# IMRF GRANT REPORT

Iowa Measurement and Research Foundation Final Report
Submitted by:
Cassie L. Barnhardt, Bryce J. Dietrich, Jielu Yao, Milad Mohebali, and Associates˜

## Title: Campus-based Mobilization: Measuring Characteristics, Processes, and Outcomes

### Overview

Since 2012 the United States has experienced a wave of contemporary protests and campus organizing on the scale that has not been seen in decades. Even with the poignancy of past campus-based activism, the last fifty years has not generated a method for taking stock of campus mobilization and activism – until now. Tracking campus mobilization is difficult on account of its spontaneity and variety of forms. However, taking stock of activism is important because campus mobilization shapes campus life, students' experiences, and reflects broader sociopolitical movements that contribute to social change (Rhoads, 2016; Rojas, 2007). With the support of an Iowa Measurement Research Foundation (IMRF) grant, we have developed a means to measure and track campus mobilization. Our approach amounts to the closest equivalent to generating a census-level measure of campus mobilization. We use machine learning algorithmic techniques to systematically scan, code, and classify mobilization events based on the movement tactics organizers utilize.

### Background for Creating a Measure

Newspapers have long been long been utilized to study social movement activity and collective mobilization and continue to be considered an appropriate and reasonable source of data (Earl, Martin, McCarthy, & Soule, 2004; Oliver & Myers, 1999). In fact, newspaper event analysis in social movement research originally gained momentum as a data collection approach as means for remedying selection bias issues that occurred when researchers sampled on the dependent variable (i.e. choosing cases by occurrence of social movement activism) (Olzak, 1989). In our contemporary era, there is a near ubiquity of electronic newspaper sources. The expanding availability of newspaper data over the past decades offers opportunities for electronic capturing on a scale not seen in prior eras.

Protest event analysis (PEA) is a content analytic technique, but it is primarily concerned with compiling comparative quantitative data regarding social movements that are dispersed over time and geographic space (Koopmans & Rucht, 2002). The method involves analyzing newspapers, media, or public accounts of protest activities. Since its early application in the 1960s, the process has become increasingly sophisticated by comparing media accounts with complementary data sources to account for any selectivity biases that may be inherent when studying a solitary account. Initially, historical sociologists used PEA to catalog various social and political indicators to conduct international comparative social movement research (Tilly, 2004; Tilly, Tilly, & Tilly, 1975). Since its beginnings, PEA has grown to emphasize specificity in source selection, coding categories, and thorough documentation of rules and procedures. Moreover, it has emerged as a means to "systematically map, analyze, and interpret the occurrence and properties of large numbers of protests by means of content analysis" (Koopmans & Rucht, 2002, p. 231). PEA has almost exclusively been performed through hand coding. Using

informatics techniques, with appropriate testing and measurement development, we make an important contribution to PEA literature by introducing machine learning techniques.

**Measurement Technique Overview**

In our study, we use supervised machine learning to not only build a data set of news articles, but also to identify the relevant characteristics. Supervised learning is a type of machine learning algorithm that uses a known dataset (i.e. training dataset) to make inference on the rules that led human coders to classify the documents as exhibiting a certain characteristic, and then applies those learnt rules to classify the remaining documents (i.e. test dataset) (Burscher et al. 2014). This approach not only reduces the costs of analyzing texts as it requires scholars to manually code only a fraction of a large corpus of documents, but also captures subtle content such as campus mobilization tactics through a concrete coding scheme and an accurate training set. In our case, whether a new article is about campus mobilization, and whether a mobilization event adopts certain tactics are considered as binary classification problems.

We rely on three popular supervised learning algorithms to classify news articles: Multinomial Naive Bayes (MNB), Support Vector Machines (SVM) and an ensemble Random Forest. MNB has been extensively used for document classification (Friedman, Hastie and Tibshirani, 2001). It implements the naive Bayes algorithm for multinomially distributed data and works extremely fast compared to other supervised learning algorithms (Zhang, 2004). A SVM is another commonly used algorithm which attempts to find the multidimensional hyperplane that best discriminates between two classes. The researcher uses a regularization parameter to make the hyperplane more or less complex and another parameter is used to define the relative influence of the points that are close/not close to the estimated boundary between two classes. To improve accuracy and robustness over an individual classifier, we also estimated an ensemble Random Forest which combine the predictions of several base learning algorithms into one “super learner.” This method has gained popularity because it avoids convergence issues surrounding local optima by running a grid search using weighted sums from many different starting points which makes the model more representative of the true data-generating process.

For each classifying problem, we applied the above three algorithms and choose the best performing one based on validation indicators. We assessed the performance of each model using the overall accuracy (AC) and the area under the receiver operating curve (AUC). The former is simply the percentage of agreement between human classifications and computer-based classifications. The latter assesses how well the model discriminates between the presence and the absence of a frame (Burscher et al., 2014). The main advantage of AUC over other evaluation methods is its insensitivity to unbalanced data sets. A random guess produces an AUC of approximately 0.5, while a perfect model scores an AUC of 1. In developing our measure, we used five-fold cross-validation to assess classification performance (Friedman, Hastie and Tibshirani, 2001). This means the training set was randomly partitioned into five groups. For each group, the model is trained on the four other groups, and then applied to the fifth group to evaluate performance. We take the average of the results from the five validation rounds to obtain a single evaluation measure for each algorithm. The classifier with the best performance was then used to classify the remaining new articles.

**Campus Mobilization Measure**

Our data consists of 7,452 articles from 2012-2017 which were gathered from Lexis-Nexis Uni database using topic-relevant search strings. Our code book was utilized PEA, but was

informed by both the literature on campus mobilization and tactics as well as Wang and Soule's (2012) prior project involving hand coding the *New York Times* for movement events between 1965-1995. Hand coders were assigned to a random sample of articles from the 10% training set (749 articles). Coders first read an article in full, then they reviewed it to determine if it met the established codebook definition of a campus mobilization event. If it met the definition, they coded the article content to log tactical characteristics. The coding scheme classified tactics into one of five conceptual clusters: conventional, expressive, disruptive, artistic, or violent.

The training and validation of machine-learning classifiers, together with subsequent content analysis of the test data, were realized in scikit-learn - an open-source machine learning library for the Python programming language. Preprocessed and quantified by transforming into a document-term matrix (Grimmer & Stewart 2013; van Zoonen & van der Meer 2016), the training set of hand-coded documents was fed to the aforementioned three machine-learning algorithms: MNB, SVM and Random Forest.

Of the 7,452 articles retrieved from Lexis Nexis, 4,199 were identified as campus-mobilization based using our algorithm (Accuracy: 0.76; AUC: 0.75). Focusing only on these articles, we then use another machine learning algorithm to predict each of the five tactic types (conventional, disruptive, expressive, artistic, and violent) using our survey-based coding scheme. Generally speaking, the supervised learning algorithms performed very well across all five types of tactics with Accuracy between 0.77 - 0.99, and AUC between 0.72 - 0.93. More specifically, respectively the Accuracy and AUC measures are: conventional tactics 0.90, 0.83, disruptive 0.78, 0.74; expressive 0.77, 0.74, artistic 0.95, 0.72, and violent 0.99, 0.93. We find that our ability to accurately classify violent tactics is especially noteworthy since campuses are especially worried about violence when campus mobilization occurs.

Collectively, our results demonstrate that machine learning can be used to accurately classify both the degree of campus mobilization in the United States and the types of tactics that are used. We think our measurement approach holds great promise for more widespread use in which campus mobilization events are assessed in real-time which would provide researchers, policy makers, and university officials and activists valuable information about the scope and characteristics of campus mobilization in the United States.

**Grant Deliverables and Dissemination**

Once we generated a reliable measure of campus mobilization, we combined this measure with other publicly available data to make sense of the character of campus mobilization. Specifically, we merged our campus mobilization data and the characteristics of the news article cases with data from the *National Center for Educational Statistics, Integrated Postsecondary Education Data System*.

We submitted a conference proposal in May of 2018. The proposal was accepted in the summer of 2018, and our first presentation of the campus mobilization measure, and accompanying analyses associated with the measure occurred in November 2018 at the Association for the Study of Higher Education annual meeting in Tampa, Florida. With the feedback offered during the conference session, we are currently refining the paper to submit it for peer review.

During the IMRF grant period, we also submitted a Letter of Intent to the Spencer Foundation. Unfortunately, we were not invited to submit a proposal. At the time of submission, we were just a few short weeks into the project, and had yet to show the efficacy of the measure at that time. Now, after having had the support of the IMRF, we are in a far better position to

seek subsequent funding for this project to leverage its utility. Indeed, we have used the funding to demonstrate a proof of concept. Currently, we have initiated outreach to potential funders to gauge their interest in further supporting our work.

Barnhardt, C. L., Dietrich, B.J., Yao, J., Mohebali, M., Young, R. L., & Waad, *A. Campus Mobilization Dynamics: Models of Methods, Counts, Coverage, and Tactics.* Paper presented at the 2018 Association for the Study of Higher Education, Tampa, FL. (November 2018).

Barnhardt, Cassie L. (Principal), Dietrich, Bryce. Letter of Intent for the Lyle Spencer Research Award, "Campus Mobilization: How Does the Exercise of Freedom of Speech and Association on U.S. Campuses Contribute to Education and Democracy?" Sponsored by The Spencer Foundation, $350,000 (Submitted October 12, 2017).

**Emerging Insights From the Data**

In contrast to prior scholarship on campus mobilization, contemporary campus activism is not a phenomenon limited to students attending elite institutions. By engaging in a field-level sample, we see that about 3% of the campus mobilization covered in the press is from the 2-year sector, likewise 57% of articles described events on campuses that are moderately selective or open access (non-elites) institutions.

With respect to the scale and type of coverage that campus mobilization receives, we highlight a few interesting patterns. The press pays more attention to campuses that use expressive tactics (holding a vigil, candle lighting, hosting a speaker or a panel of speakers, distributing informational materials about their cause like flyers). While campus newspapers are the most likely venue for campus mobilization to be reported (56% of all campus mobilization were covered in campus papers), the local and regional papers are uniquely inclined to cover violent mobilization events on campuses near them. This finding holds important implications for town-gown relationships and community policing as campuses and activists consider the likely local response from the communities where protest and dissent occur. Communities may perceive mobilization as far more violent than that which is the typical behavior pattern on campus.

Our analyses also reveal that educational trade publications are also more inclined to cover expressive tactics. Education trade publications also prioritize covering mobilization events on large campuses, and those with the largest endowments. That is, the educational field tends to highlight the activism that occurs at the big and rich institutions.

In the national press, like educational trade publications, the reporting on campus mobilization also features the big and rich campuses. National press is unique however in the sense that it tends to feature activism that occurs on public universities, and covers novel tactics like artistic approaches (skits, art installations, political theatre, wearing costumes) at a far higher rate than that which is represented in other publications (campus papers, local/regional). Additionally, national publications features activism on campuses that have smaller enrollments of African-American, Latinx, and Asian or Asian-American students.

*Notes*

~Research associates assisting with the preparation of the report include graduate students: Jielu Yao, Milad Mohebali, Ryan L. Young, and Alex Waad. Graduate students that assisted with the hand-coding process included the aforementioned research associates as well as: Christopher Patterson and Matthew Augeri

## References

Burscher, B., Odijk, D., Vliegenthart, R., De Rijke, M., and De Vreese, C. H. (2014). Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis. *Communication Methods and Measures*, 8(3):190-206.

Earl, J., Martin, A., McCarthy, J. D., & Soule, S. A. (2004). The use of newspaper data in the study of collective action. *Annual Reveiw of Sociology, 30*, 65-80.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin.

Grimmer, J. and Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, page mps028.

Koopmans, R., & Rucht, D. (2002). Protest event analysis. In B. Klandermans & S. Staggenborg (Eds.), *Methods of Social Movement Resesarch: Social Movements, Protest, and Contestation* (Vol. 16). Minneapolis: University of Minnesota Press.

Oliver, P. E., & Myers, D. J. (1999). How events enter the public sphere: Conflict, location, and sponsorship in local newspaper coverage of public events. *American Journal of Sociology, 105*(1), 38-87.

Olzak, S. (1989). Analysis of events in studies of collective action. *Annual Reveiw of Sociology, 15*, 119-141.

Rhoads, R. A. (2016). Student activism, diversity, and the struggle for a just society. *Journal of Diversity in Higher Education, 9*(3), 189-202. http://dx.doi.org/10.1037/dhe0000039

Rojas, F. (2006). Social movement tactics, organizational change and the spread of African-American studies. *Social Forces, 84*(4), 2147-2166.

Tilly, C. (2004). *Social Movements 1768-2004*. Boulder, CO: Paradigm Publishers.

Tilly, C., Tilly, L., & Tilly, R. (1975). *The Rebellious Century 1830-1930*. Cambridge: Harvard University Press.

Wang, D. J., & Soule, S. A. (2012). Social movement organizational collaboration: Networks of learning and the diffusion of protest tactics, 1960-1995. *American Journal of Sociology, 117*, 1674-1722.

van Zoonen, W., & van der Meer, T. (2015). The importance of source and credibility perception in times of crisis: Crisis communication in a socially mediated era. *Journal of Public Relations Research, 27*(5), 371-388.

Zhang, H. (2004). *The optimality of naive bayes*. AA, 1(2):3.