Political Polarization in Online News Consumption

By Garimella et al. (2021)

Chaewon Yun

Research Question & Dataset

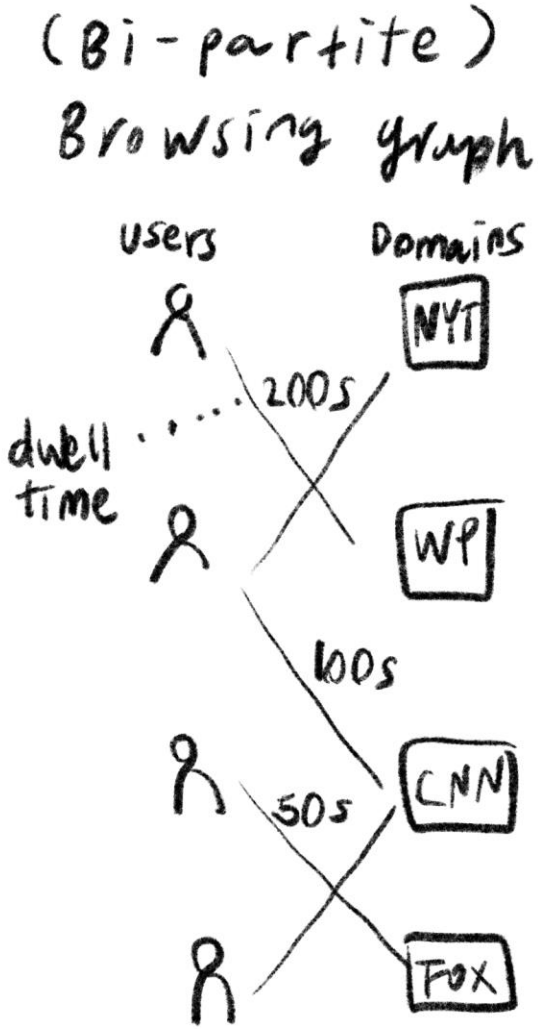
- Examine the **information-seeking behaviors** of individual Web users to understand the relationship between **mass polarization and the consumption of news content on the web**
- Dataset: Large-scale dataset consisting of detailed **browsing logs** collected via Firefox extension

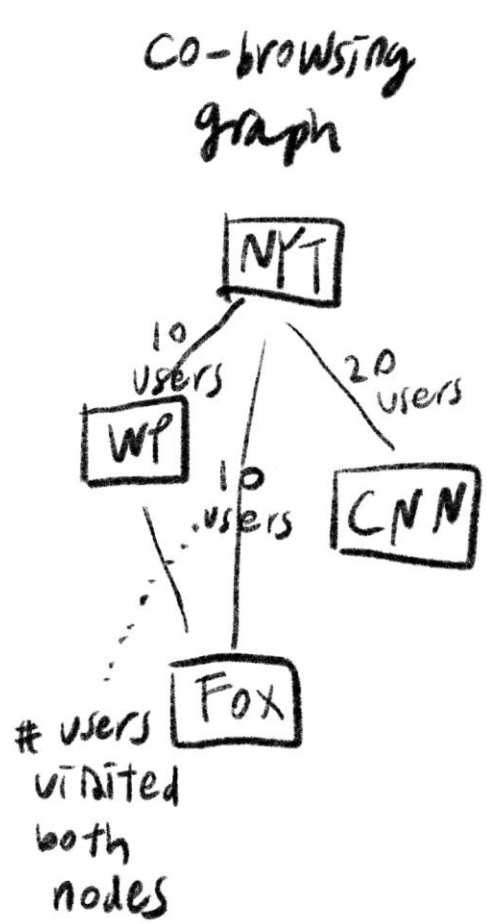
Research Method & Participants

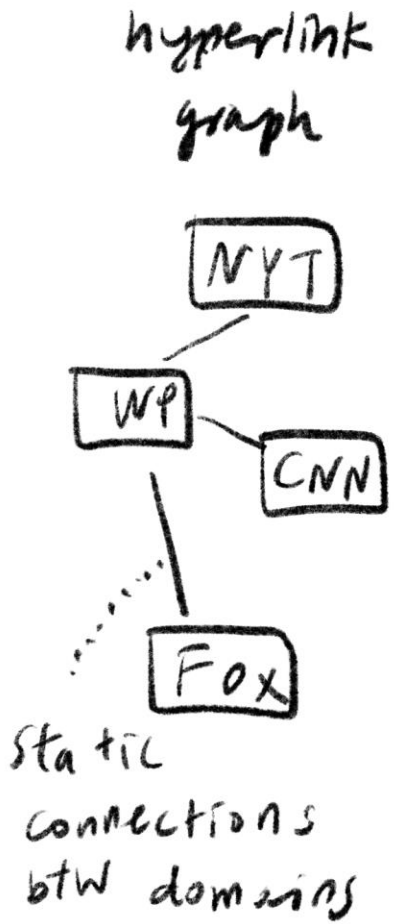
- Observational study: collect browsing log via Firefox extension
- ~~Descriptive research strategy: Looking into variable of political polarization~~
  - ~~Political polarization pattern *discovered* in online news consumption~~
- Correlational research strategy: relationship between political polarization and online news consumption
  - The association between online news consumption and political polarization
- Measure the frequency and duration of online news consumption

- Randomly recruited from active Firefox users based in the US from 5 April 2018 for three weeks.
- Opt-in participant recruitment with informed consent

Data analysis strategy







	Nodes	Edges	Avg. degree	Top PageRank
Co-browsing	1,295	176,945	273	nytimes.com, washingtonpost.com, cnn.com, theguardian.com, npr.com,
Hyperlink	1,295	323,036	498	washingtonpost.com, huffingtonpost.com, nytimes.com, businessinsider.com, theatlantic.com

Table 1: Statistics of co-browsing and hyperlink graphs.

Result 1: Polarization pattern in news browsing

1. Polarization in Dwell Times

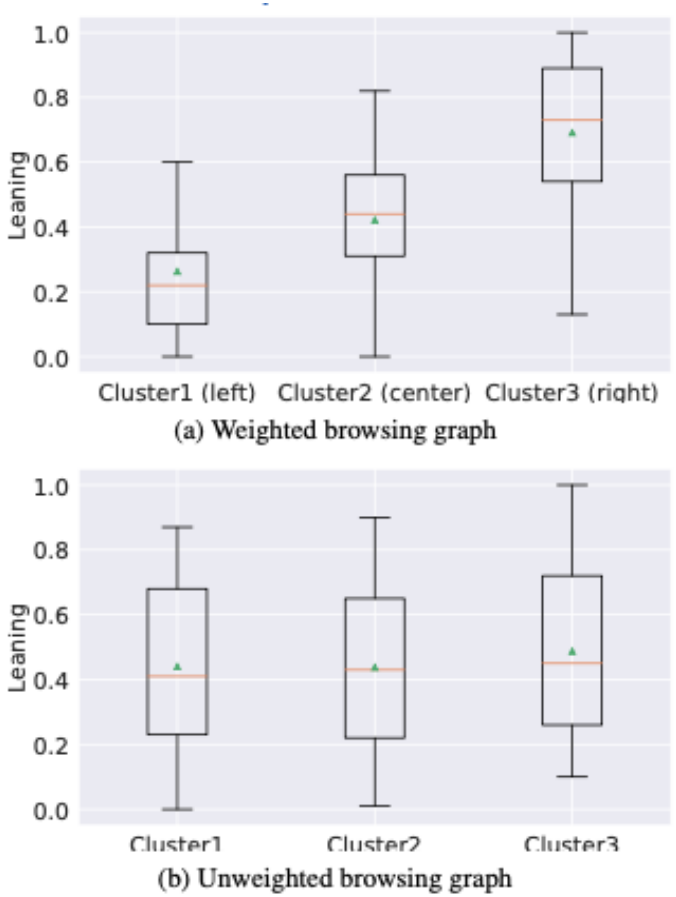


Figure 3: Boxplots of leaning scores per domain cluster, obtained by co-clustering (a) the weighted bipartite (user-by-domain) browsing graph, (b) an unweighted version of the browsing graph. Green dots are means, red lines are medians, box boundaries are quartiles.

2. Time Spent "On the Other Side"

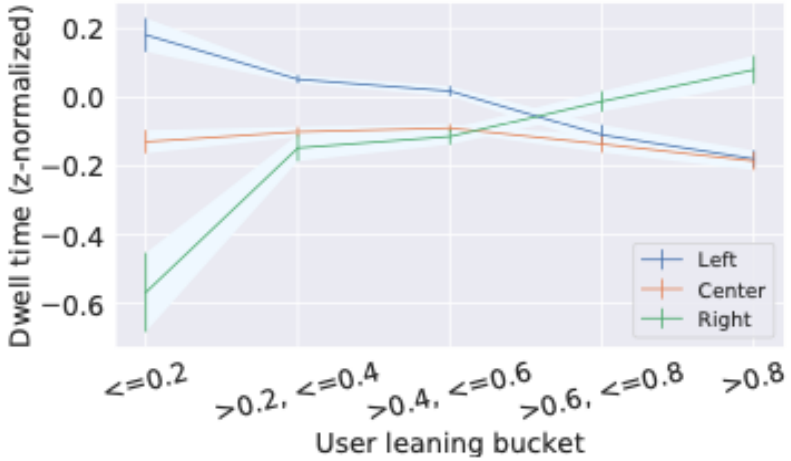


Figure 4: Average dwell time (z-score-standardized within users) spent by users of different leanings (x-axis) on domains of different leanings (three curves). Error bars show 95% confidence intervals. We see that users spend significantly more time when visiting pages on domains aligned with their own leaning.

* leaing score:
0 = left-leaning
1 = right-leaning

3. Community Structure in Co-browsing Graph

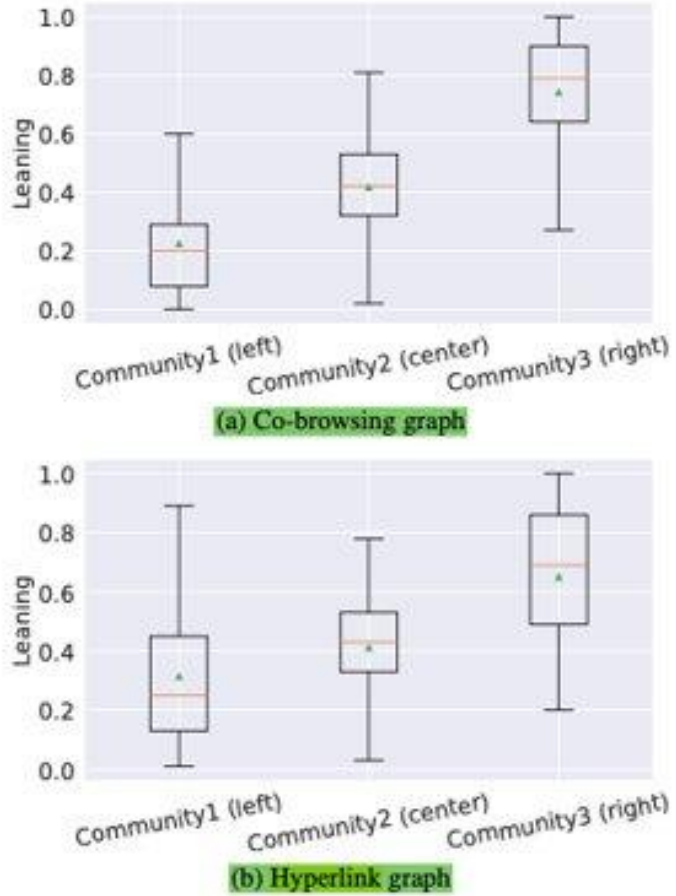


Figure 5: Boxplots of leaning scores per community of domains, obtained by running community detection on (a) the co-browsing graph, (b) the hyperlink graph. Green dots are means, red lines are medians, box boundaries are quartiles.

Result 2: Selective Exposure vs. Structure of the Web

1. Homophily in hyperlink and co-browsing graphs.

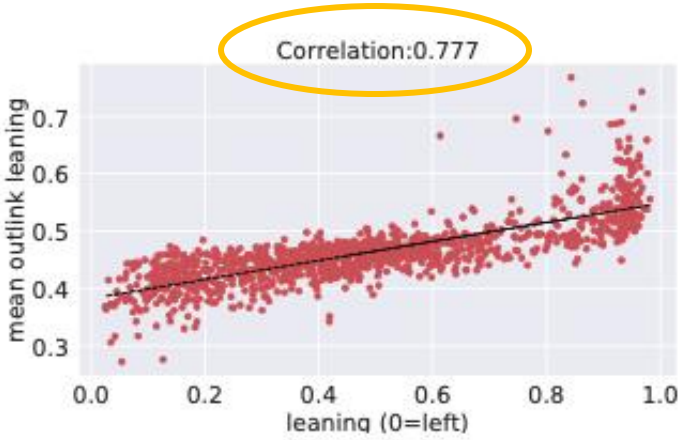
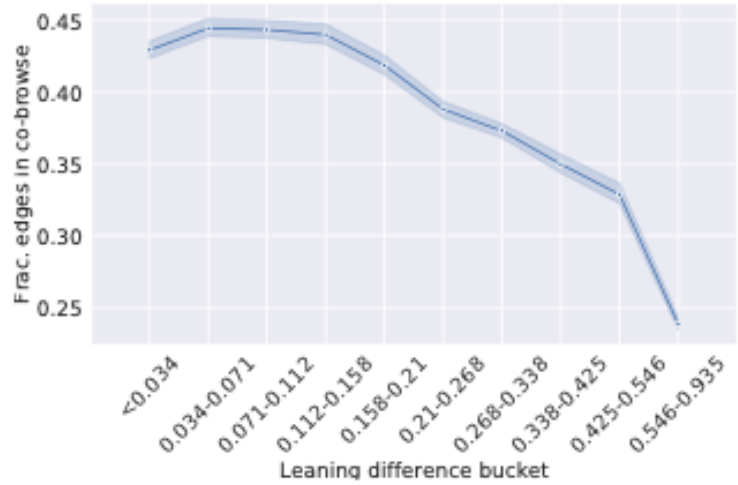


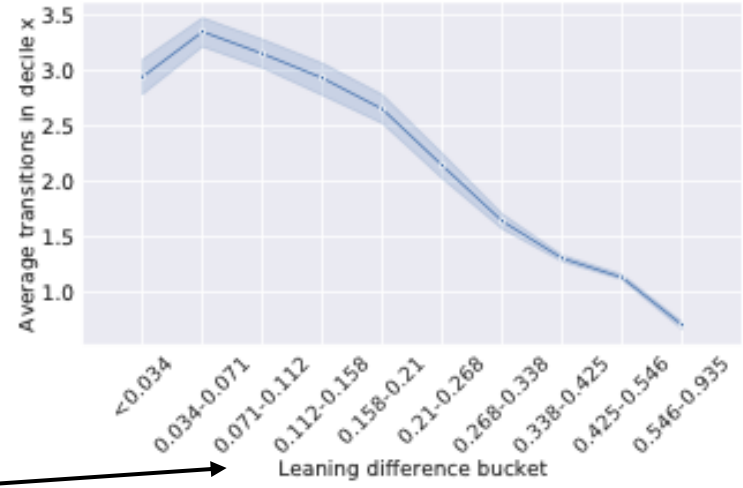
Figure 6: Leaning of each domain (x-axis) vs. average leaning of the domain's neighbors (y-axis) in (a) the co-browsing graph and (b) the hyperlink graph.

Figure 7: (a) Fraction of edges from the hyperlink graph that are also present in the co-browsing graph, when considering, for each node, only those edges that connect the node to a neighbor whose leaning difference falls into the respective decile (x-axis; deciles were computed per node over all its neighbors). (b) Analogous analysis, but with average edge weights, rather than fraction of edges, on the y-axis. The plots show that hyperlinks leading to more similar (with respect to leaning) neighbors are more likely to be chosen by users.

Pearson's correlation coefficient
Significantly lower
(0.696 vs. 0.777, $p<0.01$)



(a) Hyperlink graph vs. unweighted co-browsing graph



(b) Hyperlink graph vs. weighted co-browsing graph

Absolute difference with the node's neighbors in hyperlink graph

Result 2: Selective Exposure vs. Structure of the Web

2. Multi-hop neighborhoods: random walks

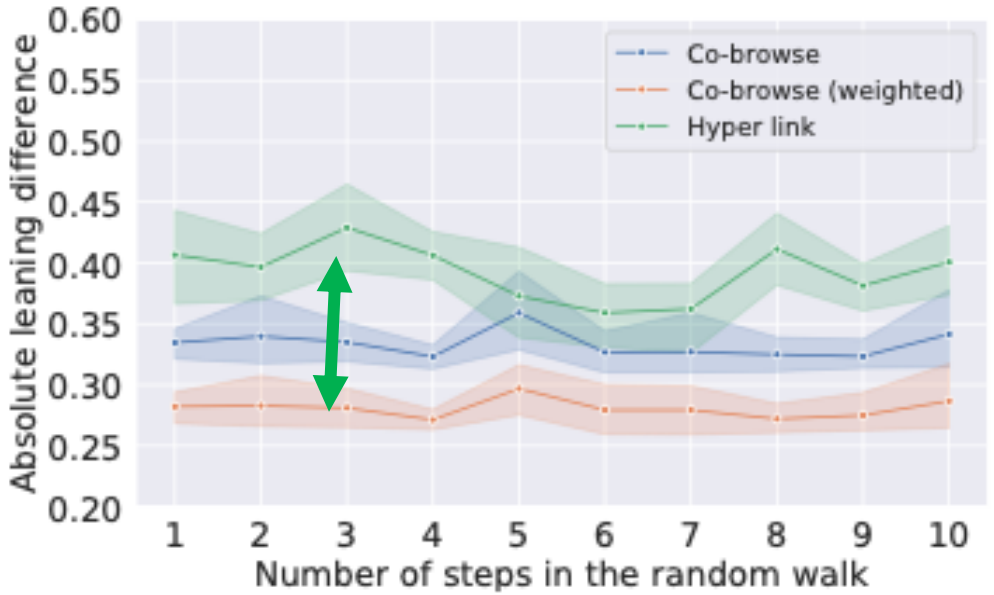


Figure 8: Leaning difference between nodes and their multi-hop neighborhoods in the weighted and unweighted co-browsing graphs and in the hyperlink graph, computed by performing random walks of length 10. Error bands show 95% confidence intervals. The weighted co-browsing graph is most homophilic, followed by the unweighted co-browsing graph and the hyperlink graph.

Strengths

1. The study created large-scale fine-grain digital observational data set to analyze online news consumption.
2. The study discovered three clusters in multiple sources (co-browsing graph, average dwell time on news domains, and hyperlink graph)
3. The study attempts to disentangle different factors in users' online news consumption - by comparing hyperlink structure and co-browsing graph.

+ informed consent for participant recruitment

? But approved by Mozilla ethical and legal board

Weakness

1. The role of algorithmic influence is absent.
   - User's own choice vs. Hyperlink?
   - Entry for news domain: social media, search engine, email newsletter, Hyperlink...?

2. The construct of political polarization is not clearly defined.
   - Loyal NYT reader who doesn't visit Fox news == polarized political left?

3. The author's interpretation of partisan audience bias score from Robertson et al. (2018) as political leaning score is misleading.
   - **Missing Democrats vs. Republican context**
   - Scaling from -1 to 1 to 0 to 1?
   - Matched to voter registration vs. Self-reported ideology

+ No shared any data, code, collection tool

? limitation of industry research with NDA?