Measuring The Influencers In The News Media's Narratives

Samuel Oliver Stern
Department of Computer Science
University College London
London, England

David Tuckett
Centre for the Study of Decision-Making Uncertainty
University College London
London, England

Robert Elliott Smith
Department of Computer Science
University College London
London, England

Rickard Nyman
Centre for the Study of Decision-Making Uncertainty
University College London
London, England

Abstract— **We present the results of a preliminary study to test the hypothesis that it is possible to automatically identify opinions, in the form of conviction narratives, as they emerge in text data, and to measure and monitor how actors in the online news media influence others in the media to adopt similar narratives to their own. Narratives are represented in the form of sentiment that online news sources express about various topics. Our results suggest that there is evidence of specific news sources acting as opinion leaders, determining the narratives that others in the online media adopt.**

Keywords—sentiment-topic modelling; opinion leader; intermedia agenda-setting; decision-making; conviction narrative theory

I. Introduction

Research in the fields of psychology, cognitive science and neuroscience has increasingly evinced the functional role that emotion plays in multiple higher-level cognitive processes such as judgment and decision-making [1]–[4]. Conviction narratives have been proposed as the mechanism that "enables actors to draw on their beliefs, causal models, and rules of thumb to identify opportunities worth acting on" [5]. As yet, little research has been undertaken that explores how conviction narratives are constructed and communicated on a societal level, how to identify conviction narratives, and how to recognise when a conviction narrative becomes dominant in society. Studying how conviction narratives can be measured and are influenced at a societal level is important for understanding public opinion formation, as well as for identifying early warning signals for when there are inconsistencies between the information people receive and the way in which they use this information to motivate actions.

This paper tests the hypothesis that it is possible to automatically identify conviction narratives as they emerge in text data and to measure and monitor how actors interact through conviction narratives by identifying those who successfully influence others to adopt similar opinions to their own.

Conviction Narrative Theory

Conviction Narrative Theory (CNT) is a socio-psychological theory of affective cognition that attempts to answer the question: In the face *radical uncertainty*, how is it that humans are nonetheless able to conjure up the conviction to make decisions? It posits that agents make decisions by constructing narratives (i.e., stories) which act as *embodied simulations* of different action outcomes. By then appraising a narrative, an agent is able to anticipate their future emotional state by mentally and physically 'experiencing' a world in which that narrative comes to fruition. The emotional response to the narrative can then give the agent the feeling of confidence required to take action (i.e., make a decision). By framing a decision outcome as a narrative, CNT argues that agents can combine cognitive and emotional elements to interpolate missing pieces of information and simultaneously assess the plausibility of an outcome by determining whether the narrative has a coherent temporal and causal structure.

Using a corpus of historical online news articles from the period May 2016 to December 2016, we approximate an actor's conviction narratives as the sentiment that they express about various news topics and show that there are statistically significant lead-lag relationships in the narratives across actors in the online news media.

II. Related Work

Previous work on CNT has shown that when the object of a conviction narrative is known *a priori*, the narratives, or more specifically, changes in the narratives, can be detected and tracked in text data [6]. It has also been demonstrated that conviction narratives can disseminate through social networks and that the networks (or node clusters within the network) can become 'fixated' on the narratives, i.e., the discourse is fixated

on one topic and there is a lack of diversity of opinions. This in turn can lead to a disconnect between the narrative and the underlying reality, resulting in systemic suboptimal decision-making [6], [7].

Similar work looking at the way in which agents in the media influence one another has been done in the context of agenda-setting theory, specifically intermedia agenda-setting. Vargo and Guo [8] examine online U.S media and show that the agendas set by the media are both homogeneous and reciprocal, and that much of the agenda-setting across media sources is generated from partisan sources. Meraz [9] similarly identifies influence in media networks as well as convergence behavior in the topics being discussed across sources. These findings suggest the existence of shared narratives across actors in the news media. However, both studies consider only the topic discussed in the news, and not the way in which it was being discussed, whereas we are concerned specifically with whether the topic evokes approach or avoidance emotions (i.e., positive or negative sentiment). Furthermore, they rely on a pre-defined set of topics.

Research focusing on intersection between sentiment and social network influence has primarily been looked at microblogging websites such as Twitter and Facebook [10]–[12]. It showed, for example, that the inclusion of semantic information, specifically sentiment and opinions, expressed by actors in social media has been proposed as a method to augment influencer detection beyond what is possible using only network topology [13], [14]. Furthermore, contagion models have been developed to study how emotion and emotive content propagates and diffuses throughout social networks; sentiment in social media–based content has also been shown to correlate with both the speed and quantity of information sharing [10], [11], [15]–[17].

III. METHODOLOGY

A conviction narrative is the subjective and imagined unfolding of an event sequence and the emotion that that event evokes in the mind of the person conjuring the narrative. Text data will never encode the author's entire mental state and will therefore always be an imperfect method for identifying conviction narrative. For the practical proposes of studying the effects of conviction narratives on aggrege levels, we argue that it is possible to identify the impact of conviction narratives with only two features: the object of the narrative, and the emotion that the narrative evokes. In particular, whether it evokes 'approach' (positive) or 'avoidance' (negative) emotions.

With this in mind, we define a proxy measure for conviction narrative from texts as a topic-sentiment pair. The topics of the narrative are assumed to be constant over time and the same across agents, however the sentiment expressed about those topics may change over time and is measured for each agent separately. Thus, the objects (i.e., topics) of the narratives discussed by the various online news sources are assumed to be the same, though the narrative (i.e., the emotion they express about the topic) may be different for each news source.

The topic-specific sentiment for each source is measured daily. This is done by inferring the sentiment of each news article and using an LDA model to assign the article a topic distribution. To determine how a particular news source feels about a particular topic on any given day, the topic probabilities for each of its articles are scaled by their sentiment, and then summed together across articles. Repeating this process for each news source daily results in a $|T|$ x $|A|$ x $|K|$ tensor, where $|K|$ is the number of topics, $|A|$ is the number of news sources, and $|T|$ is the number of days in the time interval. In other words, each news source has $|K|$ time-series that describe how their narratives on each of the topics evolve throughout time. The steps for assigning topic-sentiment timeseries is presented in algorithm 1.

---

algorithm 1: Topic-Sentiment Assignment

input: Set of articles D

output: Tensor G of size $|T|$ x $|A|$ x $|K|$

$T \leftarrow$ dates of articles in D
$K \leftarrow$ set of topics
$A \leftarrow$ set of news sources
for each day t in T **do**
    for each article d in D_t **do**
        $a \leftarrow$ source of document d
        $\varphi_d \leftarrow$ multinomial topic distribution of document d
        $s_d \leftarrow$ sentiment of document d
        $\psi_d \leftarrow s_d \times \varphi_d$
        for $\psi_{k,d}$ in ψ_d **do**
            $G_{t,a,k} \leftarrow G_{t,a,k} + \psi_{k,d}$

The sentiment of each article is determined using a set of 'excitement' and 'anxiety' sentiment keywords constructed by psychoanalysis experts and designed to identify 'approach' and 'avoidance' emotions of conviction narratives [6], [18]. Under this method, the sentiment of an article is found by taking the difference between the number of excitement keywords and the number of anxiety keywords, and then normalizing by the total number of tokens in the article. While keyword approaches to analyzing sentiment are very limited compared to the plethora of more advanced machine learning methods of measuring sentiment, they nonetheless offer several advantages compared to statistical methods. Firstly, machine learning methods are computationally more expensive, particularly when dealing with document-level text corpuses. Secondly, statistical methods generally require a large set of labelled texts which are costly to produce. While there exist labelled sentiment datasets, these are generally of shorter texts such as micro-blogs or online reviews are not guaranteed to generalize to news texts. Lastly, the keyword approach we employ has previously been shown to be sufficient powerful to capture the shift in sentiment of conviction narratives on aggregate level data [6].

IV. EXPERIMENTAL DESIGN AND RESULTS

A. Dataset & Preprocessing

The source of the news articles used in this research is the LexisNexis Newsdesk dataset, generously provided by LexisNexis. The dataset consists of 113,000,000 articles from over 34,000 news sources dating from 6 May 2016 to 31 Dec 2016. In addition to the content of the news articles, the dataset contains further metadata, including the source's country and broad topic labels such as 'Top Stories', 'Sport', and 'Fashion'. To remove irrelevant articles, only articles between 100 and 1700 words in length and tagged under 'Top Stories' are kept. Furthermore, to eliminate sources that only sporadically publish relevant content, all sources with an average publication rate lower than once per day and a coefficient of variation in the bottom 5% are ignored. This leaves 92 news sources.

Standard pre-processing is performed before performing topic modelling and sentiment extraction on the articles. First, the texts are tokenised and stopwords are removed. Next, all words are converted to lower case and are lemmatized.

Article topics are assigned using an LDA model of 200 topics and trained using the Gensim Python package [19]. A qualitative inspection of the topics learned by the LDA model suggests that the model successfully learns to cluster words based on identifiable news topics. Topics include ['europe', 'brexit', 'farage'], ['olympic', 'doping', 'russia'] and ['bank', 'rate', 'bond'].

B. Experimental Results

The first test we run looks at whether there are any actors who lead the narratives of the rest of the online news media. This is done by taking each of the LDA topics and testing whether there are any news sources whose sentiment on a particular topic is a leading indicator for the 'consensus view' being put forth by the other sources on that topic. The consensus view is expressed as the average sentiment across all news source, excluding the source that is currently being tested.

The presence of a lead-lag relationship between a particular source and the rest of the media is tested by performing a Granger causality test [20]. A source's narrative is said to lead the consensus view if, for the topic of that narrative, the source's sentiment time-series Granger-causes the consensus with a p-value of less than 0.01. The lead-lag relationship is tested for each of the news sources and for each news topic.

The results of running the test for each news source indicate that there are several cases where individual actors in the news media Granger-cause the opinions expressed by the rest of the media. An illustration is presented in figure 1, which presents, for a random subset of online news sources, the percentage of topics that the sources have been found to be opinion leaders on with a p-value 0.01 or less.
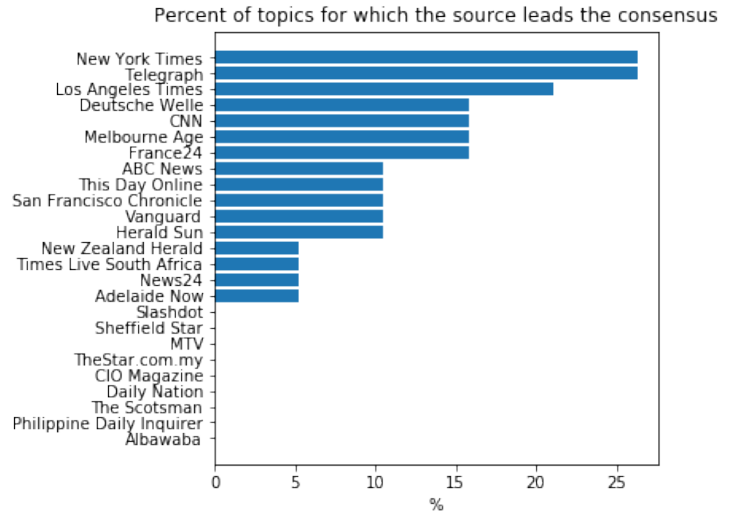

*Figure 1: **The percentage of topics for which a randomly selected subset of news sources lead the sentiment of the consensus***

Having found evidence of influencers in the online news media, we now test whether news sources are directly influenced by one another, or whether the effect of influencers is only at the aggregate level. To do this, we perform a similar test as the one described previously; however, rather than testing whether the narrative of one source Granger-causes the consensus view, we compare the narratives of pairs of sources. In other words, for each topic, we take all pairs of news sources and test whether the sentiment of one source is a leading indicator of the sentiment of the other source (and vice versa).
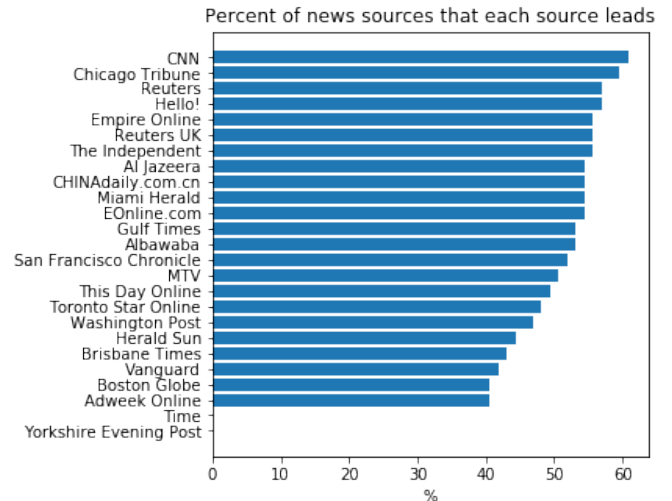

*Figure 2: **The percentage of other news sources the sources have some degree of influence over.***

As in the first test, there are statistically significant lead-lag relationships, indicating that the narratives of some news sources are dictated by certain others. What is surprising

though, is that influence does not appear to be contained to a small minority of sources, but rather that most sources influence the narratives of multiple other sources. The chart in figure 2 shows, for a randomly selected subset of news sources, the percentage of other sources for whom they lead the narrative on at least one topic. While there are some sources that don't appear to influence any of the other sources, the majority of them influence multiple other sources, suggesting a highly interconnected web of influence and influencing between the news sources.

V. CONCLUSION AND FURTHER WORK

This paper presented the results of preliminary work for identifying conviction narratives as they emerge in text data, measuring how news sources influence the narratives of others.

The results present a promising indication that influencers in the online news media can be identified and their narratives can be tracked from news data. Three areas of future research are identified.

1. Refining the methods to identify how an actor's conviction narratives can be identified and monitored over time.
2. Examining whether there are properties which determine whether one can identify the types of actors likely to be influential (i.e., mainstream vs. alternative news sources), and the types of narratives likely to dominate discussions.
3. Testing whether there is a relationship between what narratives dominate online discussions and measurable human behaviour, such as consumer spending habits and election outcomes.

REFERENCES

[1] A. Bechara, H. Damasio, D. Tranel, and A. R. Damasio, "The Iowa Gambling Task and the somatic marker hypothesis: some questions and answers," *Trends Cogn. Sci.*, vol. 9, no. 4, pp. 159–162, Apr. 2005.

[2] T. Brosch, K. R. Scherer, D. M. Grandjean, and D. Sander, "The impact of emotion on perception, attention, memory, and decision-making," *Swiss Med. Wkly.*, vol. 143, p. w13786, 2013.

[3] G. L. Clore and J. E. Palmer, "AFFECTIVE GUIDANCE OF INTELLIGENT AGENTS: How Emotion Controls Cognition," *Cogn. Syst. Res.*, vol. 10, no. 1, pp. 21–30, Mar. 2009.

[4] R. J. Davidson, K. R. Scherer, and H. H. Goldsmith, Eds., *Handbook of affective sciences*. New York: Oxford University Press, 2009.

[5] D. Tuckett and M. Nikolic, "The role of conviction and narrative in decision-making under radical uncertainty," *Theory Psychol.*, vol. 27, no. 4, pp. 501–523, Aug. 2017.

[6] D. Tuckett, R. E. Smith, and R. Nyman, "Tracking phantastic objects: A computer algorithmic investigation of narrative evolution in unstructured data sources," *Soc. Netw.*, vol. 38, pp. 121–133, Jul. 2014.

[7] D. Tuckett and R. Taffler, "Phantastic objects and the financial market's sense of reality: A psychoanalytic contribution to the understanding of stock market instability1," *Int. J. Psychoanal.*, vol. 89, no. 2, pp. 389–412, Jun. 2008.

[8] C. J. Vargo and L. Guo, "Networks, Big Data, and Intermedia Agenda Setting: An Analysis of Traditional, Partisan, and Emerging Online U.S. News," *Journal. Mass Commun. Q.*, vol. 94, no. 4, pp. 1031–1055, Dec. 2017.

[9] S. Meraz, "Using Time Series Analysis to Measure Intermedia Agenda-Setting Influence in Traditional Media and Political Blog Networks," *Journal. Mass Commun. Q.*, vol. 88, no. 1, pp. 176–194, Mar. 2011.

[10] X. Xiong et al., "An emotional contagion model for heterogeneous social media with multiple behaviors," *Phys. Stat. Mech. Its Appl.*, vol. 490, pp. 185–202, Jan. 2018.

[11] S. Stieglitz and L. Dang-Xuan, "Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior," *J. Manag. Inf. Syst.*, vol. 29, no. 4, pp. 217–248, Apr. 2013.

[12] L. Coviello et al., "Detecting Emotional Contagion in Massive Social Networks," *PLOS ONE*, vol. 9, no. 3, p. e90315, Mar. 2014.

[13] X. Song, Y. Chi, K. Hino, and B. Tseng, "Identifying Opinion Leaders in the Blogosphere," in *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, New York, NY, USA, 2007, pp. 971–974.

[14] H. Zhou, D. Zeng, and C. Zhang, "Finding leaders from opinion networks," in *2009 IEEE International Conference on Intelligence and Security Informatics*, 2009, pp. 266–268.

[15] E. Ferrara and Z. Yang, "Quantifying the effect of sentiment on information diffusion in social media," *PeerJ Comput. Sci.*, vol. 1, p. e26, Sep. 2015.

[16] W. J. Brady, J. A. Wills, J. T. Jost, J. A. Tucker, and J. J. Van Bavel, "Emotion shapes the diffusion of moralized content in social networks," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 28, pp. 7313–7318, Jul. 2017.

[17] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis, "Emotions as infectious diseases in a large social network: the SISa model," *Proc. R. Soc. Lond. B Biol. Sci.*, vol. 277, no. 1701, pp. 3827–3835, Dec. 2010.

[18] R. B. E. Nyman, "An Algorithmic Investigation of Conviction Narrative Theory: Applications in Business, Finance and Economics," Doctoral, UCL (University College London), 2016.

[19] R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, Valletta, Malta, 2010, pp. 45–50.

[20] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.