

# An Annotated Dataset of U.S. Transgender News For Determining Agenda-Setting & Information Flows

Sagar Kumar, **Alyssa Smith**, Yukun Yang, Pranav Goel

Network Science Institute, Northeastern University

# Outline

- **Motivation & Background**
- **Current Findings**
- **Annotation Method**
- **Future Work**

A red speech bubble graphic with a white outline, pointing downwards. The text "Motivation & Background" is centered inside the bubble in white. The background features a pattern of thin, light gray curved lines and dashed lines.

# Motivation & Background

## The State of U.S. News About Transgender People

- **Attacks on trans rights are increasing** (P. E. Jones and Brewer 2020; M. Fischer 2019; Carlisle 2022; T. Jones 2023)

## The State of U.S. News About Transgender People

- **Attacks on trans rights are increasing** (P. E. Jones and Brewer 2020; M. Fischer 2019; Carlisle 2022; T. Jones 2023)
- **Coverage of trans issues in U.S mainstream media is often problematic** (Cavalcante 2018; Capuzza and Spencer 2017; Osborn 2022 )

## The State of U.S. News About Transgender People

- **Attacks on trans rights are increasing** (P. E. Jones and Brewer 2020; M. Fischer 2019; Carlisle 2022; T. Jones 2023)
- **Coverage of trans issues in U.S mainstream media is often problematic** (Cavalcante 2018; Capuzza and Spencer 2017; Osborn 2022 )
- **News coverage sets the agenda for public discourses** (M. Fischer 2019; DeJong et al. 2021; Dunaway, Branton, and Abrajano 2010; Luo et al. 2019)

The State of U.S.  
News About  
Transgender  
People

FOX NEWS FLASH

# Colorado mom sues school that recruited sixth-graders for secret after-school gender and sexuality club

Erin Lee tells Harris Faulkner her daughter was told she was joining 'art' club



By Taylor Penley · Fox News

Published May 10, 2023 4:00am EDT



LOCAL NEWS

# Colorado parents sue over state law and school policy regarding transgender students

© CBS NEWS COLORADO

By Shaun Boyd  
August 9, 2024 / 6:14 PM MDT / CBS Colorado



The State of U.S.  
News About  
Transgender  
People

- **“Parental rights” framing robs queer youth of agency**
- **Tying youth education on LGBTQ+ identities to “indoctrination” feeds into harmful tropes about LGBTQ+ people.**

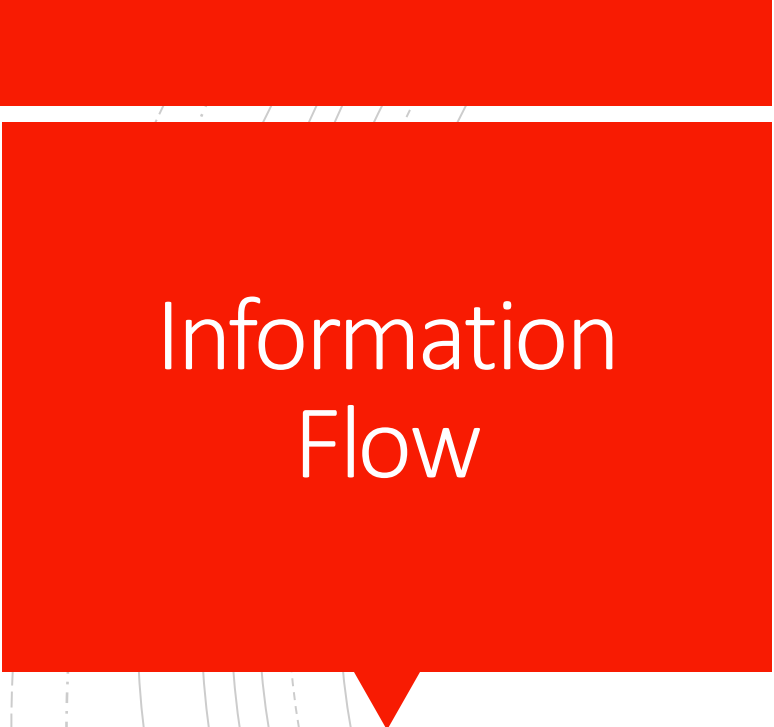


A red speech bubble with a white outline and a downward-pointing tail. The text 'Current Findings' is centered inside the bubble in white. The background features a pattern of thin, light gray curved lines, some solid and some dashed, creating a sense of motion or a stylized wave.

Current Findings

## Information Flow

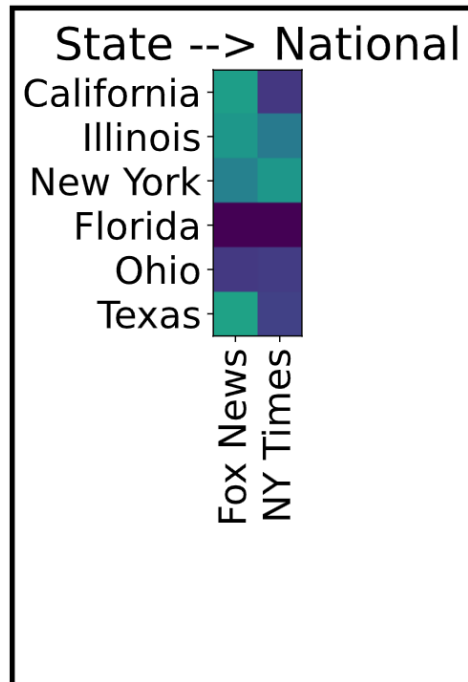
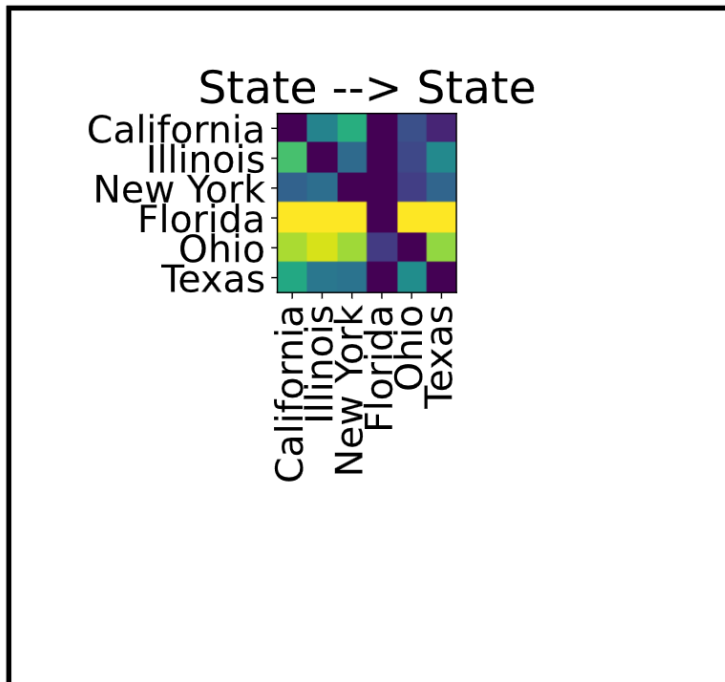
- We looked at the information flow between national outlets & local (state-level) outlets.
- We used **transfer entropy**, a measure of **drivenness**.
- Information flows from national outlets to local ones ...

A red speech bubble graphic with a white outline, containing the text 'Information Flow'. The bubble has a tail pointing downwards and to the right.

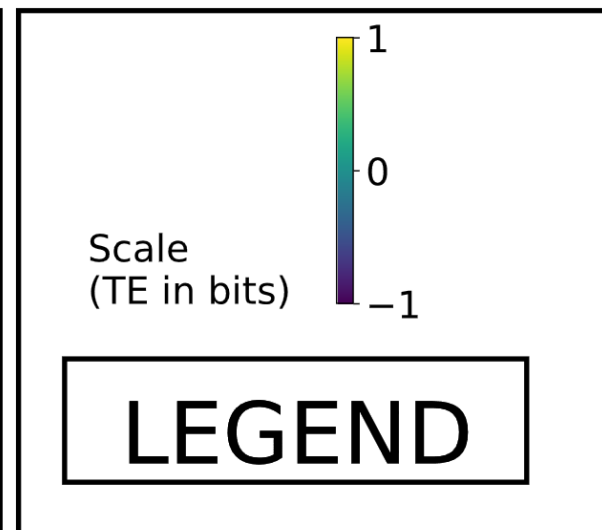
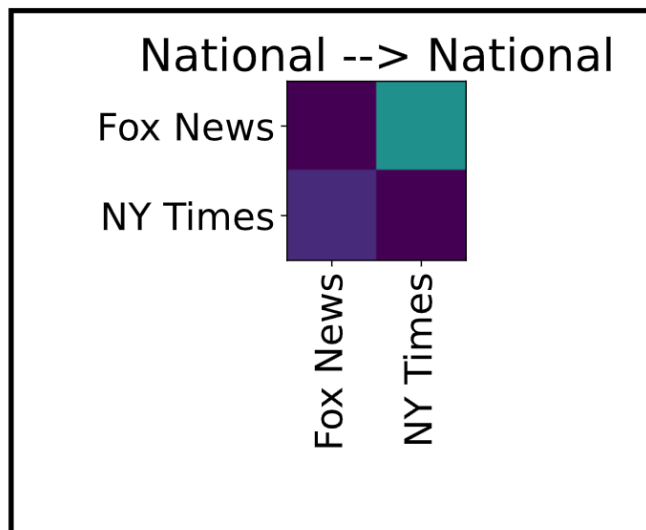
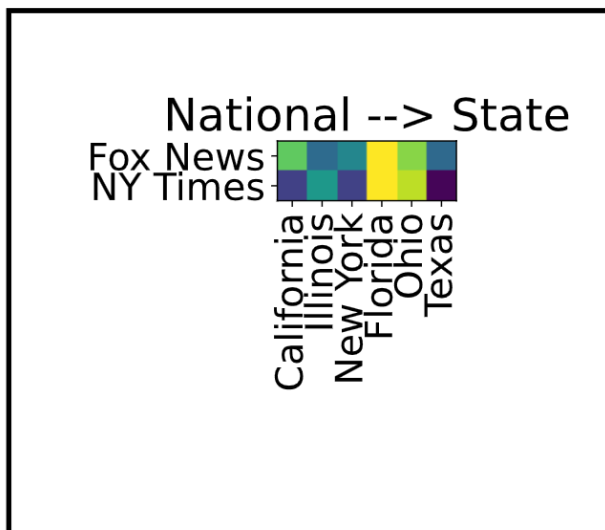
## Information Flow

- ...but it's not a direct one-step flow!
- Instead, we see one or two states mediating the information flow of a particular topic.

Transfer Entropy: K-12 school policies regarding LGBTQI+ students



Flow Pattern:  
Two-step flow  
mediated by Florida  
and Ohio



However...

- The topic model we used, LDA (Latent Dirichlet Allocation), is good for spotting broad topics
- It's not so great at spotting subtle biases and tropes.
- Humans are needed to do this work!

The background features several concentric circles of varying radii, some solid and some dashed, creating a ripple effect. A prominent red callout box is centered on the page, containing the title text. The box has a rectangular top and a pointed bottom, resembling a speech bubble or a data point marker.

# Towards A Human- Annotated Trans News Dataset

DELUGE (Bragg  
et al., 2013)

- **Efficiently produces human-generated taxonomies of labels for text datasets.**
- **Allows for probabilistic accuracy (crowdworkers can be imperfect and produce useful annotations)**

DELUGE (Bragg  
et al., 2013)

**GENERATE:** workers each suggest a label for an item

**SELECT BEST:** workers pick the best label for an item out of  $c$  choices.

**CATEGORIZE:** workers choose which label(s) out of  $s$  choices apply to an item



## Timeline

- **Sept 2024: Finish MTurk sandbox testing; finish DELUGE script writing**
- **Oct 2024: Complete the full pipeline testing; publicize the tasks**
- **Nov 2024: Finish data collection; produce basic summary statistics on collected data**
- **Dec 2024: Release dataset on the Social Media Archive (SOMAR)**

## Future Uses

- **Taxonomy of human-generated thematic labels for discourses on transgender people in U.S. mainstream media (across national and local news sources); a publicly available resource for research.**
- **An annotated dataset to help study intermedia agenda-setting in the U.S. media ecosystem around trans issues.**
- **A resource for comparing topic modeling output and human-generated labels on a dataset already filtered and focused on one broad theme; evaluation of computational labeling systems.**



Thank you!!!

(email me at [smith.alyss@northeastern.edu](mailto:smith.alyss@northeastern.edu))