Interactions in large social systems

Dennis Wilkinson
Information Dynamics Lab
HP Labs
Recent results:

- Temporal rhythm of messaging – Facebook

- Quality and cooperation – Wikipedia

- Preference and social networks – Essemble

- Dynamics of collective attention – Digg

- Collaborative tagging – Delicious
The Facebook

• A social network website for college students begun in early 2004 (now has expanded to non-students)
Facebook is Pervasive

• Currently has over 9 million users. Over 90% of U.S. college students use Facebook.

• Pervasiveness is important:
  - The technology itself has a social impact
  - The environment becomes a reasonable proxy for the whole.
Our Analysis

- 4.2 million users from 500 schools
- 284 million messages, 79 million “pokes”
- 26 months of activity

Who do users communicate with, and when? Primarily an empirical analysis.
Temporal Rhythms
Temporal Rhythms

- Compare:
  - College students in Facebook
  - email, Researchers in big IT company
  - search queries, AOL
Temporal Rhythms

- **Friend and School Rhythms**
  - Number of different-school messages plummets in the early morning.
  - Contact with non-friends skyrockets at this time.
  - Why?

**Messages + pokes to different schools** vs. **Pokes (red) and messages (blue) to non-friends**
Temporal Rhythms

- Seasonal Variation

- Number of same-school messages highest:
  - June, July, August
  - December, January
Variation by School:

Are we seeing effects of a school's:
- Academic schedule?
- Quality of nightlife?
- Academic commitment of students?
- Network effects?
Variation by school-hour by hour correlation of late night, weekday morning, and weekend hour

Fig. 9. Correlations with respect to hours 1, 9 and 135 of the week
To conclude:

• For college students, the weekend lasts from mid-Friday to mid-Sunday.

• A significant use of Facebook is supporting distant relationships.

• Individuals schools exhibit their own regularities.

S. Golder, D. Wilkinson and B. Huberman, “Rhythms of social interaction: messaging within a massive online network”. 3rd IntlConference on Communities and Technologies, June 28–30,
Wikipedia – online encyclopedia anyone* can edit
Wikipedia is a massive and successful collaborative effort

- 7.2 M articles, 7.04 M editors, 253 languages
- Our data set: the 55.2 M edits to 1.5 M articles in the English language Wikipedia, to 11/06

- Wikipedia is useful and relevant: 9th most visited website.

- Virtually no barrier to contribution leads to intense editing activity, but at what cost? Often, larger collaborations lead to stagnation or decrease in quality.
Dynamics of edit accretion on Wikipedia

Edits to a given article arrive at a rate given by:

\[
\frac{dn}{dt} = \left[ a + \xi(t) \right] n(t)
\]

where \( a \) is the average rate of accretion and \( \xi(t) \) is a mean-zero white noise process.

\[
P[N(t) = n] = \frac{1}{n \sqrt{2\pi} \sqrt{s^2 t}} \exp\left[- \frac{(\log n - at)^2}{2(s^2 t)}\right]
\]
Does editing increase article quality, on average?

Previous work: focus on number of mistakes, credibility, formality of language. Small samples of articles.

We compared the featured articles to all the rest.
Care was taken!

Topic visibility – grouped by Google pagerank.

Age – edit & editor counts normalized by age

Special considerations – eliminated the most active week of editing to account for the extra attention articles receive during the review process.
Edits (top) and editors (bottom);

Featured (red) and nonfeatured (black)

The statistic is the age-normalized number of edit counts:

\[ \chi = \frac{\log n - \mu(t)}{\sigma(t)} \]

for an article of age \( t \) with \( n \) edits.
Evidence of increased cooperation on featured articles:

More talk page activity

More edits per editor
To conclude:

- Edits accumulate in a consistent way, regardless of article age.
- Strong correlation between high editing activity and quality.
- Articles of high interest or relevance are naturally brought to the forefront of quality.

Essembly – online social network for political discussion and organization
Essem bly data set

• 1.42 M votes by 15.1 k users on 25.0 k resolves

• Voting on a 4-point scale: Agree, Lean agree, Lean against, Against

• 5358 people participate in the networking:
  friends network has 4873 nodes and 13516 links
  allies network has 3261 nodes and 15593 links
  nem eses network has 1181 nodes and 2075 links.

• Only 35.5% of the users network, but these users are responsible for 91.3% of the votes.
Essemble questions

• Correlation between the social (friends) networks and people’s votes? Preference networks (allies and nemes)?

• Effect of networking on information flow (finding resolves)?
Social or preference networks and voting: homophily

**Allies** — mean 0.75, stdev 0.14

**Friends** — mean 0.71, stdev 0.15

**Nemesis** — mean 0.44

Similarity measure is

\[
1 - \frac{1}{3} \sum_i |v_i^a - v_i^b|
\]

over resolves in common
How do votes accumulate on a resolve?

In "resolve-time": clock ticks when a new resolve is posted.
Voting rate versus resolve age

**raw**

**percentage**
Does the network have an effect?
Compare data to a simulated Essemble

Resolves are created and votes cast as in the real version...

However, the votes are cast by a randomly selected user with no regard to the network structure.

Simulated and real votes compared.
Network non-effect on voting

Each dot represents a resolve.

Blue is the actual network; other colors are different runs of the simulation.
To conclude:

- As expected, the Essembly friends network exhibits homophily.

- However, the network seems to have little effect on spreading information about resolutes.

Work in progress with Gabor Szabo, Mike Brzozowski, and Tad Hogg
Digg – online news aggregator
Distribution of final digg numbers of 29684 stories:

lognormal (Kolmogorov-Smirnov normality test on log(N))

\[ N_t = (1 + r(t)\xi_t)N_{t-1} \]

\( r(t) \) is a discount factor which accounts for the decay in novelty
The decay of novelty:

\[ r(t) \sim \exp[-0.4t^{0.4}] \]

Story “half life”: 69 minutes

Del.icio.us — social bookmarking

- all your bookmarks in one place
- bookmark things for yourself and friends
- check out what other people are bookmarking

get started

HOT NOW

The 10 Best Foods You Aren't Eating - Men’s Health
save this
first posted by bjroed
healthy food nutrition recipes cooking
tags

Through the Looking Glass - washingtonpost.com
save this
first posted by davegoodman
interview books acfi williamjibson
writing
tags

camouflage save this
first posted by spacemarkennyx

society

soccer
FootyTube.com - Broadcasting the beautiful game
Hattrick Soccer Game
David Runciman: Can an American coach and a book deliver success in English football? PT1

portfolio
-The Culture Front- illustration animals
Alberto Campana

music
Musicovery
Large variety in the size of the tag sets of different users
Overall proportions of tags for a URL:

Stable after a small number (100) of tags applied

Polya-Eggenbergerurum (1923)

Thank you!

Information Dynamics lab:
www.hpl.hp.com/research/idl

dennis.wilkinson@hp.com