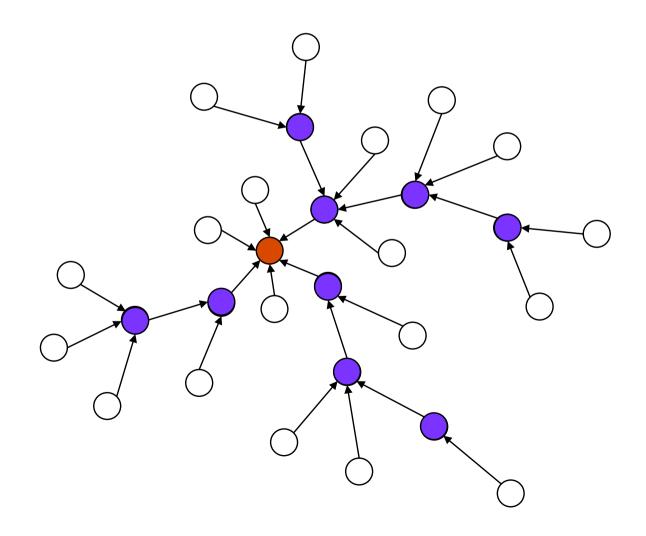
• Definition: *In-network votes* are votes coming from fans of the previous voters (including the submitter)



Patterns of network spread



 Definition: In-network votes are votes coming from fans of the previous voters (including the submitter)





Main Findings

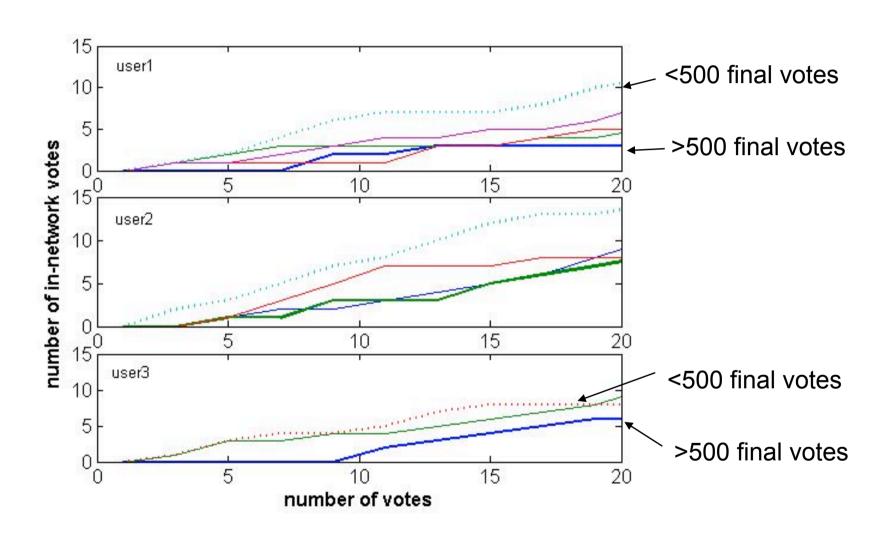


- Large number of early in-network votes is negatively correlated with the eventual popularity of the story
 - Stories receiving more in-network votes will turn out to be less popular
 - More interesting story receive fewer in-network votes



Stories submitted by the same user



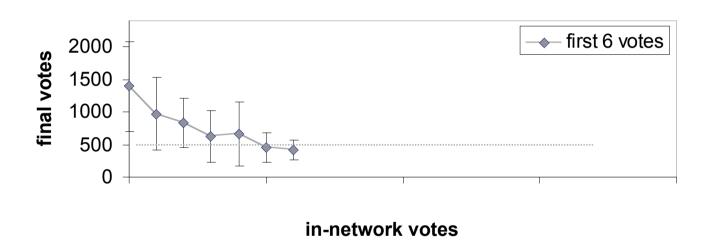




Popularity vs in-network votes



Popularity vs the number of in-network votes out of first 6

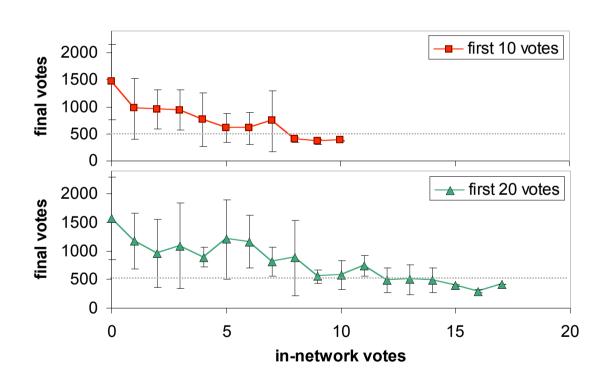


 The stories that become popular initially receive fewer innetwork votes



The trend continues



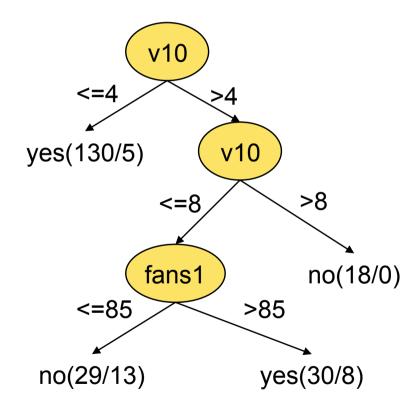




Classification: Training



- Predict how popular the story will become based on how many in-network votes it receives within the first 10 votes
- Decision tree classifier
 - Features
 - v10: Number of in-network votes within the first 10 votes
 - fans1: Number of fans of submitter
 - Story popularity
 - Yes if > 500 votes
 - No if < 500 votes





Classification: Testing



 Use the classifier to predict how popular stories will be based on the first 10 votes it received

Dataset

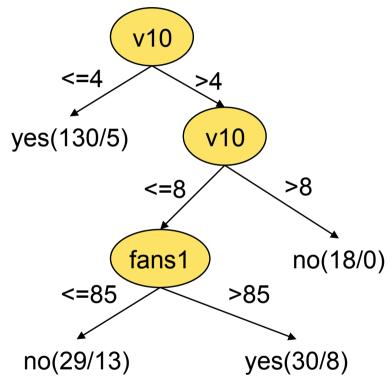
- 48 new stories submitted by top users
- Of these, 14 were promoted by Digg

Predictions

- Correctly classified 36 stories (TP=4, TN=32)
- 12 errors (FP=11, FN=1)

Compared to Digg's prediction

- Digg predicted that 14 are interesting (by promoting them)
 - Digg prediction: 5 of 14 received more than 500 votes
 - Digg prediction: Pr=0.36
 - Our prediction: 4 of 7 received more than 520 votes (Pr=0.57)
- Prediction was made after 10 votes, as opposed to Digg's 40+ votes





Summary



Social Web sites like Digg provide data for empirical study of collective user behavior

 How do social networks impact the spread of content, ideas, products?

Findings for Digg

- Patterns of voting spread on networks indicative of content quality
- Those patterns enable early prediction of eventual popularity

Future work

- More systematic and larger scale empirical studies
- Agent-based computational and mathematical models of social voting on Diggs

