Inferring “stuff” from observed networks

He had over 2000 Facebook friends. I was expecting a bigger turnout.

16.5.2012
David Stolz
Agenda

1. Structure of Approaches
2. Recommendation Network
3. Blogs
4. “Meta-Conclusion”
Structure of Approaches

1. Understand Data
2. Define Goals / Categorize
3. Method
4. Infer
5. Compare
6. Add Knowledge
Recommendation Network

- 4 Mio. Users

- 16 Mio. Recommendations
  only ~3% of purchases associated with recommendation

- 2 Years

- Monetary benefit for recommender and recommendee
Recommendation Network

• Analyze cascades

• Categorize by different product categories
  • Books, DVD, Music, Video
Recommendation Network

- Remove:
  - no-purchase nodes
  - Late recommendations

- Find all local subgraphs
  - Isomorphism test
Recommendation Network

• Most frequently observed cascade?
Recommendation Network

- Most frequently observed cascade?

- Differences: Books, DVD, Music, Video?
Recommendation Network

• Most frequently observed cascade?

• Differences:
  • Books:  70%
  • DVD:    12%
  • Music:  86.4%
  • Video:  74%
Recommendation Network

- Overall: splits = 5 * collisions

- Simple graphs sometimes *more rare* than complex graphs
Paper Conclusions

- Most cascades are small
- Underlying social networks lead to (measurably) more complex cascades
Recommendation Network
Recommendation Network
Blogs

we've been talking...
and we all think
it's time you updated
your blog
Blogs

• 4 Years  (1999 – 2002)

• 25'000 Blogs

• 750'000 Links   \((between\  blogs)\)
Blogs

• Exact notion of time

• Only actual entries
  • Filter out “Side-bars”
Blogs

• Time characteristics

• Community structure

• Bursts
Blogs

Time Graph:

• Label Edges with time
• Label Nodes with time interval

• Prefix Graph $G_t$:

• Subgraph of $G$ up to time $t$
Blogs

Community Extraction

• Two step algorithm:
  • Find new community
  • Expand it
Communities (based on Prefix Graphs)

Number

Months since Jan 1999

Dec 2001
Communities (based on Prefix Graphs)

Fraction ∈ [0,16] ?
Blogs

SCC Comparison against “Random” Graph

Observed

“Random”

Dec 2001
Blogs

Bursts

Dec 2001
Blogs

Paper Conclusions

• End of 2001:
  • #Communities: increased
  • Connectedness: increased
  • Burstyness: increased

► User behavior has changed
In another community, a blogger Dawn hosts a poll to determine the funniest and sexiest blogger. She conducts interviews with other bloggers in the community, of course listing their sites. She then becomes obsessed with one of the other bloggers Jim, which spurs comments by many others in the community.
In another community, a blogger Dawn hosts a poll to determine the funniest and sexiest blogger. She conducts interviews with other bloggers in the community, of course listing their sites. She then becomes obsessed with one of the other bloggers Jim, which spurs comments by many others in the community.
“Meta-Conclusion”

- Empirical results matter, even if they don't astonish

- *Every* step of the 4 step approach influences the result!

- Talk is silver, silence is golden.
  
  (= don't publish papers just for the sake of publishing them)
Discussion