A Relational Event Model Approach to Online Collective Action: Reddit's r/amcstock

Candidate Number: 29189

Abstract

This paper investigates the Reddit-driven AMC short squeeze of 2021 as a case study of online collective action. It leverages unique features of the movement to conduct a novel analysis of longitudinal interaction patterns among Reddit users, relating these to the movement's growth, climax, and decline over a 17-month period. Relational event models are adapted to accommodate large-scale dynamic interaction data using case-control sampling and partial likelihood estimation, developing an approach that can be broadly applied to Internet networks. A sliding window process is used to estimate the temporal variations in the model coefficients. The study finds evidence supporting the staying power of activity and popularity effects throughout the short squeeze life-cycle. It also finds that interaction predictors fluctuate substantially in short-term trends, suggesting that the underlying social formations in the network are highly temporary. The study utilizes an 'organizational agent' framework of networked collective action and explores a range of possibilities in an underdeveloped area of study. It examines interaction predictors related to clustering, activity, popularity, and other dyadic effects, advancing empirical knowledge on cases of collective action, and providing a methodological foundation for further studies in this area, particularly in analyzing dynamic network behavior.

Contents

1 Introduction

Recent decades have witnessed the increasing use of online networks for collective action. Notable examples include the use of social networking sites in the Occupy Wall Street movement of 2011 (Bennett and Segerberg, 2012), and Twitter and Telegram in the 2020 Hong Kong protests (Kow et al., 2020). It is now clear that online networks have transformed the underlying logics and mechanisms of collective action (Chen et al., 2021).

Though online media has had a substantial impact on social movements, the exact nature in which it does this has not been well studied. In fact, few studies have focused on the structural dynamics of online media in relation to collective action (Young et al., 2019). For example, little is known about the patterns of individual interactions associated with the evolution of a movement (Dolata and Schrape, 2016). This study addresses this gap and investigates the relationship between online communication networks and collective action. Given the growing complexity, specialization, and impact of online spaces in recent decades, the importance of understanding how collective action unfolds in these spaces has grown substantially.

This study focuses on a unique case of collective action, the 2021 Reddit short squeeze movement. Specifically, it focuses on the successful short squeeze of the stock of AMC Entertainment, which was coordinated and organized on the subreddit (forum) 'r/amcstock.' I contend that studying this subreddit provides a well-defined network of users who focused on achieving a *measurable* collective goal. Through this, predictors of user interaction can be measured on the evolutionary timeline of the short squeeze movement relating to the growth, climax, and decline periods. This study provides an empirical analysis that explores collective action not only in its baseline network patterns but also in how these change over time.

I use a Social Network Analysis (SNA) approach to measure the predictors of user interactions on Reddit over an observation period of 17 months. However, I deviate from traditional approaches that involve the use of static network snapshots and instead develop a method for the study of large-N, dynamic network data using Relational Event Modeling (REM). Sampling and model variation strategies are used to accomplish this. In addition, time-varying model effects are estimated using a sliding window approach.

The paper is organized as follows. I firstly form a theoretical framework on the relationship between collective action and online social networks, introducing networks as 'organizational agents.' Next, I review the literature of previous network studies on collective action, outlining research limitations and theoretical expectations. Then, I develop the background of the Reddit case study, explaining its usefulness for a longitudinal analysis of interaction behavior. After this, I develop the methodology, detailing a relational event model approach to the Reddit interaction data, adapting methods from previous network studies for large dynamic networks. The model results are then presented, detailing baseline, time-varying, weighted, and event-type findings on predictors of user interaction. I conclude with a discussion of the results in relation to collective action and Internet networks, noting extensions and limitations to the current study.

2 Theoretical Framework

2.1 Conceptualizing Online Collective Action

Collective action can be defined as "actions taken by two or more people in pursuit of the same collective good" (Marwell et al., 1988), which involves the challenge of persuading or even compelling self-interested individuals to contribute their private resources to achieve collective goals (Chen et al., 2021). As this process develops through social networks on the Internet, the functional role of communication increases in importance compared to offline settings (Dolata and Schrape, 2016). In formal terms, online collective action entails "efforts for making people cross public-private boundaries by expressing or acting on their interests in a way that is observable for those who share common interests" (Flanagin et al., 2006). Activities, including retweeting, commenting, sharing, and posting, each represent a form of visible commitment signaling in online spaces. These activities can induce others to engage themselves, which is known as mobilization (Johansson and Scaramuzzino, 2023). The growth of an online movement can be measured by its visible communication activity, which represents evidence of the strength of the movement.

2.2 Online Networks as an Organizing Agent

Traditional, pre-Internet collective action theory suggests that formal organizations or activist groups occupy the functional center of social movements, facilitating mobilization and generating momentum for the collective goal (Spier, 2017). Although evidence for the central role of formal organizations is observed in online spaces, it is not the dominant mechanism of collective action on the Internet. Movements that are born and proliferate in online networks exhibit fundamentally unique processes of evolution and mobilization.

Recent studies suggest that online technologies function as 'organizing agents' (Dolata and Schrape, 2016). This view suggests that online networks provide a free space for social movements to evolve in an organic and self-organizing process. By digitally connecting individuals on accessible platforms, online networks enable new social formations to emerge (Ahuja et al., 2018). This empowers individuals to gain broad leverage to take action and assert influence (Bennett and Segerberg, 2012). Collective strategies can emerge as information sharing structures develop and collective cultures bridge divisions, unifying network participants around common goals (Wang and Chu, 2019). Through this, networked individuals gain a substantial cumulative ability to affect change on a societal level.

Evidence for an 'organizing agent' conception of online networks is increasingly widespread. A growing number of collective action movements originate on the Internet, and many operate primarily through digital activism, including protests on social media and information sharing (Dumitrica and Felt, 2020). Notable examples include the indignados movement in Spain (Anduiza et al., 2014), the 2011 Occupy Wall Street movement (Conover et al., 2013), the 2019 Pakistan student protests (Sikandar and Fatima, 2024) and the 2020 'leaderless' Hong Kong protests (Kow et al., 2020). Online networks are no longer only tools for collective action movements, but have now become digital environments for decentralized social coordination around collective aims (Spier, 2017).

Using the conception of online networks as organizational agents, this study generates a theoretical link between online activity and observed collective action outcomes. It also contextualizes social media sites such as Reddit, Twitter, and Instagram as environments for collective social formations to generate social change. This provides a basis for making empirical predictions on interaction processes within Internet networks.

3 Literature Review

This section proceeds as follows. I first outline the challenges facing network-based studies of Internet collective action, setting out notable gaps in the current literature. Then I set out three guiding research questions which are used to review previous studies and offer theoretical expectations for predictors of online interactions. Lastly, I outline theory related to the investigation of movement-network evolution.

There are several key challenges for the research of online collective action. Firstly, the media and communication activity available on the Internet is diverse and may not generalize from one online context to another (Himelboim, 2017; Wang and Chu, 2019). Although Internet networks are usually horizontal, well connected, and decentralized in nature (Flanagin et al., 2006), there are insufficient empirical studies to reliably determine how structural and interaction dynamics change across the Internet. For instance, Reddit may exhibit unique patterns and social formations by virtue of its communication mechanisms, such as deep-nested commenting. Although empirical networks can be compared on certain measures, such as centralization and clustering, other empirical observations are likely to be context-specific. The current body of empirical literature remains too small to generate overarching conclusions.

Another challenge facing current research is the definition of user networks involved in collective action. A key observation is that social movement formations are rarely separated from larger social interaction platforms. Past studies have used identification tools such as hashtags, text analysis, or affiliation measures to identify individuals involved in a collective movement (Mancini et al., 2022; Pitcher and RufaroMuchena, 2022); however, these are prone to errors and introduce validity concerns. For example, it is difficult to determine whether a defined network includes a sufficient sample of participants in a social movement to accurately measure predictors of interaction.

In close relation to this, there is the difficulty in measuring a link between network activity and the development of collective action (Siegel, 2009). It is not clear how to empirically relate observed online activity to collective action engagement. An interplay is likely to exist, yet current methods and data are not well suited to measure this. The current literature needs further empirical and exploratory studies to make progress in this area.

The final challenge is methodological. As online networks are often found on social media sites, these networks are very large and highly dynamic, with a fluctuating user base and shifting levels of activity over time. This challenges traditional approaches in network science, such as Exponential Random Graph Models (ERGMs) and Stochastic Actor-Oriented Models (SAOMs) (Quintane et al., 2014; "The SAGE Handbook of Social Media", 2024). Newer methods, including Relational Event Modeling, are better adapted to dynamic network data (Butts, 2008); however, current research has not yet widely utilized these approaches. Because of this, longitudinal studies of collective action are few in number, and existing studies focus on more easily measured network metrics related to network clustering and centralization, rather than predictors of interaction.

Given these limitations in collective action research, my study adopts an exploratory approach and is focused on developing an appropriate methodological approach. However, I use existing theory and past findings to make general predictions to frame my findings. With this in mind, I focus on three primary research questions:

- (i) What predominant interaction dynamics are associated with online networks involved in collective action?
- (ii) How do these dynamics change over a movement's life cycle?
- (iii) What variables might explain the changes in these dynamics over time?

To examine question (i), I suggest predictors of user interaction within three areas: Centralization and clustering, user activity and engagement, and reciprocity and repetition, generating theoretical expectations for each.

3.1 Centralization and Clustering

Network centralization is an important measure of the social formation of a movement. This refers to the extent to which a network develops around a core of active, well-connected users (Scott and Carrington, 2011). Empirical findings and theoretical expectations are at odds in terms of network centralization. On the one hand, the ease of participation in online media (theoretically) enables the fluidity and reach of communication, generating decentralized and horizontal networks through features such as mentions, retweets, comments, or hashtags (Ahuja et al., 2018). However, empirical studies find that despite these decentralizing forces, key players still emerge and play dominant roles in collective action networks, generating hierarchical structures (Bennett and Segerberg, 2012; Wang and Chu, 2019; Himelboim, 2017). In addition, decentralizing forces are observed to be insufficient to prevent user clustering, a measure that characterizes how common transitive connections are within a network (Scott and Carrington, 2011). Gonzalez-Bailon and Wang (2016), for instance, in their study of online protest networks, find that influential users act as information brokers, intermediating between less connected user clusters. It follows that Internet collective action networks are expected to be largely decentralized but nonetheless develop hierarchical and stratified topographical structures. Additionally, not all users are expected to have similar structural roles, with certain individuals gaining influence through their roles as intermediaries. From this, the predictors of user interaction related to clustering are expected to have a positive effect.

3.2 User Activity and Engagement

Another key predictor is related to the engagement patterns of a user within a network. Online networks often exhibit power-law distributions in communication activity, both in sending and receiving (Himelboim, 2017). Certain users receive a disproportionate share of activity in either direction, creating heavy-tailed distributions (Himelboim, 2017). This follows from a rich-get-richer dynamic of interaction, where users are more likely to interact with users that other users have already interacted with. It can be expected that user popularity has a positive effect on user interactions.

Likewise, user activity is also important. Social movements, particularly online, rely on movement inertia to succeed and require continued, returning engagement (Marwell et al., 1988; Johansson and Scaramuzzino, 2023). It can be expected that recent user activity has a positive effect on user interactions.

3.3 Reciprocity and Repetition

Both reciprocity and repetition represent dyad-level predictors of interaction and measure repetitive or mutual interaction between users. Online interactions are often reciprocal when the number of users involved is smaller, as more attention is paid to individual instances of activity (Himelboim, 2017). However, online networks are by nature large and decentralized, reducing the relative attention paid to activity sent to an individual. From this rationale, the effect of reciprocity on user interaction is expected to be negative or neutral.

Repetition, on the other hand, is predicted to have a positive effect since online networks create users with a high degree of incoming comments (Nie et al., 2023), to whom other users are likely to send multiple communications, so repetition is expected to have a positive effect.

3.4 Temporal Change Over Movement Life-Cycle

Research questions (ii) and (iii) suggest a change in the predictors of user interaction over time, as a social movement develops. Although there are longitudinal studies of online networks, few have focused on the evolution of collective action over time (van Stekelenburg et al., 2013).

A general narrative for movement-network evolution relates to the organizational agent conception of online media (Kavada, 2015). As a network develops a self-organized social formation, it is likely to change during the growth, climax, and decline phases of the movement. It is plausible that certain structures emerge from dedicated user groups that grow resilient to exogenous change. Johansson and Scaramuzzino, 2023, for instance, finds that online organizers in an Internet-born Swedish protest movement became involved in a 'constant mobilization' process after achieving explosive movement growth. In other words, achieving the movement climax may fundamentally alter a movement's social formation. An empirical aim of the current study is to determine how dynamic underlying movement's social formation is by comparing change in interaction rate predictors over time. An exploratory approach is used to interpret change in predictors, determining if there are different movement periods that form similar trends, or if fluctuations are too sporadic or random to draw meaningful conclusions. Given a lack of substantial findings on movement-network co-evolutions, this approach is well-suited.

4 Background

The chosen case study of collective action is the Reddit-driven short squeeze of the stock of AMC Entertainment, an American movie theater company. A short squeeze is a unique form of collective action and exhibits some key attributes that make it useful for study. I begin by explaining the short squeeze and its surrounding context, before elaborating on the case study's broader connection to collective action.

Short-selling is a trading strategy in which investors borrow shares of a stock and sell them, hoping to later buy them back at a lower price, thus profiting from the difference if the stock price declines (Mancini et al., 2022). A short squeeze occurs when the price of a stock rises to such an extent that investors who have sold short buy the stock to limit their losses or meet financial margin call requirements, causing the price to increase further in a self-reinforcing squeeze (Anand and Pathak, 2021).

AMC Entertainment was subject to a high degree of short interest during 2021, due to financial difficulties related to the COVID-19 pandemic. It was well placed to become the target of a short squeeze.

It became a target as the Reddit Short Squeeze movement of 2021 officially started when GameStop, a struggling gaming retail company, was successfully squeezed in late January 2021 (Aharon et al., 2023). The price of the GameStop stock increased by 2000% over a period of one month. Melvin Capital, an institutional hedge fund, lost $6.8 billion from the event and was forced to close its business, indicating the unprecedented scale and impact of the squeeze event (Chung, 2021). More impressive is that the squeeze was achieved by retail traders who mainly coordinated on Reddit, initially on the trading strategy subreddit r/wallstreetbets (Lucchini et al., 2022). Movement coordination later advanced onto other subreddits like r/amcstock in February 2021.

Figure 1 highlights the key events in the AMC short-squeeze movement timeline. After the GameStop squeeze, the subreddit r/amcstock was created to coordinate discussion on squeezing AMC stock. It took about 5 months of build-up to eventually achieve this, with the stock increasing 30x from its starting price in June 2021. This was followed by a climax period in which the price remained elevated for several months and the hope of further squeeze was high. After this, a decline period emerged, where the price dropped to earlier levels and the hope of a further squeeze diminished, but some users remained committed to the idea.

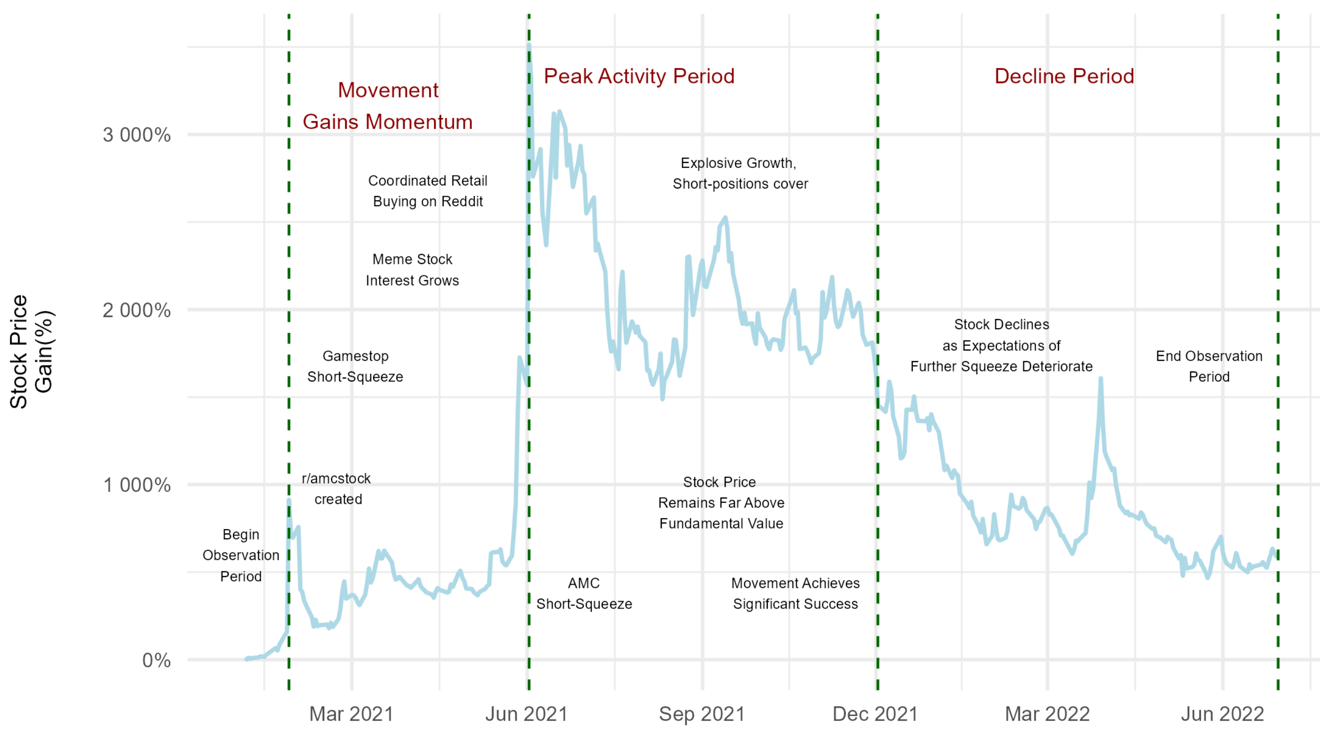


Figure 1: Timeline of AMC Short Squeeze with Key Events and Periods Listed. Stock price measured as percent increase from closing price on 04/01/2021.

The r/amcstock subreddit served as the primary hub of mobilization for the AMC short squeeze. This online community allowed users to share information, coordinate actions, and build a collective identity around the goal of squeezing short sellers. The subreddit experienced rapid growth during this period, with the number of subscribers increasing from around 50,000 in February 2021 to over 400,000 by June 2021.

Culture and sentiment on r/amcstock during the 17-month observation period were characterized by a mix of financial analysis, speculation, and meme-driven content. Users frequently shared due diligence (DD) posts, technical analysis, and theories about market manipulation. The community also developed its own jargon, with terms such as "diamond hands" (holding onto stocks despite volatility) and "to the moon" (expectations of dramatic price increases) becoming common (Boylston et al., 2021).

This online movement is unique in several aspects. First, it represents a decentralized, self-organized movement without formal leadership. Second, it combined elements of financial speculation with the dynamics of social movements, creating a hybrid form of collective action (Lucchini et al., 2022). Third, the continuous and real-time nature of interactions on the platform allowed rapid dissemination and coordination with respect to stock price change, which appears in network interaction data (Semenova and Winkler, 2023).

A short squeeze provides the stock price as an indicator of movement success and failure over time, which allows the formalization of the growth, climax, and decline periods in a longitudinal analysis (Xiaolong et al., 2021). Noting that many exogenous factors affect the price of the stock, numerous studies have found strong links between Reddit activity and fluctuations in meme stocks such as AMC (Bradley et al., 2021; Hu et al., 2021). Importantly, a Reddit short squeeze is an archetypal example of the organizing agent framework for online collective action, making it particularly relevant to understand future developments as collective action continues to unfold in Internet spaces.

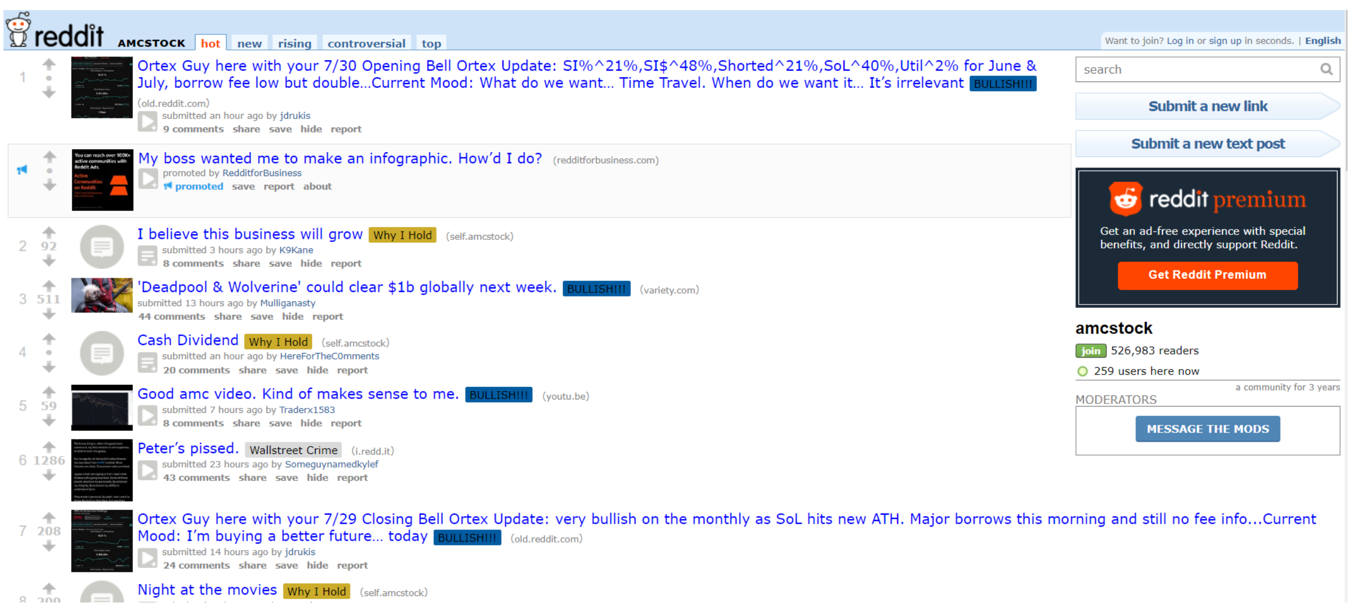


Figure 2: screenshot of subreddit r/amcstock, an online hub for posting and commenting about AMC stock and coordinating its short squeeze

5 Methodology

5.1 Data Collection and Processing

Reddit interaction data was collected using a data dump published on academic torrents. This source was used because Reddit's official API has been restricted for research purposes.

Data Filtering. Following download, missing data was filtered, removing comments and submissions where author usernames were not available. This included comments and submissions deleted by users, which is a common practice on Reddit. Next, comments with a parent comment or submission lacking an identifiable username were removed, as these lacked any relational link to the wider network. Further sources of missing data are detailed in figure 3. The raw data totaled 10,196,067 comments and 511,480 submissions, which reduced to 5,820,604 and 264,738 respectively, with a data completeness rate of 57.09% and 51.75%.

Timeline. The observation period of 17 months specified in figure 1 encompasses the entire life-cycle of the short squeeze movement. Three periods of movement evolution are signaled by two horizontal lines in time-series graphs: growth, peak and decline. These serve as approximate indicators of the stock price and by extension, the short-squeeze movement's development.

Comment Network. To create a comment network, an edge was defined as a user i commenting on another user $j's$ comment *or* submission. In effect, this is the same as i *sending* a comment to j. This formulation is used to create an *event list*, which is a list of directed interactions (comments) at discrete points in time during the period of February 1st 2021 to June 30, 2022. Note that the creation of Reddit submissions are not explicitly included as directed events given that they have no direct sender-receiver pairing (i, j). Submissions instead have an implicit influence due to the accumulation of incoming edges (comments) for users with popular submissions. An *event list* is used for the formulation of relational event models.

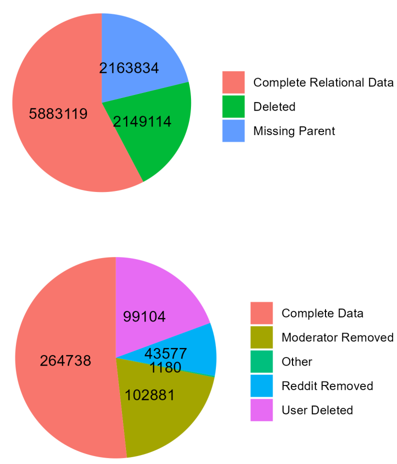


Figure 3: Counts of missing data for comments (top) and submissions (bottom)

Feature Engineering. A final processing step added three variables to each comment based on the comment type and character length. These are used in subsequent weighted and event-type models. A final event list has the form: comment time (UNIX), sender username, receiver username, sender length, receiver length, and comment type. This (anonymized) processed event list can be downloaded from Dropbox. All parsing, analysis, and visualization scripts in this paper are provided in Github Repository.

5.2 Ethics

Ethical approval was received from an LSE review board. The ethical guidelines relating to the handling of personal data were strictly followed. In particular, the usernames of all Reddit users were immediately anonymized using a hash key upon downloading the data. When possible, data was stored on the LSE server for data security best practice. Data use did not violate Reddit terms and conditions, and the use of Reddit data for research has precedent in a numerous past studies (Proferes et al., 2021).

5.3 Relational Event Modeling

Real-world relational data can often be described as dynamic. Dynamic networks are characterized by edges ordered in time, which are directed between a large number of actors, each of whom experiences different rates of interaction that vary over time (Lerner and Lomi, 2022). Internet networks are examples of large, dynamic networks. However, any social network involving singular instances of some activity, carried out by an actor at different points in time, can be classified as dynamic to varying degrees. My study seeks to develop an analysis method that models interaction predictors within large dynamic network data, overcomes limitations presented by fluctuating user-sets, and is able to measure time-varying effects.

Some network analysis methods appear to be suitable for this task. As mentioned earlier, the ERGM and SAOM frameworks are widely used for time series analysis. However, the limitation of these models is that they require that dynamic networks be collapsed into static snapshots for different points in time (Quintane et al., 2014). This results in loss of information and prevents the analysis of granular temporal interaction patterns. Transforming events into static edges also introduces concerns of validity (Brunswicker and Schecter, 2019). Moreover, longitudinal analysis with ERGM and SAOM cannot fit large networks because each network snapshot is a fundamentally new network that cannot be compared to the previous (Scott and Carrington, 2011; Meijerink-Bosman et al., 2023).

Relational Event Modeling (REM), first proposed by Butts, 2008, is designed to analyze time-stamped sequences of directed interactions between individuals in a network, preserving temporal ordering of events without loss of relational information (Scott and Carrington, 2011). REM has been used widely, such as the modeling of cabinet member meetings in government (Lerner et al., 2021), predicting user contributions on open-source software forums (Quintane et al., 2014), and predicting user interactions on Wikipedia (Lerner and Lomi, 2020). REM is flexible in accommodating changing network composition, including user-base, and includes parameters to model the variate in user activity levels over time.

Given its temporal focus and adaptability, REM is aptly suited to model the r/amcstock network data. In substantive terms, a REM predicts the rate of occurrence of comments based on a set of statistics that capture information about past comments on node, dyadic, and network levels. This enables the measurement of predictors for user interaction within the network.

5.4 Model Specification

Relational event models specify "time-varying event rates for all dyads as a function of past events on the same or other surrounding dyads" (Lerner and Lomi, 2020, p. 100).

As a foundation, a single relational event e_m is represented as a tuple in the form $e_m = (i_m, j_m, t_m)$ where i_m is the event sender and j_m is the event receiver and t_m is the event time. Each event e_m exists within the observed sequence of N total events: $E = (e_1..., e_N)$.

Importantly, for each event e_m, the sender i_m is taken from a set U_{t_i} of possible source nodes derived at event time t_m where $i_m \in U_{t_m}$. Likewise, receiver j_m is taken from the set of target nodes where $j_m \in U_{t_m}$. For each time point t, the *risk set*, defined as $R_t \in U_t \times V_t$ consists of the possible pairs of sender-receiver (dyads) pairs at time t. It follows that $(i_m, j_m) \in R_{m_t}$. The risk set R_t represents events that could have occurred at time t but did not, and is used to capture potential events at any time t to assess the relative effect of past events on event occurrence.

The core of the relational event model is the hazard rate function, which captures relative propensity for events to occur based on covariates (statistics) and their effects, defined at a time point t and a dyad (i, j) as:

$$\lambda_{ij}(t) = \exp\left(\beta^T u_{ij}(t)\right) \tag{1}$$

where $u_{ij}(t)$ is a vector of statistics (covariates) that capture information on the history of previous events before time t and β^T is the row vector of effect parameters (coefficients) for each statistic, such that:

$$\beta^T u_{ij}(t) = \beta_1 u_{1_{ij}}(t) + \beta_2 u_{2_{ij}}(t) + ... + \beta_v u_{v_{ij}}(t) \tag{2}$$

where v represents the total number of statistics. Statistics represent measures of past observed events or exogenous attributes for users (nodes), dyads, triads, or other levels of network observation at a time t. This formulation enables the hazard rate $\lambda_{ij}(t)$ to capture the differential tendencies of an individual to send or receive actions over time (Butts, 2008). In addition, it enables histories of past actions to affect future actions, which is vital to predict interaction behaviors.

To estimate statistic effects β^T, the hazard rate is incorporated into a partial likelihood L_P equation, which has the form:

$$L_P(\beta) = \prod_{l=1}^{N} \frac{\lambda_{i_m j_m}(t_m)}{\sum_{(i,j)\in R(t_m)} \lambda_{ij}(t_m)} \tag{3}$$

This is used to predict a complete order of events $p(A_t)$ based on estimations of the coefficients β^T in the hazard rate $\lambda_{ij}(t)$, using the following:

$$p(A_t) = \prod_{l=1}^{N} \left[\frac{\exp\left(\sum_h \beta_h (u_h)_{i(m)/m}\right)}{\sum_{(k)\in\Omega(m)} \exp\left(\sum_h \beta_h (u_h)_k\right)} \right] \tag{4}$$

where m is the index that goes through all events, h is an index that goes through all statistics, and k is an index that goes through all users involved in the event. The complete predicted order of events from equation 4 is used in a likelihood estimation process to derive the effects $/beta$ of statistics u by minimizing the error between the observed and predicted event sequence. Likelihood estimation is done through a Cox Proportional Hazards (Cox PH) model that originates from survival time series approaches

(Quintane et al., 2014). The resulting fitted models are log-likelihood in form. Example scripts for the Cox PH models are provided in the Appendix.

5.5 Adapting REM to Large-N Networks

An important consideration for fitting REMs on large networks is the size of the *risk set* R_t, which grows quadratically with the number of nodes (users) within the network. Given 113,060 unique users in the Reddit data, the *risk set* R_t grows to 12.7 billion. Model run times for a risk set this large are computationally unfeasible.

To fit REM to large N networks, two key adaptations are utilized. Firstly, I chose to use the partial likelihood estimation of REM, as in equations 3 and 4, rather than the full likelihood, which additionally estimates a baseline rate of event occurrence and predicts the time between events for event and nonevent dyads. My analysis only requires the ordering of events rather than their exact timing to obtain an in-depth analysis of comment interaction behaviors over time. Full-likelihood estimation is much more computationally expensive than partial estimation. A slight limitation in the model's specificity is worth substantial gains in computational feasibility. Quintane et al., 2014, for instance, describe full likelihood estimation as involving 'nuisance parameters' that add little to the outcomes of interest in the analysis of the REM network.

Secondly, I adopt a case-control sampling strategy to drastically reduce the size of the risk set for the calculation of the hazard rate in equation 1. Case-control sampling selects a set proportion of non-events from R_t for every observed event e_m at the event's time t_m. Statistics and hazard rates are calculated for these non-events and used in equation 1. The validity of case-control sampling is evidenced by Lerner and Lomi, 2020, who reliably fit partial likelihood REMs on a Wikipedia-user network with 12 million nodes and 360 million dyadic events by validating model parameters on simulated networks. They achieve this without significant error in coefficient estimates, demonstrating the robustness of control sampling from the risk set.

Following sampling guidance of Lerner and Lomi (2020), and my own experimentation on balancing model run-time with maximizing model specificity, I opted for a case-to-control ratio of 1:5, resulting in a model-ready dataset of 5,820,604 events (comments) with 25,559,402 nonevents for a total size of 31,380,006.

5.6 Network Trends and Structural Distributions

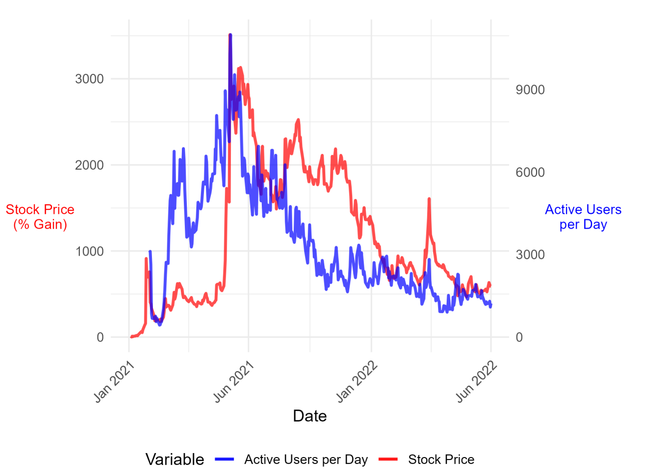


Figure 4: Observed Users per Week vs. Total Observed Users

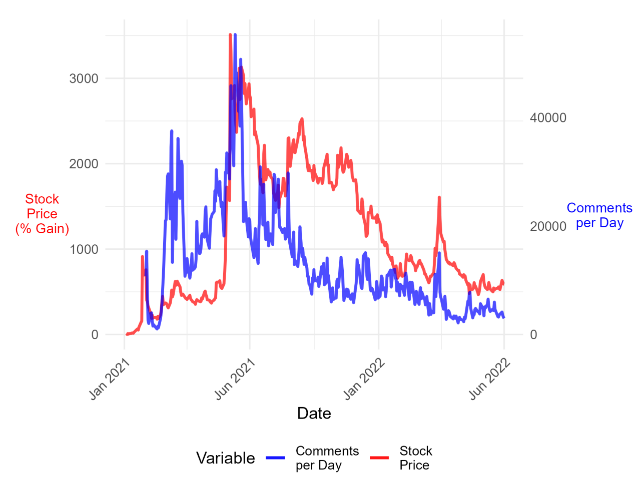


Figure 5: Daily Comment Count vs. Stock Price

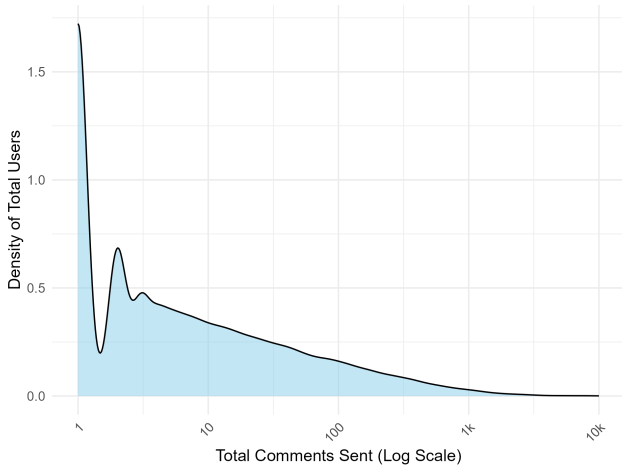


Figure 6: Density Plot of Comment Out-Degree by User

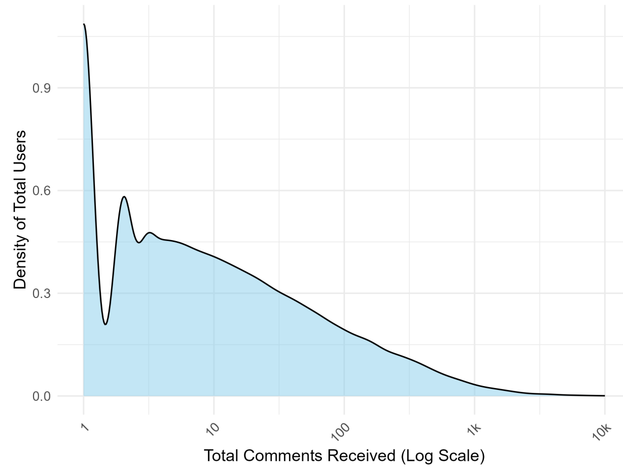


Figure 7: Density Plot of Comment In-Degree by User

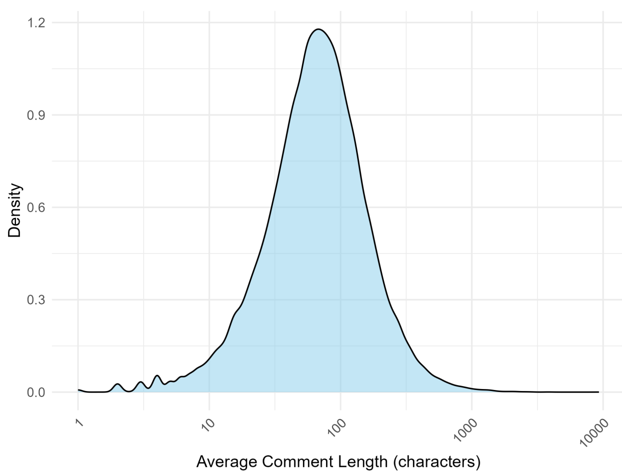


Figure 8: Density Plot of Comment Out-Degree by User

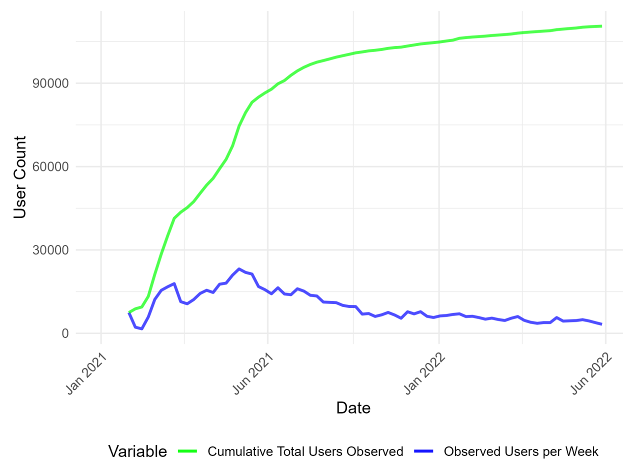


Figure 9: Observed Users per Week vs. Total Observed Users

Baseline network trends provide measurable insights to inform the creation of REM statistics and model types based on their estimated fit to the observed data.

Temporal Patterns. Figures 4 and 5 demonstrate the large variance in the network activity and active user base over the short squeeze life cycle and during the periods of price increase and decline. Structural network behaviors and interactions across time t are expected to change substantially. This motivates the creation of a time-varying model to capture interaction between time t and statistic effects β.

User Activity Distributions. Figures 6 and 7 signal a long-tailed distribution for in and out-degree comments amongst users. This is similar to distributions observed across several online settings (Panzarasa et al., 2009). Activity and popularity statistics are likely to have high explanatory potential. Additionally, it signals a non-linear effect of activity and popularity on interaction rates. Significant non-linear effects of activity and popularity have previously been observed in REM models in social interaction networks (Meijerink-Bosman et al., 2023).

Network Decay. Figure 9 demonstrates that network activity is largely short-term, as the number of active users deviates from the cumulative observed users. This suggests that the network's memory should be limited to accurately measure patterns related to recent activity rather than total observed activity. Statistics should be transformed to reflect the temporal nature of the network's recent event observations (Lerner and Lomi, 2023.

Network Attribute	Level	Representation	Half-Life (Hours)	Decay Function
Past Comments on (i, j)	Dyad	n_{ij}	336	$n_{ijt}(t) = n_{ij}(0)\left(\frac{1}{2}\right)^{\frac{t}{336}}$
Past Comments Sent by i	Node	a_{it}	336	$a_i(t) = a_i(0)\left(\frac{1}{2}\right)^{\frac{t}{336}}$
Past Comments Received by i	Node (User)	p_{it}	336	$p_i(t) = n_i(0)\left(\frac{1}{2}\right)^{\frac{t}{336}}$
Total Observed Past Comments	Network	m_t		

Table 1: Network attributes and half-life decay rates

5.7 REM Statistics and Attributes

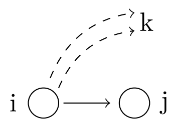
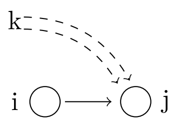
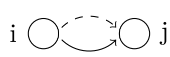
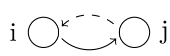
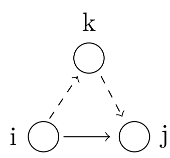
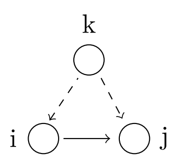
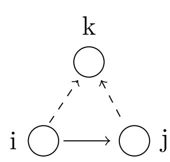
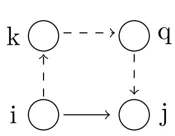
Attributes. In order to calculate an event's statistics u_{ijt}, we must first aggregate fundamental node, dyad and network attributes which encode endogenous information about observed events before time t. These attributes are defined in Table 1. Node and dyad attributes are notably subject to a half-life decay of 336 hours or 2 weeks. Attribute decay is important to specify in models of dynamic network data as the active user-base is in constant fluctuation (Meijerink-Bosman et al., 2023), as observed in figure 9. I've followed the suggestion of Hâncean et al., 2021 that the influence of past events decreases exponentially over time and the rate at which this occurs depends on a half-life parameter. Unfortunately, not much theory exists to make an informed choice on how long past interactions influence future interaction behavior, particularly in online settings (Meijerink-Bosman et al., 2023). Here, I settle on a half-life rate of 336 hours to achieve a medium-term influence of past interactions relative to a full observation period of 17 months. Note that attributes are reset to 0 if their value drops below 0.1.

Statistics. REM statistics are fully developed in Table 2 and measure specific interaction predictors over time. Each statistic undergoes a transformation $log(1 + p)$ where p is the statistic. This handles any zero values and compresses large values while maintaining the proportional differences between smaller values. This transformation adapts to the heavy-tailed comment distribution observed in Figures 6 and 7. After the computation is completed, the statistics are standardized to make their coefficient effects comparable to each other. These statistics are chosen given the theoretical expectation of their significance, which follows from the Literature Review. For example, the clustering statistics set out several possible transitive connections that had significance in past online collective action studies (Chen et al., 2021.

Network-Level Statistics. Two additional statistics, total comments observed and time (seconds) since last comment, represent global-level non-transformed measures of the network at any time t. These are included to test for a temporal and inflationary network effect on comment occurrence rate, while these not expected to have a direct explanatory effect in the partial likelihood equation 3, they nonetheless provide a good robustness check for the model.

Interactions. Seven statistic interactions are formed in Table 3. These represent theoretically motivated and interesting interactions in a large-N dynamic network. Each are possibly related to collective action processes and can indicate how specific interaction patterns change over time. While they do not have theoretic expectations, they test possibilities within an exploratory approach

Computation. Running statistics on 31 million event instances is highly intensive. It is achieved with the Java program *eventnet*, designed by Lerner and Lomi, 2020 and available on its Github. *Eventnet* also performs case control sampling in a single integrated program, outputting final data with computed statistics to be used for likelihood estimation in R with a Cox PH model. My configuration files for *eventnet* are available on this project's Github. The Appendix provides details on statistic run times and

Name	Visual Representation	Interpretation	REM Statistic
User Activity		The likelihood of i sending a comment to j at time t is a function of how many comments i has sent before time t	$u^{A}_{ijt} = \sum_{l} a_{it}$
User Popularity		The likelihood of i sending a comment to j at time t is a function of how many comments j has received before time t	$u^{P}_{ijt} = \sum_{l} p_{it}$
Repetition		The likelihood of i sending a comment to j at time t is a function of how many comments i has sent to j before time t	$u^{Rep}_{ijt} = \sum_{l} n_{ijt}$
Reciprocity		The likelihood of i sending a comment to j at time t is a function of how many comments j has sent to i before time t	$u^{Rec}_{ijt} = \sum_{l} n_{jit}$
Triad-Out-Out		The likelihood of i sending a comment to j at time t is a function of how many comments i has sent to other users k and how many times those users sent a comment to j before time t	$u^{3C}_{ijt} = \sum_{l} n_{ikt} \times n_{kjt}$
Triad-In-Out		The likelihood of i sending a comment to j at time t is a function of how many comments i has received from other users k and how many times those users sent a comment to j before time t	$u^{3io}_{ijt} = \sum_{l} n_{kit} \times n_{kjt}$
Triad-Out-In		The likelihood of i sending a comment to j at time t is a function of how many comments i has sent to other users k and how many times those users received a comment from j before time t	$u^{3oi}_{ijt} = \sum_{l} n_{ikt} \times n_{jkt}$
Four-Cycle		The likelihood of i sending a comment to j at time t is a function of how many comments i has sent to other users k who sent a comment to other users q who sent a comment to j before time t	$u^{4C}_{ijt} = \sum_{l} n_{ikt} \times n_{kqt} \times n_{qjt}$

Note: ○ user, — current (predicted) event, - - - - past event,
Users k and q represent any external user not i or j
Decay rates from table 1 are applied to network attributes used in statistic calculation at time t

Table 2: Model Statistics and Associated Effects

configuration files for replication of my findings.

5.8 Model Types

5.8.1 Baseline

A foundational baseline model predicts the rate of event occurrence throughout the observation period of 17 months. The assumptions of this model are that (i) effects are constant and do not vary over time, (ii) each comment has an equal weighting, and (iii) different comment types have the same effect.

5.8.2 Time-Varying

A time-varying model retains the assumptions of (ii) and (iii) but relaxes the assumption of (i), allowing effects to vary over time. Given the theoretical dynamics of collective action and the high variability in network activity levels over time, we can expect statistic effects to fluctuate over the observation period.

To model time-varying effects, I adapt an approach from Meijerink-Bosman et al., 2023, which utilizes a sliding-window process to isolate effects for specified time periods. In this approach, a window of set

Interaction	Interpretation	Calculation
Activity * Popularity	How user's activity level moderates the effect of their popularity	$u_{ijt}^{A} \times u_{ijt}^{P}$
Activity * Repetition	How user's activity level moderates the effect of comment repetition	$u_{ijt}^{A} \times u_{ijt}^{Rep}$
Activity * Reciprocation	How user's activity level moderates the effect of comment reciprocation	$u_{ijt}^{A} \times u_{ijt}^{Rec}$
Repetition * Reciprocation	How comment repetition interacts with comment reciprocation	$u_{ijt}^{Rep} \times u_{ijt}^{Rec}$
Popularity * Repetition	How user's popularity moderates the effect of comment repetition	$u_{ijt}^{P} \times u_{ijt}^{Rep}$
Popularity * Reciprocation	How user's popularity moderates the effect of comment reciprocation	$u_{ijt}^{P} \times u_{ijt}^{Rec}$
Activity * Four Cycle	How user's activity level moderates the effect of four-cycle network structures	$u_{ijt}^{A} \times u_{ijt}^{4C}$

Table 3: Interaction Effects Interpretation and Calculation

length slides over the observation period. In each slide, the REM is fitted to the entire observed event list, with an interaction between the baseline model predictors and a binary indicator for events within the current window. Together, the time interaction terms from these slides create a picture of how predictors of social interaction change over time (Mulder and Leenders, 2019). Importantly, this approach avoids fitting a REM on a network partition, which would introduce error into effect estimates, as the network structure is greatly altered in partitions. Instead, a sliding-window approach with time-interaction terms but fitted on the whole network enables a credible comparison across the observation period.

I opt for 35 total windows, each with a length of one month and an interval period of two weeks. These dimensions are chosen to balance the specificity of the measurement, the validity of window sample sizes, and the computational feasibility.

5.8.3 Weighted

A weighted baseline model relaxes the assumption of (ii) but retains (i) and (iii). The weighted model defines an event as $e_m = (i_m, j_m, t_m, w_m)$ where w_m is the number of characters in the body of the comment of i_m. In practice, w_m multiplies the effect of an event e_m on the model by w_m. Weight can be interpreted as a measure of a user's effort expended in a single comment, and the weighted model tests whether this dimension changes observed results. The variation in the length of the users' comments is observed in Figure 8.

I do not run a time-varying weighted model on the observation period, although this is a promising avenue for future research. Instead, a weighted model is utilized as a robustness check on the baseline model and acts as a signal for the likelihood that event weighting changes observed trends of interaction.

5.8.4 Comment-Type

An comment-type baseline model relaxes the assumption of (iii), but retains (i) and (ii). In essence, it is two models in one. It develops two dependent event rate variables, one for sub-comments and one for main-comments. It defines an event as $e_m = (i_m, j_m, t_m, s_m)$ where s_m is a binary variable that indicates whether the comment is a sub-comment or not. This model similarly acts as a robustness check on baseline models and signals whether observed trends vary depending on event type. Testing these possibilities develops the exploratory approach of this study.

5.9 Model Selection and Goodness of Fit

In order to determine which models fit the data most accurately, goodness-of-fit (GOF) measures are utilized. Unfortunately, in measuring the GOF of partial likelihood REMs with case-control sampling, traditional approaches are generally not available, and this is an ongoing area of research (Boschi and Wit, 2024). Instead, I use GOF measures from the Cox proportional hazards models: the models which compute likelihood estimation for statistic coefficients. These metrics are not ideal for the model purpose of REM; however, they measure the level of error between a predicted sequence derived from equation 4 and the observed sequence. The metrics generally convey how well the model distinguishes events from nonevents and how well it predicts observed event sequences correctly. The quality of fit metrics are shown in Table 4.

Table 4: Goodness of Fit Measures

Measure	Interpretation
Concordance	Measures the model's ability to discriminate between events and non-events. Values range from 0.5 (no better than chance) to 1 (perfect discrimination).
Likelihood Ratio Test	Compares the fit of the model to a null model. Larger values indicate the model fits the data significantly better than a model with no predictors.
Wald Test	Assesses the overall significance of the model. Large values indicate that the predictors in the model are jointly significant.
Score (Logrank) Test	Evaluates if there are differences in the probability of an event occurring at any time point. Large values suggest the predictors significantly affect the hazard of the event occurring.
Bayesian Information Criterion (BIC)	A criterion for model selection among a finite set of models; lower BIC values indicate a more parsimonious model.

6 Results

6.1 Baseline Effects

Table 5: Baseline Model Results (Entire Observation Period)

	Dependent variable				
	Event Occurrence Rate				
	Model 0	Model 1	Model 2	Model 3	Model 4
User (Node) Effects					
User Activity	0.901***	0.909***	0.909***	1.238***	1.271***
	(0.0004)	(0.0004)	(0.0004)	(0.001)	(0.001)
User Popularity	0.551***	0.745***	0.745***	1.104***	1.039***
	(0.0004)	(0.0004)	(0.0004)	(0.001)	(0.001)
Dyad Effects					
Repetition		0.082***	0.082***	-0.096***	0.637***
		(0.0003)	(0.0003)	(0.001)	(0.001)
Reciprocation		-0.191***	-0.191***	0.141***	0.145***
		(0.0002)	(0.0002)	(0.0004)	(0.001)
Network Effects					
Total Comments Observed			21.450***	16.590***	4.535*
			(1.882)	(1.882)	(1.884)
Time Since Last Comment			0.005***	0.002***	0.002***
			(0.0005)	(0.001)	(0.001)
Clustering Effects					
Triad Out-Out				-0.206***	-0.052***
				(0.0004)	(0.0005)
Triad Out-In				0.220***	0.071***
				(0.001)	(0.001)
Triad In-Out				-0.453***	-0.146***
				(0.001)	(0.001)
Four Cycle				-0.127***	0.392***
				(0.001)	(0.001)
Interactions					
Activity * Popularity					-0.224***
					(0.001)
Activity * Repetition					-0.213***
					(0.0003)
Activity * Reciprocation					0.048***
					(0.0003)
Repetition * Reciprocation					0.008***
					(0.0002)
Popularity * Repetition					-0.048***
					(0.0003)
Popularity * Reciprocation					-0.065***
					(0.0003)
Activity * Four Cycle					-0.170***
					(0.0004)
Observations	31,380,006	31,380,006	31,380,006	31,380,006	31,380,006
Number of events (N)	5,820,604	5,820,604	5,820,604	5,820,604	5,820,604
Goodness of Fit					
Concordance	0.96	0.959	0.959	0.96	0.969
Likelihood Ratio Test	15,820,305***	16,691,213***	16,691,451***	18,369,461***	20,210,590***
Wald Test	16,589,053***	15,918,372***	15,918,400***	14,054,705***	12,069,471***
Score (Logrank) Test	30,728,660***	31,895,351***	31,895,369***	35,852,906***	38,534,051***
BIC	79,217,959***	78,347,082***	78,346,876***	74,827,908***	76,398,928***
	(df = 2)	(df = 4)	(df = 6)	(df = 10)	(df = 17)

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Table 5 reports the results of 5 baseline models for the entire observation period. Each model [0-4] increases complexity and successively adds groups of event statistics. Explanatory power of statistic groupings can be measured from successive changes in the goodness of fit of each model.

An important trend to note in Table 5 is that more complex models that incorporate more statistics achieve better goodness of fit. For example, the score (logrank) test increases from 30,728,660 to 38,534,051 from model 0 to model 4.

An exception to this is the addition of the Network Statistics in Model 2, where no change in GOF is observed between Models 1 and 2. This suggests that the time since the last comment and total comments observed have no explanatory power, despite their high statistical significance (low p-value). These statistics are omitted from future models and can be reasonably ignored.

To interpret the coefficients, the most direct indicator is their positive or negative weighting, where positive values suggest a positive effect of the statistic on event occurrence rate across users and negative values suggest a negative effect. Since statistics have been standardized, the weightings of the effects can be directly compared. A challenge of interpretation is that, since statistics are subject to multiple transformations, including time decay and $log(1p)$, the interpretation of observed values is not straightforward. For example, a 1-unit increase in popularity does not signal an easily interpretable description in an observable network change. Given this, the signage of the coefficients is important to understand relative effects.

REMs are log-likelihood models and can generate fitted probabilities for dyadic combinations of statistic values. Through exponentiation, the effects in Table 5 are transformed into conditional likelihood effects on comment occurrence rate. Exponentiated coefficients of Model 4 are reported in the Appendix. For these, a zero effect is an effect equal to 1.

Models 0, 3 and 4 are selected to proceed to the time-varying effect estimation stage. These 3 models respectively represent the least complex model, the model with the best complexity-GOF balance, and the model with the best GOF overall (performance). By estimating time-varying effects across multiple models, the robustness of the findings is better supported. Note that baseline models are selected on their complexity-GOF trade-off. Model 4 is plausibly strictly the best model and is mainly used in the following analysis.

6.2 Time-Varying Effects

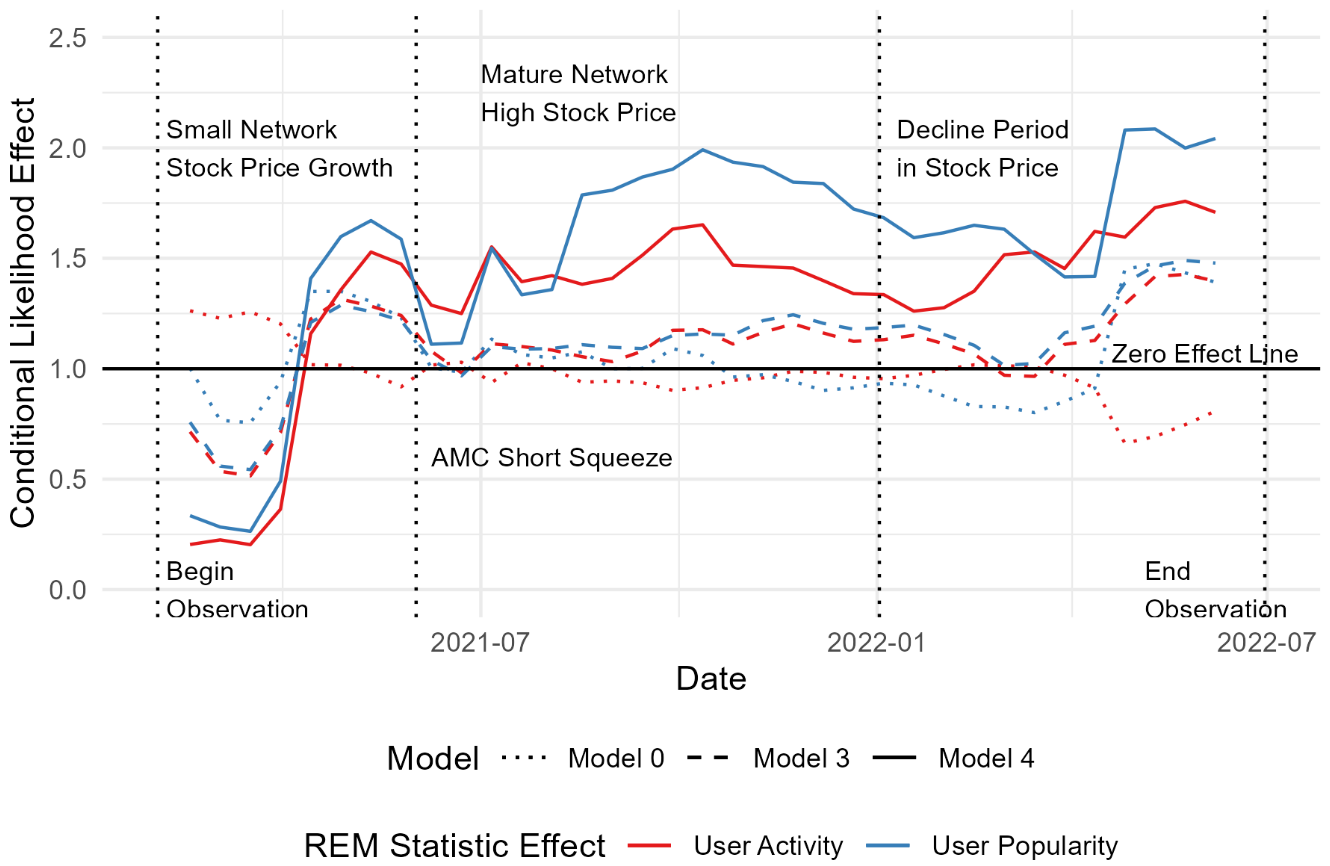


Figure 10: Model 0, 3 and 4 effects of Activity and Popularity over time

Activity and Popularity. Figure 10 reports the variation in the effects of activity and popularity statistics over time relative to their baseline values in Table 5. A notable observation is the magnitude of the temporal effects. At its highest, a one unit increase in popularity increases a user's rate of comment occurrence by approximately 200% compared to a user outside that time period, all other statistics being equal. Using the logic of fitted probability, two users each with a popularity of 1 at this time t would have a 4x greater likelihood of forming a connection than a user outside of this time period. A similar trend is observed for the activity statistic effect.

In Model 4, the effect of popularity and activity increases during the climax period of the movement and remains stable at a high level during the decline period. This indicates an active core of returning users during the decline period, who are possibly engaged in a 'constant mobilization' process after movement climax.

Figures 11 to 13 omit timeline annotations but visualize the same observation period.

Figure 11: Model 3 and 4 Effects of Reciprocation and Repetition over time

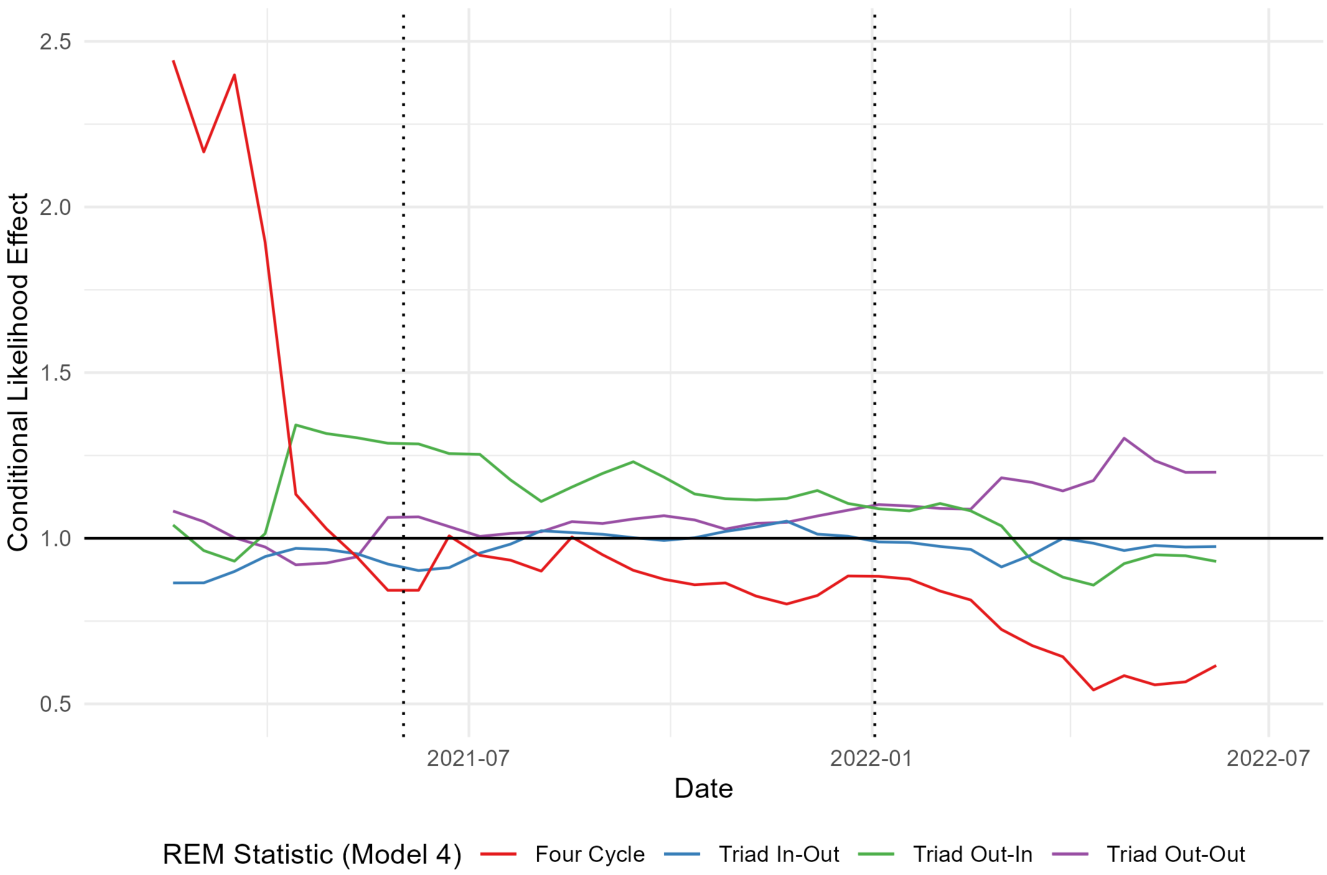


Figure 12: Model 4 Effects of Clustering Statistics over Time

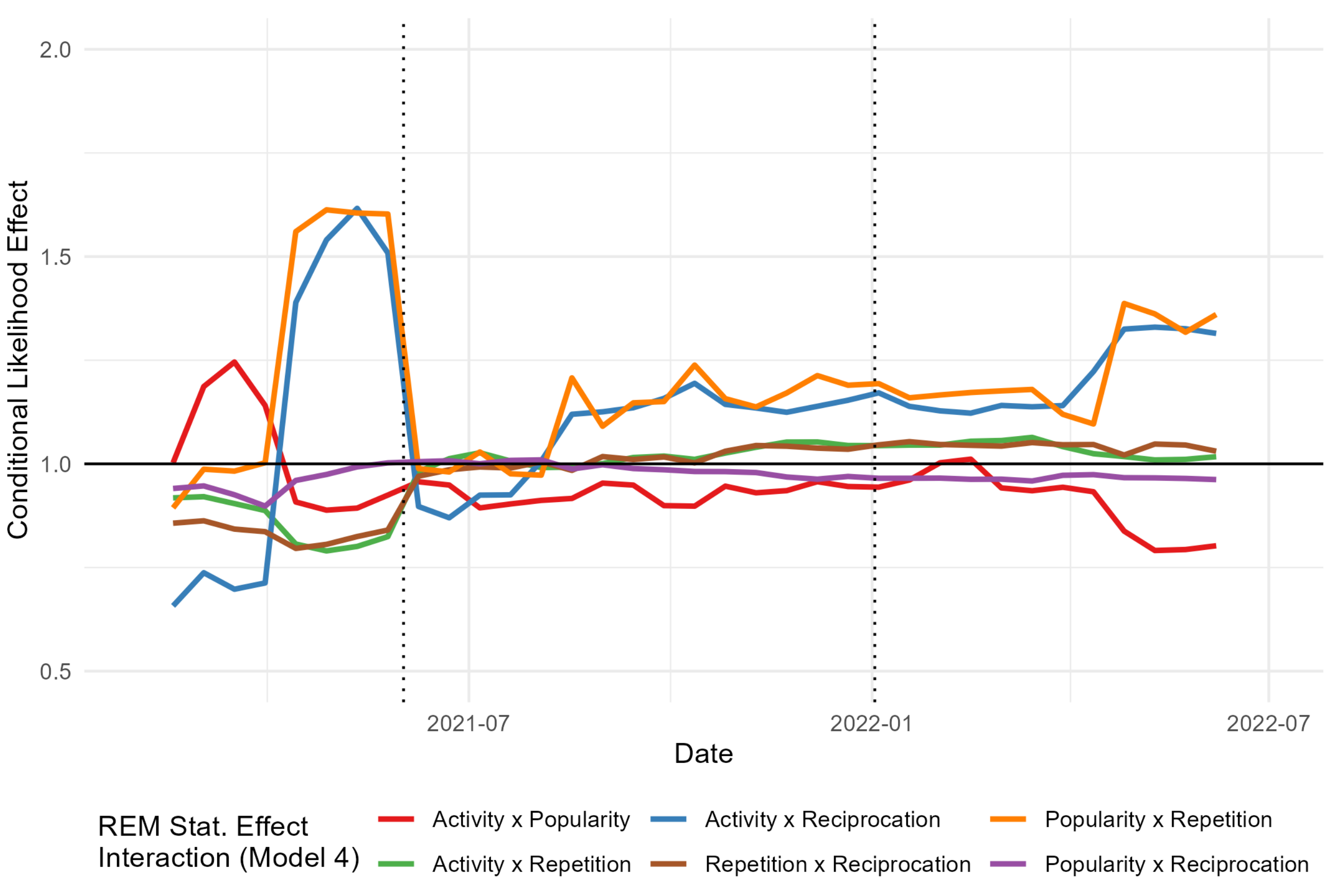


Figure 13: Model 4 Interaction Effects over time

Reciprocity and Reciprocation. Figure 11 suggests that reciprocation and repetition have a positive effect on event occurrence during the growth period, before declining to a near-zero effect, then declining further to negative in the decline period. This indicates a loss of mutual and repeated user engagement throughout the observation period, indicating that dyadic interaction patterns decline in effect.

Clustering and Centralization. Figure 12 indicates that clustering effects are inconsistent across the observation period. Though it was expected that clustering effects would be higher during the growth period, this is only true for the four cycle statistic, which is exhibits an extremely positive effect, while others are far closer to zero effect. Interestingly, four cycle declines during the climax and decline periods and becomes strongly negative, suggesting that an explosive growth in user base, followed by a gradual decline can actually reduce network clustering. Overall, clustering results do not suggest a particularly centralized network at any point, although the growth-period network is far more clustered than any other. Nonetheless, these observations deviate substantially from theoretical predictions.

Interactions. Figure 13 reports the changes in interaction coefficients over time. Several notable shifts are observed. Firstly, most interactions approach near-zero effect at the time of the short-squeeze, possibly affected by an in-flux of new users which resets previous interaction patterns.

During the growth period, popularity to repetition and activity to reciprocation exhibit strong positive effects. These suggest that frequent engagement with the network and receiving engagement from others increase the likelihood that a user reciprocates engagement and repeats engagement. This suggests that the start of the climax period led to greater dyadic engagement, as expected.

During the climax period, interactions do decline to near zero, but return to high magnitude later on, with reciprocation and repetition dyad effects further exhibiting high positive effects into the movement decline period, suggesting that these interactions involving the dyadic effects are resistant to movement decline.

6.3 Weighted Model

Table 6: Weighted Model Results (Entire Observation Period)

	Dependent variable
	Event Occurrence
Activity	1.281***
	(0.002)
Popularity	1.348***
	(0.002)
Reciprocation	0.325***
	(0.001)
Repetition	0.821***
	(0.001)
Four Cycle	0.460***
	(0.002)
Triad Out-Out	0.116***
	(0.002)
Triad In-Out	-0.064***
	(0.001)
Triad Out-In	0.196***
	(0.001)
I(Activity^2)	0.069***
	(0.001)
I(Popularity^2)	-0.037***
	(0.001)
Activity * Popularity	-0.224***
	(0.001)
Activity * Repetition	-0.331***
	(0.001)
Activity * Reciprocation	-0.030***
	(0.001)
Popularity * Repetition	-0.187***
	(0.0004)
Popularity * Reciprocation	-0.111***
	(0.001)
Activity * Triad Out-Out	-0.047***
	(0.001)
Popularity * Triad Out-Out	-0.065***
	(0.001)
Activity * Four Cycle	-0.199***
	(0.001)
Observations	31,379,985
Number of Events	5,820,604
Concordance	0.968
Wald Test	9,203,952***
	(df = 18)
Score (Logrank) Test	39,462,721***
	(df = 18)
Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$	

The most important finding of the baseline weighted model in Table 6 is its comparison with the baseline results in Table 5. In this regard, the findings are very similar. The only coefficient that switches effects from positive to negative or vice versa is Triad-Out-Out, yet it remains a near-zero, small-magnitude effect. Some interaction terms also switch signage; however, these also have small effects to begin with. This finding improves the confidence in the baseline model results, as it appears that comment weighting as measured by comment character length does not substantially alter observed interaction dynamics amongst the Reddit users. There is no evidence to suggest that time-varying results would be altered by accounting for event weighting, at least through a comment-length weighting.

6.4 Comment Type Model

Table 7 suggests that the effect of statistics strongly deviate for event occurrence rates for the sub- and main comments. Although the table presents a number of interesting coefficient results, the primary point of focus is how both models relate to the baseline Model 4 (i.e. the model with the best GOF). It appears that the rate of the main comment events is more similar to the baseline trends, suggesting that the main comments are more likely to drive the observed trends from the baseline models. The subcomment coefficients are more negatively signed than in the baseline models. This implies that the main comments drive the positive effects in the baseline models. It is not to suggest that exempting the sub comments would have made a more realistic data network. Rather, this observation reinforces the claim that even small changes in online communication settings cascade into estimation of predictors on interactions.

Table 7: Comment-Type Model Results (Entire Observation Period)

	Dependent variable: Event Occurrence	
	Main Events	Sub Events
Main Activity	1.915***	0.148***
	(0.003)	(0.002)
Sub Activity	0.137***	0.846***
	(0.002)	(0.002)
Main Popularity	3.663***	0.072***
	(0.003)	(0.0005)
Sub Popularity	-0.043***	-0.188***
	(0.001)	(0.001)
Repetition Main Out	-1.690***	0.003***
	(0.010)	(0.0003)
Repetition Sub Out	0.179***	0.130***
	(0.008)	(0.001)
Reciprocation Main	0.123***	0.071***
	(0.001)	(0.0003)
Time Since Last Comment	0.002**	0.005***
	(0.001)	(0.001)
Main 4 Cycle	0.045***	-0.173***
	(0.002)	(0.001)
Sub 4 Cycle	-0.050***	2.068***
	(0.001)	(0.002)
Main Triangle	0.005***	0.009***
	(0.001)	(0.001)
Sub Triangle	-0.058***	-0.289***
	(0.002)	(0.002)
I(Main Activity^2)	-0.262***	
	(0.001)	
I(Main Popularity^2)	-0.687***	
	(0.001)	
I(Sub Activity^2)		0.074***
		(0.001)
I(Sub Popularity^2)		0.005***
		(0.0001)
Main Activity * Main Popularity	-0.747***	
	(0.001)	
Main Activity * Repetition Main Out	0.171***	
	(0.001)	
Main Activity * Reciprocation Main	-0.020***	
	(0.0004)	
Repetition Main Out * Reciprocation Main	0.005***	
	(0.0002)	
Main Popularity * Repetition Main Out	0.551***	
	(0.001)	
Main Popularity * Reciprocation Main	-0.049***	
	(0.001)	
Main Activity * Main 4 Cycle	0.021***	
	(0.001)	
Main Activity * Sub Activity	-0.041***	-0.088***
	(0.001)	(0.001)
Main Activity * Sub Popularity	0.017***	-0.035***
	(0.0004)	(0.0004)
Main Activity * Sub Triangle	0.008***	0.072***
	(0.001)	(0.001)
Sub Activity * Sub Popularity		0.054***
		(0.001)
Sub Activity * Repetition Sub Out		-0.002***
		(0.001)
Repetition Sub Out * Reciprocation Sub		-0.003***
		(0.0001)
Sub Popularity * Repetition Sub Out		-0.001**
		(0.0002)
Sub Activity * Sub 4 Cycle		-0.572***
		(0.001)
Observations	17,505,623	15,733,253
Number of Events	3,126,258	2,756,860
Concordance	0.992	0.969
Wald Test	4,004,338***	6,396,129***
Score (Logrank) Test	23,689,857***	20,108,044***

7 Discussion

My findings present a mixed bag of evidence for theoretical predictions on user interactions in collective action networks. Considering the time-constant effects of the baseline model 4 in Table 5, the Reddit network is somewhat typical, with activity, popularity, repetition, and four-cycle clustering having strongly positive effects and other statistics near zero. This appears to align with theoretical behaviors for networks involved in collective aims, where movement inertia is derived from returning user engagement, and rich-get-richer dynamics create both popular users who engage more often and repeat interaction with the same users over time. However, without a control network to compare against, such as another similar subreddit, it is difficult to determine if these baseline effects are generated by the short-squeeze context or if they could be observed on any similar subreddit (or online forum). This issue of inference is a key limitation and motivated the exploratory approach to the network data; it also relates to the external validity of findings, as it is not yet clear if observations in this case study apply to other contexts.

However, the functional role of the baseline models is to set up the time-varying analysis, in which the findings generate significant implications for discussion. In general, the degree of variation in statistic effects over time is substantial and magnitude of effects are high, even amongst effects that appear visually close to the zero effect line, following from the large scale of the y-axes. From the conception of online networks as organizational agents, this finding suggests that online collective action networks are highly dynamic and prone to (sometimes quick) changes in underlying social formations. It is not clear what causes these changes. On one hand, they could result from an influx in new users who moderate the predictors of interaction, or they could equally result from changes in movement sentiment, relating to shifts in stock price or news developments. What is notable is that stock price did not have as a large of a correlation with shifting predictors as the theoretical approach predicted. While some large shifts were observed around the time of the short-squeeze, effects drifted away from stock price trends in the period afterwards.

In interpreting the volatility of multiple effects over time, it is tempting to conclude that the 'staying-power' of the network's underlying social formations was low, leading to the observed fluctuations; however, certain observed effects contradict this claim, notably activity and popularity. The possibility that a set of committed users kept these statistic effects high in movement decline is compelling. In fact, this a key finding of the longitudinal approach, that these effects did not decline in the face of declining confidence in the movement. However, beyond this, there is not enough evidence in the form of aligned trends over different periods to posit any kind of underlying social structure or defining network topography in different periods of the movement life-cycle.

A final discussion point is on the comment type model. It is likely that the observed increase in main event coefficients is due to their more substantive role on Reddit, as they typically act as the first comments to lengthy submission threads. Hence, they may have a more significant role in interactions related to collective aims in the network.

8 Conclusion

This paper demonstrates the powerful analytical capabilities of relational event modeling (REM) for examining fine-grained online interaction data. By applying REM to study a collective action network on

Reddit, I illustrate how this approach can uncover the dynamic predictors of user interaction in large-scale, rapidly evolving digital communities. Through sampling techniques, model formulations, and statistic transformations such as time-decay functions, I adapted REM to accommodate the challenges posed by large-N, highly dynamic networks. Building on baseline models, I used a sliding window approach with time-period interaction terms to capture the temporal evolution of predictor effects. This methodological framework, particularly when combined with specialized software such as *eventnet*, offers a versatile toolkit for analyzing any large-scale dynamic network with time-stamped relational events.

The empirical findings from the r/amcstock case study reveal the complex and fluid nature of online collective action networks. Rather than identifying a clear, stable network structure, the results highlight significant temporal fluctuations in predictor effects. However, user activity and popularity emerge as relatively consistent drivers of interaction across the movement life-cycle. Notably, shifts in stock price does somewhat correlate with substantial changes in predictor effects, underscoring the sensitivity of these networks to external factors. These observations provide valuable insights into the malleability of collective action networks in digital spaces, though further research is needed to establish the generalizability of these findings beyond the specific context of Reddit.

While this study advances our understanding of online collective action dynamics, it also faces certain limitations. A key challenge was the lack of well-suited goodness-of-fit metrics for partial likelihood REMs utilizing case-control sampling. Although progress is being made in this area, the reliance on suboptimal Cox proportional hazards metrics remains a constraint. Additionally, the results are likely influenced by various network construction parameters, such as the chosen half-life decay rate of 336 hours. While sensitivity analyses were conducted, determining the precise impact of these parameters on statistic coefficients presents an ongoing challenge.

Looking ahead, the flexibility of REMs opens up numerous avenues for future research. Incorporating exogenous statistics, such as stock price indicators, could provide deeper insights into the interplay between online dynamics and external events. Integration with text and sentiment analysis techniques offers promising opportunities to measure user archetypes and test for homophily and clustering effects, leveraging the rich metadata available on platforms like Reddit. Further development of event-type models to examine time-varying effects on different types of user interactions (e.g., main comments vs. replies) could yield more nuanced understanding of communication patterns.

Expanding this approach to other collective action contexts with measurable success metrics - such as protest turnout rates or social media engagement levels - represents a crucial next step. By operationalizing these outcomes for longitudinal analysis, researchers can deepen our understanding of the relationship between online activity and real-world movement evolution. This extension has the potential to advance future study, bridging the gap between digital interactions and tangible social outcomes. In conclusion, this study not only demonstrates the analytical power of relational event modeling for examining online collective action, but also lays the groundwork for future research exploring the complex interplay between digital network dynamics and real-world social movements. As online platforms continue to play an increasingly central role in collective action, such methodological innovations will be crucial for unraveling the mechanisms driving these important social phenomena.

9 Works Cited

References

Aharon, D. Y., Kizys, R., Umar, Z., & Zaremba, A. (2023). Did David win a battle or the war against Goliath? Dynamic return and volatility connectedness between the GameStop stock and the high short interest indices. *Research in International Business and Finance*, *64*, 101803.

Ahuja, M., Patel, P., & Suh, A. (2018). The Influence of Social Media on Collective Action in the Context of Digital Activism: An Affordance Approach.

Anand, A., & Pathak, J. (2021, December). The Role of Reddit in the GameStop Short Squeeze.

Anduiza, E., Cristancho, C., & Sabucedo, J. M. (2014). Mobilization through online social networks: The political protest of the indignados in Spain [Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2013 *Information, Communication & Society*, *17*(6), 750–764.

Bennett, W. L., & Segerberg, A. (2012). The Logic of Connective Action: Digital media and the personalization of contentious politics [Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2012.670661]. *Information, Communication & Society*, *15*(5), 739–768.

Boschi, M., & Wit, E.-J. C. (2024, July). Goodness of fit of relational event models.

Boylston, C., Palacios, B., Tassev, P., & Bruckman, A. (2021, January). *WallStreetBets: Positions or Ban.*

Bradley, D., Hanousek Jr., J., Jame, R., & Xiao, Z. (2021, March). Place Your Bets? The Market Consequences of Investment Research on Reddit's Wallstreetbets.

Brunswicker, S., & Schecter, A. (2019). Coherence or flexibility? The paradox of change for developers' digital innovation trajectory on open platforms. *Research Policy*, *48*(8), 103771.

Butts, C. T. (2008). A Relational Event Framework for Social Action. *Sociological Methodology*, *3*, 155–200.

Chen, Z., Oh, P., & Chen, A. (2021). The Role of Online Media in Mobilizing Large-Scale Collective Action. *Social Media + Society*, *7*(3), 20563051211033808.

Chung, J. (2021). WSJ News Exclusive | Melvin Capital Lost 53% in January, Hurt by GameStop and Other Bets. *Wall Street Journal*.

Conover, M. D., Ferrara, E., Menczer, F., & Flammini, A. (2013). The Digital Evolution of Occupy Wall Street. *PLOS ONE*, *8*(5), e64679.

Dolata, U., & Schrape, J.-F. (2016). Masses, Crowds, Communities, Movements: Collective Action in the Internet Age. *Social Movement Studies*, *15*(1), 1–18.

Dumitrica, D., & Felt, M. (2020). Mediated grassroots collective action: Negotiating barriers of digital activism [Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2019.1618891]. *Information, Communication & Society*, *23*(13), 1821–1837.

Flanagin, A. J., Stohl, C., & Bimber, B. (2006). Modeling the Structure of Collective Action1 This material is based upon work supported by the National Science Foundation under Grant No. 0352517. The authors are equal contributors to this article. [Publisher: Routledge _eprint: https://doi.org/10.1080/036377506 *Communication Monographs*, *73*(1), 29–54.

Hâncean, M.-G., Lerner, J., Perc, M., Ghiţă, M. C., Bunaciu, D.-A., Stoica, A. A., & Mihăilă, B.-E. (2021). The role of age in the spreading of COVID-19 across a social network in Bucharest. *Journal of Complex Networks*, *9*(4), cnab026.

Himelboim, I. (2017, August). Social Network Analysis (Social Media).

Hu, D., Jones, C. M., Li, S., Zhang, V., & Zhang, X. (2021, March). The Rise of Reddit: How Social Media Affects Retail Investors and Short-sellers' Roles in Price Discovery.

Johansson, H., & Scaramuzzino, G. (2023). Digital resource abundance: How social media shapes success and failure of online mobilisation. *Convergence*, *29*(3), 586–601.

Kavada, A. (2015). Creating the collective: Social media, the Occupy Movement and its constitution as a collective actor [Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2015.1043318]. *Information, Communication & Society*, *18*(8), 872–886.

Kow, Y. M., Nardi, B., & Cheng, W. K. (2020). Be Water: Technologies in the Leaderless Anti-ELAB Movement in Hong Kong. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12.

Lerner, J., & Lomi, A. (2020). Reliability of relational event model estimates under sampling: How to fit a relational event model to 360 million dyadic events. *Network Science*, *8*(1), 97–135.

Lerner, J., & Lomi, A. (2022). A dynamic model for the mutual constitution of individuals and events. *Journal of Complex Networks*, *10*(2), cnac004.

Lerner, J., & Lomi, A. (2023). Relational hyperevent models for polyadic interaction networks. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *186*(3), 577–600.

Lerner, J., Lomi, A., Mowbray, J., Rollings, N., & Tranmer, M. (2021). Dynamic network analysis of contact diaries. *Social Networks*, *66*, 224–236.

Lucchini, L., Aiello, L. M., Alessandretti, L., De Francisci Morales, G., Starnini, M., & Baronchelli, A. (2022). From Reddit to Wall Street: The role of committed minorities in financial collective action. *Royal Society Open Science*, *9*(4), 211488.

Mancini, A., Desiderio, A., Di Clemente, R., & Cimini, G. (2022). Self-induced consensus of Reddit users to characterise the GameStop short squeeze. *Scientific Reports*, *12*(1), 13780.

Marwell, G., Oliver, P. E., & Prahl, R. (1988). Social Networks and Collective Action: A Theory of the Critical Mass. III. *American Journal of Sociology*, *94*(3), 502–534.

Meijerink-Bosman, M., Back, M., Geukes, K., Leenders, R., & Mulder, J. (2023). Discovering trends of social interaction behavior over time: An introduction to relational event modeling. *Behavior Research Methods*, *55*(3), 997–1023.

Nie, Z., Waheed, M., Kasimon, D., & Wan Abas, W. A. B. (2023). The Role of Social Network Analysis in Social Media Research [Number: 17 Publisher: Multidisciplinary Digital Publishing Institute]. *Applied Sciences*, *13*(17), 9486.

Panzarasa, P., Opsahl, T., & Carley, K. (2009). Patterns and Dynamics of Users' Behavior and Interaction: Network Analysis of an Online Community. *JASIST*, *60*, 911–932.

Pitcher, S., & RufaroMuchena, M. (2022). *Using Netnography to Map Issue Communities on Twitter During South Africa’s #FeesMustFall Student Protests*. SAGE Publications, Ltd.

Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Social Media + Society*, *7*(2), 20563051211019004.

Quintane, E., Conaldi, G., Tonellato, M., & Lomi, A. (2014). Modeling Relational Events: A Case Study on an Open Source Software Project [Publisher: SAGE Publications Inc]. *Organizational Research Methods*, *17*(1), 23–50.

The SAGE Handbook of Social Media. (2024, June).

Scott, J., & Carrington, P. J. (2011, May). *The SAGE Handbook of Social Network Analysis* [Google-Books-ID: mWlsKkIuFNgC]. SAGE.

Semenova, V., & Winkler, J. (2023, August). Social contagion and asset prices: Reddit's self-organised bull runs [arXiv:2104.01847 [cs, econ, q-fin]].

Siegel, D. A. (2009). Social Networks and Collective Action. *American Journal of Political Science*, *53*(1), 122–138.

Sikandar, S., & Fatima, A. (2024). Mobilizing for Justice: Social Media as a Transformational Tool for Student Activism in Pakistan. In T. Saeed, R. Iyengar, M. A. Witenstein, & E. J. Byker (Eds.), *Exploring Education and Democratization in South Asia: Research, Policy, and Practice* (pp. 221–237). Springer International Publishing.

Spier, S. (2017, March). *Collective Action 2.0: The Impact of Social Media on Collective Action* [Journal Abbreviation: Collective Action 2.0: The Impact of Social Media on Collective Action Pages: 183 Publication Title: Collective Action 2.0: The Impact of Social Media on Collective Action].

van Stekelenburg, J., Roggeband, C., & Klandermans, B. (Eds.). (2013). *The Future of Social Movement Research: Dynamics, Mechanisms, and Processes*. University of Minnesota Press.

Wang, R., & Chu, K.-H. (2019). Networked publics and the organizing of collective action on Twitter: Examining the #Freebassel campaign [Publisher: SAGE Publications Ltd]. *Convergence*, *25*(3), 393–408.

Xiaolong, Z., Tian, H., Wan, Z., Wang, X., Zeng, D. D., & Wang, F.-Y. (2021). Game Starts at GameStop: Characterizing the Collective Behaviors and Social Dynamics in the Short Squeeze Episode. *IEEE Transactions on Computational Social Systems*, *PP*, 1–14.

Young, A., Selander, L., & Vaast, E. (2019). Digital organizing for social impact: Current insights and future research avenues on collective action, social movements, and digital technologies. *Information and Organization*, *29*(3), 100257.

10 Appendix

Statistic calculations were run on a computer with 16GB RAM and took approximately 2 days to finish. Time-varying effect model estimations took approximately 3 days. All code is available on the Github.

10.1 Baseline Model 4 Exponentiated Coefficients

Table 8: Model 4 Exponentiated Coefficients for Fitted Probability Estimation

Variable	Exp(Coef)	p-value
user_activity	7.6723	<2e-16
user_popularity	6.2278	<2e-16
comment_repetition	1.5577	<2e-16
comment_reciprocation	1.0001	0.8992
total.comments.observed	43.0253	0.0458
time_since_last_comment	1.0013	0.0148
triad_out_out	0.9660	<2e-16
triad_out_in	1.1281	<2e-16
triad_in_out	0.9895	<2e-16
four_cycle	0.8307	<2e-16
user_activity^2	0.7104	<2e-16
user_popularity^2	0.7664	<2e-16
user_activity * user_popularity	0.7902	<2e-16
user_activity * comment_repetition	0.8906	<2e-16
user_activity * comment_reciprocation	1.0274	<2e-16
comment_repetition * comment_reciprocation	1.0093	<2e-16
user_popularity * comment_repetition	0.8872	<2e-16
user_popularity * comment_reciprocation	1.0179	<2e-16
user_activity * four_cycle	1.0127	<2e-16

10.2 Sample Script for Cox PH Likelihood Estimation

```
event.surv <- Surv(time = rep(1,nrow(events)), event = events$IS_OBSERVED)

model.4 <- coxph(event.surv ~

                  + user_activity

                  + user_popularity

                  + comment_repetition

                  + comment_reciprocation

                  + total.comments.observed

                  + time_since_last_comment

                  + triad_out_out

                  + triad_out_in

                  + triad_in_out

                  + four_cycle

                  # Non Linear Effects

                  + I(user_activity^2)

                  + I(user_popularity^2)

                  # Interactions

                  + user_activity*user_popularity

                  + user_activity*comment_repetition

                  + user_activity*comment_reciprocation

                  + comment_repetition*comment_reciprocation

                  + user_popularity*comment_repetition

                  + user_popularity*comment_reciprocation

                  + four_cycle * user_activity

                + strata(TIME_UNIT)

                  , data = events,
```

```
control = coxph.control(iter.max = 100))
```