Wandering Words: Tracing Changes in Words Used by Teacher Tweeters Over Time

Stephen Houser
Bowdoin College
Academic Technology & Consulting
Brunswick, ME 04011, USA
houser@bowdoin.edu

Clare Bates Congdon
Bowdoin College
Department of Computer Science
Brunswick, ME 04011, USA
congdon@bowdoin.edu

Doris Santoro
Bowdoin College
Department of Education
Brunswick, ME 04011, USA
dsantoro@bowdoin.edu

Jessica Hochman
Pratt Institute
School of Information
New York, NY 10011, USA
jhochman@pratt.edu

ABSTRACT

Public school teachers in the United States are often constrained in terms of their ability to express their moral views on issues that affect their schools, classrooms, students, and teaching practices, but are able to express their ideas, concerns, and frustrations as private citizens using social media. Previously we developed the Tweet Capture and Clustering System (TCCS) in order to explore how teachers use Twitter, looking at word usage among a group of teacher tweeters, and attempting to find clusters of teachers who have similar patterns of word usage in their tweets. In the work reported here, we look at teacher tweeters across the 12 months of 2016, seeking to understand how the clusters and the words used in these clusters vary from month to month. In this initial look at the dynamics of the system, we see some evidence of word usage changing across the 12-month period. This initial work suggests that extending TCCS to have temporal topic tracing as a core capability will be a meaningful addition to of the system.

CCS CONCEPTS

• **Information systems** → **Data extraction and integration**; • **Computing methodologies** → *Cluster analysis*;

Reference format:
Stephen Houser, Clare Bates Congdon, Doris Santoro, and Jessica Hochman. 2017. Wandering Words: Tracing Changes in Words Used by Teacher Tweet-ers Over Time. In *Proceedings of Workshop on Social Media World Sensors, Prague, Czech Republic, July 2017 (SIDEWAYS 2017),* 6 pages, CEUR-WS.org.

SIDEWAYS 2017, Prague, Czech Republic

1 INTRODUCTION

The #TeacherTweets project [8] is a transdisciplinary study that combines philosophy of education, computer science, and information science to examine teachers' use of Twitter. We are aware that teachers' ability to express moral concerns or outrage is constrained in their public role (e.g., [7, 11, 12]), and seek to use computing techniqies and systems to explore the extent to which they are turning to social media as a means of expressing these moral claims and concerns.

In previous work [4], we presented the Tweet Capture and Clustering System (TCCS) and our work to build a system to understand basic patterns of communication among teacher tweeters; the initial study indicated five distinct clusters of teachers within our sample, where each cluster is defined by a "linguistic signature" of word-usage patterns across tweeters within the cluster.

In this paper, we explore the utility of an extension to TCCS that would allow us to see how these clusters and signatures change over time. This exploration traces teacher tweeters across the 2016 calendar year, and seeks to understand the extent that word usage within each cluster changes from month to month. We would also like to explore how word usage during each month differs from the overall years usage. Finally, we would like to prepare for the exploration of how clusters change or are static from month to month.

In the remainder of this paper, Section 2 explains relevant background, including an overview of the goals of the larger study of educational philosophy on Twitter, relevant computational approaches, and our previous investigation. Section 3 describes the system design of TCCS. Section 4 explains our research methodology. Section 5 presents the results; Section 6 discusses the implications of the results, and Section 7 describes future work.

2 BACKGROUND

In this section, we briefly provide background on teachers' use of Twitter, relevant machine learning concepts, and our previous results investigating teacher tweeters. We describe the educational philosophy questions the #TeacherTweets project is asking and the system we have constructed to answer them. We also provide some background on the AutoClass classification system we use.

2.1 Twitter as an outlet for moral claims

In many cases, U.S. teachers are constrained in their expression of concerns about their work in classrooms, the well-being of their students, and education policy. For example, the New York Times recently describes a New York City public school's explicit request that teachers not critique state tests [12]. While teachers were offered the option to make comments as private citizens, rather than as educators, their voices were effectively silenced within their own professional context. This example demonstrates why teachers might take to social media, such as Twitter, to voice their moral concerns as a form of civic participation and civic engagement.

Twitter is a popular social media platform that provides methods to engage, develop communities, and engage in professional learning [1]. The 140-character limit for public tweets seems, at first, to limit expression but has proven to be suitable for encapsulating concise original thoughts. The allowance for links and images allows for a deeper connection to professional literature, articles, and additional content. The popular and public nature allows the engagement of a large audience with quick and timely dissemination of ideas.

Twitter provides several mechanisims to create and maintain dynamic and persistent communities, teacher educators in our research, including "hashtags", following, and follow-lists. Topical communities connect through the use of "hashtags" that link otherwise independent tweets and ideas together creating a threaded conversation. Twitter also allows for "retweeting", repeating another account's tweet with attribution, which propagates the original tweeter's thought through the re-tweeter's network of followers. Twitter users can also "follow" other accounts and develop lists of accounts they follow for others to use in finding "thought-leaders" and to build information sharing networks. These affordances break down traditional geographical, social, and political boundaries. They allow teachers to engage other teachers and others outside their profession about their profession [10].

2.2 Clustering and the AutoClass system

The clustering approach used here is AutoClass, developed by NASA [2]. AutoClass is a Bayesian approach, meaning that it considers the distribution of values across each attribute in the full data set, and takes the distribution of these values (the "prior probabilities") across the full data set into account when describing the clusters. As an extreme example, if all of the data points have the same value for an attribute, this attribute is recognized as not being useful in describing the clusters; this concept can be applied probabilistically, so that the utility of a particular attribute in describing a particular cluster can be scaled based on the prior probabilities.

The AutoClass algorithm starts by dividing the data at random into the suggested number of clusters. It then repeatedly creates a probabilistic description of each cluster, removes the data from each cluster, and places each data point in the cluster that best describes that data. When the data no longer moves and the cluster descriptions are fixed (or a pre-specified maximum number of iterations has elapsed), the search is completed; AutoClass refers to each completed search as a "try". Since random numbers are used in the initial cluster definitions, the search is necessarily repeated for a specified number of tries. AutoClass then reports the best clustering found across the set of tries, meaning again, the clusters that have the greatest similarity within each cluster and the greatest differences across clusters.[1]

The output from AutoClass includes a set of cluster descriptions that include the expected means and variance for each attribute as well as the importance or weight of that attribute in defining the cluster. The output from AutoClass also includes a listing for each data point specifying the probability of that data point being in each of one or more clusters.

2.3 Our previous work with TCCS

In an earlier investigation [4], we looked at public twitter accounts from self-described teachers from July 16, 2015 to June 14, 2016. Using a list of educator-relevant words and hashtags and with the words pre-identified as "general" or "moral", we used TCCS to scrape Twitter. TCCS then cleaned and tokenized the collected data and computed how many times each user used each token of interest, generating a "linguistic signature" for each account. TCCS next created clusters of these teacher tweeters, based on their word usage patterns over the research period. A core question in the educational research here was: To what extent are teachers tweeting about moral things?

In that work, we found five distinct clusters of teacher tweeters. Interestingly, the five clusters were quite distinct, without fuzzy boundaries. That is, each teacher tweeter belonged to a cluster with 100% probability. Exploration of the intraclass relationships, by careful examination of word usage and a human sampling of the tweets in each cluster, allowed us to assign an understandable definition or label to each cluster with three of them clearly making moral statements [9]. The five clusters were identified as:

(1) care, with the classroom as the locus of control
(2) intersectional and institutional justice
(3) civic and democratic justice
(4) general communicative action
(5) disengaged or inactive tweeters.

These results indicated that educators are using Twitter to talk about their work and what it means to them. They are making moral claims about care in the classroom and justice (in at least two distinct ways). And lastly, they appear to form distinctive moral communities [4, 9]. These results also prompted a number of new questions, and extensions to TCCS. Some of which we begin to explore herein.

3 AN OVERVIEW OF TCCS

TCCS, as shown in Figure 1, consists of four major components; capture, extraction and translation, clustering, and post processing. Each of these components is distinct from the others in the sense that there are clear inputs and outputs separating it from the modules up and downstream and the module's purpose.

The capture module is responsible for monitoring Twitter and capturing tweets from selected accounts. It runs as two processes; one to collect tweets using Twitter's streaming interface [13], and

[1] Although the AutoClass output and documentation refers to these groupings as "classes"; AutoClass is not a classification approach. In the current terminology of machine learning, the groups would be considered "clusters", and AutoClass is a clustering approach.

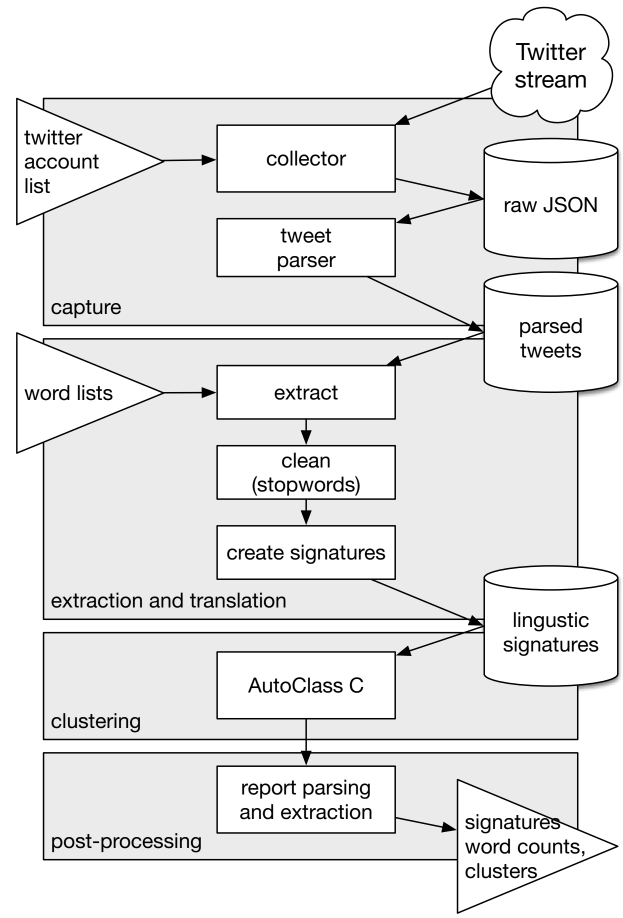

Figure 1: The TCCS architecture includes four major modules: Capture, Extraction and Translation, Clustering, and Post Processing. Inputs to the system include twitter accounts to follow and words of interest; output includes linguistic signatures for each account and the probabilistic description of each cluster.

one to parse and normalize the raw data for later use (extraction, clustering, etc.). It uses a MySQL database for both intermediate and long-term storage.

The extraction and translation module is responsible for extracting the parsed tweets collected by the capture module and creating a "linguistic signature" [5], of each account for the clustering module. This signature is a summarized description of the words used by the account (based on a set of regular expressions) relevant to our research questions. The signature thus consists of the number of times each expression was used for each captured account. Table 1 shows a small sample of account signatures.

The clustering module is AutoClass, described in Section 2. Specifically, the AutoClass C implementation [3] is used. The set of clusters that AutoClass identifies and the detailed statistics about each cluster are the two key outputs from the clustering module.

The post-processing module parses the generated reports and generates comma-separated value (CSV) files that are common, flexible, and importable into a variety of analysis software. During this processing, the post-processing module also separates the clusters, identifies which accounts are in each one, and the word usage that influences membership in each cluster. It also generates additional output files with the most common words across all the original tweets as well as the overall word influence on the clustering.

Table 1: Sample linguistic signatures for accounts. Word counts are collected for all Twitter accounts and across all signature words.

account	union	teach	test	students	…
MrTeachPhilly	1	23	1	1	…
JoyKirr	1	507	50	447	…
aceemae	5	31	30	0	…
WeLearnWeTeach	5	9	1	4	…
AJMGrandma	20	79	55	12	…
…	…	…	…	…	…

4 METHODOLOGY

The goal of the larger #TeacherTweets project is to explore how teachers use Twitter to communicate and the degree to which they are using Twitter to express moral concepts. The goal of the investigation in this paper is to explore how clusters and the signatures of word usage that define those clusters change from month to month.

Some of the questions we would like to address in this facet of our research include:

- How do the linguistic signatures change over time, across all the teacher tweeters and within each cluster?
- If we fix the clusters, how do the words used in each cluster change over time?
- How can we enable exploration of cluster changes over time?

In our initial investigation into these questions, our path is:

(1) Look at the 12-month period across calendar 2016
(2) Run TCCS on the year of data to define clusters and on each individual month
(3) Using the words most important to define and distinguish the clusters, we will investigate how the use of these words vary from month to month within each cluster.
(4) Using the 10 most frequently used words, we will investigate how the use of these words vary from month to month within each cluster.

Through these initial explorations, we hope to gain some insight into how word usage in the clusters changes over time.

5 RESULTS

Over the entire collection period (calendar year 2016) there were 956,118 tweets among the 494 accounts with a total of 12,437,458 individual tokens (words). The linguistic signature consisted of 123 root words (regular expressions) in the three cagegories of "moral words", "general words", and "hashtags" relevant to education and teaching philosophy. Table 2 shows the number of tweets, accounts, and tokens collected for each month of the collection. The number of accounts listed for each month varies as some accounts do not tweet year-round.

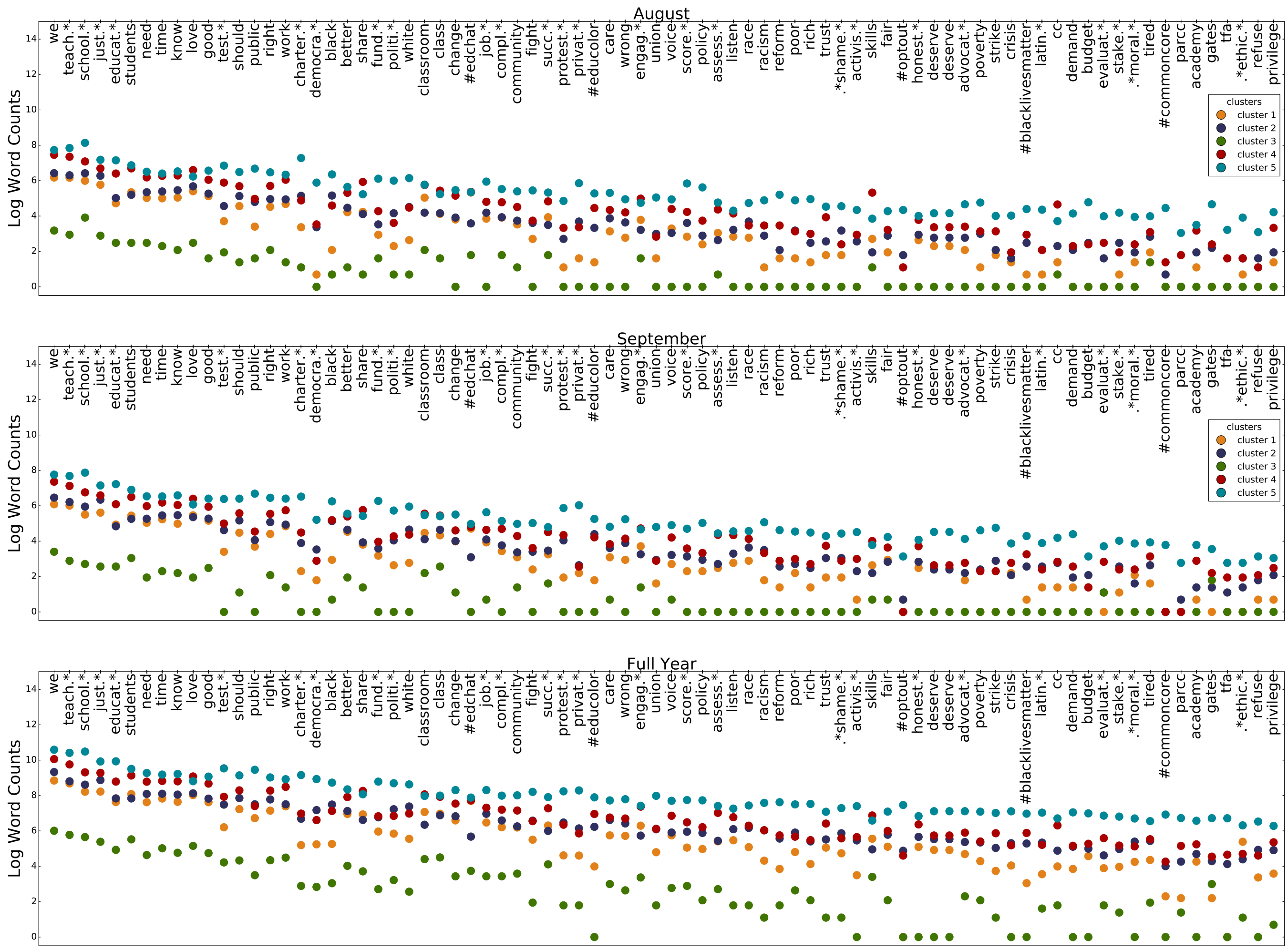

Figure 2: Example months of August and September, and the totals for the year, sorted by frequency of word use. The 80 most-used words from among the 123 predefined words of interest are shown.

Table 2: Collected Tweets, Accounts, and Tokens

Month	tweets	accounts	tokens
Year	956,118	494	12,437,458
January	100,198	427	1,277,100
February	93,245	425	1,191,735
March	77,002	425	983,499
April	95,783	425	1,223,809
May	73,527	415	939,976
June	68,494	417	886,495
July	102,050	428	1,343,956
August	71,258	418	919,116
September	63,059	401	815,137
October	78,916	415	1,055,671
November	71,401	409	975,412
December	61,185	397	825,552

We ran TCCS 13 times; once for the full year and once for each individual month, finding five distinct clusters of accounts for each time period. We have a great deal of data to explore from these runs, and report some initial findings here.

5.1 Fixing the clusters and looking at how word usage changes over time

First, we fixed the cluster membership of each individual account to those determined by the full-year run, to investigate how word usage changes over time among each group of users. The five clusters, based on full-year word usage, consisted of 158, 109, 109, 65, and 53 accounts.

Figure 2 shows the usage of the 80 most-used words from among the 123 words we looked at, sorted according to how often the words were used in the year. While we produced a separate plot for each month, two representative months, August and September, are shown, along with the plot for the full year. Each of the five clusters is represented with a different color. In looking at these plots, we get a sense of how the signatures change from month to month. For example, "skills" shows a spike in the red cluster in August, and "protest" and "private.*" show a spike in the blue cluster in September.

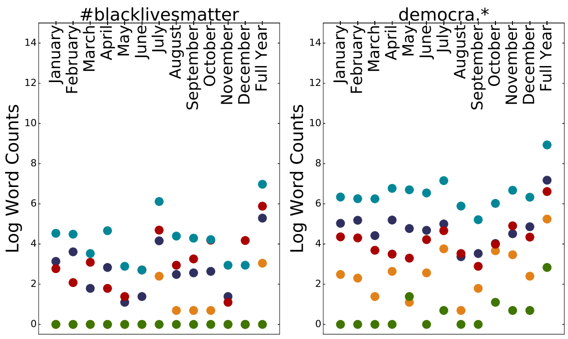

Figure 3: Two words that are used different amounts month to month are "#blacklivesmatter", and the root "democra-". Each cluster's usage is shown in a different color.

While this form of plotting helps us to understand changes in the cluster signatures over time, it is difficult to see how use of a particular word's use changes over time. Thus, we have also created plots to trace each individual word over time. Figure 3 shows two words that varied quite a bit over the 12-month period, the hashtag "#blacklivesmatter", and the root "democra-".

Figure 4 shows each of the five words most used across the year, and how their usage changes from month to month for each cluster. The five words most influential in distinguishing the clusters turn out to be not heavily used, so the illustrations of these are not interesting and have been omitted due to space concerns.

5.2 Comparing monthly runs to the yearly run

To compare monthly data to year data we too two approaches; comparing influential words in creating cluster and examining the Jaccard similarity coefficient of each month's most frequently used words to the year's most frequently used words.

To explore how the word use influenced cluster creation each month and over the entire year we examined the words identified by AutoClass as influential in determining cluster membership. For each month of the collection period, the clustering module created a new arrangement of accounts (clusters) and word frequency data about how those clusters were determined.

The influential words were determined by AutoClass in its repeated trial process. They are the words of which their use, or lack of use, were highly influential in determining which cluster an account was placed. While these words were important to understand how a cluster is created, they were not useful in examining how word use or the clusters themselves change over time.

We also utlized the Jaccard similarity coefficient [6], or intersection over union, of each set of frequently used words compared to the overall (year) frequently used words. This coefficient gives a measure of how similar two sets are to each other. A coefficient of 1.0 indicates the sets are identical, and a coefficient of 0.0 indicates they have no elements in common.

Table 3 shows the most frequently used words with the Jaccard similarity coefficient compared to the most used words across the calendar year. These data proved much more useful in understanding the differences among the months and compared to the year data.

Table 3: Frequently used Words and Similarity Coefficient. Moral words are bold and general education words are in normal font. Month-to-year similarity coefficient appears below month name.

Year	**we**, teach.*, school.*, **just.***, educat.*, students, **need**, time, **know**, **love**
January 1.0	**we**, teach.*, school.*, **just.***, educat.*, students, **need**, **love**, **know**, time
February 0.8182	**we**, teach.*, school.*, **just.***, educat.*, students, **need**, **love**, time, test.*
March 0.8182	**we**, teach.*, school.*, **just.***, students, educat.*, test.*, time, **need**, **love**
April 0.8182	**we**, teach.*, school.*, test.*, educat.*, **just.***, students, **need**, time, **know**
May 0.8182	teach.*, **we**, school.*, **just.***, students, educat.*, test.*, **need**, **love**, time
June 1.0	**we**, teach.*, school.*, **just.***, educat.*, **love**, **need**, students, time, **know**
July 1.0	**we**, teach.*, **just.***, school.*, educat.*, **need**, **know**, **love**, time, students
August 0.8182	school.*, teach.*, **we**, **just.***, students, educat.*, charter, **love**, **know**, **need**
September 1.0	**we**, teach.*, school.*, **just.***, students, educat.*, time, **know**, **love**, **need**
October 0.8182	**we**, teach.*, school.*, **just.***, educat.*, students, time, **know**, **need**, **good**
November 0.8182	**we**, teach.*, school.*, **just.***, educat.*, students, time, **need**, **know**, white
December 0.6667	**we**, school.*, teach.*, **just.***, educat.*, public, students, **know**, time, **good**

6 DISCUSSION

To understand how words change over time, we refer to Table 3 and Figure 2. These data are useful in understanding how words change from month to month and from the aggregate year of collection.

May and August stand out as the only months the word "we" is not the most-used word. In May it is second, and in August, third. "we", "teach.*", and "school.*" are always the top three frequently used words in every month, which is expected as the accounts captured are educators and partially validates our account selections. The word "we" was considered a moral word, indicating belonging, and confirms the premise that teachers are using Twitter to make moral statements. The appearance of other moral words ("just.*", "love", "need", "know") further validate this conclusion.

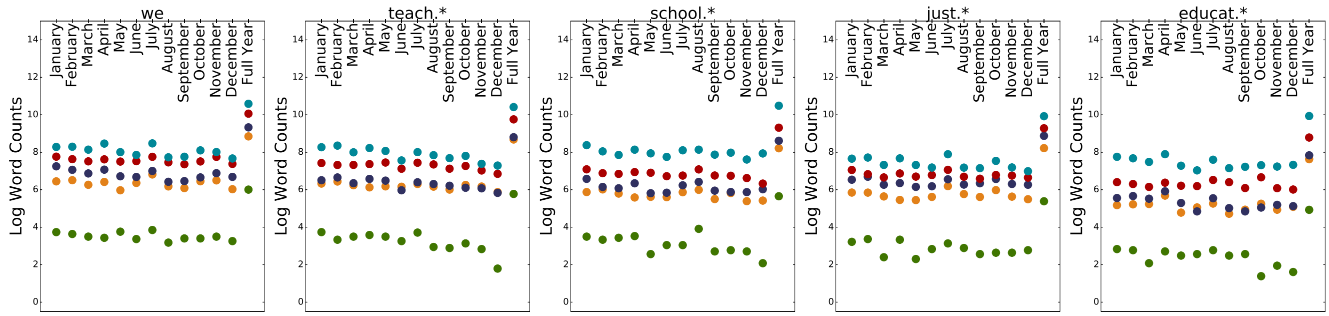

Figure 4: The top five words most used over the year and how their use varies from month to month.

Figure 3 shows two specific words that varied quite a bit over the 12-month period, the hashtag "#blacklivesmatter", and the root "de-mocra.*". The surge in July in some clusters may reflect reported police killings of African Americans that month and subsequent increases in protests (not a topic for all the teacher tweeters); the surge of "democra.*" towards the end of the year might be a reflection of the U.S. Presidential election.

To understand how the clusters change from month to month, we looked at the influential words used by AutoClass to determine monthly and yearly cluster membership. There was little correlation between the words that influence the creation of the year clusters and the monthly clusters as indicated by the set similarity coefficient and by visual inspection. While these clusterings and influential words were useful for creating clusters and seeing how the monthly clustering changes, they were not useful in understanding how word usage changed across time. That is, they were useful in creating clusters but not in understanding them.

7 FUTURE WORK

There are several areas where TCCS and this research could be continued. Herein we used pre-selected words of interest, we focused on teachers on Twitter, and we have not traced topics through clusters over time.

Our research used a set of pre-selected terms related to education and statements of morality by educators. TCCS supports alternate word lists and the use of the most frequently used words for building the linguistic signatures. This approach could provide more insight into how the words and language change over time and potentially predict emerging clusters of accounts.

There is no requirement Twitter with TCCS or educator accounts. TCCS supports any form of text input from similar social media or other platforms. While Twitter is heavily used by teachers, we may find in the future that these conversations move to other platforms. TCCS is easily adaptable to accommodate these new technologies.

We fixed the clustering module (AutoClass) to five (5) clusters and 50,000 trials to determine the clustering. These were selected based on empirical study and experience. AutoClass allows for determining the "best" number of clusters automatically. Given the resource requirements of 50,000 trials and an unlimited number of cluster possibilities, we did not attempt this configuration and leave it as a future research question.

There are other methods of looking at word usage and clusters such as Dynamic Topic Modeling and Latent Dirichlet Allocation (LDA) that we did not use in this work. These techniques could be fruitful in validating the manual topic assignment by visual inspection or be useful in other ways. Given our larger research question was about moral statements by educators, we leave these approaches to possible future work.

Lastly, in this research we made no attempt to map the themes of clusters through time. That is we did not classify the language used within a cluster during one month and attempt to find the same thematic language in a cluster in the next month. The data collected and processed via TCCS is clearly useful for this thematic mapping and tracking, we pose here that this is possible, but not yet done. This extension of TCCS, full temporal topic tracing, would add meaningful capabilities for the broader research questions about how the conversation changes over time.

REFERENCES

[1] danah boyd. 2007. Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life. In *Youth, Identity, and Digital Media*, David Buckingham (Ed.). MIT Press, Cambridge, MA.
[2] P. Cheeseman, J. Kelly, M. Self, J. Stutz, W. Taylor, and D. Freeman. 1988. AutoClass: A Bayesian Classification System. In *Proceedings of the Fifth International Conference on Machine Learning*. Morgan Kaufmann Publishers, Ann Arbor, MI, 54–64.
[3] Dr. Diane Cook, Joseph Potts, and William Taylor. 2017. AutoClass C. (2017). https://ti.arc.nasa.gov/tech/rse/synthesis-projects-applications/autoclass/autoclass-c/
[4] Stephen Houser, Clare Bates Congdon, Doris Santoro, and Jessica Hochman. 2017. Using Clustering to Explore Teacher Communication on Twitter. In *Proceedings of the 3rd International Workshop on Social Media Mining and Analysis (SMMA 2017)*. IEEE Press, Exeter, UK.
[5] Matthew L. Jockers. 2013. *Macroanalysis*. University of Illinois Press, Urbana, Illinois.
[6] Gerard Salton and Michael J. McGill. 1983. *Introductino to Modern Information Retrieval*. McGraw-Hill.
[7] Doris Santoro. 2016. 'We're Not Going to Do That Because It's Not Right': Using Pedagogical Responsibility to Reframe the Doublespeak of Fidelity. *Educational Theory* 66, 1–2 (April 2016), 263–277.
[8] Doris Santoro, Jessica Hochman, Stephen Houser, and Clare Bates Congdon. 2017. #TeacherTweets. (2017). http://www.teachertweets.org/
[9] Doris A. Santoro, Jessica L. Hochman, Stephen Houser, Clare Bates Congdon, and Bridget Gavlin. 2016. Teachers on Twitter: Distinctive moral voices in the civic sphere. In *Association for Moral Education Annual Meeting*. Cambridge, MA.
[10] Doris A. Santoro, Jessica L. Hochman, Stephen Houser, Clare Bates Congdon, and Bridget Gavlin. 2016. Teachers' Tweeting as the Practice of Democracy. In *American Educational Research Association Annual Meeting*. Washington, DC.
[11] Rebecca Smith. 2013. *The Moral Oppression of the Teaching Profession: Learning to Transcend*. Master's thesis. University of Central Florida.
[12] Kate Taylor. 2016. Teachers are warned about criticizing New York State tests. *The New York Times* (March 25 2016). http://nyti.ms/22Hp12e.
[13] Twitter, Inc. 2017. Twitter Developer Documentation. (2017). https://dev.twitter.com/docs