Congressional Tweets Corpus

- 2,813,417 tweet messages (as of Feb/27/2013, glowing) collected through Twitter search API (starting data May/2012), with the following search criteria:
  - Message sent by an official tweet account owned by a member of 112th congress ("MC account"). Currently our MC account list includes 440 handles.
  - Message sent to (one or more) MC account(s)
  - Message which mentions MC's tweet handle in the tweet text

- One message includes the following data:

<table>
<thead>
<tr>
<th>User_Id</th>
<th>Creation_time</th>
<th>From_user</th>
</tr>
</thead>
<tbody>
<tr>
<td>From_user_id</td>
<td>To_user</td>
<td>To_user_id</td>
</tr>
<tr>
<td>Location (lat/lon)</td>
<td>Source</td>
<td>Text</td>
</tr>
<tr>
<td>Retweeted</td>
<td>Retweet_id</td>
<td>Jaon</td>
</tr>
</tbody>
</table>
What has been done?

- Predicting campaign finance using topic model: **A Penny for your Tweets: Campaign Contributions and Capitol Hill Microblogs** (ICWSM 2013) Tae Yano, Dani Yogatama, Noah A. Smith.

- Predicting the general election results from the public speech: **Congressional Election Prediction Using Candidates Tweet** (Term project: Machine Learning with Large Datasets 10-605, Spring 2013) Yiwen Chen, Zhiqi Li and Yuliang Yin.
What you would like to do with this?

- Predicting MC attributes from the public speech; We have some attributes: Party (R/D/I), State, Chamber (House/Senate), District.
- Predicting committee assignment from the public speech.
- Predicting Issue Position from the public speech; Such as the "Blue Dogs", or “Defectors” for particular issues.
- Predicting popularity (e.g., retweet count) of political tweets.
- Analysis of Retweet graph. Can we guess who is going to retweet?
- Detection of major congressional events from tweet (votes on Affordable Care).
- Detection of topic specific congressional events (e.g., earmarking, hearing, committee referral).