



Incorporating Sentiment Analysis with Epistemic Network Analysis to Enhance Discourse Analysis of Twitter Data

Kamila Misiejuk¹  , Jennifer Scianna² , Rogers Kaliisa³ , Karl Vachuska² ,
and David Williamson Shaffer² 

¹ Centre for the Science of Learning and Technology, University of Bergen, Bergen, Norway
kamila.misiejuk@uib.no

² Department of Educational Psychology, University of Wisconsin, Madison, USA
{jscianna, vachuska, dws}@wisc.edu

³ Department of Education, University of Oslo, Oslo, Norway
rogers.kaliisa@iped.uio.no

Abstract. While there has been much growth in the use of microblogging platforms (e.g., Twitter) to share information on a range of topics, researchers struggle to analyze the large volumes of data produced on such platforms. Established methods such as Sentiment Analysis (SA) have been criticized over their inaccuracy and limited analytical depth. In this exploratory methodological paper, we propose a combination of SA with Epistemic Network Analysis (ENA) as an alternative approach for providing richer qualitative and quantitative insights into Twitter discourse. We illustrate the application and potential use of these approaches by visualizing the differences between tweets directed or discussing Democrats and Republicans after the COVID-19 Stimulus Package announcement in the US. SA was integrated into ENA models in two ways: as a part of the blocking variable and as a set of codes. Our results suggest that incorporating SA into ENA allowed for a better understanding of how groups viewed the components of the stimulus issue by splitting them by sentiment and enabled a meaningful inclusion of data with singular subject focus into the ENA models.

Keywords: Epistemic Network Analysis · Sentiment Analysis · Discourse Analysis

1 Introduction

The emergence of web 2.0 technologies has seen a rise in the use of microblogging platforms such as Twitter and Facebook as means for people to share information, discuss current issues, and express their opinions on almost all aspects of everyday life [1]. Consequently, this shift has captured the interest of researchers, politicians, journalists, and financial and educational organizations to increasingly seek ways to collect and make sense of the vast amount of data produced by users of microblogging platforms to understand and explain different social phenomena. However, analyzing high-volume

data from microblogging platforms such as Twitter is challenging. This is partly because conversations in such environments are characterized by extensive use of informal language, emoticons, acronyms, and message-length constraints (partly due to the imposed character limit to posts), which could pose interpretative challenges [1, 2]. This implies that the analysis of discourse based on data from microblogging platforms requires the creative use of multiple approaches to gain a richer understanding of the discourse. Sentiment Analysis (SA) is a popular method used to analyse discourse by identifying valence in text data. Another method to model discourse is Epistemic Network Analysis (ENA) that analyzes and visualizes connections among pre-defined codes. In this paper, we argue that ENA and SA, in combination, is a useful addition to the methodological toolbox for analyzing Twitter discourse. We accomplish this through a case study that visualizes the differences in the discourse between tweets directed or talking about Democrats and Republicans from the 26th of March to the 1st of April 2020 following the announcement of the Stimulus package in the USA on the 27th of March. This aid package, the largest in US history, was implemented to mitigate the economic consequences of the COVID pandemic with measures, such as one-time \$1,200 direct payments to individuals and business grants to discourage lay-offs [13, 14]. We develop two models combining SA and ENA in different ways and compare them with a model using only ENA.

2 Related Literature

2.1 Sentiment Analysis

There have been many developments in examining and interpreting data produced on microblogging platforms such as Twitter using both qualitative and quantitative approaches. One popular method is *Sentiment Analysis* (SA), also known as opinion mining, that tries to make evident what people think by providing representations, models, and algorithms that extract subjective information to create structured and actionable knowledge [3]. SA determines whether a textual corpus (e.g., document, sentence) tends towards positive, negative, or neutral [1, 4]. One of the significant early efforts for sentiment classification on Twitter data is by Barbosa and Feng [5]. They leveraged sources of noisy labels to train a model and used 1000 manually labeled tweets for tuning and another 1000 manually labeled tweets for testing. Their approach was able to capture more abstract representations of tweets and was more robust regarding biased and noisy data, a common feature of data from microblogging platforms. In another example, Agarwal and colleagues [4] used SA to build models for classifying tweets into positive, negative, and neutral sentiment. They concluded that features that combine prior polarity of words with their parts-of-speech tags are most important for the classification task. Moreover, Kouloumpis, Wilson, and Moore [6] investigated the utility of linguistic features (e.g., informal and creative language) for detecting the sentiment of Twitter messages. Their experiments indicated that part-of-speech features might not be useful for SA in the microblogging domain.

There are, however, some problems with some of the conventional approaches to analyzing Twitter data. For example, even though SA can analyze large volumes of tweets in bulk, questions may arise over its accuracy and the limited depth to the analyzed data

[3]. Further, machine learning-based sentiment classifiers can often prove less efficient in the case of tweets [5, 7], since the latter do not typically consist of representative and syntactically consistent words, due to the imposed character restriction [1]. An additional limitation is that classifiers usually distinguish sentiment into classes (positive, negative, and neutral), assigning a corresponding score to the post as a whole, even though many aspects of the same “notion” may be discussed in a single post [1]. In particular, a key area of exploration includes datasets where there is contention with how people are addressing a given subject. In these cases, one can measure the overall balance in the sentiment of a group of people who mention a single subject, but not what the sentiment is genuinely reflective of. For example, it would be possible to determine that more tweets were negative when mentioning the Supreme Court after a key decision. However, without further exploration, one would not know if the negativity is directed at the decision itself or the case that raised the decision. Therefore, SA alone might fail to provide richer qualitative insights into Twitter discourse, yet these are precisely the types of insights that can be obtained by the use of tools such as ENA which allows for the addition of connections between the subject and details of public discourse elucidating the complexity in the data.

2.2 Epistemic Network Analysis

Epistemic Network Analysis (ENA) is a quantitative ethnographic network analysis technique that analyses logfile data and other records of individual and collaborative learning [8]. ENA consists of a set of techniques that measure connections among coded data elements and represent them in dynamic network models. These models illustrate the structure of connections and measure the strength of association and changes in the composition and strength of connections in a network over time [9]. ENA has also demonstrated flexibility in its ability to combine with other methods. [10] introduced the use of social network analysis as an augmentation of the ENA projection to clarify how social and cognitive factors were influencing collaborative problem-solving. However, while ENA offers powerful mechanisms to analyze collaboration discourse and links among relevant features of collaborative learning [11], it may be challenging to visualize semantic features of different types in the same plot. For example, using the previous example concerning the Supreme Court and subsequent ruling, connections projected by ENA using codes describing topics could be enhanced by an understanding of the sentiment behind them. Thus, in this paper, we propose a novel approach that combines SA and ENA to better understand participants’ discourse as a response to the potential limitations of individual approaches to the analysis of microblogging data. This proposed approach aligns with Kontopoulos and colleagues [1], who noted that exploring various methods for visualizing the resulting sentiment is necessary to provide comprehensive insights to the users. This paper seeks to explore the following research questions:

1. Can SA and ENA be combined, and if so, how?
2. Can adding SA to ENA models provide different insights into Twitter discourse than ENA models alone?

3 Method

3.1 Twitter Dataset

COVID-19 Tweets Dataset is an open-access dataset published on the IEEE DataPort™ website. With the first tweets collected on the 20th of March 2020, this large dataset includes English tweets filtered by several corona-related keywords including keywords “corona”, “coronavirus”, “covid”, “covid19” and variants of “sarscov2”. The model itself monitors Twitter in real-time, and new datasets are published daily. Following Twitter Developer Policy, COVID-19 Tweets Dataset consists of Tweet Ids. To download the “full” tweets, open-source software that handles the Twitter API limits called DocNow’s Hydrator was used [12].

In the current study, we focused on the tweets published from the 26th of March to the 1st of April 2020, resulting in a dataset of 2,461,489 tweets. On the 27th of March 2020, President Trump signed a stimulus package. This aid package, the largest in US history, was implemented to mitigate the economic consequences of the COVID pandemic with measures, such as one-time \$1,200 direct payments to individuals and business grants to discourage lay-offs [13, 14]. The initial investigation identified key political figures being called into the conversation around COVID and the stimulus package. To reduce the dataset from all available Covid tweets to a more manageable, relevant sample, a text search filter was applied to tweet content, replies, and retweets. Tweets that were direct replies to or retweets of a politician’s Twitter handle were considered Direct Mentions of a politician. Tweets that included the politician’s Twitter handle, name, or other identifying information were labeled as Indirect Mentions. These mentions were combined with mentions of keywords related to political parties to create two groups: *Republicans* and *Democrats*. The aggregation of filter criteria can be seen in Table 1.

Since each tweet may have more than one politician or group mentioned, a function was applied that moved a given tweet further towards Democrat or Republican based on the number of keywords within the tweet. Tweets containing mentions of both parties in equal numbers were labeled as *Balanced* tweets. Any tweet that was not relevant to a political leader or party was removed from the dataset. A final filter eliminated duplicates from the dataset as we were primarily concerned with original ideas and hoped to avoid a frequently retweeted tweet skewing the sample.

3.2 Qualitative Data Coding and Validation

In order to code the tweets, we used a bottom-up approach and looked directly at the tweets to discover relevant themes [15]. After multiple iterations of 4 coders coding parts of the data, we decided on the coding scheme shown in Table 2. To validate the coding scheme, we used nCoder, a tool that helps develop automated classifiers based on regular expressions. Each code was validated against 2 raters and the automated classifier using Cohen’s Kappa and Shaffer’s rho [16]. See Table 3 for the validation scores.

Finally, we removed the non-coded tweets from our dataset, which resulted in a dataset of 4,944 tweets that were used in the ENA analysis. Moreover, SA scores were obtained with the help of the *Syuzhet* r package using an AFINN lexicon-based model, which was specifically designed for analyzing microblogs and social media [17, 18].

Table 1. Filters used to identify when tweets were directly related to a subject through response or retweet, indirectly mentioning a subject, and the political content of each tweet.

Politician/Party	Direct mentions	Indirect mentions	Political party addressed
Donald J. Trump	@realDonaldTrump, @POTUS	‘Potus45’, ‘Trump’	<i>Republican</i>
Mike Pence	@VP, @Mike_Pence	‘Pence’, ‘@vp’	
Ron DeSantis	@govRonDesantis	‘desantis’	
Republicans	N/A	‘republican’, ‘republicon’, ‘GOP’, ‘trumptard’, ‘right wing’, ‘conservative’	
Joe Biden	@JoeBiden	‘biden’	<i>Democratic</i>
Bernie Sanders	@BernieSanders	‘Feelthebern’, ‘bernie’, ‘sanders’	
Andrew Cuomo	@AndrewCuomo, @NYGovCuomo	‘cuomo’	
Democrats	N/A	‘dems’, ‘democrap’, ‘democrat’, ‘leftard’, ‘libtard’, ‘liberal’, ‘DNC’, ‘left wing’	

Every tweet was assigned one sentiment score in a range from -5 to $+5$ based on its number of positive, negative, and neutral words. Sentiment scores below zero were coded as *negative*, above zero as *positive*, and equal to zero as *neutral*.

3.3 Quantitative Modeling with Epistemic Network Analysis

Examining the benefits of SA required the construction of three separate types of models, each with different integrations of ENA and SA. Tweets were explored with the Web Tool for Epistemic Network Analysis (webENA). This tool allows users to model connections between codes within their dataset using conversation parameters, window size, and units of analysis. We defined individual tweets as both a single utterance and the entirety of the conversation (and thus window size of one) because we could not identify threads from our data. Furthermore, any attempt to link tweets by date and introduce a moving or infinite stanza would obfuscate the results because individuals came from such far-reaching places. They may have been discussing the same people, and further even possibly centering their discussions around a similar concern, but they were not actively responding to one another, a key component of true conversation. While limiting to the final dataset, it was essential to take each tweet on its own without assuming its connection to the greater body of data.

Table 2. Coding scheme.

Code	Definition	Tweet example
Vulnerable workers	Tweets referring to workers who are disproportionately affected by the COVID situation by an increased risk of infection (e.g., nurses, doctors, essential workers)	<i>@Mike_Pence @WhiteHouse @GM How are the essential workers going to be compensated?... Amazon workers working in warehouse getting Corona Virus, what about workers on front lines?... Me and my wife are in a factory making \$14 a hour putting out essential items. Military gets hazard pay</i>
High risk people	Tweets referring to groups who are disproportionately affected by the COVID situation by either decreased access to healthcare (e.g., refugees, the poor, homeless, transgender people) and increased chance of death if infected (e.g., elderly)	<i>Going by the Italian numbers is a gross miscalculation given that 12% of the death certs show corona as the direct cause lol Elderly population, and 1–3 comorbidities doesn't help</i>
Stimulus action	Tweets referring to the measures against the economic impact of COVID, especially, the stimulus package	<i>By the way, have any of you caught the Corona Virus. I hope you are taking care of yourselves better than the government is. Trump calling for the only reasonable senator's dismissal is as ridiculous as this stimulus package is</i>
Reopening	Tweets referring to the re-opening of the economy and going back to work after the lockdown	<i>@realDonaldTrump It's official!! BREAKING NEWS: OANN and USA government after further intense testing and evaluation just announced that the actual Corona virus death rate is as low as the regular flu (influenza) per Dr Anthony Fauci!! Stop the shutdowns!! Get back to school and work!!</i>
Lockdown	Tweets referring to lockdown measure (e.g., working from home, quarantine, homeschooling)	<i>The missing six weeks: how @POTUS @realDonaldTrump failed the biggest test of his life #coronavirus #CoronavirusUSA #CoronaLockdown #CoronavirusOutbreak #CoronaVillains #Corona #COVID #COVID2019 #Covid_19 #COVIDIOT #TrumpPandemic #TrumpVirus</i>

(continued)

Table 2. (continued)

Code	Definition	Tweet example
China involvement	Tweets discussing China’s involvement in the COVID spread or origin	<i>@realDonaldTrump @POTUS President Trump, don’t even talk to the reporters who keep pushing the racist narrative on the Chinese Corona Virus We’re tired of hearing this bs! ~ Trump2020</i>

Table 3. Validation scores.

Code	Rater 1 vs Classifier	Rater 1 vs Rater 2	Rater 2 vs Classifier
Vulnerable workers	Kappa: 0.97*	Kappa: 1.00*	Kappa: 1.00*
High risk people	Kappa: 1.00*	Kappa: 1.00*	Kappa: 1.00*
Lockdown	Kappa: 0.97**	Kappa: 1.00**	Kappa: 1.00**
Stimulus action	Kappa: 1.00**	Kappa: 1.00*	Kappa: 1.00*
Reopening	Kappa: 1.00*	Kappa: 1.00*	Kappa: 1.00*
China involvement	Kappa: 0.97*	Kappa: 0.97*	Kappa: 1.00*

*rho(0.9) < 0.05, **rho(0.9) < 0.01

Table 4. ENA models.

Model	Groups	Conversation	Codes	Sentiment included
1	Republican Democrat Balanced	Individual tweet	Vulnerable Workers, High Risk People, Stimulus Action, Reopening, Lockdown, China Involvement	N/A
2	Republican: Positive, Neutral, Negative Democrat: Positive, Neutral, Negative Balanced: Positive, Neutral, Negative	Individual tweet	Same as Model 1	In grouping
3	Republican Democrat Balanced	Individual tweet	Positive, Negative, Neutral + codes from Model 1	In codes

In this study, three different ENA models were developed, as seen in Table 4. Model 1 focused on comparing units based on the political party mentioned within the context of the tweet. This model served to compare Models 2 and 3 as it was the simplest delineation between groups. Model 2's comparison was based on groups defined by the sentiment directed at a party; this was identified by joining the tweet's sentiment with the political party from Model 1. We chose to append the two together to allow for nine unique groups, including all of the combinations of Positive, Negative, and Neutral with Democratic, Republican, and Balanced. Model 3 reverted to the same groups of comparison as Model 1 but added Positive, Neutral, and Negative to the codes used in Model 1.

The models were analyzed for their ability to differentiate between groups, increased accuracy, enhanced interpretability to the model. For this study, we were looking for interpretability to be enhanced through new relationships not captured by Model 1.

4 Results

Each of the three models produced originated from the same dataset. Due to the short text nature of tweets, the majority of tweets are coded with only one code (4,704 tweets); 240 tweets are coded with two or more codes. More tweets referenced *Republicans* (3,502) than *Democrats* (1,035), and fewer were *Balanced* (407), referencing both parties equally. *Republicans* dominated because there are many tweets directed at or talking about the US president, Donald J. Trump. AFINN sentiment scores classified 2,199 tweets as *Negative*, 1,239 tweets as *Neutral*, and 1,506 tweets as *Positive*. Table 5 shows the distribution of sentiment scores by party. This dataset allows us to address our two exploratory research questions as there is enough variation between the parties and sentiment to observe within an ENA model.

Table 5. Sentiment scores by political affiliation.

Sentiment	Democrat	Balanced	Republican
Negative	476	177	1,546
Neutral	184	65	990
Positive	375	165	966

4.1 RQ1: Can SA and ENA Be Combined, and if so, How?

In this study, we present two ways of incorporating SA into ENA: 1) as a blocking variable (i.e., a qualifier included as a part of the unit of analysis meant to segment the units into more refined categories), 2) as a set of additional “sentiment” codes. All models yielded statistically significant differences between the tweets referencing Democrats and those referencing Republicans on the X-axis. Table 6 shows differences in variance explained by the models and goodness of fit statistics for each model. Model 2 explains the highest

variance and Model 3 the lowest. Moreover, Model 2 had the highest co-registration correlation for both dimensions, while Model 3 had the lowest, which suggests weak goodness of fit.

Table 6. Sentiment scores by political affiliation.

	MR1	SVD2	Pearson		Spearman	
			X-axis	Y-axis	X-axis	Y-axis
Model 1	18.7%	24.0%	.96	.94	.86	.79
Model 2*	23.9%	24.0%	.94	.94	.94	.78
Model 3	8.6%	14.5%	.73	.69	.73	.50

*Model 2 statistics originate from the visualized comparison between Negative_Republicans and Negative_Democrats. All other Model 2 visualizations met or exceeded these metrics.

4.2 RQ2: Can Adding SA to ENA Models Provide Different Insights into Twitter Discourse Than ENA Models Alone?

To answer this question, we compare the three models using ENA graphs seeking to highlight differences not only in the plots themselves but also in the tweets underlying the connections.

Model 1 is the base ENA model without SA that compares Republicans with Democrats and only includes the codes from the coding scheme to which we are going to refer as *subject codes*. Model 2 integrates SA and political affiliation into one blocking variable. To visualize Model 2, we produced three graphs: 1) comparing Positive_Democrats with Positive_Republicans; 2) comparing Neutral_Democrats with Neutral_Republicans; and 3) comparing Negative_Democrats with Negative_Republicans. Model 3 incorporates SA as codes in addition to subject codes, and like Model 1 compares Democrats with Republicans. In order to improve the readability of the visualizations and highlight differences between the groups, the Balanced group, though a part of all ENA models, was hidden, and the scale for edge weights was set to 4. The models were rotated by the comparison groups - Democrats (represented by blue) and Republicans (represented by red). Means rotation refers to a reduction of dimensions in order to position both means along a common axis while maximizing the variance between the means of the two groups [19].

Model 1

Model 1, our comparison model which lacks sentiment, shows the main connections between the codes in the dataset (see Fig. 1). Tweets addressing or talking about Democrats have the strongest connections between *Stimulus Action* with *High Risk People* or *Lockdown* and, while tweets addressing or talking about Republicans have the strongest connections between *Lockdown* and *Reopening* or *China Involvement*.

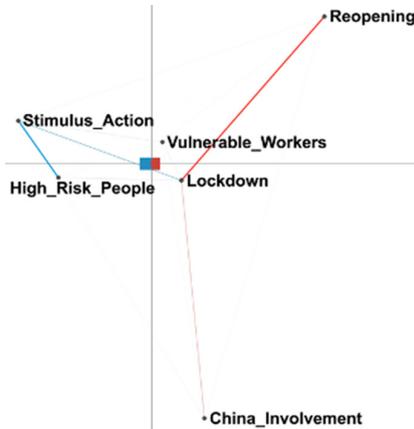


Fig. 1. Model 1 with no SA comparing tweets about or directed at Democrats (in blue) and Republicans (in red). (Color figure online)

Model 2

Model 2 used nine total groups within the model by adding SA as a blocking variable (e.g. Positive_Republican). Model 2 revealed the context in which particular connections were strong and added more nuance to Model 1 (see Fig. 2a–c). The strong connections between *Lockdown* and *Reopening* or *China Involvement* from Model 1 are only visible for positive and neutral tweets directed at or about Republicans, while negative tweets talking about or addressing Republicans have no strong connections among the codes compared with those talking about or addressing Democrats. The relationship between *Lockdown* and *Stimulus Action* seen in Model 1 for tweets directed at or about Democrats is visualized stronger for the polarized positive and negative tweets, but weaker for neutral tweets. Also similar to Model 1, *High Risk People* and *Stimulus Action* are connected strongly for neutral and negative tweets directed at or about Democrats, while only neutral tweets show a strong connection between *Lockdown* with *Vulnerable Workers*.

Adding SA helped highlight the nuances of Twitter discourse not immediately present in Model 1. For example, the sentiment expressed may tell us more about the user’s personal political alignment than the policy or subject. For example, in this tweet with negative sentiment directed toward Democrats, the connection between *Stimulus* and *Lockdown* centers around the misallocation of funding to “undeserving” *High Risk People*:

Why was any of that left in a aid package for the corona virus? This is unacceptable that Congress using their majority in the House should have even been able to add any of these packages. You liberals that support this should start remembering that this is crazy.

It is clear that the user disagrees with the political actions taken and that they do not personally align with liberal philosophies through their choice of, “You liberals”. However, in a positive tweet referring to democratic leader Andrew Cuomo, we see the

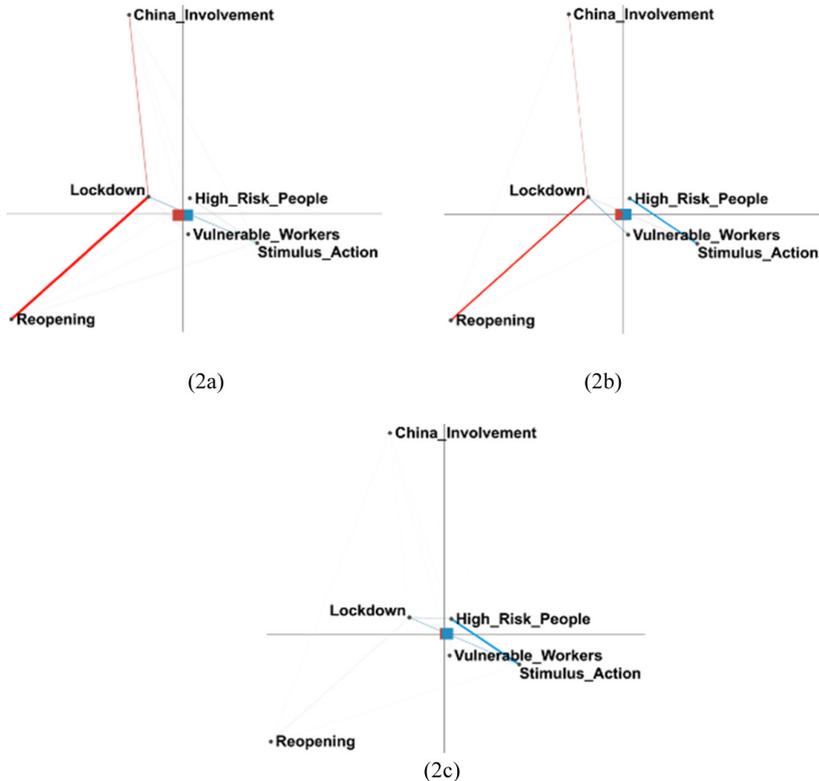


Fig. 2. ENA models comparing tweets about or directed at Democrats (in blue) and Republicans (in red): (a) Model 2a: positive sentiment integrated with the party affiliation, (b) Model 2b: neutral sentiment integrated with the party affiliation, (c) Model 2c: negative sentiment integrated with the party affiliation. (Color figure online)

same questioning of spending, but it is not motivated by political affiliation but rather selective actions.

@RepLeeZeldin I live in NY & it needs help. But this is a CORONA bill. Cuomo was unprepared. Cuomo spends like a crack addict. Cuomo shut the whole state down. Cuomo wants the federal gov't to bail out NY?

Model 3

Model 3 presents an overview of attitudes expressed in the tweets and their connection to the subject codes, however, the connections seen in Model 1's original codes became less visible by adding the SA codes (see Fig. 3). In Model 3, tweets addressing or talking about Democrats have strongest connections between *China Involvement* and *Positive*, and between *Negative* and *Stimulus Action* or *High Risk People*, while the strongest connections for tweets addressing or talking about Republicans are between *China Involvement* and *Negative* or *Neutral*. Interestingly, some topics are dominated

by political affiliation, e.g., *Stimulus Action* is connected to all three sentiments stronger for tweets directed at or about Democrats, whereas *Lockdown* for tweets directed at or about Republicans, while some are divided on the party affiliation, e.g., there is a positive sentiment in tweets that are directed at or about Democrats about *China Involvement*, in comparison to negative or neutral sentiment in tweets directed at or about Republicans.

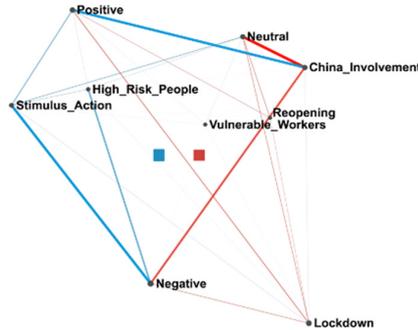


Fig. 3. Model 3 with SA as codes comparing tweets about or directed at Democrats (in blue) and Republicans (in red). (Color figure online)

Using this model, we are able to see how individual subject codes add to the sentiment connections from Model 2. While *Lockdown* is connected to *Stimulus Action* in Model 2a for positive tweets referring to Democrats, the *Lockdown* connection to *Positive* is actually dominated by the Republican group in Model 3. This comes about because there were many additional tweets solely focused on Lockdown that were able to be included in the model by allowing subject code connections to sentiment codes. Tweets express a singular frustration with the lockdown policies,

10 thousand new cases of #Corona in America in a single day... #Trump said no to lock down America!!!

Or are connected to concepts otherwise not represented in the model,

And as of latest figures, usa is now on top in number of cases of Corona virus !! Still Trump is not looking for a full lockdown ! There won't be an economy if there would be no life. #corona #coronaUS #CoronavirusOubreak #CoronaUSA #CoronaVirusUpdates.

Furthermore, the addition of SA as codes added unexpected insights to the model. Nowhere in Models 1 or 2 were tweets referencing Democrats connected to *China Involvement*. In Model 3, there is a clear *Positive - China Involvement* connection that would otherwise go unnoticed.

@thehill One great thing about the CHINESE WUHAN CORONA VIRUS is that it proved once and for all that the Liberal Socialist Fascist Demoncrats and their MSM cronies are definitely NONESSENTIAL! Remember that come November 2020 people!

The addition of sentiment codes gave greater insights into how people were talking about the subject codes. It allowed for tweets with only one subject code to be included in the model by creating a sentiment code - subject code connection. For example, a tweet solely coded as referencing *Vulnerable People* would fall out of Models 1 and 2, but by connecting *Vulnerable People* and *Positive*, the tweet is able to remain in Model 3.

Model Comparison

Social media is a place where widely disparate views can be shared with a broad audience. In the case of Twitter, users' views are broadcast to the world, leaving them open to critique, support, and overall discussion. In the case of tweets that addressed the economic stimulus package and COVID, connecting the sentiment behind a user's words with the content of their messages was essential to better understanding their intention.

When SA was included alongside the party mentioned in the group described as a blocking variable (e.g. *Neutral_Democrat*, *Positive_Republican*), it allowed for a greater amount of variance to be visualized in the ENA plots. It also allowed for larger groups to be parsed apart and visualized at once. For example, one could plot both positive and negative tweets directed at republicans and democrats parsed into four groups. This allowed for the identification of commonalities between groups at a macro level. This strategy, sentiment as a blocking variable, was less useful for determining the details of Tweet content through the plot alone, adding SA into the model as sentiment codes allowed for the model to incorporate narrowly focused tweets.

In summary, adding SA as a blocking variable produced a model with most variance explained and the highest co-registration correlation on both axes. The model using sentiment as codes produced a weaker model. However, both models provided more in-depth insights into the rich landscape of Twitter discourse that would be harder to highlight through a model limited to subject codes.

5 Discussion

In this study, we explore two ways that SA may contribute to ENA models using a case-study that included politically charged Twitter data related to Covid-19. We chose to use an external SA tool to determine the sentiment scores for individual tweets, and in doing so, we have demonstrated how SA can be a fast way to obtain information about how discourse is incorporating different subjects and ideas. In other applications, it may be possible to take a grounded approach to develop sentiment codes in a similar way in which we developed codes for subjects within the tweets. In our analysis, there were definite advantages to incorporating SA in that it (1) allowed groups to be better understood by separating the sentiment directed at different groups and (2) allowed data with a singular subject focus to be meaningfully included in the model. There are different utilities for each application, depending on the nature of the dataset. Moreover, this study highlights different ways of including new data into a network, either as a metadata that can help with data segmentation, or as a set of codes that aids exploring different narratives emerging from the data.

This study's limitations are primarily centralized around the case-study itself and secondarily around the nature of SA. The dataset used included several filters that removed

tweets that did not include COVID, political mentions, and the subject codes themselves. By working with such a reduced set, it is possible that the importance of sentiment codes was artificially constructed. The same technique may not prove as fruitful in an analysis that lacks such clear “lines in the sand.” It will consider those moving forward with integration to test the benefits of SA inclusion in their model. Furthermore, SA, in its purest form, is an automatic coding algorithm. The algorithm we utilized provides a sentiment score, but it is just a number. It can be challenging to determine where to distinguish the numeric score as to what is *Positive* and what is *Negative*. Alongside this observation of the abstraction of sentiment, there is a nuance in natural language processing that can become unwieldy in more casual forums such as Twitter. If SA is used instead of more traditional grounded coding approaches without validation as we have done, it is essential to reexamine the impact that the sentiment is having on one’s model and seek to understand how the SA algorithm is manifested in the data.

There are opportunities for the greater community of Quantitative Ethnography (QE) in this challenge to validate both the use of natural language algorithms and the algorithms themselves. Especially in the context of social media data, the amount of available data is ever-growing, allowing researchers to “see” more perspectives and include more voices in their inquiries. The acceptance of more tools that allow us to process data and provide insights into our data rapidly will challenge us to forge new collaborations across fields, integrate more fields into the work of QE, and in turn, continuously develop new methods for the advancement of the field.

References

1. Kontopoulos, E., Berberidis, C., Dergiades, T., Bassiliades, N.: Ontology-based sentiment analysis of twitter posts. *Expert Syst. Appl.* **40**(10), 4065–4074 (2013)
2. Kumar, A., Sebastian, T.M.: Sentiment analysis on twitter. *IJCSI* **9**(4), 372–378 (2012)
3. Pozzi, F.A., Fersini, E., Messina, E., Liu, B. (eds.): *Sentiment Analysis in Social Networks*. Morgan Kaufmann, Burlington (2016)
4. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.J.: Sentiment analysis of twitter data. In: *LSM 2011 Proceedings*, pp. 30–38 (2011)
5. Barbosa, L., Feng, J.: Robust sentiment detection on twitter from biased and noisy data. In: *COLING 2010 Proceedings: Poster Volume*, pp. 36–44 (2010)
6. Kouloumpis, E., Wilson, T., Moore, J.: Twitter sentiment analysis: the good the bad and the omg! In: *ICWSM 2011 Proceedings*, pp. 538–541 (2011)
7. He, Y., Lin, C., Alani, H.: Automatically extracting polarity-bearing topics for cross-domain sentiment classification. In: *HLT 2011 Proceedings* (2011)
8. Shaffer, D.W., Collier, W., Ruis, A.R.: A tutorial on epistemic network analysis: analyzing the structure of connections in cognitive, social, and interaction data. *J. Learn. Anal.* **3**(3), 9–45 (2016)
9. Shaffer, D.W., et al.: Epistemic network analysis: a prototype for 21st-century assessment of learning. *Int. J. Learn. Media.* **1**(2), 33–53 (2009)
10. Swiecki, Z., Shaffer, D.W.: iSENS: an integrated approach to combining epistemic and social network analyses. In: *LAK 2010 Proceedings*, pp. 305–313 (2020)
11. Gasevic, D., Joksimovic, S., Eagan, B.R., Shaffer, D.W.: SENS: network analytics to combine social and cognitive perspectives of collaborative learning. *Comput. Hum. Behav.* **92**, 562–577 (2019)

12. Documenting the Now: Hydrator [software]. <https://github.com/docnow/hydrator>. Accessed 27 Apr 2020
13. Kretchmer, H.: Key milestones in the spread of the coronavirus pandemic. In: World Economic Forum. <https://www.weforum.org/agenda/2020/04/coronavirus-spread-covid19-pandemic-timeline-milestones/>. Accessed 01 June 2020
14. Pramuk, J.: Trump signs \$2 trillion coronavirus relief bill as the US tries to prevent economic devastation. In: CNBC. <https://www.cnbc.com/2020/03/27/house-passes-2-trillion-coronavirus-stimulus-bill-sends-it-to-trump.html>. Accessed 01 June 2020
15. Urquhart, C.: Getting started with coding. In: Urquhart, C. (ed.) *Grounded Theory for Qualitative Research: A Practical Guide*, pp. 35–54. SAGE, London (2013)
16. Shaffer, D.W., et al.: The nCoder: a technique for improving the utility of inter-rater reliability statistics. *Epistemic Games Group Working Paper 2015-01* (2015)
17. Nielsen, F.A.: A new ANEW: evaluation of a word list for sentiment analysis in microblogs. [arXiv:1103.2903](https://arxiv.org/abs/1103.2903) (2011)
18. Jockers, M.L.: Syuzhet: Extract Sentiment and Plot Arcs from Text. <https://github.com/mjockers/syuzhet>. Accessed 05 June 2020