

# USING LARGE LANGUAGE MODELS TO ANCHOR IDEOLOGICAL POSITIONS IN ONLINE NETWORKS

work in progress / Elena Candellone<sup>1</sup>, Darja Cvetković<sup>2</sup>, Max Pellert<sup>3</sup>, Emma Fraxanet<sup>3\*</sup>  
⌚ → this mark means currently very much in progress, open to suggestions or feedback

<sup>1</sup>Utrecht University, Netherlands  
<sup>2</sup>Institute of Physics Belgrade, Serbia  
<sup>3</sup>Barcelona Supercomputing Centre, Spain

22. Mai 2026, 11:00

## Latent Ideology without public figures: Beyond Twitter data

**DATA** DERSTANDARD forum discussions (2013–2022)<sup>1</sup>



- **Focus year:** 2018
- **Structure:** Threaded discussions + user voting (signed)
- **Entities:** Anonymous users (~35k), comments (~8M), votes (~48M), articles & editorial topical tags

<sup>1</sup>Fraxanet, E., Gómez, V., Kaltenbrunner, A., Pellert, M. A Decade of News Forum Interactions: Threaded Conversations, Signed Votes, and Topical Tags. *Sci Data* (2026).

**Correspondence Analysis** (CA) of user–politician interaction networks has been widely used on Twitter/X<sup>2,3,4</sup> to project both users and politicians into a shared **latent ideological space** based on co-occurrence matrices, with results shown to approximate Bayesian Ideal Point estimation<sup>5</sup>. These approaches rely on **politicians as ideological anchors** and **follower/retweet networks** as the **interaction signal**. Our adaptation:

- **Politicians** → **Influential users:**
  - Active in discussions
  - High Page Rank centrality in a vote+reply signed network. Stronger signal and wider ideological spectrum.
- **Following/retweet** → **Positive votes:** (+ negative)
  - Including semi-active users (only voters)
- **One axis** → **Multiple issues/dimensions:**
  - Subset of votes based on editorial tags
  - Avoids conflating all topics into one dimension

## FINDING A DIRECTION....

Two-step LLM annotation pipeline<sup>6</sup> (open reasoning model: Qwen3-32B):

### ① Summarization

- For each influencer × issue, all comments (with thread and article context) are summarized into a compact stance manifesto of fixed length

### ② Scoring

- The manifesto is scored on a well-defined 7-point scale specific to each issue (e.g. Pro-EU ↔ Anti-EU)

### ⌚ Ground truth?

- Preliminary → Explicit party support as proxy (LLM selection + 2-human verification)
- Future → Human annotation set up for score validity

## FUTURE RESEARCH QUESTIONS ⌚

### RQ1: Issue alignment & ideological structure

→ Does ideological coherence emerge across issues?

### RQ2: Engagement-driven distortion

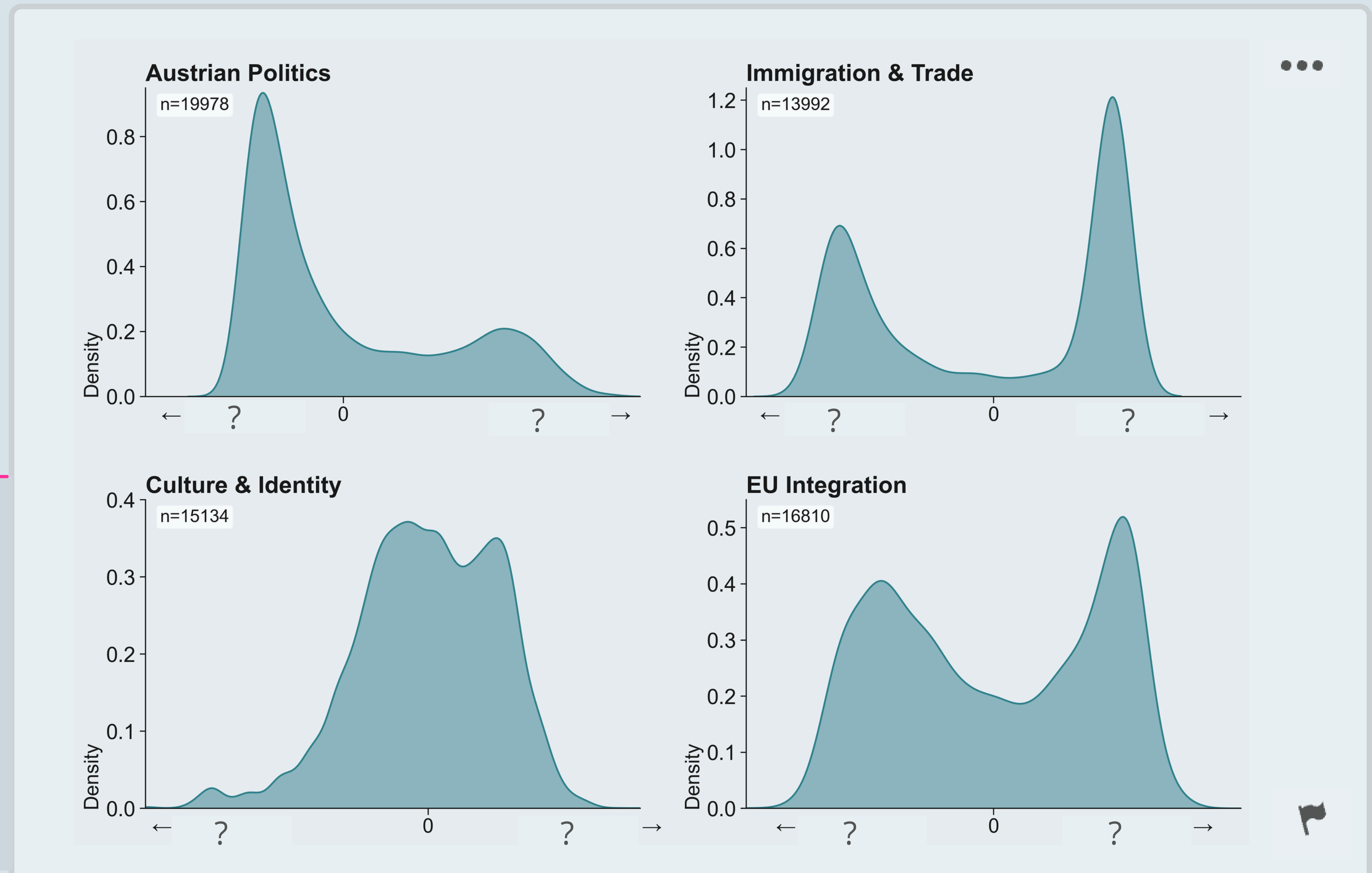
→ Does engagement distort the political signal observed in online data?

### RQ3: Issue ownership

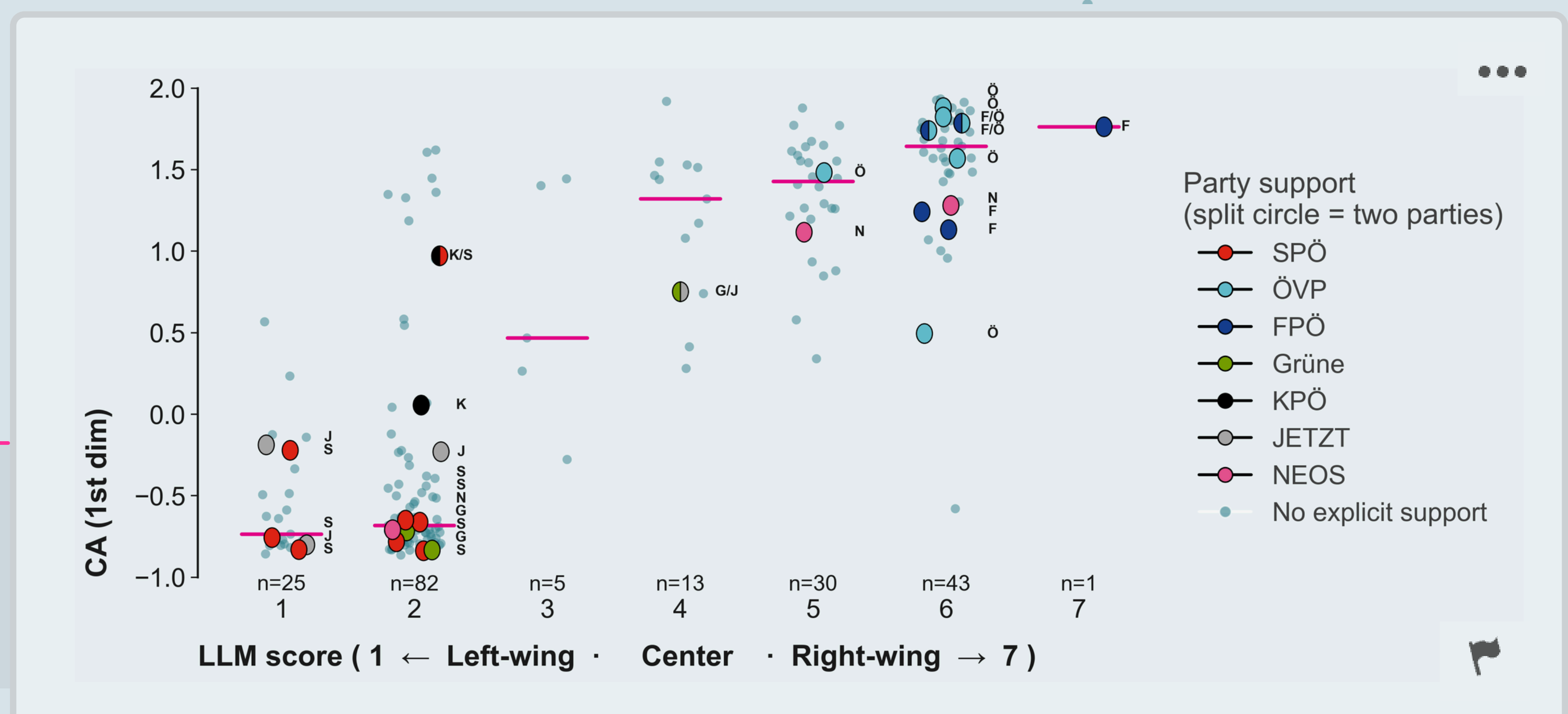
→ Are some issues dominated by one side of the political spectrum?

## Preliminary results

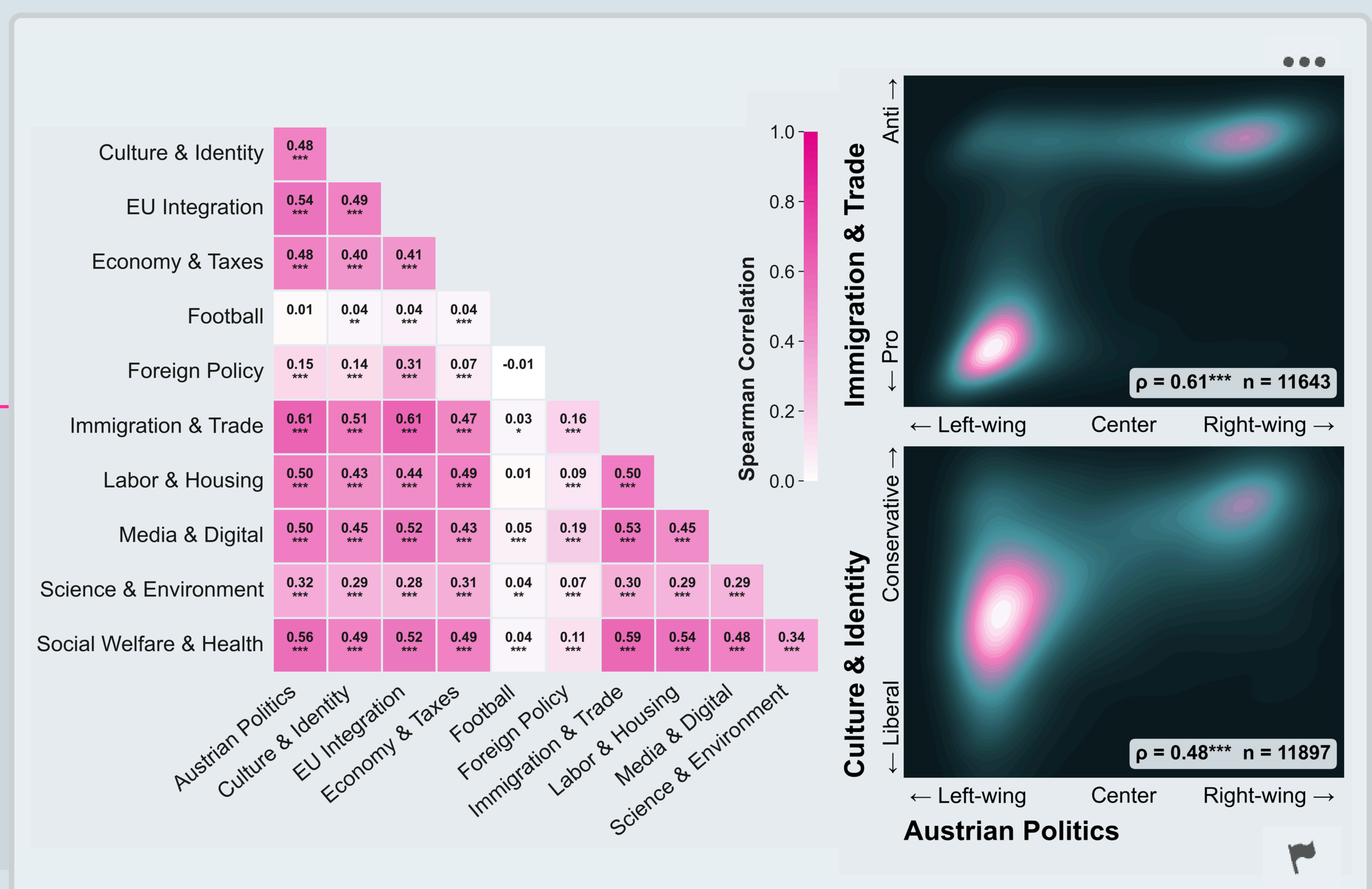
### ☐ CA (1st dim) distributions by issue · 2018



### ☐ Agreement CA · LLM Score · Explicit Party Support



### ☐ RQ1: Issue Alignment · 2018 · CA



<sup>2</sup>Barberá, et al. (2015). Tweeting from left to right: Is online political communication more than an echo chamber?. *Psychological science*.

<sup>3</sup>Falkenberg, M., et al. (2024). Patterns of partisan toxicity and engagement reveal the common structure of online political communication across countries. *Nature Communications*.

<sup>4</sup>Ramaciotti Morales, P., et al. (2022). Inferring attitudinal spaces in social networks. *Social Network Analysis and Mining*.

<sup>5</sup>Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political analysis*.

<sup>6</sup>Benoit, K., et al. (2025). Using large language models to analyze political texts through natural language understanding. *American Journal of Political Science*.

What do you think of the poster? 49 reactions

39 informative 1 helpful 2 touching 7 entertaining