Dynamiques dans les forums de discussions online

Toulouse - October 2015

Alberto Lumbreras, Marie Guegan (TCH)
Julien Velcin (ERIC)
Bertrand Jouve (CNRS)
1. Discussion Forums

2. Generative models

3. Some dynamics of discussion trees

4. Summary
1. Discussion Forums

2. Generative models

3. Some dynamics of discussion trees

4. Summary
Discussion Forums

Parser&database: https://github.com/alumbreras/reddit_parser/
Discussion Forums
Reddit, January 2015

- Power law distribution of posts/thread.

Figure 1: Posts per thread
Discussion Forums
User graphs vs Posts graphs
Discussion Forums
User graphs vs Posts graphs
Discussion Forums
User graphs vs Posts graphs
Question

Can we **model** the **dynamics** of a conversation tree?

- To understand and **summarize** its internal mechanisms.
- To **predict** the future of a conversation.
1. Discussion Forums

2. Generative models

3. Some dynamics of discussion trees

4. Summary
Generative models
Modeling graphs

- Simple generative process governed by few parameters as possible.
- Artificial graphs reproduce some properties of the real graphs.

- Equiprobable edges.
- Preferential Attachment.

Figure 2: Erdos-Renyi
n = 25
p = 0.5

Figure 3: Barabasi-Albert
n = 1000
p = 1
Generative models
Modeling online conversations

Figure 4: Gómez-Kappen-Kaltenbrunner [?]
Parameters: $\alpha_1, \alpha_c, \beta$

- Number of responses to root. ✓
- Number of children. ✓
- Number of descendants.
- Node depth. ✓
- Size vs Proportion of responses to root. ✓
- Speeds. X
- Sequences (e.g.: a-b-a-b). X
- Text. X
- ...

(a) real
(b) synthetic
Generative models
Modeling online conversations

(a) real
(b) synthetic

Figure 5: Gómez-Kappen-Kaltenbrunner [?] Parameters: $\alpha_1, \alpha_c, \beta$

$$p(\pi_t = k|\pi_{(1:t-1)}) \propto (\beta_k d_k, t)^{\alpha_k}$$

$\beta_k$: root bias  
$\alpha_k$: Preferential Attachment exponent.

- Estimate model parameters after conversations are finished (once and forever!)
Generative models

Question

Can we improve this model to simulate real time? (speed changes, bursts, slow down...)

Can we create a non-parametric growth model that continually re-evaluates its predictions on where the next post will be? (that will allow for predictions on $G_{t+1}$ given $G_{1:t}$) ("link prediction" for conversations)
1. Discussion Forums

2. Generative models

3. Some dynamics of discussion trees

4. Summary
Discussion trees
Different structures, different speeds?

- Time between posts is shorter in chains.

Figure 6: Average time between posts in chains and stars
Discussion trees
Different thread lengths, different triadic structures

<table>
<thead>
<tr>
<th>Motif</th>
<th>Motif ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>164</td>
</tr>
</tbody>
</table>
1. Discussion Forums

2. Generative models

3. Some dynamics of discussion trees

4. Summary
Conclusions

Current:
- Good graph models that explain degree distributions.
- Good thread models.
- Parametric models.

But:
- Changes of rhythm, burst, etc not modeled (time is homogeneous).
  a-b-a-b-a-b = a-b-c-d-e-f.
- Can not predict (therefore recommend posts to users).

Therefore we would like:
- Growth model that considers relation between structure and speed.
- Growth model that considers sequences (authorship).
- Link prediction for conversation trees.