Using Stochastic Models to Describe and Predict Social Dynamics of Web Users

Kristina Lerman and Tad Hogg

Yuanhui Yang

Department of Electrical Engineering and Computer Science

NORTHWESTERN UNIVERSITY

Outline

1. What is Social Dynamics of Web Users?

Outline

1. What is Social Dynamics of Web Users?
2. Why do they study Social Dynamics of Web Users?

Outline

1. What is Social Dynamics of Web Users?
2. Why do they study Social Dynamics of Web Users?
3. What are the challenges they face?

Outline

1. What is Social Dynamics of Web Users?
2. Why do they study Social Dynamics of Web Users?
3. What are the challenges they face?
4. How do they collect their data?

Outline

1. What is Social Dynamics of Web Users?
2. Why do they study Social Dynamics of Web Users?
3. What are the challenges they face?
4. How do they collect their data?
5. How do they modeling Social Dynamics of Web Users?

Outline

1. What is Social Dynamics of Web Users?
2. Why do they study Social Dynamics of Web Users?
3. What are the challenges they face?
4. How do they collect their data?
5. How do they modeling Social Dynamics of Web Users?
6. How do they solve the model?

Outline

1. What is Social Dynamics of Web Users?
2. Why do they study Social Dynamics of Web Users?
3. What are the challenges they face?
4. How do they collect their data?
5. How do they modeling Social Dynamics of Web Users?
6. How do they solve the model?
7. How is their experimental result?

Outline

1. What is Social Dynamics of Web Users?
2. Why do they study Social Dynamics of Web Users?
3. What are the challenges they face?
4. How do they collect their data?
5. How do they modeling Social Dynamics of Web Users?
6. How do they solve the model?
7. How is their experimental result?
8. How do they modify this model?

Outline

1. What is Social Dynamics of Web Users?
2. Why do they study Social Dynamics of Web Users?
3. What are the challenges they face?
4. How do they collect their data?
5. How do they modeling Social Dynamics of Web Users?
6. How do they solve the model?
7. How is their experimental result?
8. How do they modify this model?
9. Conclusion

1. What is Social Dynamics of Web Users?

Definition and Quantization

Definition

Predicting popularity of content in social media

Fact

If one user interests some social media, he/she will **VOTE** it

Quantization

- The changing rate of the number of votes at one given time point
- $$\frac{dN_{\text{vote}}(t)}{dt} \tag{1}$$

Figure: Screenshot of the front page of the social news aggregator Digg

2 Why do they study Social Dynamics of Web Users?

Value

Predicting which newly-submitted items will become popular is critically important for both hosts of social media content and its consumers

Hosts Level

Accurate and timely prediction would enable social media content hosts to maximize revenue through differential pricing for access to content or ad placement

Consumers Level

Prediction would also give consumers an important tool for filtering the ever-growing amount of content

3. What are the challenges they face?

Challenge

Technology Level

Predicting popularity of content in social media, however, is challenging due to the complex interactions between content quality and how the social media site chooses to highlight content

Privacy Level

Most social media sites also selectively present content that has been highly rated by similar users, whose similarity is indicated implicitly by their behavior or explicitly by links in a social network

4. How do they collect their data?

Data Collection

Data Collection

- They collected data for the study by scraping Digg's Web pages in May and June 2006
- structure {
  - the number of votes the stories received
  - the location of the stories on the upcoming and front pages as a function of time
  - the names of its early voters

  }

5. How do they modeling Social Dynamics of Web Users?

Prediction System

State diagram of user behavior for a single story

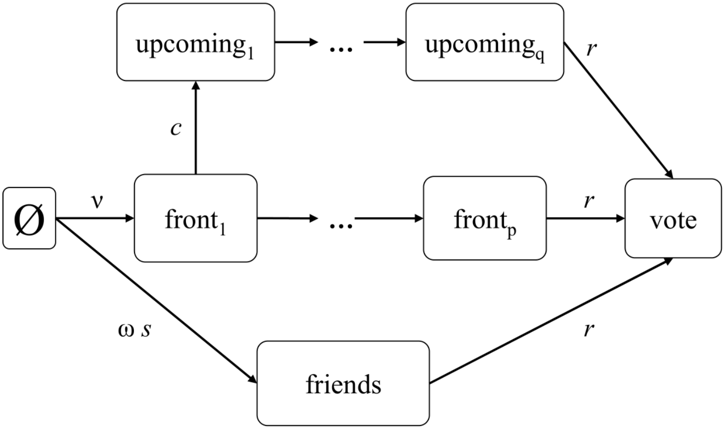

Figure: $vote = function\,(upcomings, fronts, friends)$

Governing Equation

Stochastic Master Equation

$$\begin{aligned} \frac{d\left\langle n_k \right\rangle}{dt} &= \sum_j \omega_{jk}(\langle \vec{n} \rangle)\langle n_j \rangle - \langle n_k \rangle \sum_j \omega_{kj}\left(\langle \vec{n} \rangle\right) \\ \frac{dN_{\text{vote}}\left(t\right)}{dt} &= r\left(\nu_f\left(t\right) + \nu_n\left(t\right) + \nu_{\text{friends}}\left(t\right)\right) \end{aligned} \tag{2}$$

6 How do they solve the model?

Solver

$$
\begin{aligned}
f_{\mathrm{page}}(m) &= \frac{1}{2}\left(F_m\left(-\mu\right) - \exp\left(\frac{2\lambda}{\mu}\right)F_m\left(\mu\right)\right) \\
F_m\left(x\right) &= \mathrm{erfc}\left(\alpha_m \frac{m-1+x}{\mu}\right) \\
p\left(t\right) &= k_f\left(t - T_{\mathrm{promotion}}\right) + 1 \\
q\left(t\right) &= k_t t + 1 \\
\frac{ds}{dt} &= -\omega s + a N_{\mathrm{vote}}^{-b} \frac{dN_{\mathrm{vote}}}{dt}
\end{aligned}
\tag{3}
$$

7 How is their experimental result?

Experimental Result

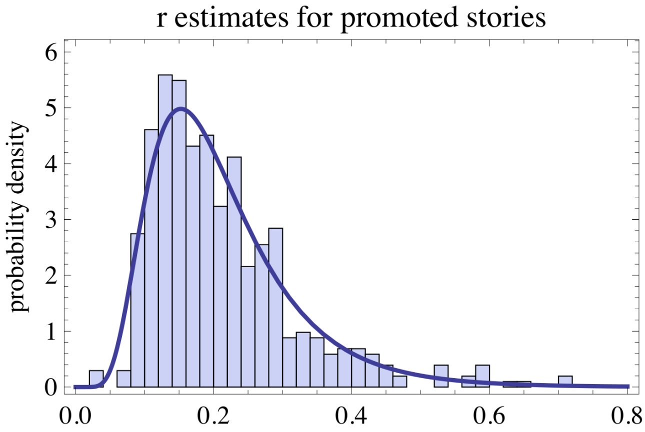

Figure: Distribution of interestingness (i.e., r values) for the promoted stories in our data set compared with the best fit lognormal distribution.

Experimental Result

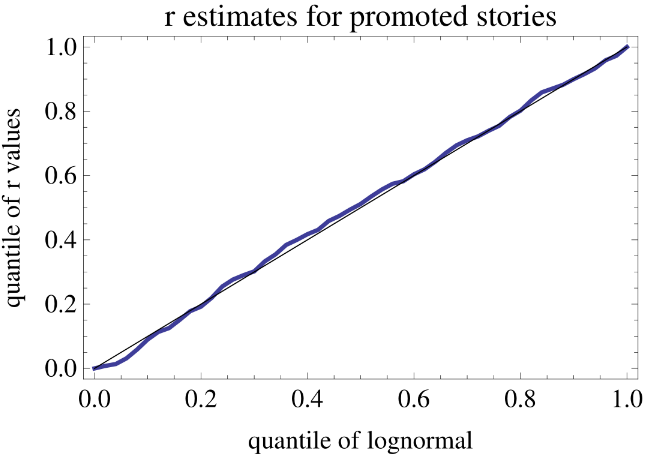

Figure: Quantile-quantile plot comparing observed distribution of r values with the lognormal distribution fit (thick curve)

8 How do they modify this model?

Modification

To investigate differences among voters with respect to the friends network, we extend the previous stochastic model to distinguish votes from fans and non-fans. The model considers the joint behavior of users and the location of the story on the web site.

A Model of Social Voting with Niche Interests

System Architecture

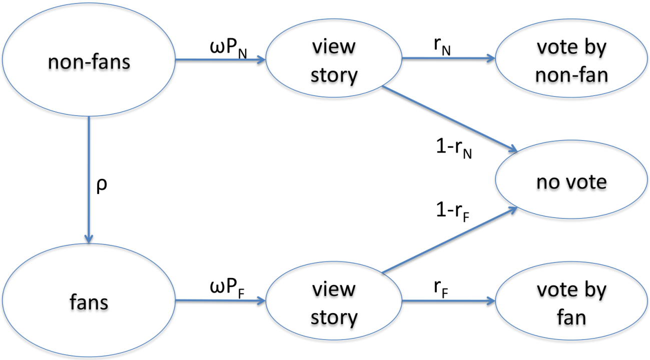

Figure: State diagram for a user. The submitter provides a story's first vote.

A Model of Social Voting with Niche Interests

System Architecture

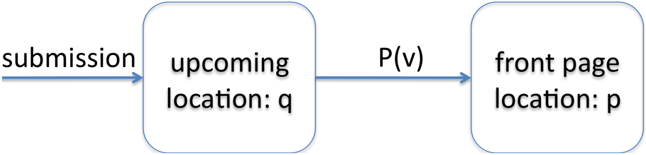

Figure: State diagram for a story

Govern Equation

$$
\begin{aligned}
\frac{dv_F}{dt} &= \omega r_F P_F F \\
\frac{dv_N}{dt} &= \omega r_N P_N N \\
\frac{dF}{dt} &= -\omega r_F P_F F + \rho N \frac{dv}{dt} \\
\frac{dB}{dt} &= -\omega r_N P_N N - \rho N \frac{dv}{dt}
\end{aligned}
\tag{4}
$$

Parameter Estimation

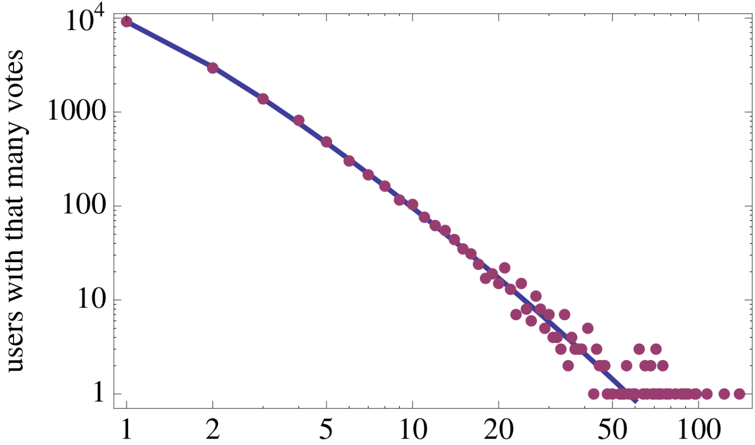

Figure: User activity distribution on logarithmic scales. The curve shows the fit to the model described in the text.

Parameter Estimation

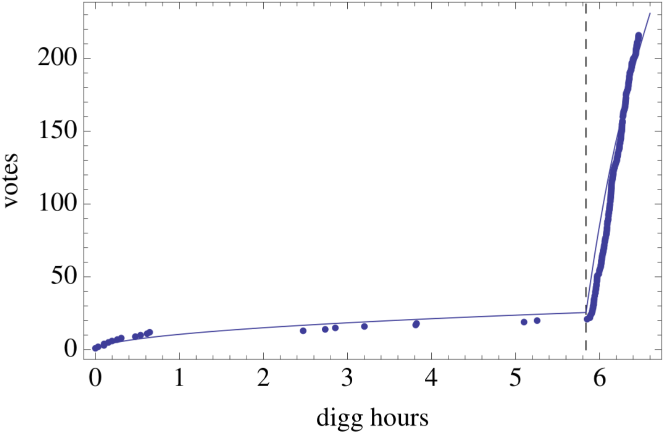

Figure: Voting behavior: the number of votes vs. time, measured in Digg hours, for a promoted story in June 2006.

Experimental Result

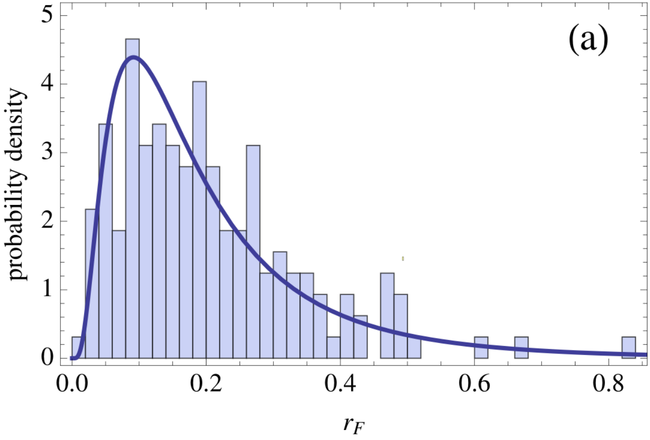

Figure: Distribution of interestingness for (a) fans

Experimental Result

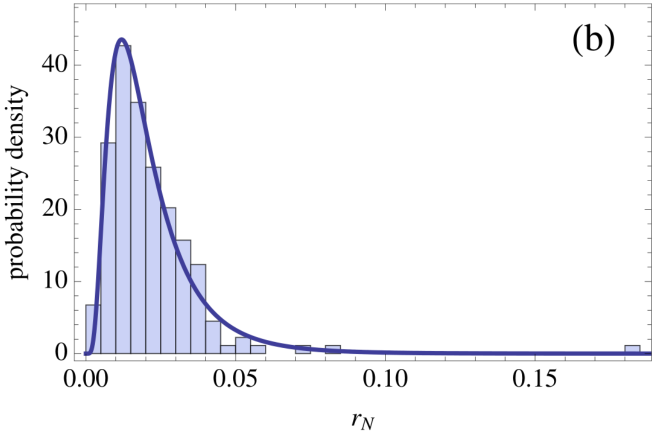

Figure: Distribution of interestingness for (b) non-fans

Experimental Result

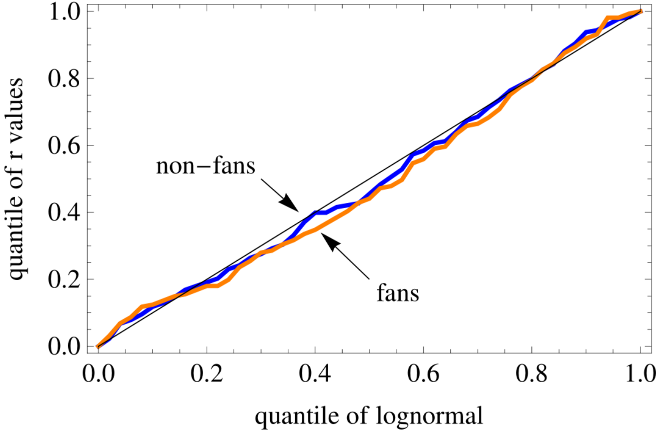

Figure: Quantile-quantile plot comparing the observed distribution for r_F (fans) and r_N (non-fans) with the corresponding lognormal distribution fits (thick curves).

Model-Based Prediction

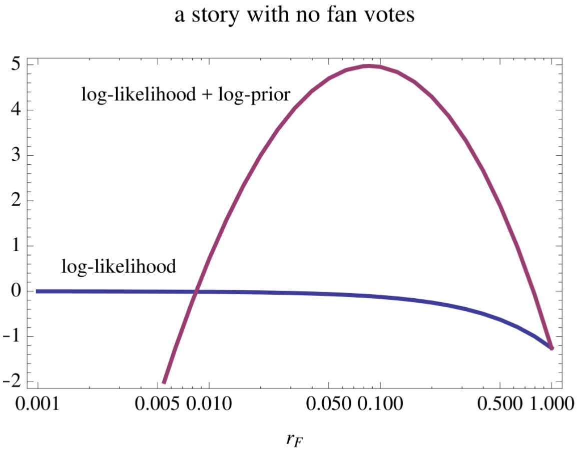

Figure: Comparisonoflog-likelihood(i.e., logP(r—votes)) and log-likelihood plus $log\left(Pprior\left(r\right)\right)$ for estimating r_F for a story with no fan votes.

9 Conclusion

Conclusion

Their solution can partially address this prediction challenge by quantitatively characterizing evolution of popularity.

Thank you