Male, metropolitan, and mainstream

Bias in algorithmic news sourcing from Twitter

Ecrea Journalism Studies Conference 2017
23-24 March 2017, Odense, Denmark
For journalists, social media can be...

- A channel for dissemination of own material
- A pointer for news and trends
  (Broersma/Graham 2013: 448f.)
  ➔ timely, sometimes obscure news
- Source for “ambient” news with the audience as both a receiver and a sender (Hermida 2010)
Problems and challenges (Craig 2014: 106)

- Multitude of accounts
  ➔ Bots: content polluters, valuable content aggregators!? (Kyumin/Eoff/Caverlee 2011; Lokot/Diakopoulos 2015; Larsson/Moe 2015)
- Large frequency of messages
- Speed with which information is spread

 ➔ “how to keep up?”
Problems and challenges (cont.)

- What's news and what’s not?
  - lots of clutter (41% on Twitter “pointless babble” in 2009 – “eating sandwich”; Pear Analytics 2009)
  - wrong/hijacked hashtags

- Verification of information
  - “s/he has a gun!”

- Who to listen to? Elites or nonelites or both (Henderson/Miller 2014)?
“The stories that show in your News Feed are influenced by your connections and activity on Facebook. This helps you to see more stories that interest you from friends you interact with the most. The number of comments and likes a post receives and what kind of story it is (ex: photo, video, status update) can also make it more likely to appear in your News Feed.”

(Facebook Help Center)
Challenges arising with algorithms choosing content:

• Not everyone sees the same
• Algorithmic accountability
  (Diakopoulos 2014; Mittelstadt 2016)
  ➔ How do the algorithms work?
  ➔ “Why do we see what we see?”
• Privacy issues (who is tracked and how?)

How are the social media posts chosen that journalists see?
How can algorithmic news sourcing help journalists?
“With the proliferation of smartphones and social media, it means that there are lot more witnesses to a lot more events. We can’t be at everything. Our tool helps shift some of the burden of witnessing and lets journalists do much more of the high value-added work.”

(Reg Chua, Executive editor of data and innovation, Reuters)

Why Twitter?

- Twitter is open to everyone: corporate, general public...
- Twitter as an especially timely source for breaking news events
• Interdisciplinary project with researchers, media companies, corporate
• Goal: application for autonomous detection, clustering, prioritising of news and infotainment from social media
• Works in real time
• Diverse material: text, images, audio and video
Twitter: Monitoring accounts via 'newshounds'

Possible newshounds are selected (initial seed: 500 journalists, then people they followed)
Monitoring accounts via 'newshounds'

- Possible newshounds are selected (initial seed: 500 journalists, then people they followed)

- Scoring system prioritises newshounds according to activity, number of followers, whether they were verified... different scoring formulas were tried

- Final newshounds database with ca. 6,000 newshounds for news lists (& a celebrity list with 846 for the soft news)
Empirical study of Twitter lists

- **Sample** (random sample, 95 % CI, 5 % MOE)
  - UK news (n=346),
  - US news (n=333)
  - ... & Celebrity list (n=260)
    (Moon/Hadley 2014: 300: Twitter is more often used as a source for soft news than for hard news by TV and newspapers)

Which characteristics do the newshounds show – and who are thus the sources of Social Sensor?
Results: Male bias

UK list – Gender (n=346)

- Male: 57.9%
- Female: 24%
- Institution: 21.1%

US list – Gender (n=329)

- Male: 49.5%
- Female: 29.8%
- Institution: 20.7%

Jessica Kunert and Neil Thurman
Results: Male bias

Celebrity list – Gender (n=258)

- Male: 33.7%
- Female: 26.4%
- Institution: 39.9%
Media cities with the big outlets, but...
UK list – Affiliation (n=343)

Out of all who named their employer:

44% from the BBC!

Are the journalists referencing themselves?
Results: Mainstream media bias

Out of all who named their employer:

38% from the New York Times!

⇒ Same here!
Results: Mainstream media bias

Even here, over 30% are media outlets or journalists!
Other tools for algorithmic news sourcing

- Reuters News Tracer (Twitter)
- NewsWhip Spike (many platforms)
- Dataminr (Twitter)
- Facebook Signal (Facebook & Instagram)
- Bloomberg’s initiative
... But: who witnesses whom and how?

• How are the algorithms programmed?

• How to make sure that not the same few people are listened to? What about the “nonelites” (Henderson/Miller 2014)?

• What about an update of the newshound list? (inactive accounts)
Thank you!

Department of Communication Studies and Media Research
Ludwig-Maximilians-Universität München


