Data-Driven In-Crisis Community Identification for Disaster Response and Management

Yudong Tao*, Renhe Jiang†, Erik Coltey‡, Chuang Yang†,
Xuan Song†, Ryosuke Shibasaki†, Mei-Ling Shyu* and Shu-Ching Chen‡
*Department of Electrical and Computer Engineering
University of Miami, Coral Gables, FL, USA
†Center for Spatial Information Science
The University of Tokyo, Tokyo, Japan
‡Knight Foundation School of Computing and Information Sciences
Florida International University, Miami, FL, USA
Email: yxt128@miami.edu, jiangrh@csis.u-tokyo.ac.jp, ecolt003@cs.fiu.edu, chuang.yang@csis.u-tokyo.ac.jp,
songxuan@csis.u-tokyo.ac.jp, shiba@csis.u-tokyo.ac.jp, shyu@miami.edu, chens@cs.fiu.edu

***Abstract*—Since 2019, the world has been seriously impacted by the global pandemic, COVID-19, with millions of people adversely affected. This is coupled with a trend in which the intensity and frequency of natural disasters such as hurricanes, wildfires, and earthquakes have increased over the past decades. Larger and more diverse communities have been negatively influenced by these disasters and they might encounter crises socially and/or economically, further exacerbated when the natural disasters and pandemics co-occurred. However, conventional disaster response and management rely on human surveys and case studies to identify these in-crisis communities and their problems, which might not be effective and efficient due to the scale of the impacted population. In this paper, we propose to utilize the data-driven techniques and recent advances in artificial intelligence to automate the in-crisis community identification and improve its scalability and efficiency. Thus, immediate assistance to the in-crisis communities can be provided by society and timely disaster response and management can be achieved. A novel framework of the in-crisis community identification has been presented, which can be divided into three subtasks: (1) community detection, (2) in-crisis status detection, and (3) community demand and problem identification. Furthermore, the open issues and challenges toward automated in-crisis community identification are discussed to motivate future research and innovations in the area.**

***Keywords*-data analysis, multimodal data, disaster management, in-crisis community identification**

I. Introduction

COVID-19 has seriously impacted the world and led to tremendous losses in human lives and continued economic instability. As of November 2021, there are more than 47 million confirmed cases and 773 thousand deaths in the United States; while 258 millions confirmed cases and 5 million deaths have been recorded worldwide [1]. Meanwhile, it has been witnessed in the past decades that the frequency and severity of extreme weather and climate disasters have increased globally [2]. For example, the Pacific Typhoon Season and the Atlantic Hurricane Season have caused the highest losses on record in 2019 and 2020, specifically. The risks of co-occurring natural disasters and pandemics appear to be trending upward, with downstream negative effects in areas such as healthcare, finance, and economy [3], [4].

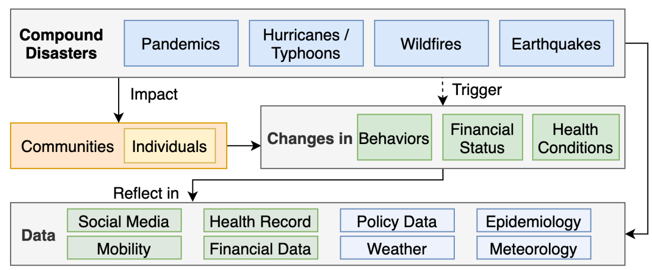

Figure 1. The impacts of disasters reflected in various types of data.

Massive population has been affected by these disaster events socially and economically and diverse communities have faced various challenges and crises caused or induced by the disasters. However, the conventional disaster response and management rely on human surveys and case studies to identify in-crisis communities and their problems [5], [6], [7]. It takes time to properly identify the in-crisis communities, understand their demands, and provide appropriate assistance to them. When the scale of the disasters is large and major communities and problems are prioritized, smaller, more diverse but potentially vulnerable communities might not obtain aid in time [8]. Therefore, it is critical to develop efficient and effective methods for in-crisis community identification to address such emerging challenges in disaster responses and management.

Due to the recent advances in social media and remote sensing technologies, more and more disaster-related data become available. For example, a large number of datasets [9], [10], [11], [12] and dashboard applications [1], [13], covering data in epidemiology, policies, mobility, so-

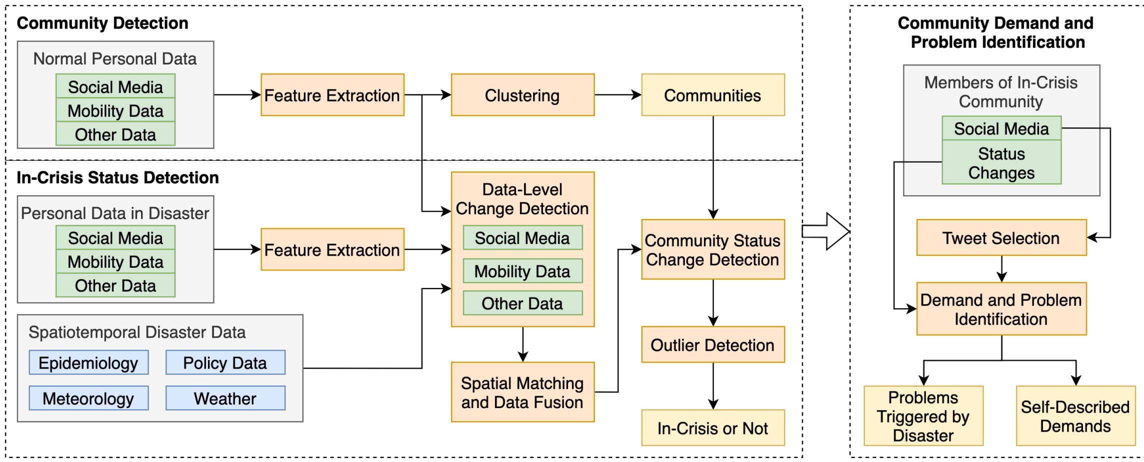

Figure 2. The proposed data-driven framework for in-crisis community identification. This framework separates the problem into three subtasks: (1) community detection, (2) in-crisis status detection, and (3) community demand and problem identification.

cial media, health, weather, etc., have been developed to assist in the disaster responses and management for COVID-19. Depending on the spatial resolution of the data, information about the impacts of disasters on different people and communities has become accessible [14]. Figure 1 depicts how the impacts of the (compound) disaster events are reflected in the data. Furthermore, recent advances in data science and machine learning have shown promising performance to extract information and knowledge from the raw unstructured data for many critical applications [15]. Therefore, due to the increased accessibility of disaster-related data and the advanced data analysis algorithms, data-driven approaches can now be utilized to tackle the in-crisis community identification tasks at scale. However, there remain many challenges and research problems in data analysis and machine learning that need to be addressed to develop a successful solution to in-crisis community identification.

In this paper, we propose a novel data-driven framework to address the in-crisis community identification problem for effective and timely disaster responses and management. In particular, the proposed framework that divides the in-crisis community identification problem into three subtasks and an overview of the related work to each subtask are presented. Open questions and challenges for addressing in-crisis community identification problem are also discussed.

The rest of this paper is organized as follows. The proposed conceptual framework for in-crisis community identification is presented in Section II, along with the existing work closely related to the problem. Then, Section III discusses the challenges to address the in-crisis community identification problem. In the end, Section IV concludes this vision paper.

II. In-Crisis Community Identification

In this section, the proposed data-driven framework for in-crisis community identification is presented. As shown in Figure 2, the in-crisis community identification problem is solved by separating it into three subtasks: (1) community detection, (2) in-crisis status detection, and (3) community demand and problem identification.

A. *Framework Design*

The overall goals of the proposed framework are (1) detecting the communities that encounter crises due to the disasters and (2) for those detected in-crisis communities, identifying their problems and demands. Thus, three subtasks are defined to address the in-crisis community identification problem. Specifically, "community detection" aims to detect the communities based on the personal data during a normal period, "in-crisis status detection" aims to determine whether each community detected in the first subtask is considered an in-crisis community, and "community demand and problem identification" aims to leverage natural language processing (NLP) and other status change information to identify the problems faced by the community due to the unanticipated changes caused by the disasters as well as the demands to bring the community back to normal.

In the following, the designed processing steps of each subtask and the existing work closely related to the subtask are described.

B. *Community Detection*

There are lots of previous work in the field of detecting communities or user subgroups in social networks [16], [17], [18], [19] and human mobility data [20], [21], [22], where the main approaches are based on comparing and clustering the properties and characteristics of the contents, the network structure, or a mixture of both. To detect communities based on such data, the features characterizing each individual need to be extracted and data clustering algorithms should be properly designed to group these individuals to the communities they belong to.

However, each individual in a society can possibly fill different roles and thus in fact can belong to multiple communities at once. During disasters, individuals may encounter a crisis in only one of the communities they belong to. Therefore, soft clustering approaches are considered better methods to solve the problem. Most community detection methods are typically based on either social media and human mobility, and some of these works are introduced in the following.

Community Detection Based on Social Networks. Due to the high-dimensional nature of typical social network data, Deep Learning (DL) methods are the preferred approach in this domain. For example, DL methods can be used to embed the very high dimensional network data common in social networks [23]. This low-dimensional representation could thereafter be used with standard clustering algorithms to detect communities. Furthermore, General Adversarial Networks (GANs) can help generate communities with overlapping vertices, which is very common in a social network where a single user may belong to multiple communities [24]. The increasing interest in Graph Neural Networks (GNNs) and Graph Convolutional Networks (GCNs) has also led to the creation of novel community detection algorithms [25], [26], leveraging their power in finding relationships in graph data.

Community Detection Based on Mobility Data. Nowadays, with the rapid development of Internet of Things (IoT) and Cyber-Physical Systems (CPS) technologies, big mobility data is being generated from mobile phones, car navigation systems, WLAN networks, and location-based social networks. Such data has been widely used to address urban challenges, including intelligent transportation systems, event crowd management, and disaster management/response, etc. In particular, a series of studies [20], [21], [22], [27], [28], [29] have been proposed for community detection by utilizing human mobility data. The first step in each of these efforts is to build the graph from the mobility data and then apply the community detection algorithms to find communities. According to utilization scenarios, it can be divided into two types: communities in transportation networks [22], [28] and movement agents [20], [21], [27], [29]. Our vision falls on the second one. By treating each trajectory as a node and the pass count as edges, Jiang *et al.* proposed a variant of the bottom-up community detection mechanism [30] to detect non-overlapping communities from football players' trajectories [20]. Similarly, treating each trajectory as a node, TODMIS [29] and STCCD [21] created multiple similarity metrics considering supplementary semantic information with raw trajectory data to construct the graph. Besides, the similar construction process is also used to find and analyze the mobility communities in the opportunistic network by measuring the length and frequencies of contact for a pair of nodes [27]. However, none of the existing methods considered behavioral change, which is crucial under crisis.

C. In-Crisis Status Detection

Since each disaster event is unique and distinct from the previous ones, it is very difficult to predict which communities will be impacted by future disasters and in what way. Therefore, the in-crisis status detection should be conducted in an ad-hoc manner. Moreover, it is expected that the significant and abrupt changes in the data patterns can be observed for people in the communities, which can be utilized to identify which communities or subgroups of a community are in crisis. For example, for social media data, the changes in content, posting frequency, the post sentiment, and other language or behavioral patterns during disasters can be observed [31], [32]. The differences in these features can be measured at the data-level in the "data-level change detection" step in Figure 2.

However, the social media data, mobility data, and other data modalities are usually collected from various data sources without personal identifiable information to link the data together. Therefore, if it is desired to integrate all the data from various modalities to determine whether the communities are in crisis or not, specific data fusion methods at community-level are needed. Specifically, since geographic locations are embedded in some social media data or can be inferred from the contents of such data, an integral analysis of the identified in-crisis communities using social media, mobility data, and other data modalities with geographic location information can be conducted. The patterns in geographic locations can be compared using the same distance metrics for detection based on mobility data, so the communities identified by social media data and by mobility data can be matched.

While all the communities will be impacted by the disasters, only a small proportion of these communities would encounter severe adverse effects from a disaster compared with other communities. Therefore, once the status changes at the community-level can be obtained by integrating changes from all the data, the communities in crisis can be detected by applying the local outlier factor detection algorithm [33] to the status changes in data patterns. The outlier communities whose changes are significantly different from the others can thus be suspected as in-crisis community.

D. Community Demand and Problem Identification

In addition to detecting the in-crisis communities, understanding the causes of the crisis and the aid necessary to help a community suspected to be in crisis is another important task. Solving this allows the stakeholders to respond immediately to effectively and efficiently provide relevant aid to the identified communities. While it is always important to understand a community's needs via in-person communication, manual communication is inefficient. In our proposed framework, the social media data from the members of detected in-crisis communities and their status changes after the disasters are utilized to perform "community demand and problem identification." Specifically, the social media data that are related to the disasters will be first selected, and NLP techniques will be utilized to identify the problems and demands represented in that data. Essentially, this can be regarded as a "Reading Comprehension Question Answering" task, where a set of social media data from the communities is analyzed to answer two specific questions: (1) What are the problems faced by the members in these communities, and (2) What aid do they need the most urgently. Therefore, the question answering NLP models can be adapted to generate these answers. In addition, structural status changes can be used to characterize the problems of members in these communities and assist with the problem identification.

NLP Models for Reading Comprehension Question Answering. Since the state-of-the-art question answering models are mostly based on deep learning techniques, the existing work on deep-learning-based question answering models is summarized. Compared to conventional techniques such as TF-IDF, deep-learning-based methods learn to encode questions and documents (here, we have a set of social media data from the identified in-crisis community) into dense question and document representations where the semantic information in the texts are measured [34], [35]. In particular, [36] and [37] train custom encoders to convert each document and question into dense representations, and compute the similarity scores between each pair of documents using the inner product of their vectors. Once the relevant tweets are retrieved, the final answers can be extracted based on machine reading comprehension models. BiDAF [38] learns a question-aware document representation from the raw text data based on the bidirectional attention flow mechanism and various embedding layers. QANet [39] leverages convolutional neural networks and transformer models to learn both the short- and long-distance interactions. In addition to two-stage approaches, end-to-end models [34], [40], [41] have been proposed to directly extract answers for the given question.

III. Open Questions and Challenges

A. Mobility-based Detection

Nowadays, researchers are performing spatio-temporal data analysis with various human mobility data, that includes: (1) GPS trajectory data collected from smartphones or taxis; (2) check-in data collected from apps such as SafeGraph[1], Foursquare, Gowalla or Twitter; (3) taxi/bike trip record data collected from car-hailing/bike-sharing systems[2][3]; and (4) call detail record (CDR) data collected from telephone systems.

The GPS trajectory data is typically structured as *(id, timestamp, latitude, longitude)*, through which we can know a person's location according to the timestamp. When a major disaster like a hurricane or earthquake happens, the most straightforward method for mobility detection is to extract the affected ids as a coarse in-crisis community, given the specific spatial and temporal information of the disaster (i.e., disaster location and time). Moreover, as mentioned above, clustering-based methods that consider spatio-temporal proximity can also be applied to discover the communities. Apart from this, it is more significant for us to attach the semantic information to these communities, namely the "in-what-trouble community" because people may need different aid (e.g., medical, emotional, financial, or psychological [42]) after a disaster happens. Accurate identification of such fine-grained needs will help the implementation of counterpart support. Such latent information is possibly reflected in the transformation of human movement patterns, like the change of periodic moving distance and visits to Region of Interests (ROIs): near-stopped movement may symbolize health problems; stopping visiting the previous workplace simultaneously with less and less shopping behavior may indicate the need for some financial support. Many techniques have been proposed to help detect these potential changes [43], [44]. However, to the best of our knowledge, mapping from changed movement patterns to in-crisis communities are understudied. Detecting these communities is quite challenging because many other factors also positively impact the crisis [42], such as the rise of online shopping and video conferencing. A promising solution is to integrate some social media data to infer problems (e.g., twitter data to infer health conditions [45]) while human mobility data mainly provides the possible location where these problems occur. Through the integration of location information, the two complement each other to achieve precise identification of the location and semantics of an in-crisis community. However, this leads to a different set of challenges. The key reason is that GPS trajectory data has a high sampling rate which means the time interval between GPS records is usually several minutes or even seconds, while check-in data has a low sampling rate. Evidence shows that most of twitter users tweet several times a month [46]. How to effectively fuse and align the sparse mobility data (i.e., check-in data) and the dense mobility data (i.e., GPS data) is a brand-new technical challenge for us. On the other hand, human mobility data has different levels of accessibility, sampling rate, and semantic information [43]. Usually large-scale GPS trajectory data

[1]https://www.safegraph.com/

[2]https://ride.citibikenyc.com/system-data

[3]https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

is the most difficult to obtain, while the check-in data on Location-Based Network Service (LBSN) like Twitter and Foursquare is relatively easy to collect. To design a model with wide applicability, the accessibility of each type of the data should be considered.

B. Social-Network-based Detection

The evolution and adoption of different social media platforms with unique primary modalities and functionality has been absolutely explosive in the past decade. Data on people's experiences at a given moment, especially during a disaster, are becoming increasingly available. In theory, such data could be used to detect in-crisis communities, but challenges still exist towards actually using this data effectively.

A social network is typically considered a variant of a general network, and community detection is analogous to clustering [47]. Many general algorithms work on arbitrary data, such as the K-means, Spectral, and DBSCAN clustering algorithms. More advanced methods such as those described in [24], [48], [49] specifically mention applications in community detection using social media. There has also been research on how social networks are formed and change throughout disaster situations, such as in [50], [51]. Some of the primary takeaways include the exponential increase and decrease of activity during a disaster and the very high degree of activity concentrated in only a few social network users during that period of activity. These core users are typically individuals disseminating news and local government officials and organizations.

Many of the challenges in social network-based detection come from bridging the gap between the available data and community detection algorithms. Different platforms (TikTok, Twitter, Facebook, Instagram) all present information in different primary modalities. For example, Twitter and Facebook primarily use short text-based information, while TikTok displays all content through video and Instagram through images. While it is already possible to do topic detection with examples including Twitter's trending topics functionality as well as existing research works [52], [53], [54], detecting in-crisis communities requires more granular analysis. The need for such analysis provides an opportunity to find novel ways to use the dynamics of different social networks and multimodal data to help detect in-crisis communities, even those which may be more challenging to detect due to downstream effects of different kinds of disasters.

C. Bridging Semantic Gap between Mobility and Social Media

Mobility data and social media data are fundamentally very different, but with key common links, such as the availability of geotags and timestamps. Many studies have tried to simply aggregate social media information using such georeferencing information, however it is likely that more refined methods are necessary [55]. These links can sometimes be very noisy with many outliers, not exist, or not useful at all in creating a consistent representation of a disaster situation. Many edge cases come with a naive integration of the two data types, such as social media posts coming from entirely different locations than those of the disaster or available mobility data not granular enough to help understand the dynamics of smaller communities.

Bridging this semantic gap can allow relevant social media information and mobility data to have a shared representation, providing valuable insights to disaster managers. For example, being able to match mobility data to social media during a disaster using a shared embedding similar to multimodal captioning or language translation as described in [56], [57] can help generate meaningful networks based on both data types. Similarly, integrating web search query and mobility data has shown to be effective to predict the destination of users [58], [59]. To help automate the detection of in-crisis communities, the resulting embeddings could be fed into a more general community detection algorithm.

Bridging this gap will require significant research in understanding the interplay between social media posts and movement. For example, someone consistently tweeting about an incoming disaster may have several distinct mobility outcomes, be it leaving the area entirely, staying at home, or going to a neighboring location. A large amount of such ambiguous information could affect the final shared representation and make the results significantly less helpful for identifying in-crisis communities. Other issues also include a lack of data in smaller or rural communities or even being able to validate the trustworthiness of social media data. Social media can sometimes propagate falsehoods or be out of date with factual information on an event, which could create a large gap with current mobility data.

D. Spatio-Temporal Adaptability.

Depending on the type of disaster/emergency situation, the timescales may be significantly different. An earthquake is seen as a short-term disaster, whereas a hurricane is a mid-term one and a pandemic like COVID-19 is a long-term one. Similarly, from a spatial point of view, disasters can happen at a district, city, state, or even global level. Accordingly, the in-crisis communities should be discovered at different spatio-temporal resolutions and scales. For a pandemic like COVID-19, discovering in-crisis communities for all districts over the entire country and monitoring how they are dynamically changing are important. For an earthquake, we have to make a real-time response for the affected states or cities and keep updating the list of in-crisis communities at an hour level. These will involve vast amounts of streaming data and heterogeneous components for preprocessing, storing, and feature extraction [60]. Developing algorithms and systems with high spatio-temporal adaptability (i.e., resolution, scale, and response time) is a big challenge for us. Streaming computing engines [61] for massive data mining could be necessary.

E. Lack of Benchmark Datasets

An essential component to testing the validity of any framework is the application of that framework onto benchmark datasets. Separate datasets exist for social media or mobility data, often focusing on specific locations or events. For example, data was collected on the Origin-Destination (OD) flows of mobility data in the US throughout the COVID-19 pandemic, derived from detailed mobility data provided by Safegraph [62]. Meanwhile, many social media datasets exist, collecting billions of records from sites such as Facebook, Twitter, Reddit, and Flickr. However, based on literature review, very few, if any, such benchmarking datasets exist which merge social network and mobility data for a comprehensive view of the dynamics of an area over a specified period. A free and fully prepared benchmark properly aggregating both mobility and social network data on crises in specific areas and periods could promote rapid developments in research in this space. A good benchmark dataset based on existing benchmarks avoids many data preprocessing steps, which can create friction for researchers. A good dataset with a possible reference model can be a catalyst to promote disaster response research among the greater community. It would now be possible to quickly test out a variety of novel architectures and ideas on analysis-ready data.

A good benchmark would also allow researchers to compare their methods with those of others. A well-designed benchmark consisting of various disaster events and types can give researchers a quantitative goal by improving upon previously proposed methods. A high-quality dataset can ensure that such findings are beneficial to both the research community and disaster managers, with fast and well-performing models possibly being deployed in the field. However, for in-crisis community detection, there is no existing benchmark dataset that is ready for machine learning and data analysis. More importantly, due to the highly unstructured nature of the data collected from a community, generating a high-quality annotated dataset for in-crisis community identification can be extremely time-consuming and costly.

F. Fairness and Disparity

Low-income communities might have less social network access and fewer mobile phones that provide reliable GPS data [63]. Compared to other communities, low-income communities might suffer a data poverty problem in the proposed data-driven approaches. Similarly, other minority communities might suffer from the same problem due to the small population, community size, limited technology access, etc. To this end, for the community of concerns, additional information that are able to provide insights about their behaviors and living status might be used. Meanwhile, this can be treated as a specific data imbalance issue, where the data from these communities are considered a minority. Data sampling, augmentation and other techniques [64], [65] to improve model performance on imbalanced data can be leveraged to mitigate the issue of fairness and disparity.

IV. Conclusion

In this paper, a novel data-driven framework for in-crisis community identification is presented to improve immediate and in-time disaster response and management. Thereafter, we envisioned and designed three subtasks in the proposed framework to address the in-crisis community identification problem. For each subtask, the potential approaches and related works are explained. In the end, we discussed the open questions and challenges relevant to this problem.

Acknowledgment

This work was partially supported by National Science Foundation (NSF), Grant Number CNS-1952089 and CNS-2125165, and JST SICORP Grant Number JPMJSC2002 and JPMJSC2104.

References

[1] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track covid-19 in real time," *The Lancet infectious diseases*, vol. 20, no. 5, pp. 533–534, 2020.

[2] D. G. Huber and J. Gulledge, *Extreme weather and climate change: Understanding the link, managing the risk*. Pew Center on Global Climate Change Arlington, 2011.

[3] H. Lambert, J. Gupte, H. Fletcher, L. Hammond, N. Lowe, M. Pelling, N. Raina, T. Shahid, and K. Shanks, "Covid-19 as a global challenge: towards an inclusive and sustainable future," *The Lancet Planetary Health*, vol. 4, no. 8, pp. e312–e314, 2020.

[4] I. Monasterolo, M. Billio, and S. Battiston, "The importance of compound risk in the nexus of covid-19, climate change and finance," *Climate Change and Finance (June 20, 2020)*, 2020.

[5] C. M. L. Leong, S. L. Pan, P. Ractham, and L. Kaewkitipong, "Ict-enabled community empowerment in crisis response: Social media in thailand flooding 2011," *Journal of the Association for Information Systems*, vol. 16, no. 3, p. 1, 2015.

[6] R. C. Runyan, "Small business in the face of crisis: identifying barriers to recovery from a natural disaster 1," *Journal of Contingencies and crisis management*, vol. 14, no. 1, pp. 12–26, 2006.

[7] I. H. S. Sawalha, "Collaboration in crisis and emergency management: Identifying the gaps in the case of storm alexa," *Journal of Business Continuity & Emergency Planning*, vol. 7, no. 4, pp. 312–323, 2014.

[8] M. Shokoohi, M. Osooli, and S. Stranges, "Covid-19 pandemic: What can the west learn from the east?" *International Journal of Health Policy and Management*, vol. 9, no. 10, p. 436, 2020.

[9] T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar, and H. Tatlow, "A global panel database of pandemic policies (oxford covid-19 government response tracker)," *Nature Human Behaviour*, vol. 5, no. 4, pp. 529–538, 2021.

[10] S. Gao, J. Rao, Y. Kang, Y. Liang, and J. Kruse, "Mapping county-level mobility pattern changes in the united states in response to covid-19," *SIGSpatial Special*, vol. 12, no. 1, pp. 16–26, 2020.

[11] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for covid-19," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 4–15, 2020.

[12] A. Mahdi, P. Błaszczyk, P. Dłotko, D. Salvi, T.-S. Chan, J. Harvey, D. Gurnari, Y. Wu, A. Farhat, N. Hellmer, A. Zarebski, B. Hogan, and LionelTarassenko, "Oxcovid19 database, a multimodal data repository for better understanding the global impact of covid-19," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.

[13] J. Hasell, E. Mathieu, D. Beltekian, B. Macdonald, C. Giattino, E. Ortiz-Ospina, M. Roser, and H. Ritchie, "A cross-country database of covid-19 testing," *Scientific Data*, vol. 7, no. 1, pp. 1–7, 2020.

[14] T. Li, N. Xie, C. Zeng, W. Zhou, L. Zheng, Y. Jiang, Y. Yang, H.-Y. Ha, W. Xue, Y. Huang *et al.*, "Data-driven techniques in disaster information management," *ACM Computing Surveys (CSUR)*, vol. 50, no. 1, pp. 1–45, 2017.

[15] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–36, 2018.

[16] O. Fraisier, G. Cabanac, Y. Pitarch, R. Besançon, and M. Boughanem, "Uncovering like-minded political communities on twitter," in *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, 2017, pp. 261–264.

[17] M. C. Benigni, K. Joseph, and K. M. Carley, "Online extremism and the communities that sustain it: Detecting the isis supporting community on twitter," *PloS one*, vol. 12, no. 12, p. e0181405, 2017.

[18] A. Irsyad and N. A. Rakhmawati, "Community detection in twitter based on tweets similarities in indonesian using cosine similarity and louvain algorithms," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 6, no. 1, pp. 22–31, 2020.

[19] W. Silva, Á. Santana, F. Lobato, and M. Pinheiro, "A methodology for community detection in twitter," in *Proceedings of the International Conference on Web Intelligence*, 2017, pp. 1006–1009.

[20] Z. Jiang, X. Chen, B. Dong, J. Zhang, J. Gong, H. Yan, Z. Zhang, J. Ma, and S. Y. Philip, "Trajectory-based community detection," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 6, pp. 1139–1143, 2019.

[21] C. Liu and C. Guo, "Stccd: Semantic trajectory clustering based on community detection in networks," *Expert Systems with Applications*, vol. 162, p. 113689, 2020.

[22] J. D. Lamosa, L. R. Tomás, M. G. Quiles, L. R. Londe, L. B. Santos, and E. E. Macau, "Topological indexes and community structure for urban mobility networks: Variations in a business day," *Plos one*, vol. 16, no. 3, p. e0248126, 2021.

[23] S. Xue, J. Lu, and G. Zhang, "Cross-domain network representations," *Pattern Recognition*, vol. 94, pp. 135–148, 2019.

[24] Y. Jia, Q. Zhang, W. Zhang, and X. Wang, "Communitygan: Community detection with generative adversarial nets," in *The World Wide Web Conference*, 2019, pp. 784–794.

[25] D. Jin, Z. Liu, W. Li, D. He, and W. Zhang, "Graph convolutional networks meet markov random fields: Semi-supervised community detection in attribute networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 152–159.

[26] Z. Chen, L. Li, and J. Bruna, "Supervised community detection with line graph neural networks," in *International Conference on Learning Representations*, 2019.

[27] T. Hossmann, T. Spyropoulos, and F. Legendre, "A complex network analysis of human mobility," in *2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS)*. IEEE, 2011, pp. 876–881.

[28] M. Yildirimoglu and J. Kim, "Identification of communities in urban mobility networks using multi-layer graphs of network traffic," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 254–267, 2018.

[29] S. Liu, S. Wang, K. Jayarajah, A. Misra, and R. Krishnan, "Todmis: Mining communities from trajectories," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 2013, pp. 2109–2118.

[30] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[31] M. Mäkinen and M. Wangu Kuira, "Social media and postelection crisis in kenya," *The international journal of press/politics*, vol. 13, no. 3, pp. 328–335, 2008.

[32] R. Goolsby, "Social media as crisis platform: The future of community maps/crisis maps," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 1, no. 1, pp. 1–11, 2010.

[33] M. Al Hasan, V. Chaoji, S. Salem, and M. J. Zaki, "Robust partitional clustering by outlier and density insensitive seeding," *Pattern Recognition Letters*, vol. 30, no. 11, pp. 994–1002, 2009.

[34] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M. Chang, "REALM: retrieval-augmented language model pre-training," *CoRR*, vol. abs/2002.08909, 2020.

[35] Y. Qu, Y. Ding, J. Liu, K. Liu, R. Ren, W. X. Zhao, D. Dong, H. Wu, and H. Wang, "Rocketqa: An optimized training approach to dense passage retrieval for open-domain question answering," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tür, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, and Y. Zhou, Eds. Association for Computational Linguistics, 2021, pp. 5835–5847.

[36] R. Das, S. Dhuliawala, M. Zaheer, and A. McCallum, "Multi-step retriever-reader interaction for scalable open-domain question answering," in *7th International Conference on Learning Representations*. OpenReview.net, 2019.

[37] Y. Feldman and R. El-Yaniv, "Multi-hop paragraph retrieval for open-domain question answering," in *Proceedings of the 57th Conference of the Association for Computational Linguistics*, A. Korhonen, D. R. Traum, and L. Màrquez, Eds., 2019, pp. 2296–2309.

[38] M. J. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi, "Bidirectional attention flow for machine comprehension," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

[39] A. W. Yu, D. Dohan, M. Luong, R. Zhao, K. Chen, M. Norouzi, and Q. V. Le, "Qanet: Combining local convolution with global self-attention for reading comprehension," in *6th International Conference on Learning Representations*. OpenReview.net, 2018.

[40] K. Lee, M. Chang, and K. Toutanova, "Latent retrieval for weakly supervised open domain question answering," in *Proceedings of the 57th Conference of the Association for Computational Linguistics*, A. Korhonen, D. R. Traum, and L. Màrquez, Eds. Association for Computational Linguistics, 2019, pp. 6086–6096.

[41] G. Izacard and E. Grave, "Leveraging passage retrieval with generative models for open domain question answering," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, P. Merlo, J. Tiedemann, and R. Tsarfaty, Eds. Association for Computational Linguistics, 2021, pp. 874–880.

[42] J. J. Van Bavel, K. Baicker, P. S. Boggio, V. Capraro, A. Cichocka, M. Cikara, M. J. Crockett, A. J. Crum, K. M. Douglas, J. N. Druckman *et al.*, "Using social and behavioural science to support covid-19 pandemic response," *Nature human behaviour*, vol. 4, no. 5, pp. 460–471, 2020.

[43] C. Parent, S. Spaccapietra, C. Renso, G. L. Andrienko, N. V. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. A. F. de Macêdo, N. Pelekis, Y. Theodoridis, and Z. Yan, "Semantic trajectories modeling and analysis," *ACM Comput. Surv.*, vol. 45, pp. 42:1–42:32, 2013.

[44] R. Jiang, J. Zhao, T. Dong, Y. Ishikawa, C. Xiao, and Y. Sasaki, "A density-based approach for mining movement patterns from semantic trajectories," in *TENCON 2015-2015 IEEE Region 10 Conference*. IEEE, 2015, pp. 1–6.

[45] V. M. Prieto, S. Matos, M. Alvarez, F. Cacheda, and J. L. Oliveira, "Twitter: a good place to detect health conditions," *PloS one*, vol. 9, no. 1, p. e86191, 2014.

[46] S. Wojcik and A. Hughes, "Sizing up twitter users," *Pew Research Center*, vol. 24, 2019.

[47] L. Peel, D. B. Larremore, and A. Clauset, "The ground truth about metadata and community detection in networks," *Science advances*, vol. 3, no. 5, p. e1602548, 2017.

[48] J. Majmudar and S. Vavasis, "Provable overlapping community detection in weighted graphs," *arXiv preprint arXiv:2004.07150*, 2020.

[49] B. Rozemberczki, R. Davies, R. Sarkar, and C. Sutton, "Gemsec: Graph embedding with self clustering," in *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*, 2019, pp. 65–72.

[50] J. Kim and M. Hastak, "Social network analysis: Characteristics of online social networks after a disaster," *International Journal of Information Management*, vol. 38, no. 1, pp. 86–96, 2018.

[51] J. K. Kavota, J. R. K. Kamdjoug, and S. F. Wamba, "Social media and disaster management: Case of the north and south kivu regions in the democratic republic of the congo," *International Journal of Information Management*, vol. 52, p. 102068, 2020.

[52] C. Comito, A. Forestiero, and C. Pizzuti, "Word embedding based clustering to detect topics in social media," in *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 2019, pp. 192–199.

[53] L. Chen, H. Zhang, J. M. Jose, H. Yu, Y. Moshfeghi, and P. Triantafillou, "Topic detection and tracking on heterogeneous information," *Journal of Intelligent Information Systems*, vol. 51, no. 1, pp. 115–137, 2018.

[54] H.-J. Choi and C. H. Park, "Emerging topic detection in twitter stream based on high utility pattern mining," *Expert Systems with Applications*, vol. 115, pp. 27–36, 2019.

[55] Z. Wang and X. Ye, "Social media analytics for natural disaster management," *International Journal of Geographical Information Science*, vol. 32, no. 1, pp. 49–72, 2018.

[56] I. Laina, C. Rupprecht, and N. Navab, "Towards unsupervised image captioning with shared multimodal embeddings," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7414–7424.

[57] A. Conneau, G. Lample, M. Ranzato, L. Denoyer, and H. Jégou, "Word translation without parallel data," *arXiv preprint arXiv:1710.04087*, 2017.

[58] R. Jiang, Q. Chen, Z. Cai, Z. Fan, X. Song, K. Tsubouchi, and R. Shibasaki, "Will you go where you search? a deep learning framework for estimating user search-and-go behavior," *Neurocomputing*, 2020.

[59] Q. Chen, R. Jiang, C. Yang, Z. Cai, Z. Fan, K. Tsubouchi, R. Shibasaki, and X. Song, "Dualsin: Dual sequential interaction network for human intentional mobility prediction," in *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, 2020, pp. 283–292.

[60] J. J. Lin and D. V. Ryaboy, "Scaling big data mining infrastructure: the twitter experience," *SIGKDD Explor.*, vol. 14, pp. 6–19, 2013.

[61] S. Chintapalli, D. Dagit, B. Evans, R. Farivar, T. Graves, M. Holderbaugh, Z. Liu, K. Nusbaum, K. Patil, B. J. Peng *et al.*, "Benchmarking streaming computation engines: Storm, flink and spark streaming," in *2016 IEEE international parallel and distributed processing symposium workshops (IPDPSW)*. IEEE, 2016, pp. 1789–1792.

[62] Y. Kang, S. Gao, Y. Liang, M. Li, J. Rao, and J. Kruse, "Multiscale dynamic human mobility flow dataset in the us during the covid-19 epidemic," *Scientific data*, vol. 7, no. 1, pp. 1–13, 2020.

[63] T. Schmidt, J. Kerr, Y. Kestens, and J. Schipperijn, "Challenges in using wearable gps devices in low-income older adults: Can map-based interviews help with assessments of mobility?" *Translational behavioral medicine*, vol. 9, no. 1, pp. 99–109, 2019.

[64] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, pp. 1–54, 2019.

[65] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, "A survey on addressing high-class imbalance in big data," *Journal of Big Data*, vol. 5, no. 1, pp. 1–30, 2018.