Network Approaches to Identifying Online Echo Chambers

Ella Guest
Mitchell Centre
University of Manchester

What are echo chambers?

"An echo chamber comes into being where a group of participants choose to preferentially connect with each other, to the exclusion of outsiders"

- Axel Bruns. 2017. Echo chamber? What echo chamber? Reviewing the evidence. In 6th Biennial Future of Journalism Conference

- ● Choice homophily
- ● Natural in topic oriented communities
- ● But taken to the extreme of exclusion (active or passive)

Reddit...?

- Subreddits
  - topic-oriented message-board style communities
  - norms develop independently
  - limited platform moderation

- Redditors
  - pseudonymous
  - varying levels of engagements
  - skews North American, college-educated, tech literate, male

- Natural two mode structure:
  - Subreddits connected by co-participants
  - Redditors connected by co-commenting in subreddits

(Meta)-Echo Chambers on Reddit

- Previous work on echo chambers determined that for individual subreddits that may qualitatively appear to be 'echo chambers', it's very difficult to meaningfully quantify echo chamberness

- Politically-oriented subreddits more discursive? More active contributors?

- For example, comment authors in The_Donald

r/The_Donald

"Trump Supporters ONLY – This sub is for supporters of Donald J. Trump ONLY. This is not a place for you to debate with us about Donald Trump, or to ask us to convince you to like Donald Trump. **This is not a neutral place – we are 100% in support of Donald J. Trump.** Moderators reserve the right to ban non-supporters as we see fit."

→ self-categorisation fits 'echo chamber' definition

(Meta)-Echo Chambers on Reddit

- Previous work on echo chambers determined that for individual subreddits that may qualitatively appear to be 'echo chambers', it's very difficult to meaningfully quantify echo chamberness

- Politically-oriented subreddits more discursive? More active contributors?

- For example, comment authors in The_Donald
  - comment in (relatively) *a lot* of other subreddits
  - spend (relatively) *a lot* of time outside The_Donald

- ➔ Network approach
  - (meta)-echo chambers of highly connected subreddits, bounded by shared views

Questions

- Bounding problem:
  - Selecting subreddits under observation
  - Just political? How to identify?

- Still be able to compare to general distribution
  - Need to understand overall community structures

- Reddit allows us to map the complete network
  - What's a 'normal' level of similarity between subreddits?
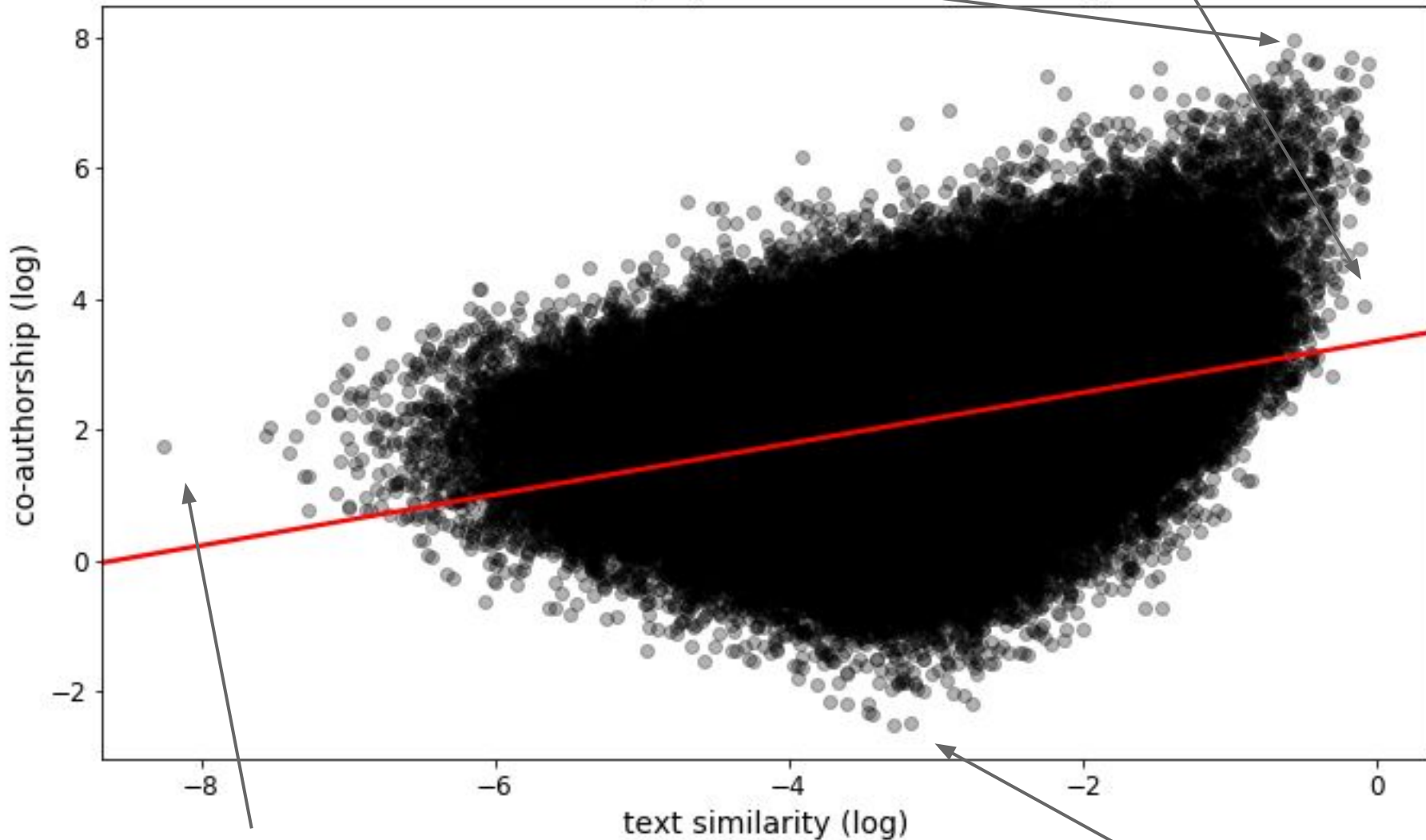
- Then how to define similarity?

The Data

- Complete monthly datasets of comments, by Jason Baumgartner
  - January 2019
- Selected the top 1000 subreddits by number of unique comment authors
  - min 818 authors
- Co-authorship
  - Observed: Number of authors who commented in both subreddit i and subreddit j
  - Weighted: observed/expected at random
- Text similarity
  - Bag of words from all comments per subreddit
  - Weighted by term frequency - inverse document frequency (tf-idf)
  - Cosine similarity taken for all pairs of subreddits

correlation - 0.33

Deltarune - Undertale

Vaping - electronic_cigarettes
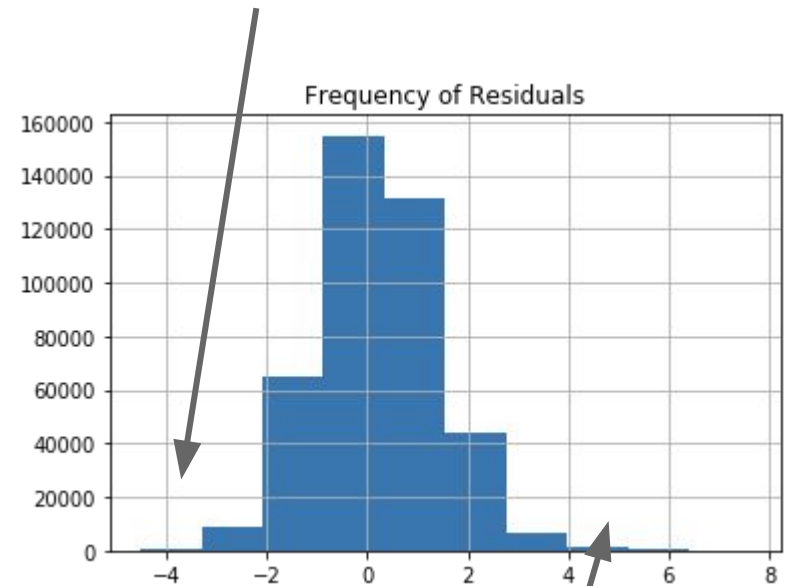
Text similiarity by Co-authorship Per Edge

co-authorship (log)

text similarity (log)

curlyhair - ik_ihe

dankmemes - orangetheory

9

Combined Similarity

- Simple linear regression of co-authorship on text similarity

- Higher residuals -- higher co-authorship after controlling for text similarity

- Topic-based pattern appears to emerge

dankmemes - orangetheory
PewdiepieSubmissions - xxfitness
beyondthebump - dankmemes

Frequency of Residuals

Market76 - fo76bazaar
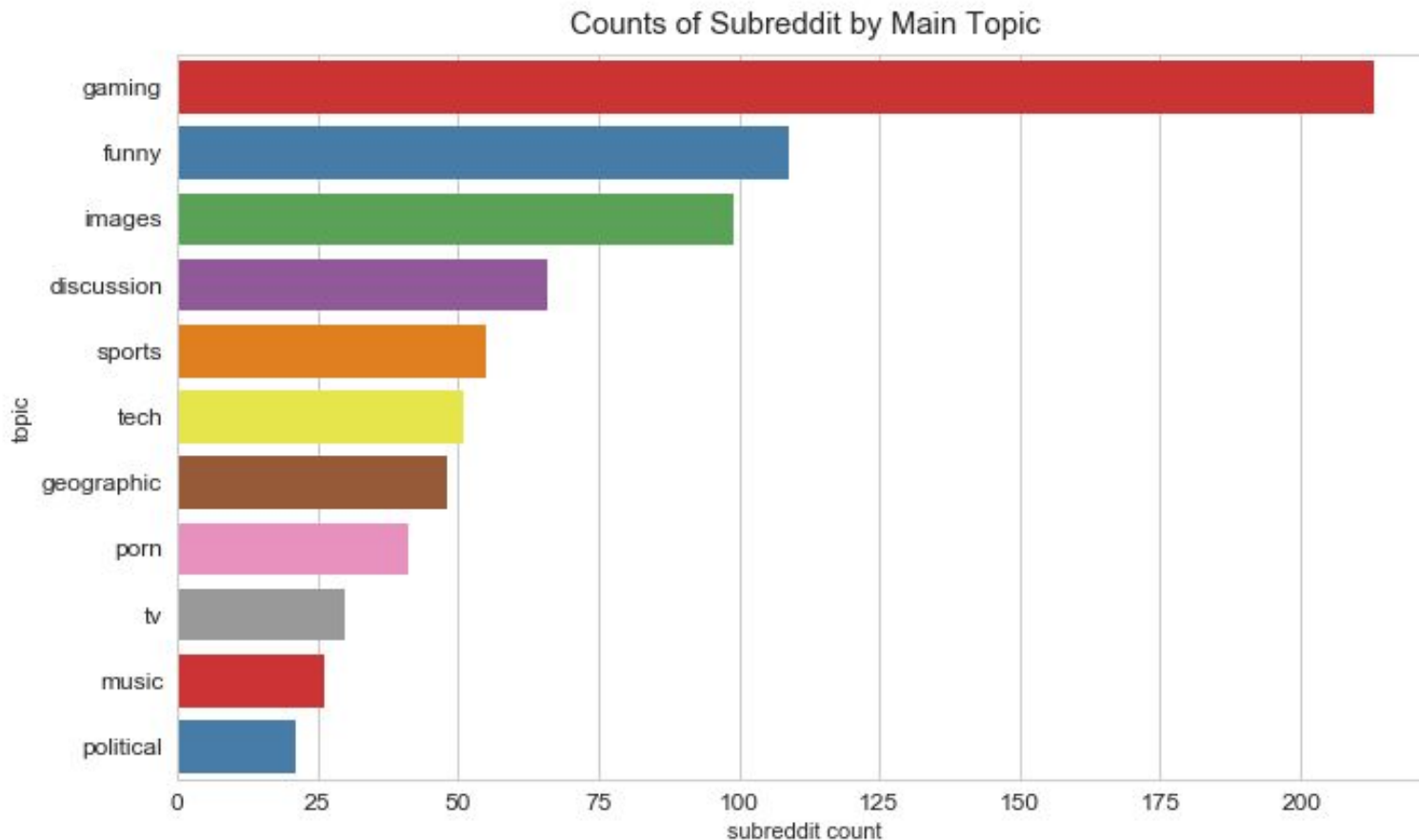DankMemesFromSite19 - SCP
Deltarune - Undertale

Community detection

- Residuals as edge weights, subset top 5%
- Louvain with community and networkx packages in python

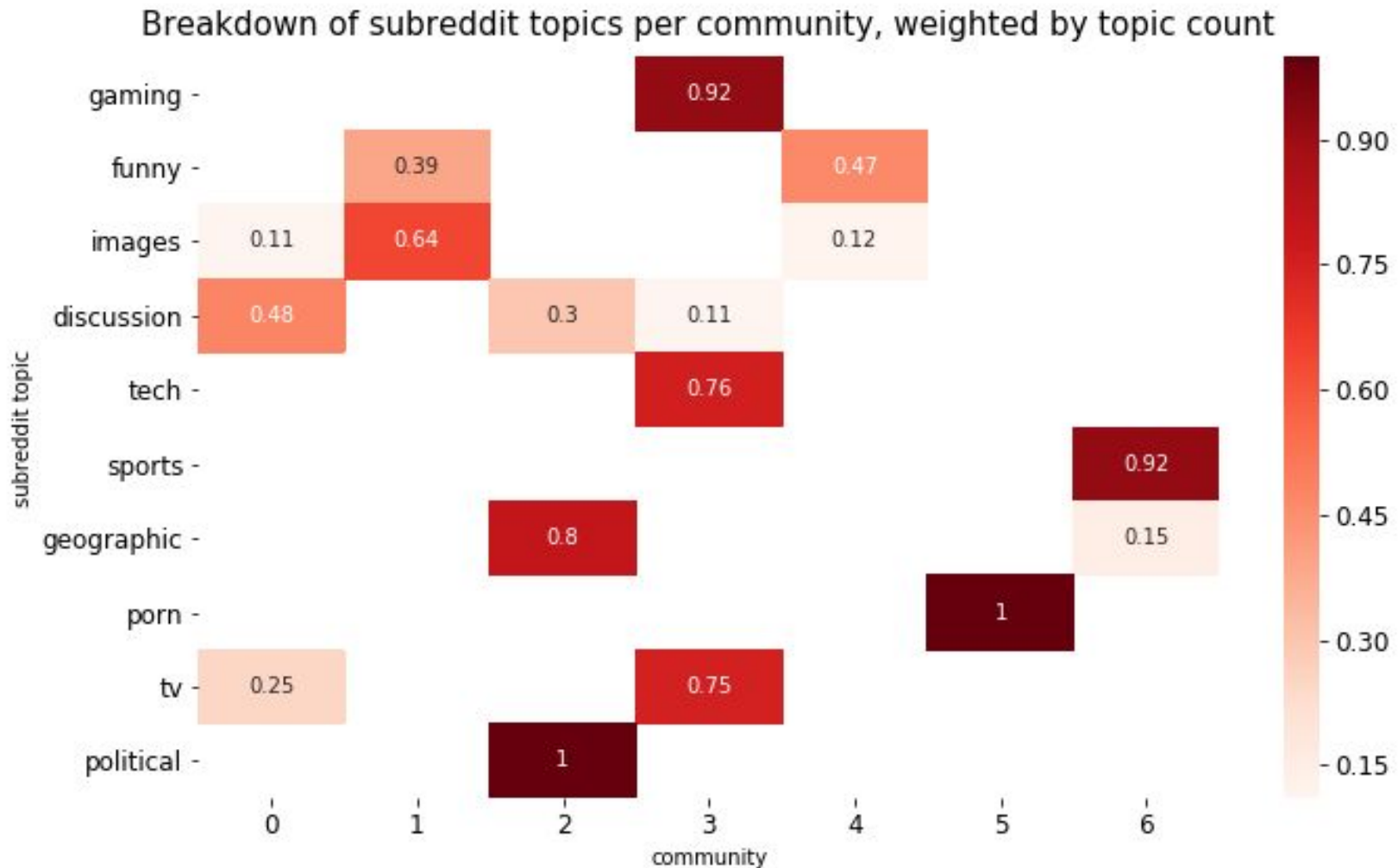| Community	0	1	2	3	4	5	6	7
N	166	145	153	315	116	42	56	3
EI index	2.3	2.5	1.6	5.6	1.3	21.1	12.7	

- All communities have more internal edges and external
- 5, 6, and 3 especially high

Topic Labels
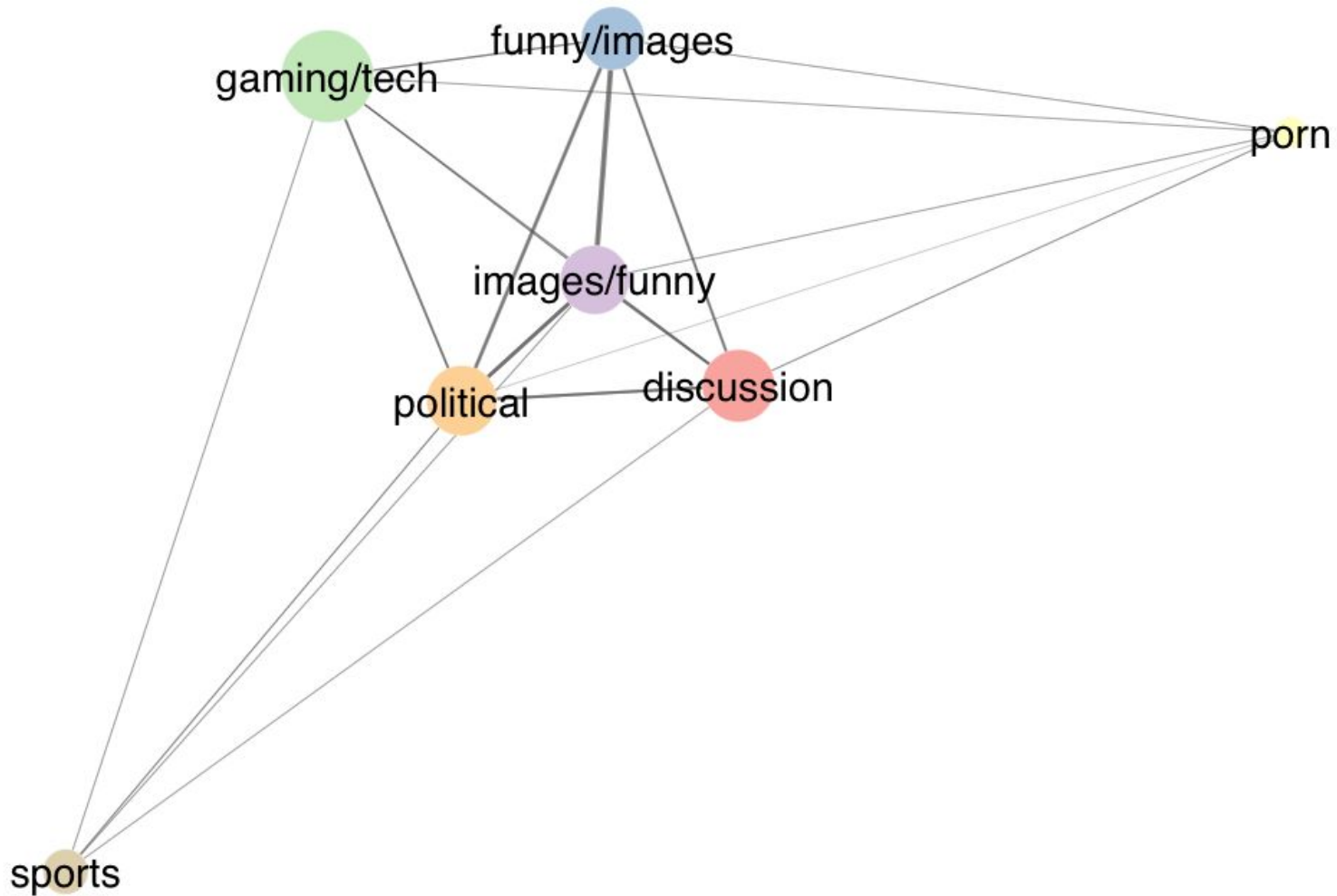
- Manually tagged subreddits
- Multiple, ordered tags
- Automatic labelling? (eg Google Cloud's Natural Language API)

Counts of Subreddit by Main Topic

Topic Breakdown Among Communities

Breakdown of subreddit topics per community, weighted by topic count
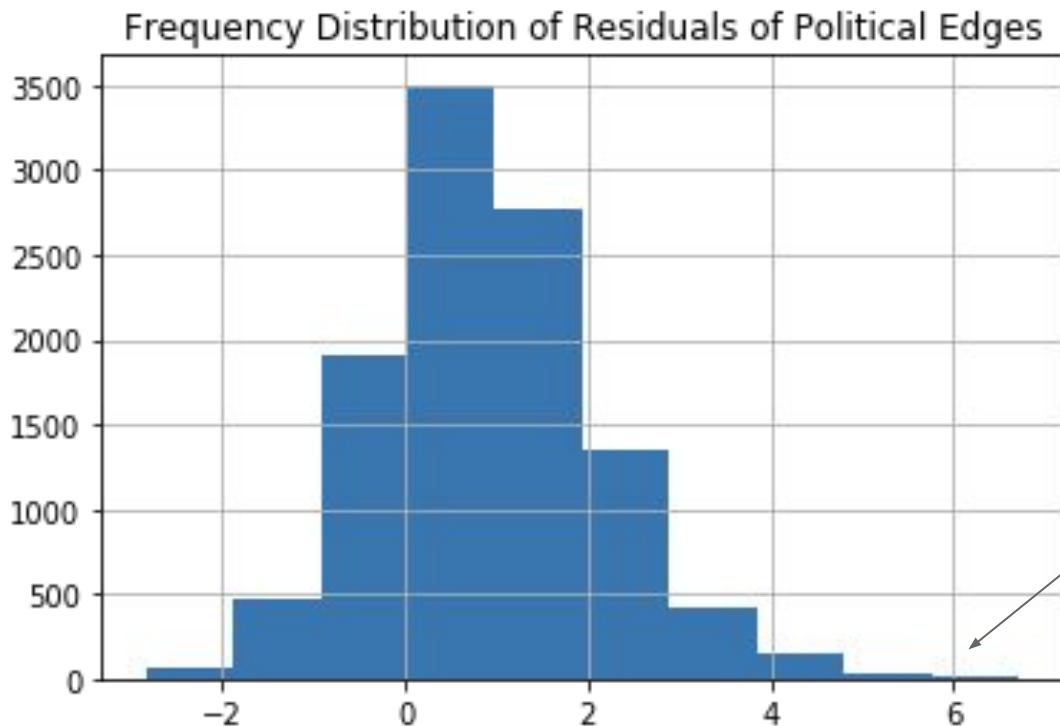
Community Graph

Political (et al) Community

- All political subreddits in one community
- 152 subreddits total
  - also geographic and discussion subreddits

The most 'echo-y'?

ukpolitics - unitedkingdom

COMPLETEANARCHY - socialism

COMPLETEANARCHY - ChapoTrapHouse

LateStageCapitalism - socialism

Fuckthealtright - beholdthemasterrace

Frequency Distribution of Residuals of Political Edges

Summary

- Co-authorship, controlling for text similarity highlights latent topic communities

- Porn commenters very insulated (multiple profile maintenance?)

- Sports commenters also quite insulated (coming to Reddit for a purpose?)

- Relative to these, political subreddits may not be as insular

- However, left-wing subreddits might be more insular than right-wing

- Going forward:

  - Longitudinal comparison - before alt right subreddits banned

  - Two mode analysis - subreddits and authors