

# Analysis of Social Voting Patterns on Digg

**Kristina Lerman**

**Aram Galstyan**

**USC Information Sciences Institute**

**{lerman,galstyan}@isi.edu**



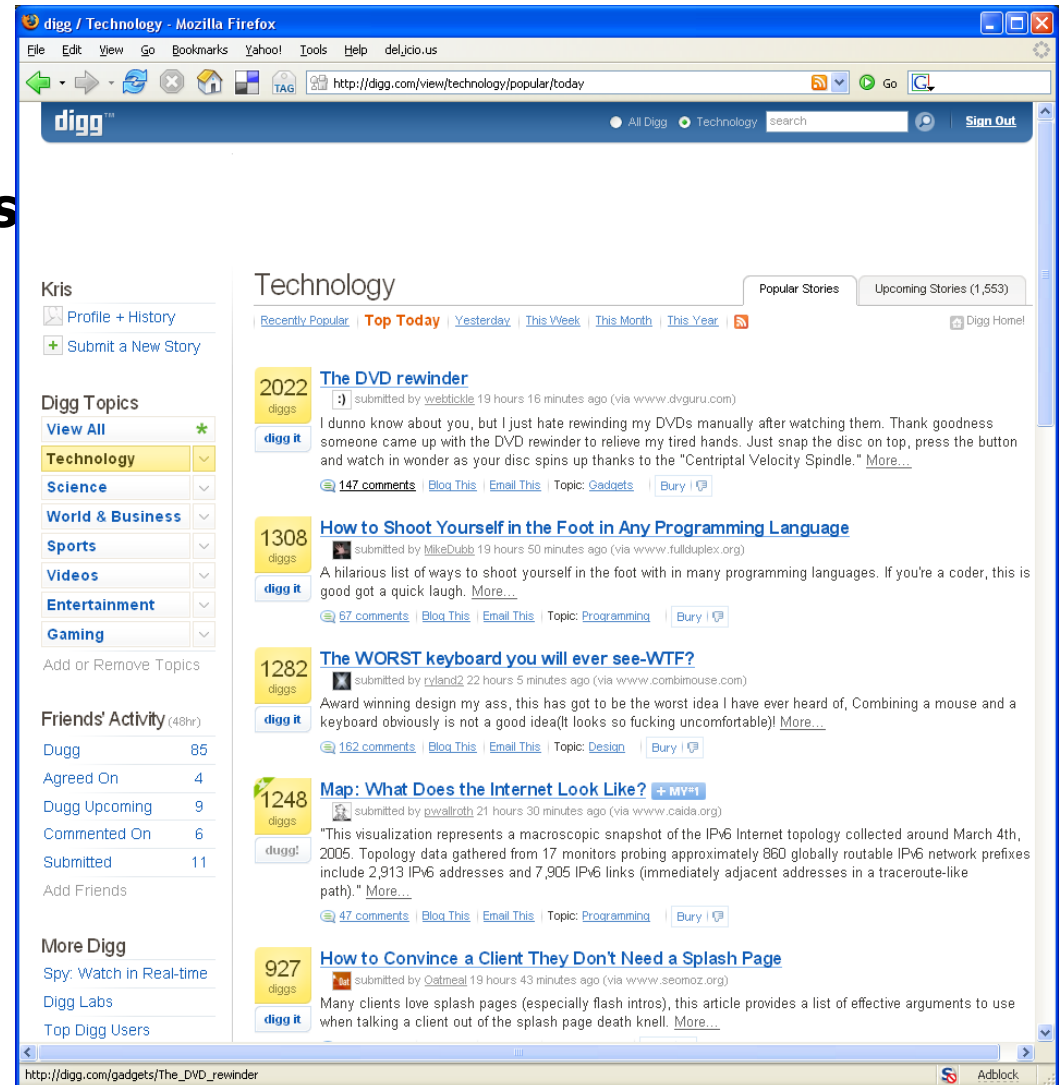
- **Explosion of user-generated content**
  - 2G/day of “authored” content
  - 10-15G/day of user generated content
- **How do users/consumers find relevant content?**
- **How do producers promote their content to potential consumers?**

- **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*

- **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**



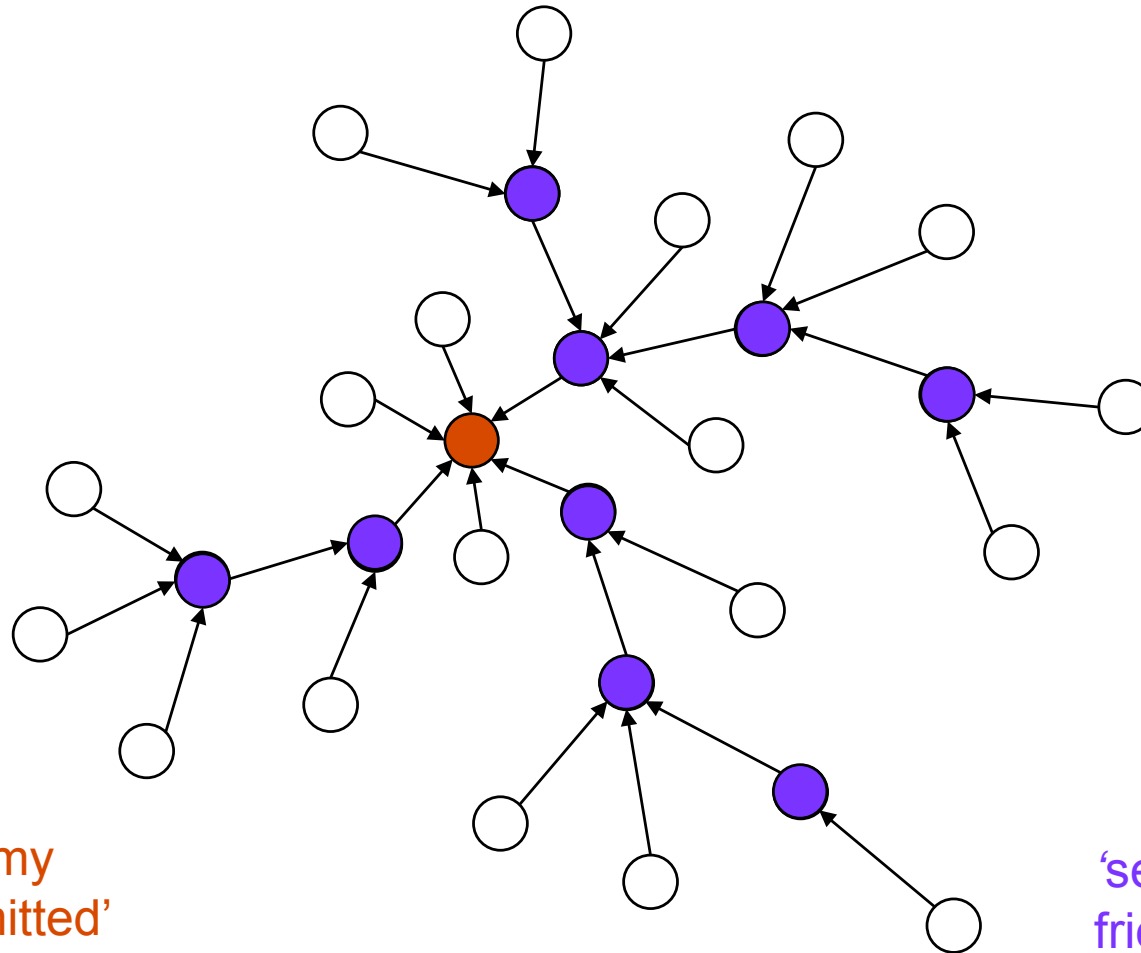
# Lifecycle of a story

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,
  - ...

The screenshot shows the Digg website interface. The top navigation bar includes 'All Digg', 'Technology', a search bar, and a 'Sign Out' button. The main content area is titled 'Technology' and features a list of stories. The 'Upcoming Stories' queue is circled in orange, and the 'Popular Stories' section is also circled. A story titled 'The DVD rewinder' is highlighted with a yellow box and a '2022 diggs' badge. The 'Friends' activity table is also circled in orange. Arrows point from the text on the left to these elements on the screenshot.

Activity	Count
Dugg	85
Agreed On	4
Dugg Upcoming	9
Commented On	6
Submitted	11

# How the Friends Interface works



'see stories my  
friends submitted'

'see stories my  
friends dugg'

- **What are the patterns of “vote diffusion” on the Digg network?**
- **Can these patterns in early dynamics predict story’s eventual popularity?**



- **Stories**

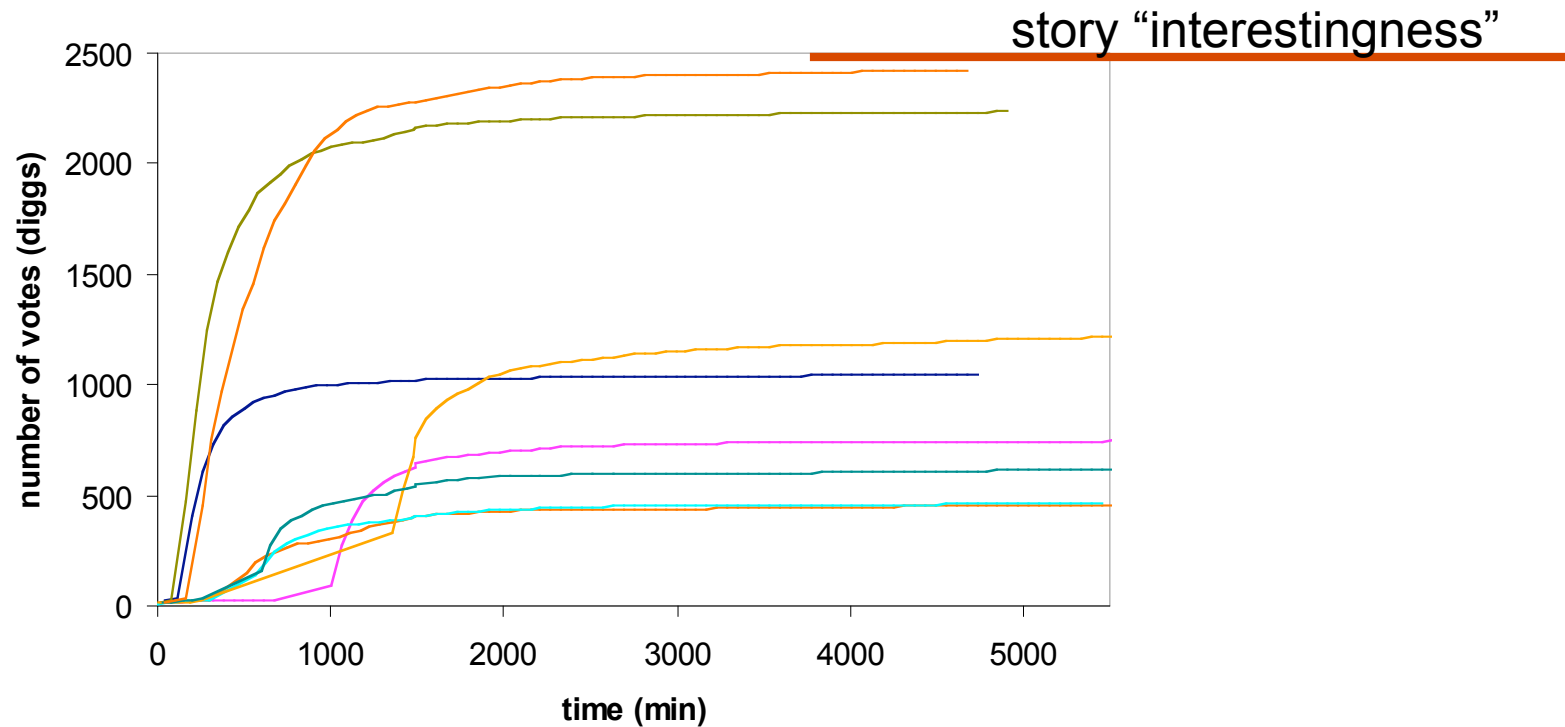
Collected by scraping Digg ... now available through the API

- ~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  $A \rightarrow B := B$  is a **friend** of A
- Fans: incoming links  $A \rightarrow B := A$  is a **fan** of B
- Enables to reconstruct the diffusion process

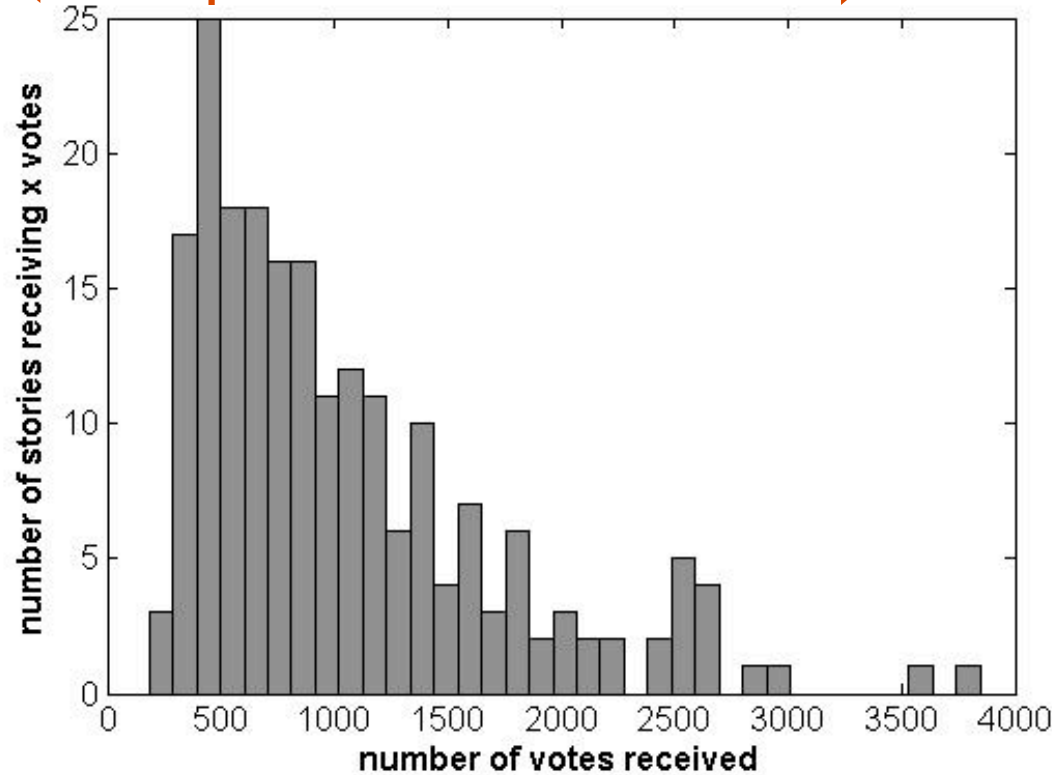
# Dynamics of votes



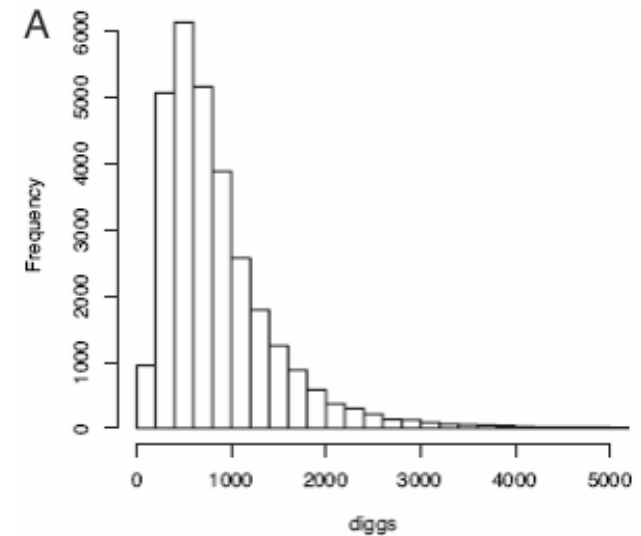
- Shape of the curves (votes vs time) is qualitatively similar
- Large spread in the final number of votes
  - Implicitly defines the "interestingness", or popularity, of a story

# Distribution of votes

not interesting ← | Interesting (popular) →



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



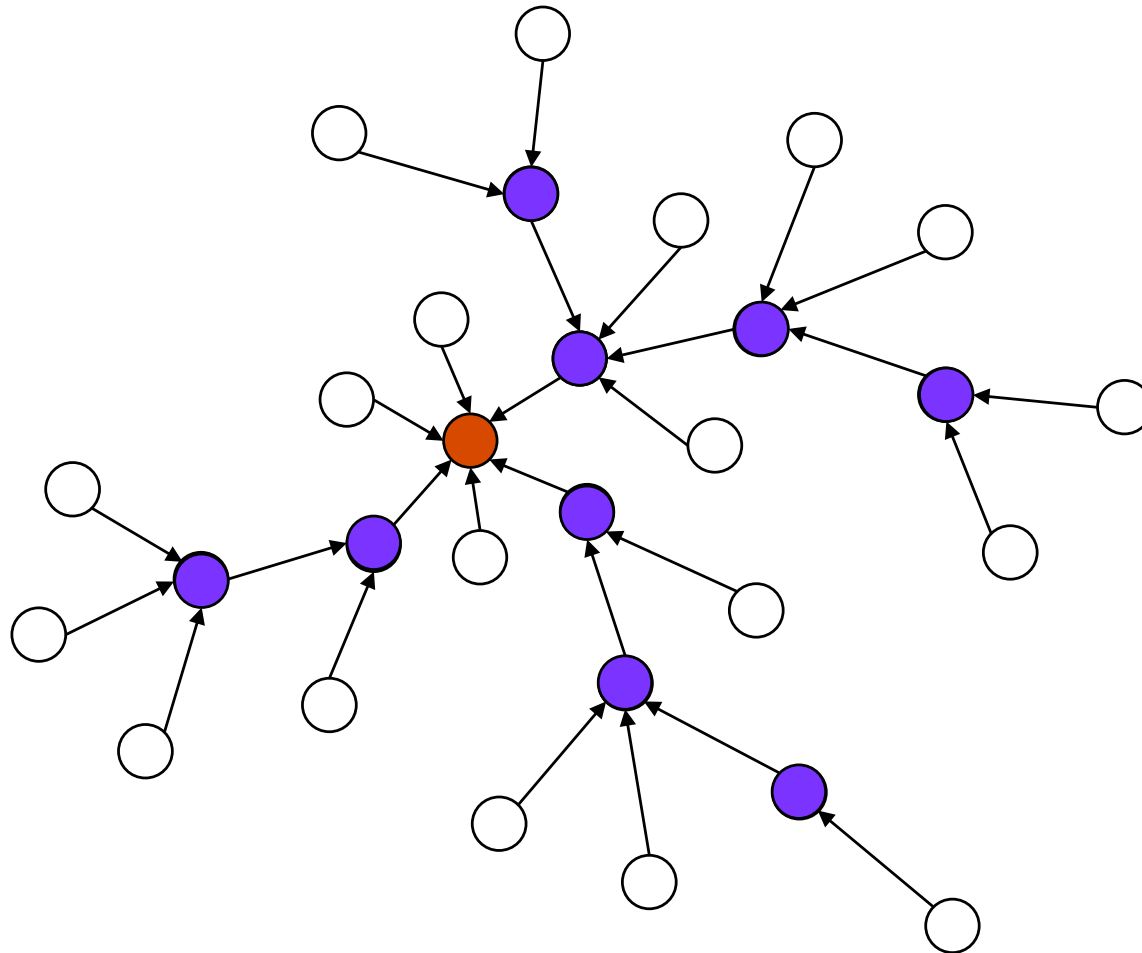
Wu & Huberman, 2007

~30,000 front page stories submitted in 2006

- **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?**

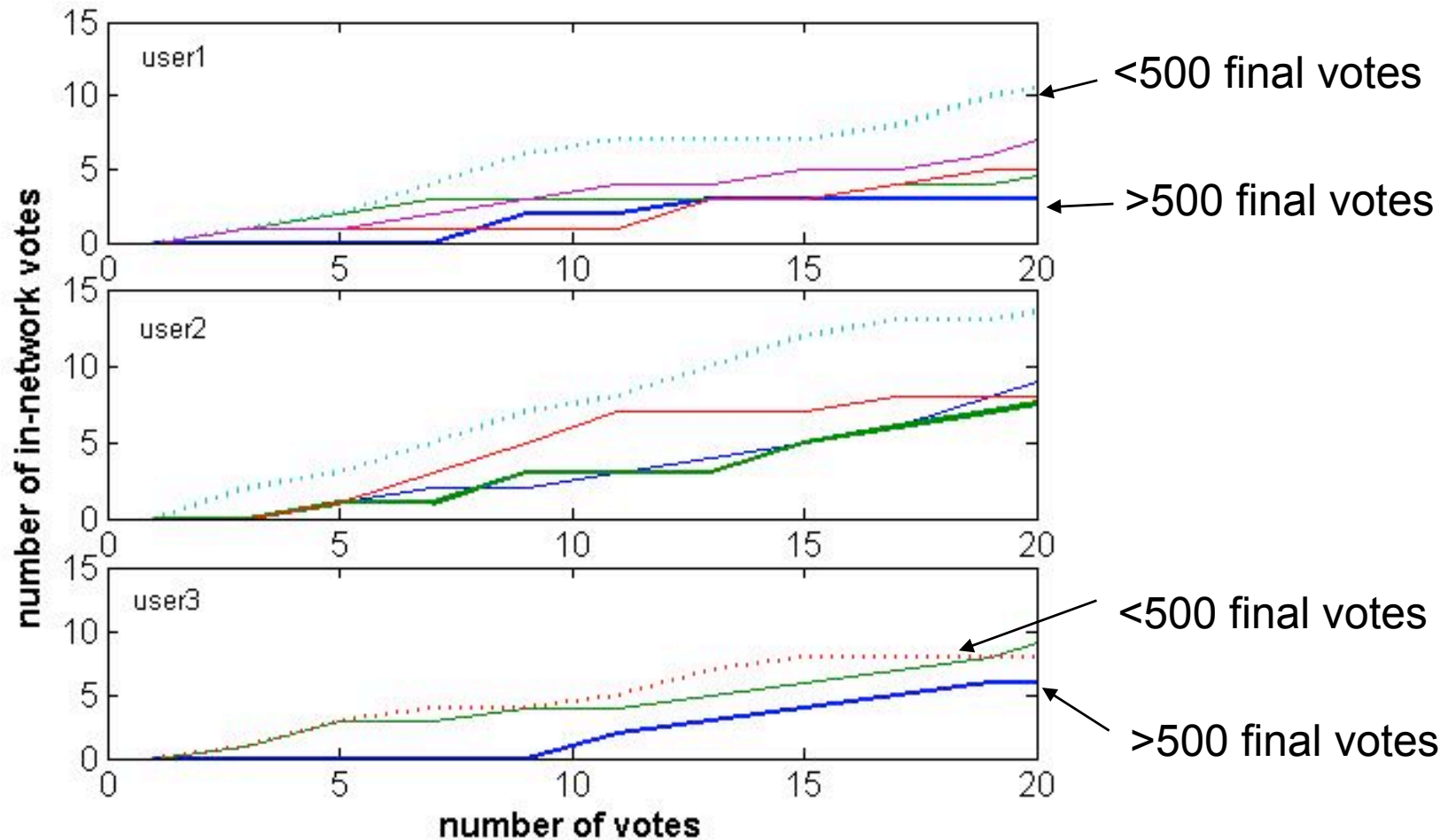
- **Definition:** *In-network votes* are votes coming from fans of the previous voters (including the submitter)

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



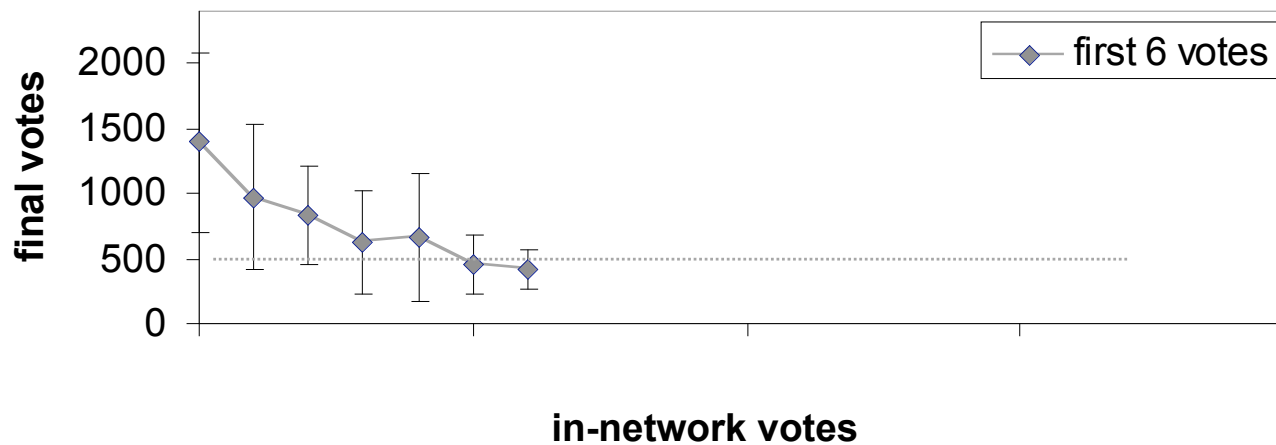
- **Large number of early in-network votes is negatively correlated with the eventual popularity of the story**
  - Stories receiving more in-network votes will turn out to be less popular
  - More interesting story receive fewer in-network votes

# Stories submitted by the same user



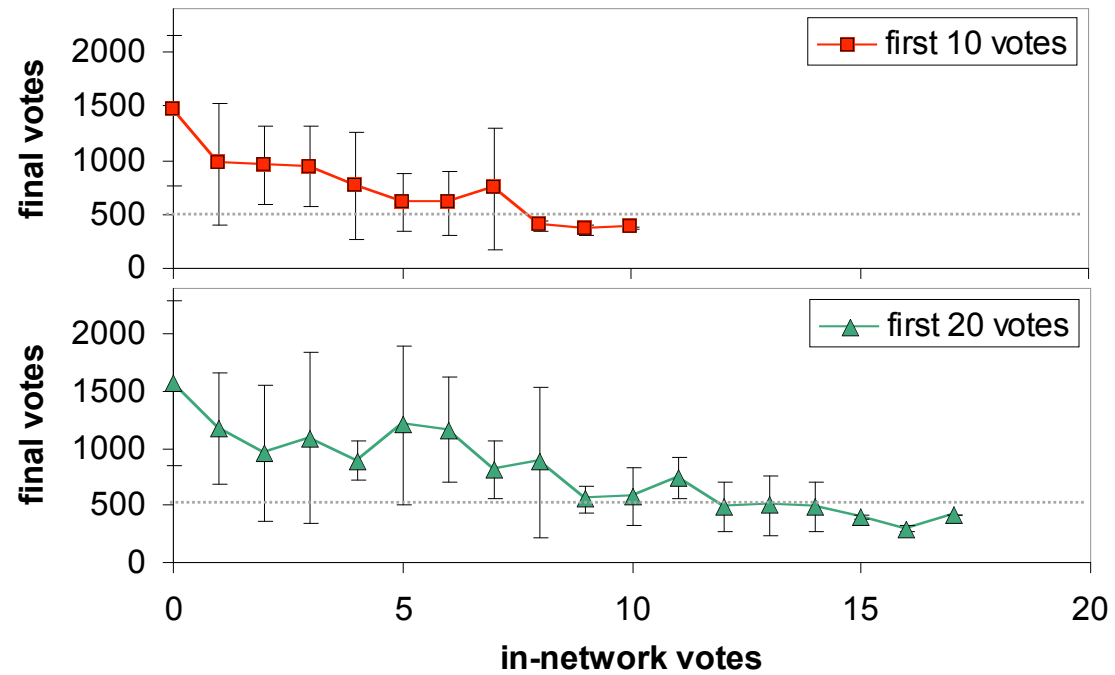


## Popularity vs the number of in-network votes out of first 6

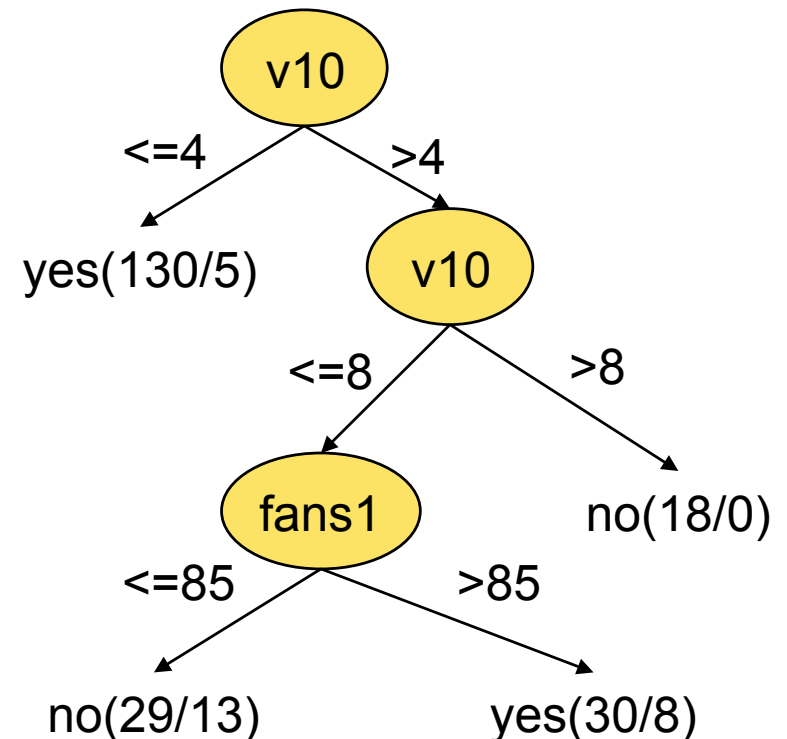


- **The stories that become popular initially receive fewer in-network votes**

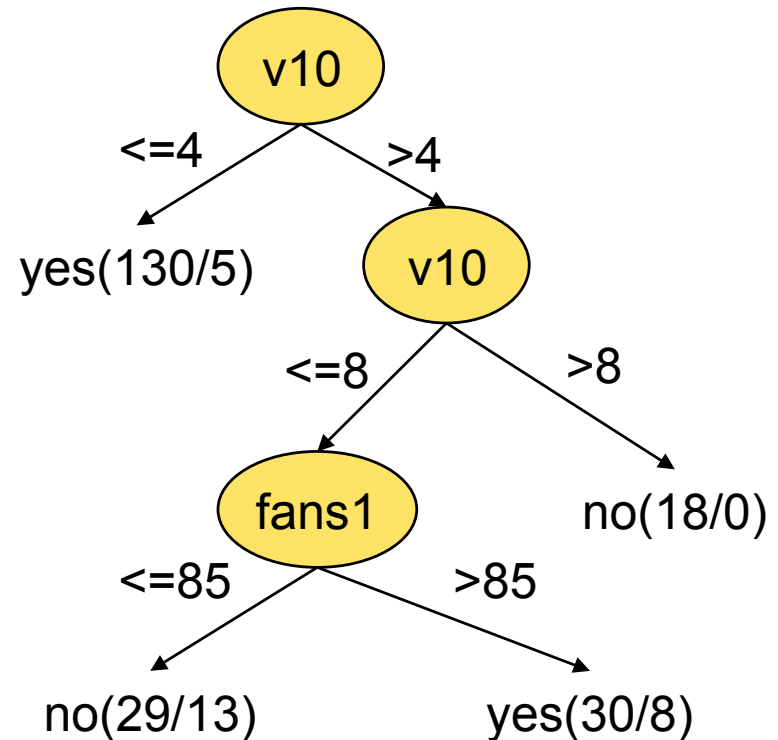
# The trend continues



- **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



- 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



- **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 Digg