Characterizing Social Movement Narratives in Online Communities: The 2021 Cuban Protests on Reddit

Brian Keith Norambuena*
briankeithn@vt.edu
Virginia Tech

Tanushree Mitra
tmitra@uw.edu
University of Washington

Chris North
north@vt.edu
Virginia Tech

ABSTRACT

Social movements are dominated by storytelling, as narratives play a key role in how communities involved in these movements shape their identities. Thus, recognizing the accepted narratives of different communities is central to understanding social movements. In this context, journalists face the challenge of making sense of these emerging narratives in social media when they seek to report social protests. Thus, they would benefit from support tools that allow them to identify and explore such narratives. In this work, we propose a narrative extraction algorithm from social media that incorporates the concept of community acceptance. Using our method, we study the 2021 Cuban protests and characterize five relevant communities. The extracted narratives differ in both structure and content across communities. Our work has implications in the study of social movements, intelligence analysis, computational journalism, and misinformation research.

CCS CONCEPTS

• **Applied computing** → *Publishing*; • **Computing methodologies** → **Information extraction**.

KEYWORDS

narrative extraction, online communities, narrative maps

1 INTRODUCTION

Social movements are dominated by storytelling [12]. In fact, narratives play a key role in social change and social movements [7], as people use them to learn and exercise agency, shape identity, and motivate actions [13, 17]. Social movements are made possible by the construction of collective narratives within free spaces [10]. In particular, online communities in social media provide free spaces for people to develop such narratives [27]. Thus, studying how narratives evolve in these online communities is key to understanding social movements.

Journalists face unique challenges when reporting social movements [31] and making sense of the emerging narratives in social media. In particular, the fluid nature and rapid evolution of social movements [25] makes it hard to keep track of its goals and motivations. In this context, identifying the driving narratives in the different communities involved in a social movement would aid journalists in their sensemaking process. The accepted narratives of a community provide a *cognitive roadmap* by which community members are to live and define the collective identity of the group and their relations with other groups [22, 23, 28], providing insight towards their motivations and goals.

In this work, we propose a narrative extraction algorithm that incorporates the concept of *community acceptance* to model and extract the accepted narratives from social media communities. Our proposed algorithm leverages the *narrative maps* approach of Keith and Mitra [24] to construct a complex narrative representation via optimization. The extracted narratives could provide journalists with an overview of the evolving narratives surrounding the social movement, giving them a window into these communities. In particular, we seek to answer the following research question:

- **RQ**: How can we model community acceptance in social media and extract the accepted narratives of a social media community?

To test our method, we study the start of the *2021 Cuban protests*—the biggest protests in decades in Cuba [33]—and the narratives surrounding this movement that emerged in online communities. Cuba has a unique sociopolitical, economic, and cultural landscape [14]. Its long-standing communist regime faces unique challenges [4]. The continued US embargo [40] has undermined the Cuban economy [21] and its health care system [2]. The protests were triggered by food and medicine shortages [16], and the inefficient handling of the COVID-19 pandemic by the government [15].

Our approach should be able to capture these relevant events and storylines, allowing journalists to easily identify key elements of the narratives espoused by different communities in social media. In particular, we focus on a specific social media platform—Reddit—to explore how narratives change across communities. Specifically, we characterize the accepted narratives in a number of communities with a high level of activity related to the 2021 Cuban protests (e.g., r/cuba and r/politics). Thus, we make the following contributions:

(1) A narrative **extraction algorithm** that models *community acceptance* in online communities, which could be used by journalists and other practitioners to explore social media narratives.

(2) An **analysis of the accepted narratives** espoused by online communities surrounding the 2021 Cuban protests to showcase the application of our method.

We believe our approach has widespread implications, both for the study of social movement narratives online and narratives in general. At its core, our methods allow users to understand the accepted narratives of online communities. Furthermore, expert users, such as intelligence analysts, journalists, and misinformation researchers could use our methods to explore the narrative landscape of social movements emerging in online communities.

2 RELATED WORK

Narratives are systems of stories with coherent themes [18], where each story is represented as a sequence of events [1]. Most narrative

---
*Also with Universidad Católica del Norte.

extraction methods use an event-based representation, where each event is associated with a single document in a data set [24, 34]. We follow this approach, as event-based representations have strong theoretical foundations in narratology [1]. Narrative extraction approaches usually work by optimizing different criteria [34, 39]. We use a narrative extraction algorithm based on the criteria of coherence maximization through linear programming, an extension of previous work in [24]. However, unlike previous models, we integrate the concept of *community acceptance* in our optimization model to select important events and storylines.

Narratives are key to the formation of social movements [7, 12]. In this context, social media plays a key role in shaping the narrative [20, 27] and mobilizing people [29]. Thus, several works have sought to model these narratives. Computational methods to study social movement narratives in social media usually rely on keyword and hashtag analysis [41], topic modeling [36], and user network analysis [38]. Other approaches include qualitative research, focusing on interviews and the experiences of people [11, 26]. However, none of these approaches consider the underlying structure of the narrative, which depends on temporal and causal relationships between events. In contrast, our approach extracts the narrative structure, allowing for a detailed model of the social movement narrative via narrative maps—an event-based directed acyclic graph narrative representation.

3 MATERIALS AND METHODS

Data Set Description. We used Pushshift [6] to extract Reddit data. In particular, we extracted all submissions that contained the word "Cuba" from 07/07/2021 to 07/17/2021, covering a few days before the protests started on July 11th and the first week of the protests, allowing us to capture any useful signals on social media prior to the start of protests and the early narratives that emerged just after they started. We retrieved 4810 submissions. We focused on the top five communities according to their number of posts (see Table 1). We stopped at five because the sixth community (r/FreeKarma4U) was a *karma farming* group [3] unrelated to the protests that did not have a cohesive ideology or group identity and the number of posts rapidly declined in all subsequent communities. We manually analyzed all the submissions to ensure that they were relevant to the protests, finding that roughly 95% out of 550 were relevant.

Community	Description	Posts	Subscribers
r/cuba	The biggest Cuban community in Reddit.	168	36k
r/Conservative	US-centric right-leaning community focused on political issues.	126	833k
r/GenZedong	Pro-communist community focused on humor and world news.	99	33k
r/worldnews	Community focused on world news (excluding US news).	79	26.6M
r/politics	US-centric left-leaning community focused on political issues.	78	7.6M

Table 1: Summary of the analyzed Reddit communities, including the number of subscribers and submissions retrieved.

Narrative Extraction Method. We use narrative maps [24] as our representation (see Figure 2). Narrative maps provide a generic foundation to encode different types of narratives extracted from data, requiring only the existence of a total ordering (e.g., timestamps) and text representation of the event (e.g., a news headline or post title). Based on a *route map* metaphor, they consist of a directed acyclic graph of event nodes, where each path represents a route connecting events. These nodes are called *landmarks*, based

on the idea of route maps having a series of landmarks to highlight important elements. Certain events are highlighted—called *representative landmarks*—at the point of highest width in the graph to capture all parallel storylines. Finally, by finding the path of maximum coherence we find the *main route* that contains the core events of the narrative. Narrative maps are extracted by optimizing coherence relations between events—based on event similarity and topical consistency—subject to coverage constraints.

In this work, we propose a new optimization model that accounts for *community acceptance* in addition to coherence and story coverage. We include new parameters to account for the **submission score** (i.e., the difference between upvotes and downvotes) and **upvote ratio** as a way to measure **community acceptance** of a submission. In particular, posts that have a high score and upvote ratio in a subreddit are considered accepted by that community [8, 30] and should appear in the narrative. We model this with a constraint that requires the average score percentile to be above a certain threshold. Note that this does not necessarily discard all the low score submissions, as their presence might be necessary to construct a coherent narrative. Furthermore, we incorporate **community acceptance** into the optimization objective itself, extracting maps by maximizing both coherence and acceptance.

We show the optimization model in Figure 1. Our goal is to maximize the minimum edge strength (minedge), which depends on coherence—a measure of how much sense it makes to join two events together, based on event similarity and topical similarity—and acceptance—a measure of community acceptance that depends on the upvote ratios and submission score in the community. Table 3 in the appendix presents the details of our model.

$$\max \quad \text{minedge}$$
$$\text{s.t.} \quad \text{minedge} \leq \text{coherence}_{ij} + \text{acceptance}_{ij} - \text{edge}_{ij} \quad \forall_{ij}$$
$$\sum_i \text{event}_i = K \qquad \text{edge}_{ij} = 0 \quad \forall_{i \geq j}$$
$$\sum_{ij} \text{edge}_{ij} = K - 1 \qquad \text{cluster}_k = \sum_{ij} \text{edge}_{ij} \cdot \text{membership}_{ijk} \quad \forall_k$$
$$\sum_{i<j} \text{edge}_{ij} \geq \text{event}_j \quad \forall_j \quad \text{mincover} \leq \frac{1}{\text{num-clusters}} \cdot \sum_k \text{cluster}_k$$
$$\sum_{j>i} \text{edge}_{ij} \leq \text{event}_i \quad \forall_i \quad \text{minscore} \leq \frac{1}{K} \cdot \sum_i \text{event}_i \cdot \text{score-percentile}_i.$$

Figure 1: Linear program formulation. Parts highlighted in red are our contributions to the model.

For the parameters, we set mincover = 0.5, ensuring that we cover at least half of the stories without excessively restricting the narrative. We set $K = 8$ to get main stories with around 8 events, based on [24, 34]. We set minscore = 0.85, ensuring that the average community acceptance is at least 85%—intuitively, this means that roughly 85% of the community regards the selected submissions positively. Lower thresholds allowed too many low-acceptance events to be on the map. We encode submissions with Multilingual USE embeddings [42], as some communities use English and Spanish. The rest of the implementation follows [24]. We note that all parameters can be easily modified, allowing journalists and other users to explore their influence on the narrative.

4 RESULTS

Now we present the accepted narratives extracted from each community. Due to space limits, we offer details of only one community

(r/cuba) while briefly mentioning the rest. We also analyze how the narratives espoused by each community differ from each other and manually categorize the relevant storylines and events.

r/cuba: Figure 2 shows the narrative map extracted from r/cuba. The main storyline starts with a submission criticizing the standing of Cuba as a leader in health care. Moreover, it highlights economical issues and political repression. Next, a submission discusses the state of the pandemic in Cuba on the same day that the protests start. Overall, the main storyline gives the impression that the protests started due to health care issues, alongside other catalysts.

Afterward, we see a simple submission with the rallying call "Viva Cuba libre!" (translated as "long live free Cuba") with an image about tyranny in its internal contents. We note that this submission was locked by the moderators of the community. The next day there is a post criticizing the moderators for taking up anything about the protests. Then, we see a rallying call with a video showing the protests. After this, there are two submissions seeking information about the protests, asking about counter-protests and whether the internet has been shut down. Finally, the story ends with reports of abuse by Cuban police and a call against the regime.

Second, we analyze the representative landmarks highlighted in green and their storylines. Storyline ①and its representative landmark are focused on the US. This story discusses the US embargo, a potential intervention, and a migratory crisis. The second representative landmark is a rallying call in the main storyline, which makes sense as a key narrative event that highlights the protests. The third representative landmark is another rallying call with a photo of protesters carrying a Cuban flag. The rest of the storyline ②focuses mostly on repression in the protests, with Spain demanding the release of a reporter, and the police and military shooting protesters. However, it ends on a better note with changes in customs restrictions by the Cuban government due to the protests. The final representative landmark and side storyline ③focus on the counter-protests and support for the regime, with people mobilizing in defense of the communist revolution.

Overall, the side storylines provide three key points that complement the main storyline and point towards potential future events. First, the US is likely to be a key player in this narrative, although it is too early at the time of this writing how it might respond. Second, the protests have at least some limited success so far, such as the lifting of customs restrictions. Finally, there is a strong response by the government and its forces, with police and military forces being used, as well as a counter-movement that backs the government.

r/Conservative: We present the narrative of r/Conservative—a right-leaning community centered on US politics [37]—in Figure 3. The main story starts with the protests and how Cubans seek to end the regime. However, the rest of the narrative focuses mostly on US political actors. The narrative of r/Conservative uses the protests as a tool to criticize its political opponents (left-leaning media, Democrats, Black Lives Matter, etc.), rather than focusing on the protests themselves. Regarding the representative landmarks, the first one is a satirical article criticizing communist supporters and their views on the regime after the protests. The second one is part of the main story and highlights how celebrities are also stepping in to support the protests against communism.

r/GenZedong: We present the narrative of r/GenZedong—a pro-communist community—in Figure 4. Due to the linear nature of this narrative, extracting representative landmarks is not useful, as they rely on the existence of parallel stories [24]. The storyline starts with an unrelated submission criticizing the presence of an anti-communist Cuban exile in a documentary about Cuba. Although not related to the protests, it aligns with the pro-communist stance of the community. The rest of the timeline supports the regime in Cuba with mostly satirical renditions of anti-communist and pro-US submissions. Overall, this narrative supports the Cuban regime and criticizes its opponents.

r/worldnews: Figure 5 shows the narrative of r/worldnews—a neutral community focused on global news [5, 19]. The main storyline starts with an unrelated event on Cuba developing effective vaccines. This is an interesting parallel with the r/cuba narrative that also started with a health care-related post but in a negative light. The next event is the Mexican president calling for an end to the US embargo—a representative landmark of this map. Next, the storyline focuses on how the protests developed in Cuba. Then, the Miami mayor is discussing bombing Cuba. Immediately after this, we see Cuba lifting the customs restrictions, heavily implying that fear of US intervention led to removing the restrictions, unlike the r/cuba narrative that presented this as a result of the protests. The side story on the left focuses on the international response (e.g., Russia, China, Iran, and the US). Overall, this narrative has two main themes: the protests themselves and the responses by international actors. This makes sense as the purpose of the community is to discuss world news. Moreover, this external perspective is an interesting complement to the r/cuba narrative, as it provides a bigger picture view rather than an emotional one. For example, note the lack of rallying call submissions. This makes sense as this community is neutral and should have a more objective view.

r/politics: We present the narrative of r/politics—a left-leaning community centered on US politics [35, 37]— in Figure 6. The narrative starts with a report on the protests, then it mostly focuses on calls to lift the embargo. The storyline then shifts towards the protests in Miami, with a focus on the anti-riot law. The narrative uses this to criticize the political opponents of the community, in a similar way to r/Conservative.

Relevant Storylines	r/cuba	r/Conservative	r/GenZedong	r/worldnews	r/politics
Causes of the Protest	×			×	
Cuban Government Response	×			×	×
Police Brutality and Repression	×			×	
Supporting the Protests	×	×			×
Supporting the Regime (counter-protests)	×		×		
International Response				×	
US Response and Potential Intervention	×		×	×	×
Ending the US Trade Embargo	×		×	×	×
Criticizing Political Opponents	×	×	×		×

Table 2: Relevant storylines & events in each narrative map.

Narrative Comparison: We show the relevant storylines and events in each narrative in Table 2. We note that r/cuba covers almost all the aspects of the protest except the international response, save for the US, given its relevance in the historical and geopolitical context of Cuba. In contrast, the other communities tend to focus on specific storylines based on their agenda.

5 DISCUSSION

Social Movement Narratives. Our work unravels the different narratives espoused by online communities pertaining to the 2021
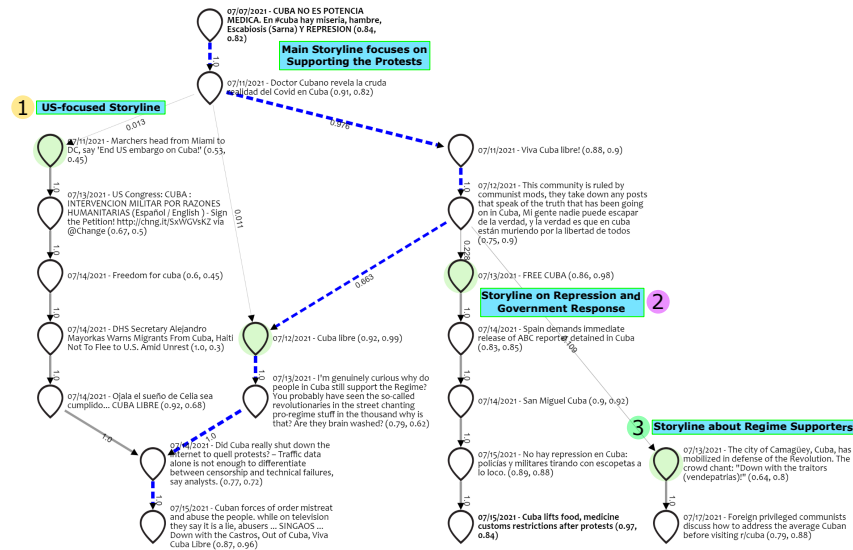
Figure 2: Narrative map for r/cuba. The numbers beside each submission represent the upvote ratio and score percentile. The main storyline is represented by the blue dashed lines. Highlighted nodes correspond to representative landmarks ⬭. Edge weights represent coherence. Manual annotations describing specific storylines have been added with markers 1 , 2 , and 3 .

Cuban protest movement. In general, our work provides a computational framework to analyze the accepted narratives that emerge in online communities in the context of social movements, allowing analysts and social scientists to understand the motivations and relations between different groups in the context of a social movement. In particular, we found that r/cuba had the most complex narrative with a tree-like structure and diverse content covering multiple storylines. In contrast, r/GenZedong had the simplest narrative—a single linear storyline—and had pro-communist content. Meanwhile, both the US-based political communities—r/politics and r/Conservative—used the protests to criticize their opponents and further their own agendas, rather than focusing on the protests themselves. Finally, r/worldnews showed the most neutral presentation, focusing on factual content rather than emotional content, with news discussing the international response and the events of Cuba itself.

Implications. In the context of computational journalism [9]—which provides tools for various computational needs in journalism—our approach could serve as a support tool for journalists. For example, narrative maps could be part of a narrative exploration suite, which could also include text summaries, data plots, and references to trustworthy sources. Such a tool could aid analysts, journalists, and other users in trying to understand the narratives espoused by online communities [24]. Our work could also be applied by social scientists interested in a narrative analysis of how social movements are evolving, as the accepted narratives provide a window into what different communities are thinking and their views.

Furthermore, our community acceptance approach could be enhanced with additional attributes. In particular, we could model the credibility and political bias of Reddit submissions (e.g., by analyzing the source they are linking). These attributes could aid in modeling how misinformation spreads during social movements. Another possible extension would be using sentiment analysis [32] in conjunction with our narrative model to obtain more nuanced

narratives that explicitly model how communities feel with respect to specific storylines and events. Finally, we note that our methods are grounded in theoretical constructs of narratology [24] and could be applied beyond social movement narratives.

Ethical Considerations. There are two stakeholder groups in this research: members of the studied communities and potential users of the system (e.g., analysts, journalists, and researchers). As with any system that could be used to monitor social movements, there is a risk that malicious actors leverage the system for nefarious uses, such as tracking activists involved in organizing or participating in these movements or designing targeted information operations to counter or support specific narratives. Regarding tracking issues, the system works at an aggregate level, without user-level details, and the system uses non-confidential and publicly available data and does not require any specific user information. Thus, it would be hard to perform user re-identification. Next, we note that the system does not distinguish between the truth and misinformation, which could allow users to explore how misinformation spreads in online communities, but also lead to incorrect interpretations of the narratives. We could reduce the amount of potential misinformation in maps by giving control to the analysts over the information sources used.

Nevertheless, we note that the narrative extraction tool is still in the prototyping phase. We expect to address any issues that arise as we work with stakeholders in user studies and pilot testing. Through this process, we will ensure that the system is resistant to these issues. Finally, we note that our approach does not involve any direct interaction with the studied communities, it only analyzes publicly available content from these communities. Thus, we do not anticipate any direct harm resulting from this research.

Limitations. Our work is not without limitations. The current embedding model only considers the text of a submission, disregarding images and videos. Thus, future work could consider multimedia

embeddings to model these elements. Moreover, we note that our data spans a relatively small period of time. Thus, the analysis does not capture all the narratives surrounding the movement.

6 CONCLUSIONS

This paper presents an algorithmic approach to extract the accepted narratives of online communities in the context of social movements, specifically the 2021 Cuban Protests during its first few days. Building upon existing narrative extraction methods based on linear optimization, we designed a novel approach that integrates the concept of *community acceptance* with pre-existing definitions of coherence and coverage. Our technique allows us to identify the major themes surrounding the protests in five different online communities. Recognizing the accepted narratives of different communities is central to understanding social movements. Thus, our method could aid journalists in making sense of emerging narratives in social media in the context of reporting social movements. In general, our work has potential applications in the study of social movements, intelligence analysis, computational journalism, and misinformation research. Moreover, it opens up multiple avenues for future research, such as adapting the method to find rejected narratives or modeling acceptance in other social networks.

ACKNOWLEDGMENTS

REFERENCES

[1] H Porter Abbott. 2008. *The Cambridge introduction to narrative.* Cambridge University Press, NY, USA.
[2] Narges Akbarpour Roshan and Mohsen Abbasi. 2018. The Impact of the US Economic Sanctions on Health in Cuba. *Int J of Resistive Econ* 6, 2 (2018), 17–36.
[3] Meysam Alizadeh, Jacob N. Shapiro, Cody Buntain, and Joshua A. Tucker. 2020. Content-based features predict social media influence operations. *Science Advances* 6, 30 (2020), 13 pages.
[4] Antón L Allahar and Nelson P Valdés. 2013. The Bureaucratic Imperative: Economic and Political Challenges to Cuban Socialism in the Early 21st Century. *The CLR James Journal* 19, 1/2 (2013), 392–422.
[5] Mohamed Bahgat, Steve Wilson, and Walid Magdy. 2020. Towards Using Word Embedding Vector Space for Better Cohort Analysis. *Proc. of the Intl. AAAI Conf. on Web and Social Media* 14, 1 (May 2020), 919–923.
[6] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The Pushshift Reddit Dataset. *Proc. of the Intl. AAAI Conf. on Web and Social Media* 14, 1 (May 2020), 830–839.
[7] Lee Anne Bell. 2010. *Storytelling for social justice: Connecting narrative and the arts in antiracist teaching.* Routledge, NY, USA.
[8] Tolga Buz and Gerard de Melo. 2021. Should You Take Investment Advice From WallStreetBets? A Data-Driven Approach.
[9] Sarah Cohen, James T Hamilton, and Fred Turner. 2011. Computational journalism. *Commun. ACM* 54, 10 (2011), 66–71.
[10] Richard A Couto. 1993. Narrative, free space, and political leadership in social movements. *The Journal of Politics* 55, 1 (1993), 57–79.
[11] Clara Crivellaro, Rob Comber, John Bowers, Peter C. Wright, and Patrick Olivier. 2014. A Pool of Dreams: Facebook, Politics and the Emergence of a Social Movement. In *Proc. of CHI 2014.* ACM, NY, USA, 3573–3582.
[12] Joseph E Davis. 2002. *Stories of change: Narrative and social movements.* SUNY Press, Albany, NY.
[13] Jill P. Dimond, Michaelanne Dye, Daphne Larose, and Amy S. Bruckman. 2013. Hollaback! The Role of Storytelling Online in a Social Movement Organization. In *Proc. of the 2013 Conf. on CSCW.* ACM, NY, USA, 477–490.
[14] Michaelanne Dye, David Nemer, Josiah Mangiameli, Amy S. Bruckman, and Neha Kumar. 2018. El Paquete Semanal: The Week's Internet in Havana. In *Proc. of the 2018 CHI Conf. on Human Factors in Computing Systems.* ACM, NY, USA, 1–12.

[15] Anthony Faiola. 2021. Cubans take to the streets for the biggest anti-government protests in decades. https://web.archive.org/web/20210712033912if_/https://www.washingtonpost.com/world/2021/07/11/cuba-protests/
[16] Pascal Fletcher. 2021. Cuba protests: Frustration at government runs deep. https://web.archive.org/web/20210714110213/https://www.bbc.com/news/world-latin-america-57823130
[17] Marshall Louis Ganz. 2001. The power of story in social movements. In *Proc. of the Annual Meeting of the ASA.* American Sociological Association, CA, USA.
[18] Jeffry Halverson, Steven Corman, and H Lloyd Goodall. 2011. *Master narratives of Islamist extremism.* Springer, NY, USA.
[19] Benjamin D. Horne, Sibel Adali, and Sujoy Sikdar. 2017. Identifying the Social Signals That Drive Online Discussions: A Case Study of Reddit Communities. In *Proc. of the ICCCN 2017.* IEEE, Vancouver, BC, Canada, 1–9.
[20] Joseph Kahne, Erica Hodgin, and Elyse Eidman-Aadahl. 2016. Redesigning civic education for the digital age: Participatory politics and the pursuit of democratic engagement. *Theory & Research in Social Education* 44, 1 (2016), 1–35.
[21] Antoni Kapcia. 2020. *A short history of revolutionary Cuba: revolution, power, authority and the state from 1959 to the present day.* Bloomsbury, London, UK.
[22] Kerry Kartchner and Jennie Johnson. 2009. *Strategic culture and weapons of mass destruction: culturally based insights into comparative national security policymaking.* Springer, NY, USA.
[23] Stuart J Kaufman. 2009. Narratives and symbols in violent mobilization: the Palestinian-Israeli case. *Security Studies* 18, 3 (2009), 400–434.
[24] Brian Keith and Tanushree Mitra. 2021. Narrative Maps: An Algorithmic Approach to Represent and Extract Information Narratives. *Proc. of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–33.
[25] Olessia Y. Koltsova and Galina Selivanova. 2019. Explaining Offline Participation in a Social Movement with Online Date: The Case of Observers for Fair Elections. *Mobilization: An International Quarterly* 24, 1 (03 2019), 77–94.
[26] Yubo Kou, Yong Ming Kow, Xinning Gui, and Waikuen Cheng. 2017. One social movement, two social media sites: A comparative study of public discourses. *Computer Supported Cooperative Work* 26, 4 (2017), 807–836.
[27] Yong Ming Kow, Yubo Kou, Bryan Semaan, and Waikuen Cheng. 2016. Mediating the Undercurrents: Using Social Media to Sustain a Social Movement. In *Proc. of CHI 2016.* ACM, NY, USA, 3883–3894.
[28] Jeffrey A Larsen and Jeannie L Johnson. 2006. *Comparative strategic cultures curriculum project: assessing strategic culture as a methodological approach to understanding WMD decision-making by states and non-state actors.* Technical Report. Science Applications International Corp, McLean, VA.
[29] Reinoud Leenders and Steven Heydemann. 2012. Popular mobilization in Syria: opportunity and threat, and the social networks of the early risers. *Mediterranean politics* 17, 2 (2012), 139–159.
[30] Vicki Li. 2020. The Race Talk on Reddit:"China Virus" and its Effect on Asian-Americans.
[31] Alice Mattoni. 2013. *Journalism and Social Movements.* Wiley, NJ, USA, 339–340.
[32] Kumar Ravi and Vadlamani Ravi. 2015. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowl-based Syst* 89 (2015), 14–46.
[33] Frances Robles. 2021. Cubans Denounce 'Misery' in Biggest Protests in Decades. https://web.archive.org/web/20210711212619/https://www.nytimes.com/2021/07/11/world/americas/cuba-crisis-protests.html
[34] Dafna Shahaf and Carlos Guestrin. 2010. Connecting the Dots between News Articles. In *Proc. of the 2010 SIGKDD Conf.* ACM, NY, USA, 623–632.
[35] Ryan P Shepherd. 2020. Gaming Reddit's algorithm: R/the_donald, amplification, and the rhetoric of sorting. *Computers and Composition* 56 (2020), 102572.
[36] Anna S. Smoliarova, Svetlana S. Bodrunova, Alexandr V. Yakunin, Ivan Blekanov, and Alexey Maksimov. 2019. Detecting Pivotal Points in Social Conflicts via Topic Modeling of Twitter Content. In *Internet Science,* Svetlana S. Bodrunova, Olessia Koltsova, Asbjørn Følstad, Harry Halpin, Polina Kolozaridi, Leonid Yuldashev, Anna Smoliarova, and Heiko Niedermayer (Eds.). Springer, Cham, 61–71.
[37] Katherine Van Koevering, Austin R. Benson, and Jon Kleinberg. 2020. Frozen Binomials on the Web: Word Ordering and Language Conventions in Online Text. In *Proc. of The Web Conf. 2020.* ACM, NY, USA, 606–616.
[38] Stefania Vicari. 2017. Twitter and Non-Elites: Interpreting Power Dynamics in the Life Story of the (#)BRCA Twitter Stream. *Social Media + Society* 3, 3 (2017), 14 pages.
[39] Lu Wang, Claire Cardie, and Galen Marchetti. 2015. Socially-Informed Timeline Generation for Complex Events. In *Proc. of the Conf. of the North American Chapter of the ACL: Human Language Technologies.* ACL, Denver, Colorado, 1055–1065.
[40] Nigel D White. 2019. Ending the US embargo of Cuba: international law in dispute. *Journal of Latin American Studies* 51, 1 (2019), 163–186.
[41] Ying Xiong, Moonhee Cho, and Brandon Boatwright. 2019. Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement. *Public relations review* 45, 1 (2019), 10–23.
[42] Yinfei Yang, Daniel Cer, Amin Ahmad, Mandy Guo, Jax Law, Noah Constant, Gustavo Hernandez Abrego, Steve Yuan, Chris Tar, Yun-hsuan Sung, Brian Strope, and Ray Kurzweil. 2020. Multilingual Universal Sentence Encoder for Semantic Retrieval. In *Proc. of the 58th Annual Meeting of the ACL.* ACL, Online, 87–94.

APPENDIX A: LINEAR MODEL DETAILS

Type	Name	Description
Variables	minedge	Represents the lowest coherence value over all active edges.
	$event_i$	Represents the presence of event i in the narrative map.
	$edge_{ij}$	Represents the strength of the connection between events i and j in the narrative map.
	$cluster_k$	Represents the average presence of cluster k in the narrative map.
Data-based Parameters	num-clusters	Represents the number of clusters of events.
	$membership_{ijk}$	Represents whether the connection of events i and j belongs to cluster k. Defined as $\sqrt{membership_{ik} \cdot membership_{jk}}$—the geometric mean between the probabilities of the events belonging to cluster k.
	$coherence_{ij}$	Represents the coherence score of the connection between events i and j. Defined as $\sqrt{angular\text{-}similarity_{ij} \cdot cluster\text{-}similarity_{ij}}$—the geometric mean between the angular similarity of the underlying event embeddings and the clustering similarity based on comparing the cluster membership probability distributions of each event using Jensen-Shannon divergence.
	$score\text{-}percentile_i$ (*)	Represents the score percentile of event (submission) i in the community.
	$upvote\text{-}ratio_i$ (*)	Represents the upvote ratio of event (submission) i in the community.
	$acceptance_{ij}$ (*)	Represents the acceptance score of the connection between events i and j. Defined as $\sqrt{score\text{-}percentile_i \cdot score\text{-}percentile_j} \cdot \sqrt{upvote\text{-}ratio_i \cdot upvote\text{-}ratio_j}$—the product between the geometric means of submission score percentiles and upvote ratios.
User-defined Parameters	K	Defines the expected length of the main storyline in the narrative map.
	mincover	Defines the minimum average coverage over all clusters in the narrative map.
	minscore (*)	Defines the minimum average acceptance over all events (submissions) in the narrative map.

Table 3: Description of the variables and parameters of the proposed extraction algorithm. All elements take values between 0 **and** 1, **except** K **and** num-clusters, **which are positive integers. Most elements of the model are based on the extraction algorithm of Keith and Mitra [24]. The (*) symbol represents new additions to the linear problem formulation designed to incorporate** *community acceptance* **in social media into the model.**

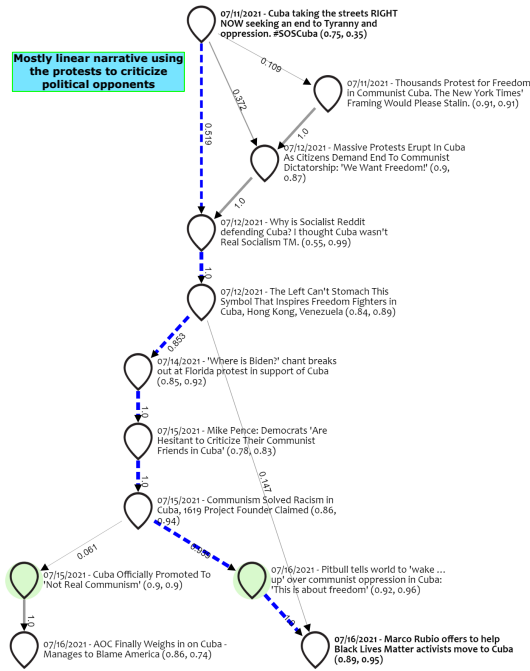## APPENDIX B: NARRATIVE MAP FIGURES

Figure 3: Narrative map for r/Conservative.

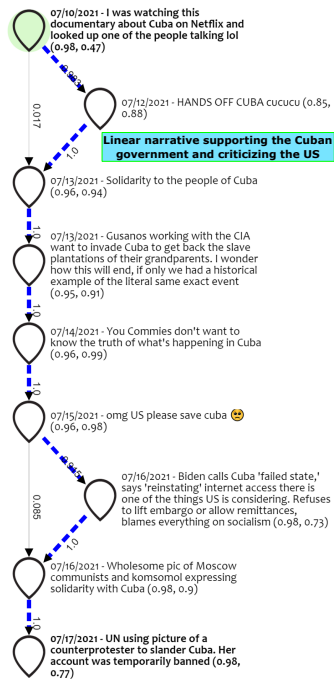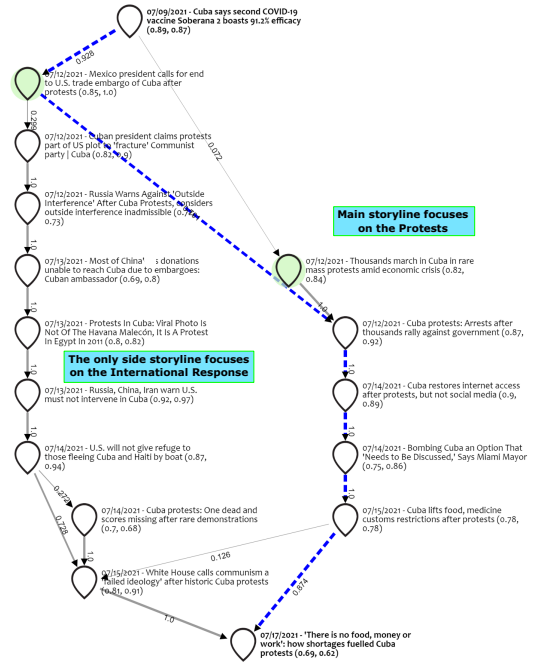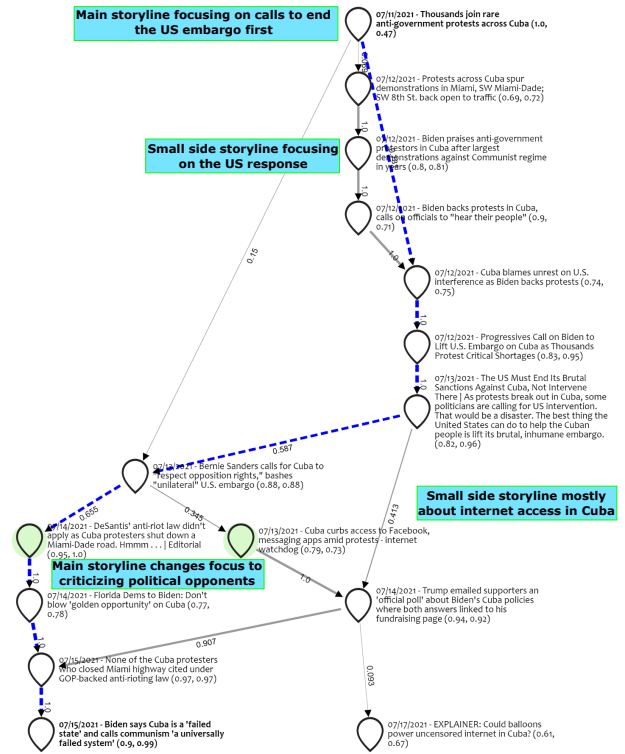**Figure 4: Narrative map for r/GenZedong.**

Figure 5: Narrative map for r/worldnews.

Figure 6: Narrative map for r/politics.