BLOGS ARE ECHO CHAMBERS:

Eric Gilbert | Tony Bergstrom | Karrie Karahalios | University of Illinois
Are blogs **echo chambers**?

*Do blogs cut readers off from dissenting opinions?*
“It’s hardly possible to overstate the value, in the present state of human improvement, of placing human beings in contact with other persons dissimilar to themselves, and with modes of thought and action unlike those with which they are familiar.”

John Stuart Mill

*Principles of Political Economy, 1848*
112 M blogs

as measured by Technorati, Jun 2009
Are blogs *echo chambers*?

*Do blogs cut readers off from dissenting opinions?*
RELATED WORK

Political bloggers link to like-minded bloggers.

Adamic & Glance 2005  
Hargittai & Gallo 2008

Groups often adopt extreme viewpoints.

Baron & Hoppe 1996  
Sechrist & Stangor 2001
OUR DATASET

5 blog genres
33 top blogs
1,094 blog comments
<table>
<thead>
<tr>
<th>Technology</th>
<th>Political</th>
<th>Entertainment</th>
<th>Lifestyle</th>
<th>Meta</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 TechCrunch</td>
<td>1 Huffington Post</td>
<td>5 Boing Boing</td>
<td>6 Life Hacker</td>
<td>10 ReadWriteWeb</td>
</tr>
<tr>
<td>3 Gizmodo</td>
<td>11 Daily Kos</td>
<td>13 Gawker</td>
<td>29 Consumerist</td>
<td>18 Dosh Dosh</td>
</tr>
<tr>
<td>4 Engadget</td>
<td>26 Think Progress</td>
<td>20 Perez Hilton</td>
<td>32 uthink</td>
<td>21 ProBlogger</td>
</tr>
<tr>
<td>23 Kotaku</td>
<td>41 Crooks &amp; Liars</td>
<td>31 Valleywag</td>
<td>45 Zenhabits</td>
<td>27 Copyblogger</td>
</tr>
<tr>
<td>30 Scobelizer</td>
<td>58 NewsBusters</td>
<td>36 Neatorama</td>
<td>38 Dooce</td>
<td>34 ShoeMoney</td>
</tr>
<tr>
<td>35 Gigaom</td>
<td></td>
<td>42 Slashfilm</td>
<td>53 Sartorialist</td>
<td>43 Daily Blog Tips</td>
</tr>
<tr>
<td>37 TUAW</td>
<td></td>
<td></td>
<td></td>
<td>71 Matt Cutts</td>
</tr>
<tr>
<td>44 Joystiq</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 Threat Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# Technorati rank
Apr 08
“Great post and I really like the video. This is extremely similar to the approach I use in writing almost anything…”

ProBlogger
Inter-rater reliability

Cohen's $\kappa = 0.71$
pointwise $= 0.74$
Proportions of agreement

- neither: 49.4%
- agree: 39.2%
- disagree: 11.1%

Wald method
p < 0.05
Agreement proportions by genre

<table>
<thead>
<tr>
<th>Genre</th>
<th>Neither</th>
<th>Agree</th>
<th>Meta</th>
</tr>
</thead>
<tbody>
<tr>
<td>tech</td>
<td>63%</td>
<td>24%</td>
<td>13%</td>
</tr>
<tr>
<td>enter.</td>
<td>63%</td>
<td>24%</td>
<td>13%</td>
</tr>
<tr>
<td>lifestyle</td>
<td>44%</td>
<td>51%</td>
<td>5%</td>
</tr>
<tr>
<td>politics</td>
<td>40%</td>
<td>47%</td>
<td>13%</td>
</tr>
<tr>
<td>meta</td>
<td>33%</td>
<td>55%</td>
<td>12%</td>
</tr>
</tbody>
</table>

$\chi^2(8, N=979) = 86.3$

$p < 0.001$
ALGORITHMIC PREDICTION OF AGREEMENT
Algorithmic prediction:
Lexical features

This feels like an echo chamber within an echo chamber!
Algorithmic prediction:
Lexical features

This feels like an echo chamber within an echo chamber!
Algorithmic prediction:
POS features

This feels like an echo chamber within an echo chamber!
Algorithmic prediction:
Semantic features

Cosine similarity to post
WordNet similarity to post
Explicit Semantic Analysis
Algorithmic prediction:
Sentiment features

LIWC positive, negative, anger, ...

Rotten Tomatoes +/- classifier

Congressional floor +/- classifier
Algorithmic prediction:
The actual algorithm

Bagged Complement Naive Bayes

AGREE/DISAGREE/NEITHER

- LEXICAL
  - uni/bi/trigams
  - TFIDF

- POS
  - raw tags
  - combo lexical

- SENTIMENT
  - congressional floor
  - rotten tomatoes
  - LIWC

- SEMANTIC
  - sim to post
  - ESA

- NAMED ENTITY
  - organizations
  - people

Bagged comp. naive bayes
Learning an NLP-based model of commenter agreement

Model accuracy

\[
\chi^2(1, N=196) = 7.6 \\
p = 0.006
\]

Rotten Tomatoes classifier: 90%
Congressional floor classifier: 71%
<table>
<thead>
<tr>
<th>Feature</th>
<th>Polarity</th>
<th>Info Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIWC pos. emotion words</td>
<td>agree</td>
<td>0.079</td>
</tr>
<tr>
<td>LIWC affect words</td>
<td>agree</td>
<td>0.049</td>
</tr>
<tr>
<td>exclamations</td>
<td>agree</td>
<td>0.043</td>
</tr>
<tr>
<td>adjectives</td>
<td>agree</td>
<td>0.041</td>
</tr>
<tr>
<td>@</td>
<td>neither</td>
<td>0.041</td>
</tr>
<tr>
<td>ellipsis</td>
<td>!disagree</td>
<td>0.038</td>
</tr>
<tr>
<td>great</td>
<td>agree</td>
<td>0.035</td>
</tr>
<tr>
<td>is tech blog</td>
<td>neither</td>
<td>0.034</td>
</tr>
<tr>
<td>cosine similarity to post</td>
<td>!disagree</td>
<td>0.034</td>
</tr>
<tr>
<td>great [noun]</td>
<td>agree</td>
<td>0.03</td>
</tr>
<tr>
<td>personal pronouns</td>
<td>!disagree</td>
<td>0.028</td>
</tr>
<tr>
<td>present tense verbs</td>
<td>neither</td>
<td>0.026</td>
</tr>
<tr>
<td>[prepos] [poss pronoun]</td>
<td>agree</td>
<td>0.026</td>
</tr>
<tr>
<td>tf-idf dot product with post</td>
<td>!neither</td>
<td>0.026</td>
</tr>
<tr>
<td>coordinating conjunctions</td>
<td>agree</td>
<td>0.026</td>
</tr>
</tbody>
</table>
Blogs are echo chambers.

77.9% of opinionated commenters agree with the blog author.
An algorithmic approach could help.

Consider an echo index.