Detecting and Analyzing Framing of Controversial Issues in U.S. Social Media

Introduction. Language plays an important role in forming perspectives and shaping opinions. This often occurs through framing: how a speaker defines an issue and relates it to other problems, people, or values [1]. For example, discourse about same-sex marriage invoke topic-level frames such as *morality*, *civil rights,* and *democracy* [2]. Framing also occurs in specific phrases; while *gay marriage* centers an affected group identity, *marriage equality* emphasizes commonalities between all marriages. An issue's framing has vast implications for public opinion, policy changes, and social movements [3]. Since millions of people contribute to online discourses about controversial issues, it is crucial to be able to automatically detect framing on a large scale. I will develop computational methods to detect and analyze the framing of controversial issues on social media, and study how framing influences the public towards particular perspectives.

Prior NLP research has identified topic-like frames in mass media [4,5,6]. A related area is stance detection: an author's stance for or against an issue may impact readers' opinions. Cognitive science work has shown that word choice affects how issues are perceived [7]. I thus consider 3 "dimensions" of framing: topic-level, stance, and lexical. NLP framing work has not addressed lexical framing or stance, and computational approaches have yet to be developed for social media. I will fill these gaps in my proposed research, which will be conducted in 3 phases. In Phase 1, I detect framing on Twitter and analyze relationships between dimensions. In Phase 2, I study how framing influences public opinion in Twitter and the "real world". I use immigration as a case study because many topical and lexical frames have been identified for this issue [7]. I validate my approach for a range of issues on multiple social media platforms in Phase 3.

Data and Methods. I have access to tweets from 2012 onwards through the Twitter Decahose and a full collection of Reddit data since 2007. I will use other publicly available data, including speeches from the U.S Congress [8], opinion polls, and lists of controversial issues. This project will require GPUs for deep learning, which I can access through my university.

Phase 1. I will develop new NLP methods to detect framing along three dimensions: topic-level, lexical, and stance. **Topic-level.** I will develop semi-supervised and unsupervised methods to detect topic-level frames. First, I will annotate a subset of tweets with frames from social science codebooks [4], and use label propagation to label the remaining tweets. Second, I will jointly embed tweets and words, and perform clustering to assign tweets to frames, which can be interpreted by nearby words [9]. While semi-supervised learning leverages experts' codebooks for mass media, unsupervised learning may find more relevant frames for social media. I will validate both approaches with a "tweet intrusion" task [9]. **Lexical**. Though straightforward to detect a phrase's presence, identifying implicit meaning differences between phrases is more challenging. I will analyze lexical framing by training word/phrase embeddings on tweets and identifying words/phrases near issue keywords via cosine similarity. I previously used this method to reveal connotational differences between the definitionally-similar words *gay* and *homosexual*. **Stance**. While stance detection is often a classification task, I will obtain continuous scores for a finer-grained perspective. Using contextual embeddings as features in a regression fit to a real-valued sentiment lexicon, I will predict stance directly from keywords [10]. For example, *immigrant* may be positive in one tweet and negative in another, reflecting opposing stances towards the issue.

I will then investigate relationships between framing strategies. If a user has positive stance towards immigration with a *morality* frame, are they more likely to say *undocumented people* over *illegals*? I will compare frame co-occurrence frequencies and use high-dimensional stratified propensity score analysis to understand the effect of a given frame on the use of other frames [11]. This technique enables causality estimation while accounting for confounding variables.

Phase 2. I will better understand how sociopolitical information diffuses across a network by investigating how frames spread on social media and change over time. Using information diffusion techniques, I will identify which Twitter users create and propagate frames. I will also combine NLP and network analysis to study this phenomenon on a group level. This can answer a central question in political communication theory: do frames spread "top-down", originating among elites and later reaching the public, or "bottom-up"? I will label Twitter users as politicians, journalists, and members of the public. Manually curating full lists of journalists and politicians is labor intensive, so I will apply community detection along with incomplete lists to uncover a more complete set of individuals. By calculating frame frequencies over time and Granger causality, I will determine which group changes narratives surrounding controversial issues.

I will also show how topic-level, lexical, and stance framing impact a tweet's influence by using them as features in a supervised model to predict likes and retweets. I will then use deep learning techniques to analyze how framing on Twitter impacts and predicts real-world outcomes, such as public opinion polls and politicians' framing in official statements.

Phase 3. I will ensure that my methods generalize across controversial issues, which is challenging due to the substantial ways in which they differ. Some are partisan (immigration), others are not (online privacy). Some evenly split public opinion (abortion), others do not (vaccinations). Using polling data and lists of issues, I will create an ontology to systematically address such factors and replicate my analysis from Phases 1 and 2 for each issue in this ontology. Comparing across issues will help me determine how these factors impact framing. For example, politicians may shape narratives around partisan issues, but not others. This approach will help me identify universal aspects of framing on Twitter and better understand our complex social media landscape. I will also generalize to other social media platforms, especially Reddit, which has value-oriented communities (e.g. r/sustainability) across which I can compare framing strategies.

Intellectual Merit. The proposed work is a deeply interdisciplinary endeavor that will contribute to NLP and social science research. My project will advance NLP technology for recognizing speakers' implicit attitudes from subtle linguistic signals, which cannot be fully understood with existing methods. Thus, this work has implications for improving socially-aware NLP models, such as machine translation or text summarization, that can capture the subtleties of framing and stance. Moreover, my project can transform framing research by facilitating large-scale analysis and providing re-usable tools for social scientists to conduct similar studies.

Broader Impact. I will encourage the wider academic community to build upon this work by sharing my findings through publications and releasing datasets. This proposal has implications far beyond academia. Framing can influence public opinion, policy, and political campaigns, all of which deeply impact people's lives. Finally, by providing a nuanced view of people's perspectives towards controversial issues, this project could lay foundations for more productive and cooperative discourse, and help erode stark divisions in our society.

References. [1] Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of communication*. [2] Hull, K. E. (2001). The political limits of the rights frame: The case of same-sex marriage in Hawaii. *Sociological Perspectives*. [3] Chong, D., & Druckman, J. N. (2007). Framing theory. *Annu. Rev. Polit. Sci.*. [4] Card, D. et al. (2015). The media frames corpus: Annotations of frames across issues. *ACL*. [5] Tsur, O. et al. (2015). A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. *ACL*. [6] Field, A. et al. (2018). Framing and Agenda-setting in Russian News: a Computational Analysis of Intricate Political Strategies. *EMNLP*. [7] Lakoff, G. (2006). The framing of immigration. [8] Gentzkow, M. et al. Congressional Record for the 43rd-114th Congresses: Parsed Speeches and Phrase Counts. [9] Demszky, D. et al. (2019). Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings. *NAACL*. [10] Field, A., & Tsvetkov, Y. (2019). Entity-Centric Contextual Affective Analysis. *ACL*. [11] Olteanu, A. et al. (2017). Distilling the outcomes of personal experiences: A propensity-scored analysis of social media. *CSCW*.