

A Computational Study of Protest Narratives in Journalism

By Peter Ball, BA Cultural Studies & Computer Science

What?

The project examines the narrative structure of articles about American protests in the past 6 years, mapping narratological theory from the humanities onto the classic Hidden Markov model from computer science.

Why?

- Public concern over divisive and coercive powers of media narratives
- Currently major efforts to model media narratives at scale, but most work so far has been on document-scale content, not form and structure
- Opportunity to return to narrative theory's root in linguistics during a renaissance in both fields

The Model

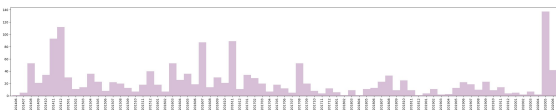
- HMM detects structure in a symbol sequence by learning an underlying system of **hidden states** with distributions over the symbols, and that transition to another state at each time step with some probability. This system is said to **have produced** the sequence.
- In our project, the symbols are **actants**, and the sequence contains their **interactant relationships**. The actants are protesters and police, and system captures a grammar of their interactions: when a journalist writes an article, they use their "narrative competence" to navigate this graph.
- Reduce article to sequence of symbols where each symbol is an instance of the actant as either subject or object:

	Protester	Police
Subject	P	S
Object	Q	T

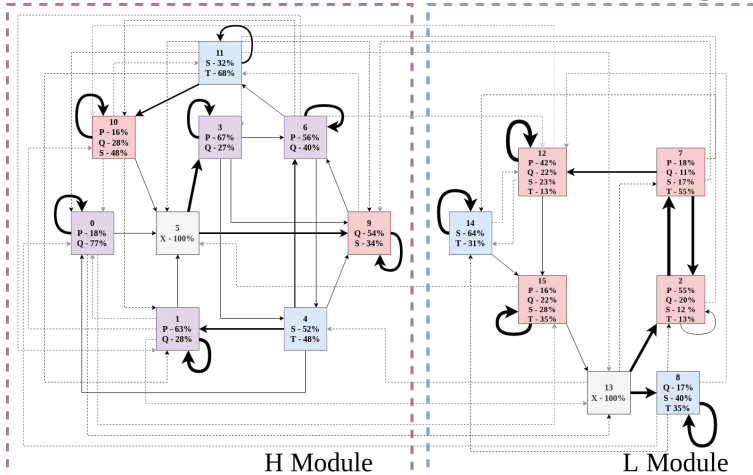
"Around 11:15 a.m., officers told the protesters to leave the area or be arrested. Officers backed the protesters down the street, and the protesters had dispersed from east main street around 11:30 a.m."

→ S Q S Q P → 10 10 10 10 10

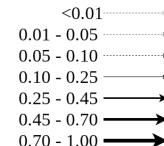
Data Time Series:



Protester-State Interactions in Protest Coverage

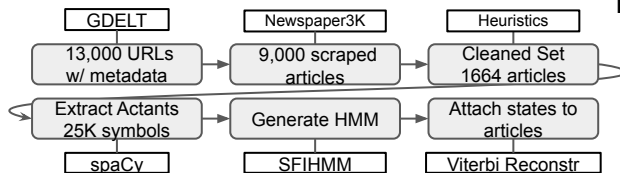


Transition Probabilities



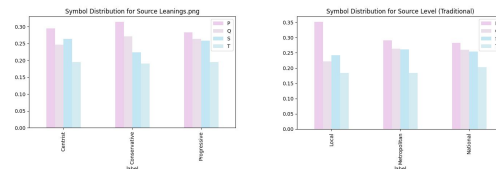
Symbol Set

P - Protester as Subject
 Q - Protester as Object
 S - Police as Subject
 T - Police as Object
 X - End of Article



Dataset:

The final set has 1664 articles with metadata including source and date. The set is comprised of articles published by US news sources between 2014-06-01 and 2020-06-01 covering contemporary protest. The top 100 sources have been tagged for political leaning and distribution level.



Results:

- **Motifs:** The most common motifs for H, PPPP and L, SSSS, diverged in actant but were similarly uninterested in interaction. However when the most distinctive motif was calculated (ratio of probabilities), we found QQQQ was 15 times more probable in H, and TPTS was 48 times more probable in L; H was interested in protester objectivity, while L focused on interaction.
- **Classifying on module:** political leaning and distribution level of source were ineffective. But given "before/after election," the classifier achieved 65% accuracy, with L being twice as common after the election. This suggests that the election had a greater impact on protest narratives than even political bias.
- **Paths:** 3 most common paths (collapsing self-loops) are [2 7 12 15 13] (p=0.15), [8 4 15 13] (p=0.09), and [3 4 1 5] (p=0.09).
 - LA Times and Fox News and RawStory were similar in path usage
 - Fringe conservative publications Breitbart, Bizpacreview, and Daily Caller diverged heavily from the norm, opting for [9, 6, 4, 1, 5] despite it's low overall usage (p=0.03).
 This suggests that newer, online conservative publications have developed a separate narrative space from the norm.

Acknowledgements:

A big thank you to my fantastic supervisors Richard Jean So and Simon DeDeo, and to the AUS for sponsoring my award.