Social Media Data Mining of Stakeholder Value Systems on Community Resilience in Florida

Hang Ren, S.M.ASCE[1]; Lu Zhang, Ph.D., A.M.ASCE[2];
Arif Mohaimin Sadri, Ph.D., A.M.ASCE[3]; N. Emel Ganapati, Ph.D.[4];
and Travis A. Whetsell, Ph.D.[5]

[1]Graduate Student, Myers-Lawson School of Construction, Virginia Tech, Blacksburg, VA. Email: hangr@vt.edu
[2]Associate Professor, Myers-Lawson School of Construction, Virginia Tech, Blacksburg, VA (corresponding author). Email: luzhang@vt.edu
[3]Assistant Professor, School of Civil Engineering and Environmental Science, Univ. of Oklahoma, Norman, OK. Email: sadri@ou.edu
[4]Professor, Dept. of Public Policy and Administration, Florida International Univ., Miami, FL. Email: ganapat@fiu.edu
[5]Assistant Professor, School of Public Policy, Georgia Institute of Technology, Atlanta, GA. Email: travis.whetsell@gatech.edu

ABSTRACT

Building disaster-resilient communities requires a better understanding of what community stakeholders value regarding resilience. Communities with various characteristics (e.g., geographic, demographic, social vulnerabilities) are expected to prioritize resilience strategies to align with the value systems of their stakeholders. Stakeholder values include the aspects that are of importance, merit, and provide utilities to them, and stakeholders from different communities may hold values with varying degrees of importance, forming their distinct value systems. Despite the importance of involving stakeholders in resilience planning, there is limited awareness of stakeholder value systems with respect to resilience planning. To address this gap, this study focused on identifying stakeholder value systems and analyzing how the value systems vary across different communities (i.e., coastal vs. inland; metropolitan vs. non-metropolitan; more vs. less socially vulnerable). A semi-supervised learning technique—anchored correlation explanation (anchored CorEx)—was used to analyze the social media data (i.e., Twitter) collected within the Florida communities. The results demonstrate that communities with various characteristics have significantly different priorities regarding several stakeholder values. The findings of this study could inform policymakers to better plan for community resilience by aligning their strategies with stakeholder value systems, thereby increasing the effectiveness and efficiency of implementing resilience practices.

INTRODUCTION

There is a growing number of studies that proposed “stakeholder-centered” resilience planning in addressing climate or disaster-induced challenges in recent years. Such planning approach focuses on integrating the needs, concerns, and perspectives of stakeholders in developing resilience practices to mitigate, adapt to, and recover from potential disasters (Tompkins et al. 2008, Ashmawy 2021, Gosain et al. 2022). However, achieving community resilience is not an easy task as different communities have their unique vulnerabilities depending on their characteristics. For instance, as compared to inland communities, coastal communities are typically located in areas with lower elevations; and, thus, are more susceptible to threats such as sea-level rise, coastal erosion, and flooding (NOAA 2022). Thus, stakeholders from coastal communities may prioritize different resilience strategies as compared to inland communities. In the disaster domain, stakeholders could be any persons, organizations, or groups that show an interest or concern in disaster planning (Pathak et al. 2020). Stakeholder values are defined as the things that are of importance, merit, and utilities to them. Various stakeholders usually hold numerous values with varying degrees of importance, forming their distinct value systems (Gosain et al. 2022). To tackle the unique vulnerabilities of various communities, policymakers are expected to develop stakeholder value-centered strategies that are aligned with the specific needs of stakeholders from different communities.

To facilitate stakeholder-centered resilience planning, prior studies have emphasized the importance of engaging diverse stakeholders in resilience planning (e.g., Pyke et al. 2018, Taeby and Zhang 2019, Ren et al. 2022) and have proposed various approaches to facilitate stakeholder engagement (e.g., Desportes et al. 2016, Sitas et al. 2016). However, limited research has been conducted to identify and analyze stakeholder values in resilience planning. Although some researchers in the disaster domain have recently explored stakeholder value systems across multiple sectors (Gosain et al. 2022) and throughout different disaster phases (Pathak et al. 2020), they did not specifically compare stakeholder value systems across communities with varying characteristics.

To address this knowledge gap, this study aims to examine the variations of stakeholder value systems across different communities (i.e., coastal vs. inland; metropolitan vs. non-metropolitan; more vs. less socially vulnerable) in Florida through social media data analysis. The results indicate that stakeholders from different communities had significant variations in their priorities regarding some values. The findings of this study could help policymakers improve resilience planning by aligning resilience practices with the value systems of community stakeholders. The remaining sections of this paper introduce the research methodology, present and discuss the results, and summarize the work with future recommendations.

METHODOLOGY

Data Collection. We collected two sets of Twitter data through the Twitter Streaming Application Programming Interface (API). The first set of data was collected within the geographic boundary of Florida. The second set of data includes those tweets that are related to community resilience. To collect the second set of data, we used a list of 144 keywords in data collection, such as disaster, resilience, community, power, food, infrastructure, and safe, among others. We identified those keywords through a comprehensive review of literature in the community resilience and disaster management domains (e.g., Taeby and Zhang 2019, Li et al. 2021, Gosain et al. 2022).

Data Preprocessing, and Filtering. We then preprocessed the collected tweets through the following steps: (1) excluding the tweets that are not in English; (2) converting the letters to lowercase; (3) removing the user mentions (e.g., @user); (4) removing the invalid symbols and URL links; (5) removing the punctuations and digits; (6) expanding the word contractions (e.g., converting “didn’t” to “did not”); (7) lemmatizing words into their root form (e.g., converting “playing” to “play”); and (8) removing stop words (e.g., “are”, “that”) that are less meaningful in the analysis.

To ensure the first set of data was related to community resilience, we filtered data using the same 144 relevant keywords. We then filtered the second set of data to ensure the tweets were either generated in Florida or were relevant to Florida stakeholders. To achieve that, we created a place name dictionary of Florida, which includes the names of the counties, cities, and towns in Florida. In this process, we excluded those place names that might cause confusion or add noise to the analysis (e.g., "baker", "orange"). We filtered the relevant tweets by determining if the tweet fields (i.e., username, user description, user location, or tweet place) contains the texts from the place name dictionary. Lastly, we combined these two filtered datasets and removed the duplicate tweets.

Stakeholder Value Classification. This step aims to classify the collected tweets into 20 stakeholder value topics that were predefined by Gosain et al. (2022). Figure 1 shows the hierarchy of those values (see Gosain et al. (2022) for the detailed definition of each value). We leveraged the Anchored Correlation Explanation (Anchored CorEx) algorithm to classify the tweets into 20 stakeholder values. Compared to unsupervised learning methods (e.g., Latent Dirichlet Allocation), Anchored CorEx allows users to integrate their domain knowledge into analysis, and it is powerful for analyzing short texts (e.g., tweets) (Gallagher et al. 2017). Therefore, users can define a set of "anchor words" that encourage the model to search for a potential topic. Anchored CorEx focuses on maximizing the total correlation among a group of words when conditioning on latent topics through Eq. 1 (Gallagher et al., 2017).

$$\max_{\alpha_{i.j}, p(y_i|x)} \sum_{j=1}^{m} (\sum_{i=1}^{n} \alpha_{i.,j} I(X_i : Y_j) - I(X : Y_j)) \tag{1}$$

where X and Y are two discrete random variables that take on a finite number of words and topics; X_i represents a subcollection of words where $i \subseteq \{1, \ldots, n\}$; Y_j represents the latent topics that correspond to the 20 stakeholder values where $j \subseteq \{1, \ldots, m\}$; m is the number of topics; n is the number of words corresponding to the topic; I calculates the mutual information of the included variables; and $\alpha_{i,j}$ controls the anchoring strength of words to the topics.

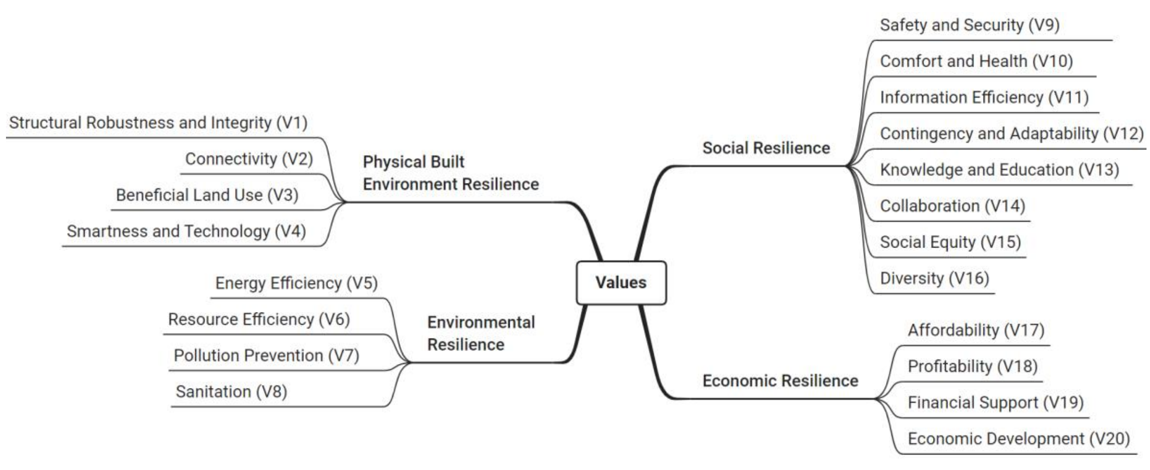

Figure 1. Stakeholder values hierarchy (Source: Gosain et al. 2022).

We used a combination of deductive and inductive methods to select anchor words for each stakeholder value. For the deductive approach, we identified the anchor words based on a comprehensive literature review for each stakeholder value (e.g., Desportes et al. 2016, Taeby and Zhang 2019, Li et al. 2021, Gosain et al. 2022). For the inductive approach, we identified the anchor words based on the empirical observation of the tweets. The top frequently shown words were manually examined to see if they were relevant to each stakeholder value and/or could complement the deductive approach.

Model Performance Evaluation. We randomly selected 500 tweets and manually labeled them to compare the performance of the Anchored CorEx with four supervised machine learning models (i.e., Linear Support Vector Classifier (SVC), Logistic Regression, Multinomial Naive Bayes, and Random Forest Classifier) that are popularly used for text classification (Kotsiantis 2007). The accuracy, precision, recall, and F1 score were used for performance evaluation, and the results are shown in Table 1. As per Table 1, the Anchored CorEx model even outperformed the Multinomial Naive Bayes and Random Forest Classifier. The performance of Anchored CorEx is only slightly lower than those of the Linear SVC and Logistic Regression. These results indicate that Anchored CorEx has a relatively reliable performance in stakeholder value classification.

Table 1. Comparison of model performance.

Model	Accuracy	Precision	Recall	F1 score
Anchored CorEx	0.90	0.92	0.90	0.90
Linear SVC	0.92	0.93	0.92	0.92
LogisticRegression	0.92	0.93	0.92	0.93
Multinomial Naive Bayes	0.88	0.88	0.88	0.87
Random Forest	0.83	0.91	0.83	0.85

Determining Stakeholder Value Priorities. To measure the importance or priorities of each stakeholder value, we calculated the Ratio index (RI) of each stakeholder value through Eq. 2. The RI measures the ratio of the number of tweets that are related to a specific stakeholder value topic to the total number of tweets. The RI ranges from 0 to 1, where a high RI indicates the value is highly discussed by the stakeholders.

$$RI_k = \frac{n_k}{N} \tag{2}$$

where RI_k is the ratio index of stakeholder value k; n_k is the total number of tweets related to the value k; and N is the total number of tweets.

Comparing Stakeholder Value Priorities across Communities. We investigated the differences in stakeholder value priorities across communities with different geographic, demographic, and social vulnerability characteristics. For geographic characteristics, we classified Florida counties into coastal and inland groups. The coastal group contains those counties that are adjacent to the water classified as either coastal water or territorial sea (USCB 2021). For demographic characteristics, we classified Florida counties into metropolitan and non-metropolitan groups based on the condition of whether they are within the metropolitan statistical areas as defined by the United States Office of Management and Budget (2022). For social vulnerability characteristics, we classified Florida counties into the more socially vulnerable and less socially vulnerable groups using the Social Vulnerability Index (SVI) defined by ATSDR (2018). SVI measures the vulnerability of each county, and it ranges from 0 to 1, where a higher SVI (greater than or equal to 0.5) indicates that the county is more socially vulnerable. We calculated the stakeholder value priority of each group by taking the average of the ratio indexes. We then performed the Mann-Whitney U test to determine if there were significant distribution differences between groups.

RESULTS AND DISCUSSION

We collected the Twitter data from September 2021 to November 2021, during which there were no disasters that could substantially affect stakeholders' discussion on community resilience. We gathered 544,968 tweets for the first set of data, all of which were generated within Florida's geographic boundaries. Within this dataset, 31,567 tweets were related to community resilience. For the second dataset collected based on community resilience keywords, there were a total of 12,550,000 tweets. Within the dataset, 70,854 tweets originated in Florida or were related to Florida stakeholders. We ended up with a total of 98,725 tweets for analysis after combining these two datasets and removing the duplicates.

Comparisons of Stakeholder Value Priorities across Communities. Tables 2 show the difference of ratio indexes between different groups and the p-value results of the Mann-Whitney U test. A p-value less than 0.05 indicates a significant difference in the distribution of ratio indexes between the two types of communities under analysis.

Table 2. Comparison of stakeholder value priorities across communities.

Stakeholder Value	Coastal vs. Inland	Metropolitan vs. Non-Metropolitan	More vs. Less Socially Vulnerable
Structural Robustness and Integrity (V1)	-0.003	0.001	0.008*
Connectivity (V2)	-0.005	0.007	0.007
Beneficial Land Use (V3)	0.002	0.005*	-0.002
Smartness and technology (V4)	0.001	0.004	-0.001
Energy Efficiency (V5)	0.006*	0.003	0.013*
Resource Efficiency (V6)	0.010*	-0.007	0.009*
Pollution prevention (V7)	-0.010*	-0.063*	-0.012*
Sanitation (V8)	0.007*	0.015*	0.001
Safety and security (V9)	-0.003	-0.004	-0.002
Comfort and health (V10)	-0.004	0.006	-0.004
Information efficiency (V11)	0.014*	-0.003	-0.009*
Contingency and adaptability (V12)	0.000	-0.002	-0.005
Knowledge and education (V13)	-0.001	-0.001	-0.001
Collaboration (V14)	0.005*	-0.004	0.005*
Social equity (V15)	-0.001	0.000	0.004
Diversity (V16)	-0.004*	0.003	-0.001
Affordability (V17)	0.002	-0.001	-0.004
Profitability (V18)	-0.001	-0.006*	-0.002
Financial support (V19)	0.005*	-0.004	0.005
Economic development (V20)	-0.006*	-0.002	0.001

*The p-value is significant at 0.05 level.

Coastal vs. Inland Communities. As per Table 2, there are significant differences in the distribution of ratio indexes regarding nine stakeholder values between the coastal and inland communities. The ratio indexes of energy efficiency (V5), resource efficiency (V6), sanitation (V8), information efficiency (V11), collaboration (V14), and financial support (V19) of the coastal group are significantly higher than those of the inland group. On the contrary, the ratio indexes of pollution prevention (V7), diversity (V16), and economic development (V20) of the inland group are significantly higher than those of the coastal group. These discrepancies may be due to the different levels of disaster threats faced by coastal and inland communities. The coastal communities often suffered from more serious damage than inland communities even exposed to the same disaster (Bathi and Das 2016). Thus, the coastal communities seem to prioritize those values that are related to addressing the immediate needs in a disaster. For instance, energy, resources, information, and financial assistance are all essentially needed in disaster preparedness, response, and recovery. In contrast, inland communities seem to attach higher importance to the values that are related to addressing long-term challenges faced by their communities, such as environmental pollution, lack of diversity, and economic development.

Metropolitan vs. Non-Metropolitan Communities. Table 2 also shows that significant differences exist in the distributions of ratio indexes of four values between the metropolitan and non-metropolitan communities. To be specific, the ratio indexes of beneficial land use (V3) and sanitation (V8) in metropolitan communities are significantly higher than those in non-metropolitan communities. On the other hand, the ratio indexes of pollution prevention (V7) and profitability (V18) in non-metropolitan communities are significantly higher than those in metropolitan communities. Such differences may be caused by the varying levels of urbanization and populations. For instance, the rapid development and population growth in Florida metropolitan communities have created major challenges in land planning and use; many existing structures are built on lands that are highly vulnerable to natural hazards. For example, in Hurricane Ian, the cities that were hit the hardest were built on vulnerable lands that were created by destroying natural ecosystems (e.g., draining swamps, cutting down mangroves) (Mahadawi 2022). Avoiding new construction in high-risk areas as well as relocating the existing structures that are in high-risk locations are critical for metropolitan communities (EESI 2020). The non-metropolitan communities, on the other hand, are less populated, less developed, and may have more agricultural acreage, and stakeholders in these communities may tend to prioritize those resilience practices that help create a healthy environment for agriculture, which may help with their long-term economic profits (Bayabil et al. 2022).

More vs. Less Socially Vulnerable Communities. Table 2 shows that there are significant differences in the distribution of ratio indexes of six values between the more socially vulnerable and the less socially vulnerable communities. More specifically, the more socially vulnerable communities have significantly higher ratio indexes in structural robustness and integrity (V1), energy efficiency (V5), resource efficiency (V6), and collaboration (V14), while less socially vulnerable communities have significantly higher ratio indexes in pollution prevention (V7) and information efficiency (V11). Overall, the results may suggest that the more socially vulnerable communities prioritize those values that are related to addressing their basic needs in a disaster, such as having a safe living environment, gaining access to resources and energy, and obtaining support from others. On the other hand, the less socially vulnerable communities tend to place higher priorities on long-term development-related values, such as protecting the natural environment and efficiently managing data and information. For instance, natural environmental protection can facilitate stable development of tourism and create more job opportunities, which would eventually sustain the long-term resilience development.

CONCLUSIONS

This study aimed to analyze stakeholder value systems across communities with varying geographic, demographic, and social vulnerability characteristics. Using a semi-supervised topic modeling approach named Anchored CorEx, we analyzed Twitter data and found significant differences in the priorities of stakeholder values across various types of Florida communities. The research offers both theoretical and practical contributions. From a theoretical perspective, it offers a holistic understanding of stakeholder value systems in community resilience planning through analyzing big social data. The semi-supervised topic modeling and classification approach used in this study facilitates rapid and efficient quantitative comparisons of stakeholder value systems across different communities. From a practical perspective, the findings of this study may assist policymakers in determining the most appropriate resilience strategies based on the value systems of the community stakeholders. In the future, we plan to further analyze the community stakeholders' sentiment on the values and identify those values that may require more attention. In addition, we acknowledge that social media data is inherently biased (Chandrasekaran et al. 2020). It is thus recommended to utilize other research methods (e.g., surveys, interviews) to collect data from stakeholders who have limited access to social media platforms.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 1933345. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

REFERENCES

Ashmawy, I. K. I. M. (2020). "Stakeholder involvement in community resilience: evidence from Egypt." *Environ. Dev. Sustain.*, 23 (5): 7996–8011.

ATSDR (Agency for Toxic Substances and Disease Registry). (2022). "CDC SVI Documentation 2020." <https://www.atsdr.cdc.gov/>(Jan. 10, 2023).

Bathi, J., and H. Das. (2016). "Vulnerability of Coastal Communities from Storm Surge and Flood Disasters." *Int. J. Environ. Res. Public Health*., 13 (2): 239.

Bayabil, H. K., Y. Li, J. Crane, B. Schaffer, A. Smyth, S. Zhang, E. A. Evans, and T. Blare. (2022). "Saltwater Intrusion and Flooding: Risks to South Florida's Agriculture and Potential Management Practices." *EDIS*., 2022(3).

Chandrasekaran, R., V. Mehta, T. Valkunde, and E. Moustakas. (2020). "Topics, Trends, and Sentiments of Tweets About the COVID-19 Pandemic: Temporal Infoveillance Study." *J. Med. Internet Res.*, 22 (10): e22624.

Desportes, I., J. Waddell, and M. Hordijk. (2016). "Improving flood risk governance through multi-stakeholder collaboration: A case study of Sweet Home informal settlement, Cape Town." *S. Afr. Geog. J.*, 98 (1): 61–83.

EESI (Environmental and Energy Study Institute). (2020). *A Resilient Future for Coastal Communities: Federal Policy Recommendations from Solutions in Practice*. Washington, D.C.: EESI.

Gallagher, R. J., K. Reing, D. Kale, and G. Ver Steeg. (2017). "Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge." *Trans. Assoc. Comput. Linguist.*, 5 (Dec): 529–542.

Gosain, P., L. Zhang, and N. E. Ganapati. (2022). "Understanding multisector stakeholder value systems on housing resilience in the City of Miami." *Int. J. Disaster Risk Reduct.*, 77 (Jul): 103061.

Kotsiantis, S. (2007). "Supervised Machine Learning: A Review of Classification Techniques." *Inform. Slov.*, 31(3): 249–268.

Laerd Statistics. (2022). "Laerd statistics." <https:// statistics.laerd.com/> (Mar. 10, 2023).

Li, Q., M. Roy, A. Mostafavi, and P. Berke. (2021). "A plan evaluation framework for examining stakeholder policy preferences in resilience planning and management of urban systems." *Environ. Sci. Policy.*, 124 (Oct): 125–134.

Mahdawi, A. (2022). "Hurricane Ian was less a natural disaster than a human-made one. We must stop building on swamps." <https://www.theguardian.com/commentisfree/2022/oct/05/hurricane-ian-was-less-a-natural-disaster-than-a-human-made-one-we-must-stop-building-on-swamps> (Mar. 15, 2023).

NOAA (National Oceanic and Atmospheric Administration). (2022). "Hurricanes in History." <https://www.nhc.noaa.gov/outreach/history/>(Jan. 20, 2023).

Pathak, A., L. Zhang, and N. E. Ganapati. (2020). "Understanding multisector stakeholder value dynamics in Hurricane Michael: Toward collaborative decision-making in disaster contexts." *Nat. Hazards Rev.*, 21 (3): 04020032.

Pyke, J., A. Law, M. Jiang, and T. de Lacy. (2018). "Learning from the locals: the role of stakeholder engagement in building tourism and community resilience." *J. Ecotourism.*, 17 (3): 206–219.

Ren, H., L. Zhang, T. A. Whetsell, and N. E. Ganapati. (2023). "Analyzing Multisector Stakeholder Collaboration and Engagement in Housing Resilience Planning in Greater Miami and the Beaches through Social Network Analysis." *Nat. Hazards Rev.*, 24 (1): 04022036.

Sitas, N., B. Reyers, G. Cundill, H. E. Prozesky, J. L. Nel, and K. J. Esler. (2016). "Fostering collaboration for knowledge and action in disaster management in South Africa." *Curr. Opin. Environ. Sustainability.*, 19 (Apr): 94–102.

Taeby, M., and L. Zhang. (2019). "Exploring stakeholder views on disaster resilience practices of residential communities in South Florida." *Nat. Hazards Rev.*, 20 (1): 04018028.

Tompkins, E. L., R. Few, and K. Brown. (2008). "Scenario-based stakeholder engagement: Incorporating stakeholders preferences into coastal planning for climate change." *J. Environ. Manage.*, 88 (4): 1580–1592.

USCB (United States Census Bureau). (2021). "Emergency Management Coastal Areas." <https://www.census.gov/>(Jan. 24, 2023).

United States Office of Management and Budget. (2010). "2010 Standards for Delineating Metropolitan and Micropolitan Statistical Areas". <https://www.govinfo.gov/> (Jan. 10, 2023).