Replicating “Who Leads? Who Follows?” (APSR 2020)

November 2025

Presented by

Emily Zhao

Yi Dou

Contents

Introduction – What the Paper Does

Research question:Who sets the political agenda — legislators or the public?

Data & Methods

- Twitter data (2013–2014, 113th U.S. Congress)
- Tweets from Congress members, citizens, and media
- Citizens divided into: general public, attentive public, party supporters
- LDA topic model (≈50 political issues)
- Vector autoregression (VAR) to test "who leads whom"

Introduction – What the Paper Does

Research question:Who sets the political agenda — legislators or the public?

Main Findings

- Legislators follow public attention more than they lead it
- Strongest responsiveness to party supporters
- Moderate response to attentive citizens
- Little reaction to the general public
- Media amplifies partisan agendas

Methods – Data & Text-as-Data

Two main tools: LDA topic model + VAR time-series model

LDA (Latent Dirichlet Allocation)

- Each document = one day of tweets by Congress members
- Chooses number of topics k = 100
- Defines ≈53 political topics (46 distinct issues)
- Measures daily issue attention for each group (politicians, public, media)

Methods – Data & Text-as-Data

Two main tools: LDA topic model + VAR time-series model

VAR (Vector Autoregression)

- Analyzes temporal influence among multiple time-series
- Variables: daily attention levels for each group
- Includes 7-day lags to test who reacts to whom
- Produces Impulse Response Functions (IRFs) → shows how a 10-point rise in attention by one group affects others over 15 days

Results – What We Replicated

Replication 1 : LDA topic models

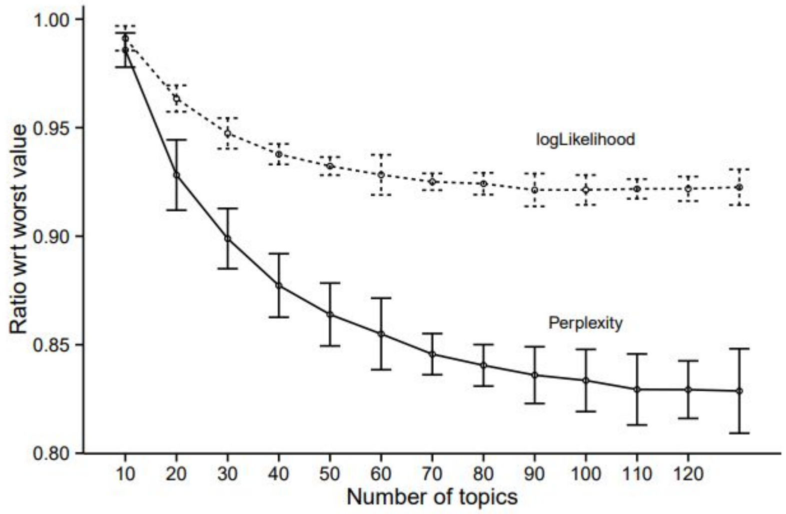
- We don't have the raw data, so we applied the authors' codes to a similar congress data set
- Created document term matrix(DTM)
- Ran our model multiple times with different values of the number of topics (K)
- Fit a LDA model based on the K we chose

Results – What We Replicated

Replication 1 : LDA topic models

Quoted from the paper
"...We find that K=100 fits the data best. A higher value of K would minimize the log likelihood and the perplexity measures, but we choose a conservative K in order to avoid overfitting."

FIGURE A12. LDA model fit with different number of topics

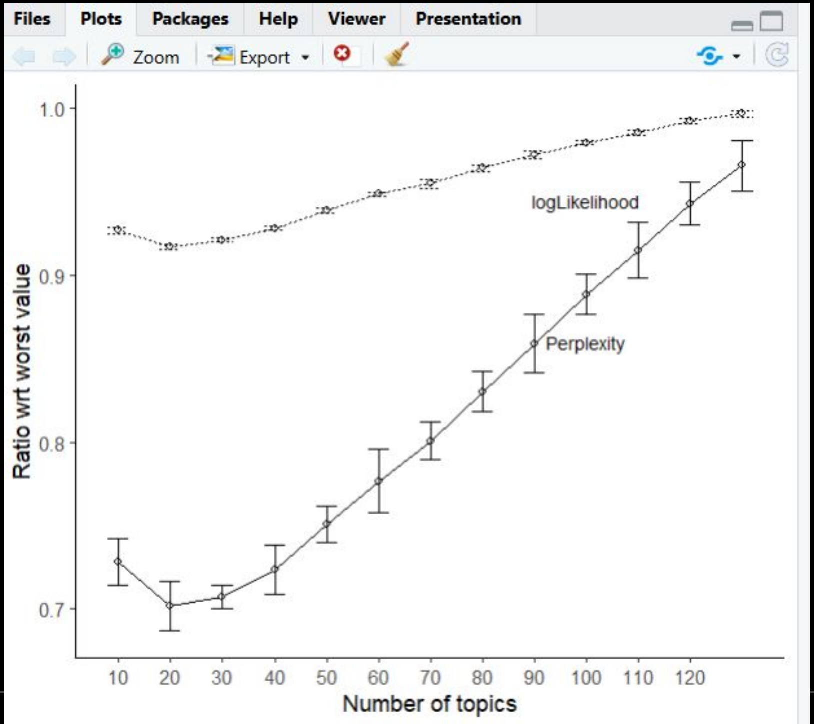


Note: This figure shows the cross-validated log likelihood and estimated perplexity after running our topic model with different numbers of topics. We find that 100 topics yields the best performance.

Results – What We Replicated

Replication 1 : LDA topic models

Quoted from the paper
"...We find that K=100 fits the data best. A higher value of K would minimize the log likelihood and the perplexity measures, but we choose a conservative K in order to avoid overfitting."



Replication 2 - main-time-series.csv

The original main-time-series.csv cannot be regenerated exactly, because the **Twitter API is now restricted**,

We reconstructed ONLY one piece (Replication 04):

- Rebuilt DEM/REP Congressional attention series
- Used tweets_congress.csv (preprocessed)
- Applied authors' LDA model
- Aggregated daily topic probabilities
- Some NAs appear due to sampling & missing tweets

Everything else came directly from the authors' released data (Replication 05):

- Public (Dem, Rep, Attentive) attention
- Random user attention
- Media attention
- US-random user attention
- Pre-merged PRE.csv for all groups except S-random

Reproduce only the transformations:

Merged sub-topics → macro-topics
Added US-random series
Filtered political issues
Saved final main-time-series.csv

Replication 3 - Intercoder Reliability of Topic Labels

Five human coders labeled each of the 100 LDA topics as political or non-political.

Used authors' coder file (topic-list-classification.csv) to classify 100 LDA topics.

Recomputed reliability metrics:

- APIR: 0.83
- Cronbach's α: 0.92
- Consensus labels match the authors.

Confirms the political–topic list we use in all later analyses.

Ensures our pipeline starts from the same validated set of political issues

Metric	Value
Average pairwise agreement (APIR)	0.83
Cronbach's alpha	0.92
Topics labeled political (consensus)	53 of 100

Replication 4 - Table 3

Table 3 reports correlations between issue-attention distributions for members of Congress and public/media groups.

Data used

– Our replicated time-series: data_replication/main-time-series.csv

– Filtered to **political topics only** (3, 7, 9, …, 101–104)

How we computed it

– For each MC group (dem, rep), correlated with pubdem, pubrep, public, random_us, media

– Dropped rows with missing values, used **Pearson correlation**, rounded to 2 decimals

Output

– 5×2 matrix with public/media groups as rows, MC parties as columns

– Saved as images_replication/table3_replication.png for the slides

Result: Our replicated table **matches the published Table 3 exactly**

Group	Democrats in Congress	Republicans in Congress
Democratic supporters	0.69	0.51
Republican supporters	0.41	0.77
Attentive public	0.49	0.52
General public	0.38	0.34
Media	0.52	0.63

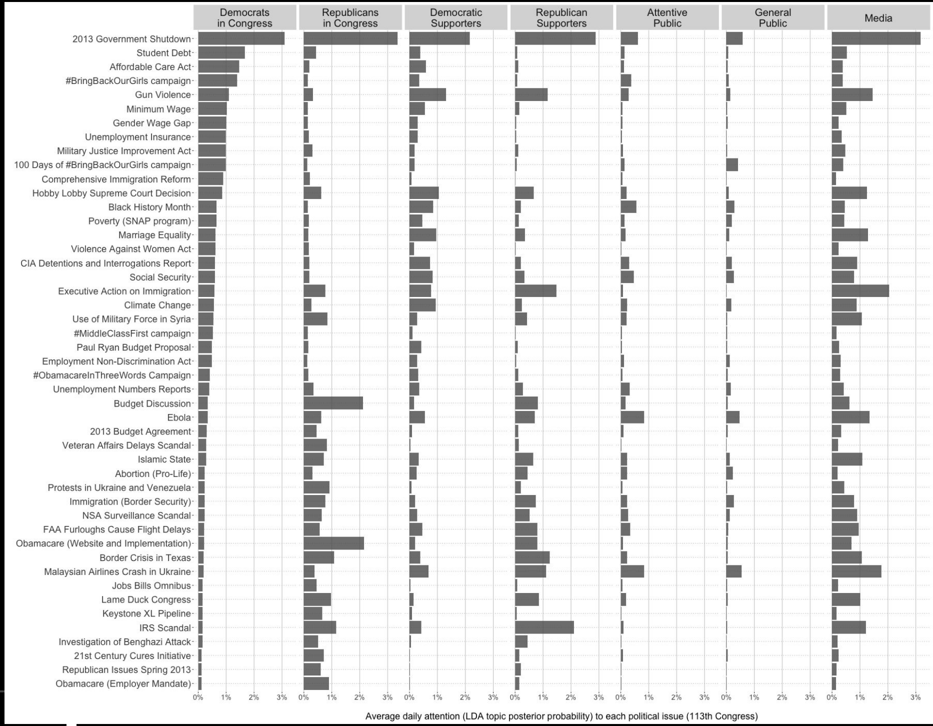
Group	Democrats.in.Congress	Republicans.in.Congress
Democratic supporters	0.69	0.51
Republican supporters	0.41	0.77
Attentive public	0.49	0.52
General public	0.38	0.34
Media	0.52	0.63

Replication 5 - Figure 1

Figure 1 displays the average daily LDA topic probability for each political issue, with separate panels showing issue attention for Democrats and Republicans in Congress, party supporters, the attentive public, the general public, and the media.

How we built it (replication pipeline)

- Start from our reconstructed main time-series
- Keep only topics coded as political (including merged topics 101–104)
- Rename random_us → random and merge in human-readable topic labels from pa2our_topics_crosswalk_merged_subissues.csv
- Reshape to long format and compute mean attention by issue × group over the 113th Congress



Replication 5 - Figure 1

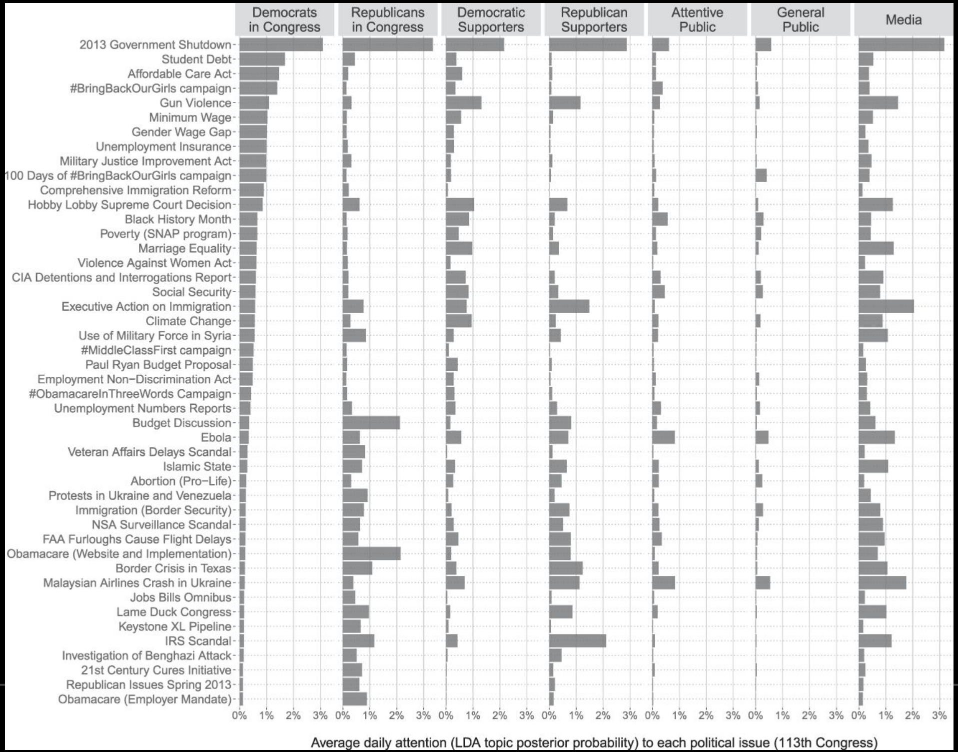
Figure 1 displays the average daily LDA topic probability for each political issue, with separate panels showing issue attention for Democrats and Republicans in Congress, party supporters, the attentive public, the general public, and the media

Plot construction

- One bar per (issue, group) showing mean topic probability
- Facet by group (columns), issues ordered within each panel by attention level
- Y-axis: 0–3%+ of daily attention, shown as percentage

Replication assessment

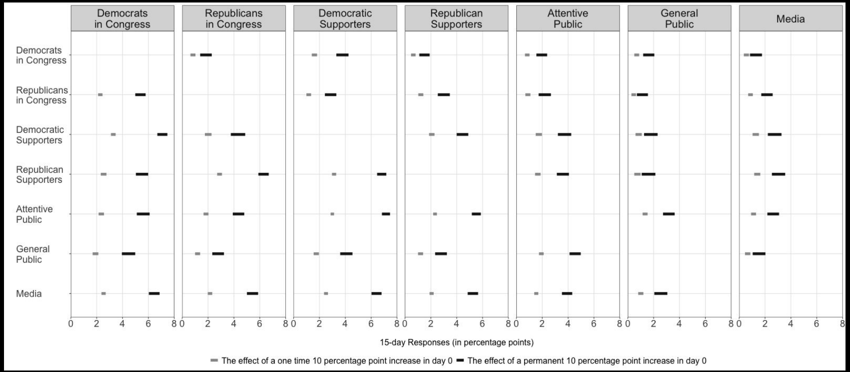
- Rankings of issues and relative bar heights match the original Figure 1
- Only minor cosmetic differences (fonts, gridlines, resolution) due to our ggplot/export settings

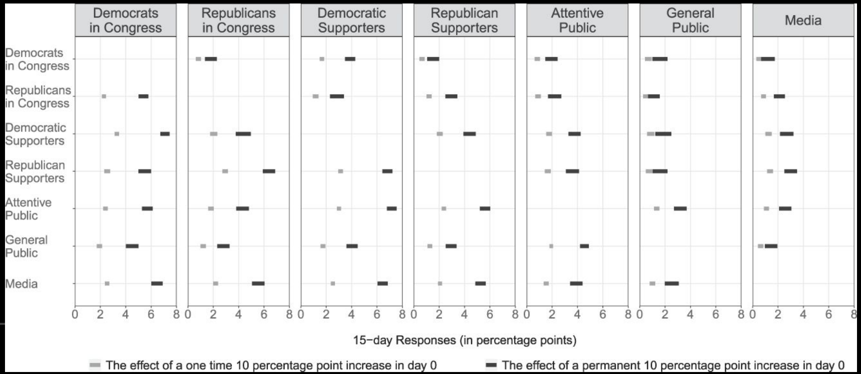


Replication 6 - Figure 2

15-Day Agenda-Setting Responses Across Groups

- VAR(7) on daily topic attention for 7 actors (MCs, party supporters, publics, media).
- Plot shows 15-day cumulative response to a 10-point increase in attention by the covariate group.
- Gray bars: one-time shock at day 0; black bars: permanent shock starting at day 0.
- Our replication reproduces the authors' pattern of who responds to whom (co-partisans > general public).





Replication 7 - VAR-based Figures (3–6)

Figures 3–6 are based on VAR models + impulse response functions (IRFs).

Figure 3

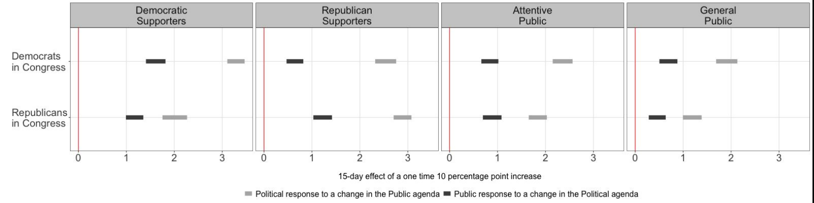


Figure 4

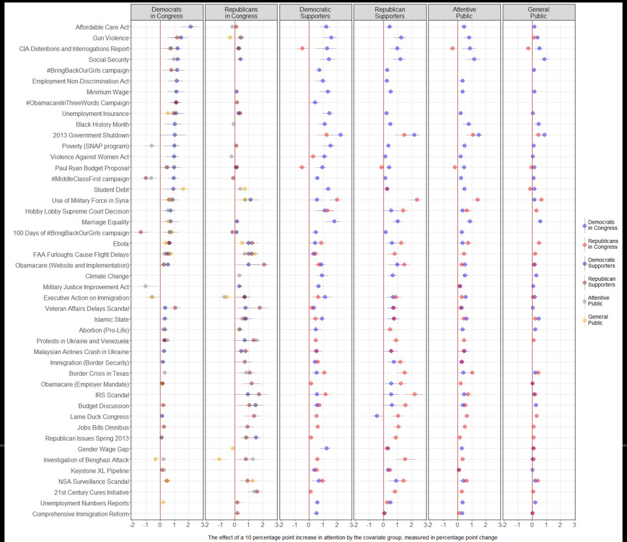


Figure 5

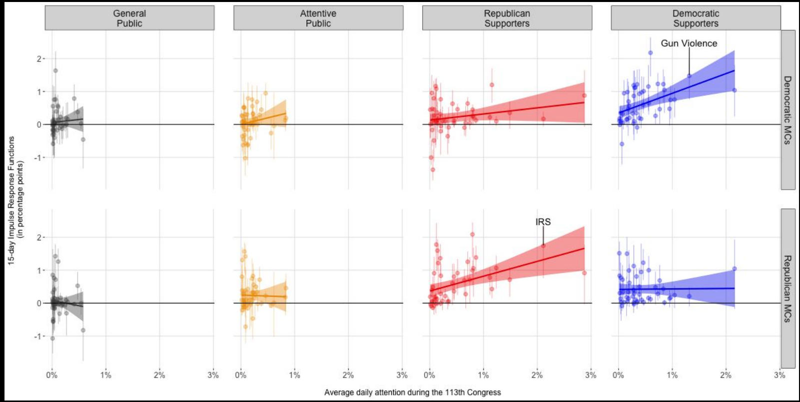
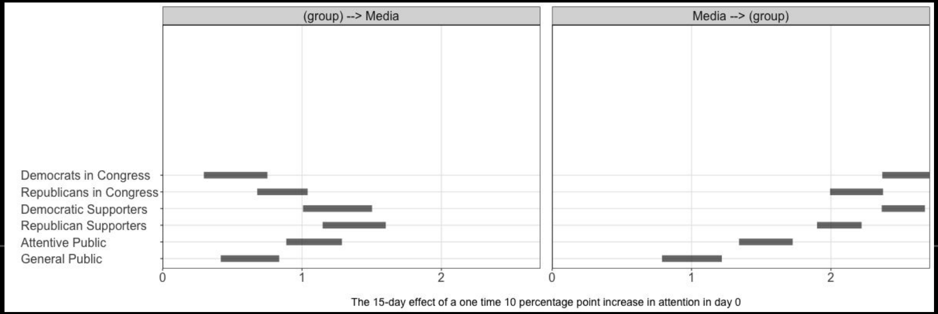


Figure 6



Differences – Data

Authors had full access to:

Full congressional tweet firehose
Original LDA-ready corpora and daily DFMs
Their complete preprocessing pipeline

We could NOT replicate these because:

Twitter/X API is restricted in 2025
Original tweet corpus is no longer available
Several preprocessing scripts rely on deprecated API endpoints

Impact on replication:

- Cannot rebuild the original LDA model
- Cannot match the authors' exact document–topic probabilities

Our workaround:

- Used the authors' **preliminary file to create dataset** for downstream analysis
- Used **tweets_congress.csv** instead of regenerating DFMs
- Able to fully replicate results that rely on **aggregated topic attention**, not raw text

Differences – Technical

Folder structure differs:
· Authors: ./data/, ./var/, ./images/
· Ours: data_replication/, images_replication/, var_replication/, and custom filenames

Updated R environment:
· Newer versions of ggplot2 (size → linewidth), boot::inv.logit, etc.
· Required small fixes and recoding steps

None of these change results—only organization and package behavior.

Differences – Minor Output & Formatting

Figure 4 (Issue-Level IRFs) — Small Numerical Differences

- Authors **did not set a random seed** for VAR bootstrapping
- IRF confidence intervals come from stochastic bootstrap draws
- Even with the same dataset, bootstrapped IRFs will vary run-to-run
- **Our replication uses:**
    - Newer vars and boot package version
    - Modern random number generator behavior
- **Result:**
    - Same issues flagged as significant
    - Same direction and substantive patterns
    - But CI widths and exact point magnitudes differ slightly

Autopsy – What Worked Well

Reproducible downstream results

Once we had main-time-series.csv, nearly all outputs matched the APSR article.
Table 3 correlations identical; Figures 1, 2, 5, 6 matched except for formatting.
Intercoder reliability replicated exactly.

Text-as-data pipeline worked

Reproduced topic-attention series using authors' LDA gamma matrix.
Topic merging (101–104) and crosswalk labels matched the original.

VAR results stable

Main VAR model (Figure 2) ran cleanly.
IRFs showed the same qualitative patterns as the authors.

Autopsy – What Didn't Work Well

Could not reproduce authors' text-processing pipeline

- Original tweet corpus unavailable due to API restrictions.
- Could not rebuild DFMs or retrain LDA.
- Had to rely on tweets_congress.csv + authors' gamma matrix.

Figure 4 not identical

- Issue-level IRFs differ slightly because no seed was set.
  VAR/boot packages updated → minor numeric changes.
  Substantive patterns match, but exact values don't.

Script and environment issues

- Authors' code used deprecated functions; required fixes.
- Folder paths needed restructuring.
- Large VAR loops occasionally crashed R.

Crosswalk inconsistencies

- Topic labels mismatched until correct *pa2our* version was identified.
- Required manual alignment.

Extensions & Innovations

If We Wrote This Today

- Move beyond Twitter (due to API restrictions)
  - use *multi-platform public data*: Reddit, YouTube comments, TikTok political content, news comments, Threads.
- Replace LDA with transformer embeddings
  - Use SBERT/MPNet/GTE embeddings + clustering instead of LDA topics
- Use real-time agenda tracking
  - Streaming embeddings → detect shifts in public and political attention within hours.
- Apply modern causal ML instead of classic VAR
  - VAR-LASSO, Granger-causal forests, Bayesian VARs
- Improve public representativeness
  - Demographic inference + weighting (age, region, ideology) to avoid Twitter biases
- Hybrid human–AI topic validation
  - Use LLM-assisted topic labeling + human adjudication to strengthen validity

Thank you