OUTLINE for Lec 4 on "Surface" Structure of conversations

- **Interesting**: Snow, Sandefur
- **Controversial**: Glass, Sandefur
- **Vote (link)** → Speed dates
- **Replies**: Prabhakaran
- **Live**
  - [Enron] vs. [thread]

**Top** vs. **Bottom**

- **Dimensions explicitly correspond to conv. aspects** (Fig 1)
- **Single measure**
  - *not anon. usage*

**What if you add participants explicitly?**

- Email might not even be trees
- [Prabhakaran: Rainbow '14] - add: drag are actually features!

1. **Kumar**: Fig 4, section 4.3
   - avg. # of unique author/thread length (log-log)

2. **Backstrom et al.**
   - distribution of # unique authors
   - **Count length**
   - Topics explicitly:
     - Pick mention of topic models
     - Show Regnik slides (for Nguyen et al. MLJ '14)

- **Topic shifts**
  - Show why for Nguyen et al. paper of shifts
    - Relate to polls → "Power of confidence"
    - [Enron] org-chart
In our 1st lecture, we distinguished between 2 sets of social interaction:

- conversation
- broadcast: social effect

We've spent the last two lectures on broadcast; social interaction,

doing some preliminary exploration of,

Now I'd like to turn to the conversation paradigm.

- repeat -

Thread structure

Let's assume we're dealing w/ a post: reply setting.

A basic statistic:

Post count

A basic statistic: length

Analyzing: mining [Siersdorfer et al. 2014]:

- Histogram of lengths, as expected, not much to talk about.
  - Table 5 above: "inactivity" vs. "seeds" (and edges)
  - race to get a convo going (>50? posts; start a comment
    - note: highlight: control for politics category STELLA thread)

- What does a long thread correspond to?

  Fig 9: diff. sign of denou: for how "good" initiator of long thread is.

  enum to make new post

  > 

For Yahoo, is there that men length = more intinit >>

- does Yahoo show that men + ratings? (i.e. + bias toward + attention?

- no: YouTube is the other way. (flame war)

-qq really need error bars on fig 9, especially right hand bars have little data.

Now what if we consider conversations as trees (explicitly, or, via thread induction, like what Elsner/Charniak did)? Then we have another "dimension" besides length:

- or Wang/Josh/Ros

2nd member view from above: Also LiWang
Size, Depth, other measures of tree "shape"

Kumar, Mahdian, McGlohon KDD 2010, Dynamics of Conversations.

Twitter

"Usenet group: Yahoo groups. \( \text{graphs for Usenet (other are qualitatively similar)} \)

q: "Why Usenet? <which is email, btw> scan y. to part before 8.3."

While Usenet is declining in pop., public, easy to crawl, obv. thread structure... some gaps still active. This is the material.

remark about "footnote in reaction to reviewer." e.g.

"Recall that x is not y." A social interaction!

\( q: \text{about corpus of reviewer-author interaction?} \)

\( \text{fig. 1b: empirical size vs. depth fits a} \ \sqrt{ \text{law (given the exponent value in the key).} } \)

\( \text{Conversation trees are "deeper" than a "rich-get-richer" would predict (log-depth) } \)

\( \text{counter-claim: the interfaces for these settings don't rank by branching, so an 'attention bias' towards the rich-get-richer wouldn't happen by affordance.} \)

\( \rightarrow A/B test with HackerNews vs. Slashdot? See Wang/Joshi/Rek:} \)

\( \text{fig. 3: do diff. levels have diff. or same branching?} \)

\( \text{Lines slope can you expect fewer things to have long branching.} \)

\( \text{B: Point is diff. then lines: higher levels (towards tree top) have more branches.} \)

reminder:

... all of this: about what these trees look like.
Depth = other measure of tree shape
Kaltenbrunner, Banks = "stability of political" = on Slashdot!
Gonzalez-Bailon = "for structure what participants explicitly label"
abused the term "branching factor" a # of times. (cf. "branching")
max = branching

h-index = # measures of depth: "weight in controversy of certain command" [engagement of most active users; not relevant for lessextensive]

# of tiers \( X \rightarrow X \) layers w/ \( @ \) least \( X \) comments in that layer

Fig 1: a general theory about (useful discussions), Fig 1.
vertical = "argumentation level", is practically "literally" depth of the tree
better discussion have participants involved (nominal - can't have negative depth!)

horizontal: democratic discussion involves more people
\( \Rightarrow \) more people pile in at a particular post, so correlate it w/ branching
[let's also explicitly consider # of unique participants in the tree]

Scan them: note highlights on "why SlashDot?" (remark: "human people really like Slashdot")
"had enough time to evolve, consolidate, overcoming the problems associated to spam or misbehavior, and proving its robustness"
also about handling of anon posts ("Anonymous Coward" is not a single person)
- don't remove = would break the tree structure.

Fig 2: examples of 2x2 plane (presumably real Slashdot trees)
- highlights: quadrant I is the "good" one.

Fig 3: plot political v. non-political on that plane.
"interaction" of dotted lines = mean width = depth,
Conservatve shows the political discussions do indeed "look diff" - in Quad I -
from non-political
(Fig 4: divide by category. Not in color, unfortunately.
living near center = anom: very low participation (Quad III): few people, less active category?)

An h-index measure to try to summarize shape in a single #.
See highlighted text under Fig 4.

Also, here's translation...
(note, layers w/ big branching don't have to be consecutive)
Now add: participant ID to conversation structure

Kumar et al. again, fig 4 (and section 4.3):

On a log-log, we see that the avg # of unique authors (red) vs. size:

a polynomial relationship.

But, what about the distribution of # of unique authors?

(End of what is a long thread)

<Backstrom, Kleinberg, Lee, Pan, Ning, Muslum WSDM '13> : (display the WSDM poster)

Guestbook vs. involved:

Show limit may on poster:

[Graph]

# unique participants/thread

Thread Length

matters b/c you may want to rate posts by type of thread, not just length of thread

- If it's a thread someone's likely to re-enter, key them off

Also since it came up before, note that social-connectivities of 1st few posters has interesting effects on the length of a reply thread vs. a "like" thread

- In the setting of email, you may not even have trees, because of addition of people's identities

Envron data set (post ref)

Forwarding vs. replying

<need to clear out the "prior email text or make sure you know who wrote which part"

Dropping/adding recipients

Power relationships known

(Actually used as features in Prabhakaran? Rambha '14)

Note also the UMass work on the North CarolinaInn corpus (Freedom of Info request?)

Note also the W3C data (Glasgow)
First talked about adding recipients, what about race adding context to the structure of case?

Next slide: a shift on the topic shifts & related to poll standings.

From that, Palin is okay shifting to a diff. topic.

Next slide: some criticism of Fill for not managing the conversation as the new shift. Moderator: Palin changing topics to a diff. topic.

Some more Fill, then to get back to a diff. topic.

I want to talk about, again, my usual approach.

Did Fill real-time correct or conversational?