

# Detecting and Annotating Narratives in Social Media: A Vision Paper

Mack Blackburn, Ning Yu, Alex Memory, W. Graham Mueller

Leidos, Inc.

4001 N. Fairfax Drive, Suite 800

Arlington, Virginia 22203

{mack.blackburn, ning.yu, alexander.c.memory, william.g.mueller}@leidos.com

## 1 Introduction

Understanding how narratives spread and influence target audiences in dynamic information environments like social media has a broad range of applications, e.g., to anticipate public reaction to policy changes or disinformation campaigns. In this paper we describe our ongoing efforts to define, collect, detect, and annotate narratives online in order to analyze and model the spread of these narratives. We also share lessons learned and our vision for how to improve these sub-tasks. We define *narrative* as a collection of statements expressing a point of view on some topic. These statements may explain events, interpret the motives of actors, or emphasize specific perspectives. We believe this minimally-restricted definition helps our findings apply across disciplines.

## 2 Narrative in Social Media

Narrative plays an important role in online interaction and has been studied in the fields of literature, philosophy, communication, marketing, and more recently, computational social science. While past research has often used lengthy documents such as news articles, there is rising demand to study narratives in social media due to the proliferation of data and its broad author/reader base, which leads to diverse perspectives and evidence of influence campaigns (Sadler 2018).

**Characteristics** Unlike narratives in traditional media, those in social media are often written by multiple authors—sometimes in coordination—and they evolve over time through online interactions between social media users. For those reasons, detecting narratives in social media is a unique challenge. Additionally, narratives may emerge in social media then spread to traditional media, or vice versa. Therefore, detecting narratives in social media may also require analysis of traditional media. Both traditional and social media have a mix of readers, but transparent interactions between readers and authors on social media provide an opportunity to study the influence of narratives on readers.

**Understanding Spread of Narratives on Social Media**  
An important application of social media narrative detection

is modeling the spread of narratives online. Using technologies such as simulation, we can predict how quickly and widely a narrative will spread or whom it will reach. Early narrative simulation efforts used simple surrogate information such as hashtags, URLs, and topics. However, detecting and using narratives themselves provides greater information and context to the simulations in the form of ground truth.

## 3 Narrative Detection and Annotation

We have constructed two narrative datasets related to the White Helmets (a Syrian Civil Defence group, SETA), and the 2019 Venezuelan presidential crisis (SETB) from Twitter, YouTube, and web pages linked from tweets and video captions. See (Blackburn et al. 2020) for details. In this section we briefly describe our approach and lessons learned.

**Detection Strategies** We applied topic modeling to facilitate the understanding of possible narratives within our data, then detected and defined narratives via manual review of topics and sampled text. Our preliminary results suggest that topic modeling is promising as a method to explore narratives in social media. It was also useful for tuning our data collection to avoid irrelevant texts. For example, in SETA this revealed tweets about *white helmets* that were actually about sports. We also found that Markov clustering (MCL) (Van Dongen 2000) on graphs of tweets and re-tweets from the datasets helped identify communities of users associated with narratives. Figure 1 contains an example for SETB, where user influence was calculated by PageRank.

In both datasets, stance groups were visible and often corresponded to clusters found by MCL. Our results on SETA suggest that careful keyword selection can help detect narratives. However, we found narratives in SETB that could not be captured with keywords. Depending on the availability of manual annotation, either method may be useful.

**Annotation Methods** We chose different methods to annotate the datasets due to data volume and narrative complexity. For SETA, we simply associated narrative annotations with texts based on which keywords they contained. For SETB, we used a supervised annotation method. A team of eight SMEs first annotated roughly 11,000 texts with three stance groups (support, against, neutral) and 49 narratives, follow-

