2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

SOCIAL MEDIA PUBLIC RESPONSE TO CRITICAL INFRASTRUCTURE CYBER DISRUPTIONS: A REDDIT BASED MULTI-DIMENSIONAL SENTIMENT ANALYSIS

Bharadwaj R. K. Mantha[1,2], Balaji Kesavan[3], and Johan Ninan[4]
[1]University of Sharjah, Sharjah, United Arab Emirates
[2]Indian Institute of Technology Madras, Chennai, India
[3]Independent Freelancer, Beaverton, United States of America
[4] Delft University of Technology, Delft, The Netherlands

Abstract

This study examines social media public response and discourse patterns to critical infrastructure cyber disruptions through a reddit based multi-dimensional sentiment analysis of the Florida water treatment plant incident. Results from 300 most relevant posts indicate that discourse is highly time-sensitive (e.g., 90% of total engagement occurred within first six hours). Security concern emerged as the dominant sentiment (~50%), while humor and sarcasm (~26%) reflected social coping dynamics. The findings highlight how online community signals can inform crisis communication and public trust strategies, rather than technical cybersecurity management, to support more transparent, trust-oriented CI protection efforts.

Introduction

Critical Infrastructure (CI) constituting of facilities such as roads, bridges, power grids, nuclear facilities, water and wastewater distribution systems, and flood protection mechanisms are vital for any country's economy, security, and societal well-being. These are considered critical because of their functional significance, whose disruption could have severe consequences on public safety, health, and economic prosperity. For instance, the U.S. Department of Homeland Security (DHS, 2020) defines CI as "*the physical and cyber systems and assets so vital that their incapacitation or destruction would have a debilitating impact on our physical or economic security or public health or safety*". Similarly, the European Programme for Critical Infrastructure Protection (EPCIP) highlights the need for uniform protection levels, minimal single points of failure, and rapid recovery arrangements across the European Union (EC, 2006).

Traditionally, these systems and processes were operated, maintained, monitored, and inspected within their respective functional capabilities with little to no overlap among different CIs. For example, transportation network had no direct linkages on their operational synergies when compared to flood protection mechanisms. Technological advancements enabled synergistic, strategic, and efficient integration of different CIs. The growing interdependencies and technological integrations among CIs amplify their complexity, as disruptions in one sector can cascade into others, exacerbating the impact of failures (Mantha et al., 2024). Moreover, the increasing reliance on cyber-physical integration for real-time monitoring and control, where physical systems are interconnected with cyber systems, introduces vulnerabilities that make these infrastructures susceptible to cyberattacks. In an era of human-centred digital transformation, social media platforms serve as a crucial medium for connecting end users with policymakers, offering a dynamic space to share concerns, disseminate information, and respond to crises in real time. Such vulnerabilities can compromise the physical systems they govern, potentially leading to catastrophic failures (Palleti et al., 2021). This study critically examines these vulnerabilities of CIs to cyberattacks and explores the implications to ensure the resilience of such essential systems. Understanding public responses through social media can provide actionable insights that incorporate public feedback, fostering a more adaptive and resilient approach to managing cyber crises. Cyberattacks on these CI assets pose a severe threat to physical systems, potentially rendering them inoperable, delegating control to unauthorized entities, or compromising sensitive data (Pacheco et al., 2019). The interconnected nature of CI, while essential for meeting performance and design specifications, exacerbates these risks. Interdependencies among infrastructures mean that faults or attacks on one CI can cascade into others, leading to complex and unforeseen disruptions (Rinaldi et al., 2002). Such escalation can disrupt operations, initiate feedback loops, and amplify disturbances across interconnected systems (Palletti et al., 2021).

Despite significant advancements in securing networked control systems, existing technologies remain inadequate to address the fast-evolving threat landscape (Stellios et al., 2018). Furthermore, as technology advances, malicious actors continue to exploit vulnerabilities, using increasingly sophisticated tools to compromise systems (Tounsi and Rais, 2018). These disruptions inconvenience the public, disrupting daily life and eroding trust in essential services. Efforts to address these cybersecurity challenges vary internationally. For instance, the European Union Agency for Cybersecurity (ENISA) promotes a collaborative, multi-stakeholder approach to securing CI, with an emphasis on public-private partnerships and citizen engagement (Bossong and Wagner, 2017). In contrast, countries such as Japan and South Korea emphasize top-down regulatory frameworks with centralized response mechanisms (Aggarwal and

Reddie, 2018). These diverse models of policy and engagement reflect contextual differences in governance, digital infrastructure maturity, and cultural attitudes toward public trust in technology (Krishna et al., 2023). However, across these diverse international approaches it is important to understand community responses to cyberattacks necessitating the need for this research.

Social media analysis can serve as an early detection mechanism, offering valuable insights into public sentiment, concerns, and evolving perceptions of cybersecurity incidents, thus playing a pivotal role in shaping adaptive cybersecurity frameworks within the architecture engineering construction and facility management (AECFM) industry. In this context, understanding public reactions to cyberattacks can offer valuable insights for enhancing cybersecurity strategies and future infrastructure resilience. This study seeks to answer two research questions, 1) How does the public respond on social media during a cyberattack on critical infrastructure assets, and what patterns emerge in their engagement? 2) How can insights from non-expert public responses inform crisis communication and public trust strategies for future infrastructure projects?

To address these research questions, literature review section critically explores existing research, methodology section details the research methodology and case selection strategy, in which a case study approach is used focusing on the Florida water plant hack (CISA, 2021). Specifically, this case study was chosen due to its public involvement and reaction which is critical in the context of this study. Following this, the results and discussion section discusses the findings on how people react to the identified cyberattack. Finally, the conclusions and future work section details the contributions to theory, practice, study limitations and future research directions.

Literature Review

Cybersecurity Challenges of Critical Infrastructure

Critical infrastructure (CI) encompasses systems and functions essential for society's uninterrupted functioning, such as energy grids, transportation networks, water supply systems, and smart city technologies (Lehto, 2022). In recent years, cyberattacks targeting CI and critical information infrastructures have become increasingly frequent, complex, and targeted, as attackers grow more professional and sophisticated (Tounsi and Rais, 2018). These cyberattacks can disrupt or damage physical infrastructure by infiltrating digital systems controlling critical processes, leading to catastrophic consequences without requiring a physical assault. Such attacks may not only damage physical systems but also pollute environments, disrupt vital services, and compromise public safety.

The evolving nature of cyber threats compounds the challenges of defending against these attacks. Identifying the source, motive, and trajectory of an attack is notoriously difficult due to the global scope of cyberspace and blurred boundaries between national, international, public, and private interests (Lehto, 2013). Rapid technological advancements further complicate this landscape (Garcia de Soto et al., 2022), making it difficult to anticipate and mitigate new forms of threats. While existing literature emphasizes intrusion detection and prevention methods (Baykara and Das, 2018), these measures are often inadequate to address the dynamic and pervasive nature of modern cyber threats.

Human factors play a significant role in cyberattacks, often serving as the weakest link in cybersecurity. Reports indicate that employee errors account for a majority of data breaches, frequently exploited through social engineering tactics (Romanosky, 2016). At the same time, attackers leverage sophisticated techniques to exploit cyber-physical systems in critical sectors such as energy, water, and transportation, causing disruptions with far-reaching implications for public safety, economic stability, and trust (Stellios et al., 2018). For instance, cyberattacks have the potential to contaminate water supplies, disrupt electricity grids, or compromise transportation systems, highlighting the severe societal impacts of such crises (George et al., 2024).

Technical vulnerabilities in CI systems commonly stemming from legacy hardware and software, insufficient authentication protocols, and poor segmentation between IT and operational technology networks provide additional context for public responses (Sonkor and Garcia de Soto, 2021). Systems such as programmable logic controllers, SCADA platforms, and human–machine interfaces often lack robust cybersecurity measures because they were not originally designed with digital threats in mind (Stellios et al., 2018). These technical weaknesses can be considered important contextual factors affecting public perception and emotional response. Thus, the growing overlap of technical and social vulnerabilities, increased attack frequency, and severity of cyber incidents necessitates framing cyberattacks on CI as socio-technical crises.

Cyber Attacks as Socio-Technical Crises

Crisis, as a concept, encapsulates human reactions to significant disruptions or misfortunes, ranging from personal losses to large-scale societal events. Multiple frameworks help understand crises, including the Protective Action Decision Model (Heath et al., 2018), sensemaking theory (Maitlis and Christianson, 2014), and Situational Crisis Communication Theory (Coombs, 2007). Prior work has applied structured emotional trajectory models to interpret public reactions during disruptive events. While qualitative interpretations of emotional evolution remain useful, recent research increasingly treats crises as socio-technical phenomena observable through large-scale digital traces and measurable public discourse patterns.

The growing availability of social media data has enabled quantitative analysis of public responses to crises using natural language processing (NLP), network analysis, and engagement metrics. Social media platforms such as Reddit function as real-time participatory environments where users interpret events, assign accountability, and construct risk narratives during unfolding crises.

Comparative platform studies show that different social media environments generate distinct thematic and emotional patterns due to their structural affordances and communication norms (Kwon and Park, 2023). Reddit supports longer-form, focused and threaded discussions that enable users to express nuanced opinions and collectively interpret uncertain or controversial events, like group discussions and focused interviews. Quantitative social media research demonstrates that online discussions can reveal hidden facets of societal reactions, identify topic-aligned communities, and track the evolution of narratives (Goel and Sharma, 2020).

Network-based and linguistic analyses further show that language characteristics and engagement patterns influence how cohesive online communities form and how discourse evolves over time (Arazzi et al., 2023). Prior studies indicate that discussions of vulnerabilities and threats often emerge earlier or with deeper engagement on Reddit compared with other platforms, highlighting its potential as an early indicator of public awareness and concern (Horawalavithana et al., 2019).

These advances position social media as both a communication channel and a measurable socio-technical sensor for crisis dynamics. Quantitative analyses using sentiment scoring, topic modelling, and engagement metrics allow researchers to capture temporal shifts in public perception and identify dominant narratives surrounding risk, trust, and accountability. In addition, systematic reviews of Reddit-based research emphasize the importance of considering platform-specific norms, subreddit cultures, and ethical considerations when interpreting findings (Proferes et al., 2021).

Within this study, cyberattacks on CI are conceptualized as socio-technical crises that unfold through both technical disruption and public interpretation across decentralized online communities. Quantitative, thread-level analysis of Reddit discussions provides a means to observe how users collectively interpret infrastructure vulnerabilities, assign responsibility, and express trust or concern in real time. This framing integrates crisis theory with computational social media analytics, enabling the measurement of public sentiment trajectories and engagement patterns during cyber incidents.

Research Gaps and Contributions

The following are the three main research gaps based on the aforementioned discussion.

1)Human trust dynamics gap: There is limited evidence-based understanding of how decentralized online communities shape public trust, risk perception, and accountability narratives following cyberattacks on CI.

2)Methodological gap: Most existing studies focus on technical threat detection or rely on single-platform, polarity-based sentiment analyses, offering only partial insights into how public discourse and crisis narratives evolve in real time.

3)Analytical gap: Quantitative social media research has largely emphasized topic detection, vulnerability mentions, or information diffusion patterns, with limited use of multi-dimensional sentiment, engagement-aware metrics, and thread-level analysis to capture nuanced public reactions.

This study therefore contributes to advancing empirical understanding of public response dynamics during CI cyber disruptions by treating social media discourse as a measurable feedback signal for cybersecurity communication and resilience. The aims are to: (1) quantify key sentiment distributions in early-stage discussions, (2) examine thread-level and engagement dynamics shaping trust and risk narratives through the lens of Situational Crisis Communication Theory (SCCT), and (3) identify dominant discourse patterns that can inform more transparent and trust-oriented cybersecurity communication.

Methodology

The following interconnected subsections provide a coherent approach for quantifying public response patterns and addressing the study's objectives.

Research Design and Case Context

This study adopts a quantitative social media analysis approach to examine public responses to a CI cyber disruption through Reddit discourse. Cyberattacks on infrastructure systems often generate rapid and emotionally varied public reactions that unfold in real time across digital platforms. Social media platforms, particularly Reddit, provide a valuable data source for capturing these reactions because of their threaded discussion structure, relative anonymity, and capacity for detailed user commentary (Proferes et al., 2021). Unlike survey-based approaches that capture retrospective perceptions, Reddit discussions enable observation of emergent public sentiment during the early stages of a crisis, making them suitable for examining evolving trust, accountability, and risk narratives.

Building on prior qualitative work that examined emotional trajectories in cyber crisis discourse, this study extends the analysis through a quantitative, thread-level, multi-dimensional sentiment framework. The goal is to empirically capture how decentralized online communities interpret and respond to cyber disruptions affecting CI, with a particular focus on measurable sentiment patterns and engagement dynamics. While this study relies on a single-incident design, this approach aligns with recent sentiment literature that leverages single-case events to establish deep, thread-level analytical baselines (Goel & Sharma, 2020; Kahlawi et al., 2025; Ninan et al., 2025). Furthermore, it has to be noted that Reddit discussions enable the observation of emergent public sentiment during the early stages of a crisis as a first step, making them suitable for examining evolving trust, accountability, and risk narratives.

Data Source and Collection

Reddit was selected as the primary data source because its subreddit-based structure facilitates focused, topic-specific discussions and supports longer-form discourse compared with microblogging platforms. Previous

research shows that Reddit often hosts deeper and more sustained conversations on cybersecurity and infrastructure-related issues, making it a suitable platform for capturing nuanced public reactions (Horawalavithana et al., 2019). Furthermore, Reddit's threaded architecture enables analysis of conversational dynamics across original posts, comments, and replies.

Data were collected from relevant public subreddits discussing the selected cyber incident within the first week following the event. Figure 1 shows the pseudo code developed to achieve the same. The collection window was chosen to capture early public reactions, which are often critical in shaping trust and risk perception narratives during crises. Prior research suggests that the initial hours and days following a disruptive event are particularly influential in establishing dominant public narratives (Maitlis & Christianson, 2014).

The final reviewed, filtered, and cleaned dataset consisted of 300 posts spanned across 15 total subreddits, 20 original posts (OP), 66 comments, 214 replies. Data cleaning involved removing duplicate posts, irrelevant discussions, spam, and posts lacking substantive textual content. Only posts directly related to the incident and public reaction were retained. All data were drawn from publicly accessible Reddit threads and anonymized prior to analysis.

```
BEGIN PROCEDURE
  Load structured CSV containing Reddit posts and comments

  FOR EACH post/comment in dataset:
    Build thread context (chain parent comments for replies)

    Analyze relevance using OpenAI gpt-5.2-2025-12-11
      // Outputs: Relevant_to_attack, Relevant_to_thread + explanations

    Analyze sentiment tags using OpenAI gpt-5.2-2025-12-11
      // Outputs: 20 boolean tags + explanation
      // Tags: Denial, Anger, Bargaining, Acceptance, Anxiety, etc.

    Record results (relevance + tags + explanations) to dataset
  END FOR

  Save enhanced CSV with analysis results
END PROCEDURE
```

Figure 1: Pseudo code developed to pull the Redditt Data

Unit of Analysis and Data Preparation

The unit of analysis in this study is the individual Reddit post, defined as any original post (OP), comment, or reply within a discussion thread. This approach allows for capturing both primary narratives and conversational responses, providing a more comprehensive understanding of discourse dynamics. Previous research highlights that analyzing threaded discussions rather than isolated posts yields richer insights into collective sensemaking processes during crises (Kahlawi et al., 2025).

Each post was assigned a unique identifier and categorized by type (OP, comment, reply), subreddit, and time of posting. Basic engagement metrics including upvotes and interaction structure were recorded to support engagement-weighted analysis. Textual data were preprocessed by removing URLs, emojis where necessary for clarity, and non-relevant formatting while preserving semantic meaning. Posts were retained in their original wording to maintain contextual integrity for sentiment tagging.

Multi-Dimensional Sentiment Framework

To capture the emotional and interpretive dimensions of public response, this study applies a multi-dimensional sentiment framework consisting of eight categories namely 1) Denial, 2) Anger, 3) Bargaining, 4) Depression, 5) Acceptance, 6) Security Concern, 7) Sarcasm, and 8) Humor. Five of these categories (i.e., 1 to 5) were derived from the Kübler-Ross model of emotional response to crisis (Kübler-Ross, 1969). This model has been widely applied beyond grief contexts to understand reactions to disruptive societal events and organizational crises (Ninan et al., 2025). Its structured progression provides a useful lens for examining how individuals process shock, assign blame, and eventually move toward adaptation following a disruptive event.

However, applying the Kübler-Ross model alone is insufficient for analyzing social media discourse. To address this, these methods were calibrated for a sentiment analysis of a crisis response which augments the baseline emotional framework with interpretive and coping mechanisms such as humor and sarcasm, which function as social responses to uncertainty and stress. Including humor, sarcasm, and security concern allows for capturing these communicative behaviors that may not fit neatly within traditional emotional stages, making the framework defensible for digital discourse. Humor has been shown to act as a psychological coping strategy during crises, enabling individuals to process fear and anxiety collectively (Martin and Ford, 2018). Similarly, sarcasm is frequently used in online discourse to express skepticism, critique institutions, or signal distrust (Filatova, 2012). Including humor and sarcasm as distinct categories allows for capturing these communicative behaviors that may not fit neatly within traditional emotional stages. The category security concern was added to explicitly capture expressions of risk awareness, vulnerability perception, and calls for cybersecurity improvements. Prior research indicates that public discourse following cyber incidents often centers on perceived system weaknesses and demands for accountability, making this a distinct and analytically relevant category (Tounsi & Rais, 2018). Together, these eight categories provide a comprehensive framework for analyzing both emotional and interpretive dimensions of public response.

Sentiment Tagging Procedure

Sentiment tagging was conducted using a large language model (LLM) assisted classification process supported by manual verification through the algorithm logic represented as pseudocode in Figure 2. Each post was assigned sentiment categories based on its tone and intent. Where posts contained multiple sentiments, the most salient sentiment was selected to maintain consistency, but the others were preserved and allocated in quantitative aggregation. To improve tagging reliability, a subset of posts (e.g., 10% randomly sampled) were manually reviewed to confirm alignment between automated classification and contextual interpretation. While identifying sarcasm computationally has historically been

a significant challenge for traditional algorithms, modern Large Language Models have demonstrated high accuracy in emulating human experts' evaluation of complex public sentiments on social media (Kim & Kim, 2025). Furthermore, recent NLP literature confirms that advanced LLMs are highly reliable in detecting contextual sarcasm and irony in digital discourse. Consequently, verification threshold is also sufficient to confirm alignment. This hybrid approach effectively combines computational efficiency with human interpretive oversight, which is the recommended standard in current social media sentiment analysis research (Kim & Kim, 2025). Tagged data were then further analyzed to identify overall distribution patterns and dominant narrative themes

To improve tagging reliability, a subset of posts (e.g., about 10% randomly sampled) were manually reviewed to confirm alignment between automated classification and contextual interpretation. This hybrid approach combines computational efficiency with human interpretive oversight, which is recommended in social media sentiment analysis research (Kim & Kim, 2025). Tagged data were then further analyzed to identify overall distribution patterns and dominant narrative themes.

Engagement Metrics and Quantitative Analysis

Quantitative analysis focused on both frequency based and engagement-weighted metrics. Frequency counts were used to determine the proportion of posts associated with each sentiment category. Engagement-weighted metrics considered upvotes and reply structures to account for the visibility and influence of posts within discussions (Fottner et al., 2022). Engagement weighting is important because highly endorsed posts often shape perceived consensus within online communities (Arazzi et al., 2023). By examining both raw counts and engagement-adjusted measures, the analysis captures not only what sentiments are expressed but also which sentiments gain traction within the community. Descriptive statistics were used to summarize sentiment distributions, while cross tabulations examined variations across post types such as OP and comments.

Temporal and Thread-Level Analysis

Temporal analysis examined how sentiment distribution evolved during the initial response window. Posts were grouped by time intervals to identify early narrative formation and shifts in dominant sentiments. Previous research suggests that early discourse patterns can strongly influence long-term public perception of crises (Coombs, 2007). Thread-level analysis examined how sentiments propagated within discussions. This included identifying whether certain sentiments were more likely to appear in original posts versus replies, and whether humor or sarcasm functioned as responses to expressions of concern or anger.

Reliability and Ethical Considerations

As discussed earlier, to ensure analytical consistency, a subset of posts were manually reviewed to validate sentiment classifications. Tagging procedures and aggregation methods were documented to support reproducibility. Results were compared using both raw and engagement-weighted counts to compare and contrast stability of findings. All data analyzed were publicly available Reddit posts. Usernames were not recorded or reported, and findings are presented in aggregate form. No interaction with users occurred, and the study adheres to standard ethical practices for research using publicly accessible social media data (Proferes et al., 2021).

```
BEGIN PROCEDURE
  Load structured CSV containing Reddit posts and comments

  FOR EACH post/comment in dataset:
    Build thread context (chain parent comments for replies)

    Analyze relevance using OpenAI gpt-5.2-2025-12-11
      // Outputs: Relevant_to_attack, Relevant_to_thread + explanations

    Analyze sentiment tags using OpenAI gpt-5.2-2025-12-11
      // Outputs: 20 boolean tags + explanation
      // Tags: Denial, Anger, Bargaining, Acceptance, Anxiety, etc.

    Record results (relevance + tags + explanations) to dataset
  END FOR

  Save enhanced CSV with analysis results
END PROCEDURE
```

Figure 2: Pseudo code developed to conduct multi-dimensional sentiment tagging and analysis

Results and Discussion

This section presents the quantitative findings specifically, sentiment distribution and engagement patterns summarized in first, while thread-level dynamics and interpretation in the second. Lastly, implications and limitations connecting the results to broader cybersecurity and resilience considerations are detailed.

Sentiment Distribution and Engagement Patterns

The final filtered most relevant dataset comprised 300 posts across 15 subreddits within the early-response window following the cyber incident, including 20 OPs (6.7%), 66 comments (i.e., response made to an OP) (22%), and 214 replies (i.e., response made to a comment) (71.3%). The initial extraction yielded a larger candidate set, including 30 OPs after automated relevance tagging and manual contextual verification, this was refined to 20 original posts retained for analysis. A similar two-stage filtering process was applied to comments and replies, where non-relevant or context-detached items were removed. This hybrid human-in-the-loop screening ensured that only posts directly tied to the incident and its interpretation were included in the quantitative analysis.

Across the dataset, as can be seen in Figure 3, security concern appeared in almost half (50%) of all posts, emerging as the dominant sentiment category and reflecting widespread public attention to infrastructure vulnerability, system reliability, and accountability. Expressions of Anger (~20–25%) and Denial (<5%) were also present, particularly in original posts where users reacted to initial reports or questioned the severity and legitimacy of the incident. At the same time, Humor (~20%) and Sarcasm (~25–30%) appeared consistently across comments and replies, suggesting the use of coping and critique mechanisms common in online discourse during disruptive events. The remaining Kübler-Ross–

derived categories namely bargaining, depression, and acceptance (<15%) appeared in smaller but notable proportions, indicating a gradual shift in tone as discussions progressed from initial shock and skepticism toward reflection and normalization.

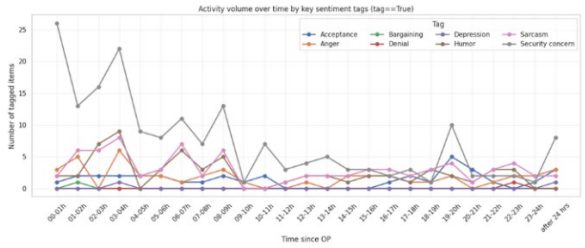

Figure 3: Temporal distribution of activity across different sentiments

Figure 4 reveals from the engagement patterns that a large proportion of total interactions occurred within the earliest hours of discussion. Nearly 50% posts were generated within the first six hours, while approximately 90% of cumulative engagement score was concentrated in this same early period, underscoring the rapid formation of public narratives. High-engagement posts were not always those expressing the most frequent sentiment, rather, posts combining concern with critique or humor often attracted the most interaction. The highest average endorsement per item occurred within the first hour after posting, suggesting that early contributions disproportionately shaped perceived credibility and importance. While original posts frequently introduced concerns or questions about the incident, replies tended to amplify or reinterpret these narratives, often through humor or sarcasm. This distribution highlights how public reaction to cyber disruptions unfolds through layered conversational dynamics rather than isolated expressions.

The analytical workflow used to generate relevance and sentiment tags involved automated LLM-assisted tagging followed by targeted human validation across sampled posts, ensuring alignment between computational classification and contextual interpretation. This process confirmed consistency between automated outputs and manual review, supporting the reliability of the final curated dataset and subsequent quantitative aggregation.

Thread-Level Dynamics and Interpretation

Thread-level analysis as shown in Figures 5 and 6 demonstrates that Reddit discussions function as decentralized spaces for collective sensemaking during infrastructure-related crises. For example, left to right represented from 0 (OP) to 1 (comment) to 2 (replies) and so on. Initial posts typically framed the event in terms of risk and vulnerability while comments and replies elaborated on these themes through personal interpretations, technical speculation, or institutional critique. For instance, security concern score – grey dropped from 73% to 47% while anger – orange increased to 23% and other sentiments peaked in the latter depths of discussions. Notably, denial is the least with about 1.6% which corroborates what was observed through the anger sentiment statistics observed. Activity volume peaked at early reply layers with most interaction occurring within the first two levels of response (layers 0,1 score > 1600 while layer 4 score ~ 400), indicating that collective interpretation emerges quickly once discussions begin. In several threads, expressions of concern were followed by humorous or sarcastic remarks that reframed the incident in more accessible or socially resonant terms (e.g., also evident from color tones in layers 1 and more but nonexistent in 0). Such patterns suggest that humor and sarcasm operate not merely as distractions but as mechanisms through which users negotiate uncertainty and express skepticism or distrust.

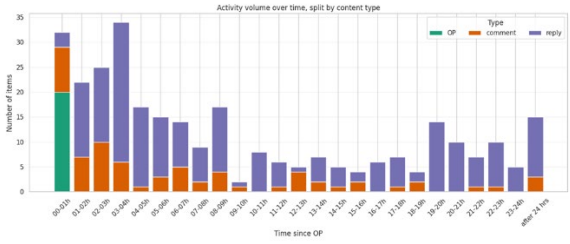

Figure 4: Distribution of activity type and volume over time

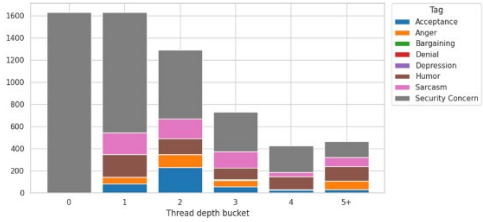

Figure 5: Thread depth weighted scores across sentiments

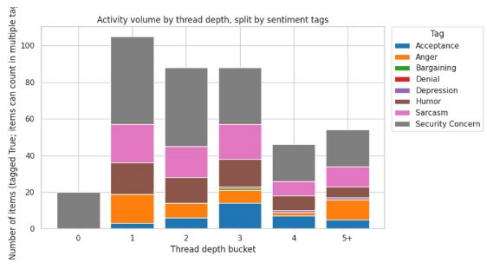

Figure 6: Thread depth activity across sentiments

The presence of Denial (rare at deeper levels) and Bargaining (moderate mid-thread presence) sentiments indicates attempts by some users to minimize perceived risk or reinterpret the incident's implications. Over time, however, these sentiments often gave way to Acceptance-oriented content (blue) in later discussion layers, reflecting normalization of systemic vulnerability discussions. This progression reflects the dynamic, non-linear nature of public response to disruptive events and supports the use of a multi-dimensional sentiment framework that captures both emotional and interpretive reactions.

From an analytical standpoint, these findings address previously identified gaps in understanding how decentralized online communities shape trust and risk narratives following cyber incidents. As guided by Situational Crisis Communication Theory (SCCT), the non-expert nature of this data is not a limitation but the

focal point of the study. The sample reflects public speculation, uninformed opinion, and casual commentary, which provides a direct, measurable lens into how sentiment evolves and which narratives gain traction to help agencies tailor their messaging and manage trust. The results demonstrate that public discourse is not only measurable but also structured by engagement dynamics that influence which narratives gain prominence during a crisis response. The iterative reduction from the initial candidate dataset including the refinement from 30 to 20 original posts and corresponding filtering across comments and replies also highlights the importance of combining automated tagging with human contextual verification when interpreting crisis-related social media discourse. By examining thread-level interactions rather than isolated posts, the analysis highlights how trust, skepticism, and concern are negotiated collectively within online communities. This supports the view of social media platforms as socio-technical environments where public perception and crisis interpretation co-evolve.

Implications and Study Limitations

The findings suggest that early-stage social media discourse can serve as a valuable feedback signal for cybersecurity communication and infrastructure governance, particularly within the architecture, engineering, construction, and facility management (AECFM) industry. The prominence of security concern across posts indicates strong public awareness of infrastructure vulnerabilities, while the prevalence of humor and sarcasm highlights the need for communication strategies that acknowledge and address public skepticism. Monitoring sentiment distributions and engagement patterns may help agencies and infrastructure developers identify emerging trust deficits and adjust messaging accordingly, aligning closely with SCCT principles. For AECFM practitioners, while based on a single-incident design, this approach offers an essential first step for integrating public perception into adaptive communication and lifecycle management strategies, thereby supporting more transparent and trust-oriented responses to cyber disruptions.

Several limitations should be acknowledged. First, the analysis focuses on a single incident within a limited early-response window, which may restrict the generalizability of findings. Second, Reddit users represent a specific subset of the broader population, and their discourse may not reflect wider public opinion. Third, while the multi-dimensional sentiment framework captures a range of emotional and interpretive responses, categorizing complex posts into single dominant sentiments may overlook nuance. Finally, LLM-assisted tagging introduces potential classification variability despite manual verification. These limitations highlight the need for broader comparative analysis across multiple incidents and longer timeframes. Expanding the dataset to include additional cyber events and cross-sector infrastructure contexts will enable more robust identification of recurring patterns and improve the generalizability of findings.

Conclusions and Ongoing Work

This study contributes to the existing body of knowledge by providing empirical, quantitative evidence on how decentralized online communities respond to CI cyber disruptions. By analyzing 300 Reddit posts across 15 subreddits, the study identifies dominant sentiment patterns, engagement dynamics, and thread-level interactions that shape public interpretations of cyber incidents. The findings show that security concern emerges as a primary narrative driver, while humor and sarcasm play significant roles in mediating public reaction and critique. These observations demonstrate how early-stage online discourse forms rapidly, concentrates engagement, and amplifies particular narratives that influence public perception. Such insights are relevant for cybersecurity agencies, infrastructure operators, policymakers, and specifically AEC professionals seeking to embed human-centered resilience into future infrastructure projects. By addressing AEC infrastructure vulnerabilities directly, these findings provide actionable suggestions for the industry to proactively incorporate public feedback into lean-inspired cybersecurity frameworks and facility management crisis response strategies.

However, the study is subject to several limitations. It focuses on a single incident and a short early-response window, which may not capture longer-term discourse evolution or variations across sectors and regions. The dataset is limited to Reddit users, whose demographic and behavioral characteristics may differ from the broader population. Therefore, it is worth noting that presented analysis and discussion serves as an initial methodological first step rather than a generalized conclusion. To address these limitations, ongoing work is expanding the dataset to include multiple cyber incidents across different infrastructure sectors and geographic contexts to validate these pattern claims and improve generalizability. This extended analysis will support the development of more robust, evidence-based frameworks for integrating social media feedback into cybersecurity communication and resilience planning, ultimately contributing to more adaptive and human-centred approaches to infrastructure protection.

References

Aggarwal, V. K., and Reddie, A. W. (2018). Comparative industrial policy and cybersecurity: a framework for analysis. Journal of Cyber Policy, 3(3), 291-305.

Arazzi, M., Nicolazzo, S., Nocera, A., and Zippo, M. (2023). The importance of language for the evolution of online communities. ESA, 222, 119847.

Baykara, M., and Das, R. (2018). A novel honeypot based security approach for real-time intrusion detection and prevention systems. JISA, 41, 103-116.

Bossong, R., and Wagner, B. (2017). A typology of cybersecurity and public-private partnerships in the context of the EU. Crime, Law and Social Change, 67, 265-288.

CISA (Cybersecurity and Infrastructure Security Agency) (2021). Compromise of U.S. Water Treatment Facility. Available at https://www.cisa.gov Accessed-2026/2/6.

Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. Corporate reputation review, 10(3), 163-176.

DHS (2020). Critical infrastructure security. U.S. Department of Homeland Security, www.dhs.gov/topic/critical-infrastructure-security

EC (2006). On a European programme for critical infrastructure protection. Communication from the Commission, COM(2006) 786 final, EC.

Filatova, E. (2012). Irony and sarcasm: Corpus generation and analysis using crowdsourcing, LREC.

Fottner A, Okhrin Y, Pfahler J, Wustl J. Reddit financial image post sentiment dataset. Data Brief. 2022 Nov 17;45:108759. doi: 10.1016/j.dib.2022.108759.

Garcia de Soto, B., Georgescu, A., Mantha, B., Turk, Z., Maciel, A., and Sonkor, M. S. (2022). Construction cybersecurity and critical infrastructure protection: new horizons for Construction 4.0. J. Inf. Technol. Constr., 27, 571-594.

George, A. S., Baskar, T., and Srikaanth, P. B. (2024). Cyber threats to critical infrastructure: assessing vulnerabilities across key sectors. Partners Universal International Innovation Journal, 2(1), 51-75, DOI:10.5281/zenodo.10639463.

Goel, R., & Sharma, R. (2020). Understanding the MeToo movement through the lens of Twitter. In S. Aref et al. (Eds.), Social Informatics (Lecture Notes in Computer Science, Vol. 12467, pp. 67–80). Springer.

Heath, R. L., Lee, J., Palenchar, M. J., and Lemon, L. L. (2018). Risk communication emergency response preparedness: Contextual assessment of the protective action decision model. Risk analysis, 38(2), 333-344.

Horawalavithana, S., Bhattacharjee, A., Liu, R., Choudhury, N., Hall, L. O., & Iamnitchi, A. (2019). Mentions of security vulnerabilities on Reddit, Twitter and GitHub. WI ’19 Proceedings.

Kahlawi, A., Masri, F., Ahmed, W., & Vidal-Alaball, J. (2025). Cross-cultural sense-making of global health crises: a text mining study of public opinions on social media related to the COVID-19 pandemic in developed and developing economies. JMIR, 27, e58656.

Kim, K., & Kim, S. (2025). Large language models’ accuracy in emulating human experts’ evaluation of public sentiments about heated tobacco products on social media. JMIR, https://doi.org/10.2196/63631.

Krishna, B., Krishnan, S., and Sebastian, M. P. (2023). Examining the relationship between national cybersecurity commitment, culture, and digital payment usage: an institutional trust theory perspective. ISF, 25(5), 1713-1741.

Kübler-Ross, E. (1969). On death and dying. Macmillan.

Kwon, S., & Park, A. (2023). Examining thematic and emotional differences across Twitter, Reddit, and YouTube: The case of COVID-19 vaccine side effects. Computers in Human Behavior, 144, 107734.

Lehto M. (2013). The cyberspace threats and cyber security objectives in the cyber security strategies. IJCWT, 3(3):1-18.

Lehto, M. (2022). Cyber-attacks against critical infrastructure. In Cyber security: Critical infrastructure protection (pp. 3-42). Cham: Springer.

Maitlis, S., and Christianson, M. (2014). Sensemaking in organizations: Taking stock and moving forward. Academy of management annals, 8(1), 57-125.

Mantha, B. R., Sonkor, M. S., and Garcia de Soto, B. (2024). Investigation of the cyber vulnerabilities of construction networks using an agent-based model. Developments in the Built Environment, 18, 100452.

Martin, Rod A., and Thomas Ford. 2nd Ed. (2018) The psychology of humor: An integrative approach. Academic press, ISBN-13: 978-0128121436

Ninan J, Mantha BRK, Kesavan B (2025), "Human-centred cybersecurity for critical infrastructure: the case of the Florida water plant hack" ECAM, 32(13).

Pacheco, J., Benitez, V. H., and Pan, Z. (2019). Security framework for IoT end nodes with neural networks. IJMLC, 9(4), 381-386.

Palleti, V. R., Adepu, S., Mishra, V. K., and Mathur, A. (2021). Cascading effects of cyber-attacks on interconnected critical infrastructure. Cybersecurity, 4, 1-19.

Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit. Social Media + Society, 7(2).

Rinaldi, S. M., Peerenboom, J. P., and Kelly, T. K. (2001). Identifying, understanding, and analyzing critical infrastructure interdependencies. IEEE CSM,21(6),11-25.

Romanosky, S. (2016). Examining the costs and causes of cyber incidents. Jrn. of Cybersecurity, 2(2), 121-135.

Sonkor, M. S., and García de Soto, B. (2021). Operational technology on construction sites: A review from the cybersecurity perspective. JCEM, 147(12), 04021172.

Stellios, I., Kotzanikolaou, P., Psarakis, M., Alcaraz, C., and Lopez, J. (2018). A survey of iot-enabled cyberattacks: Assessing attack paths to critical infrastructures and services. IEEE-CST,20(4),3453-95.

Tounsi, W., and Rais, H. (2018). A survey on technical threat intelligence in the age of sophisticated cyber attacks. Computers and security, 72, 212-233.