2022

Not All Influence is Born Equal: On the Effects of Various Types of Behavioral Influence Relationships on Social Media

Chathurani Senevirathna
University of Central Florida

NOT ALL INFLUENCE IS BORN EQUAL: ON THE EFFECTS OF VARIOUS TYPES OF BEHAVIORAL INFLUENCE RELATIONSHIPS ON SOCIAL MEDIA

by

CHATHURANI SENEVIRATHNA
B.S. University of Peradeniya, 2011
M.S. University of Central Florida, 2017

A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2022

Major Professor: Ivan Garibay

ABSTRACT

Typically, online social influence is analyzed using a single metric approach. However, social influence is not monolithic; different users exercise different influences in different ways, and influence is correlated with the user and content-specific attributes. One such attribute could be whether the action is an initiation of a new post, a contribution to a post, or a sharing of an existing post. Thus, this dissertation uses this platform-independent action classification and models the influence as multiple entities and examines social networks through the perspective of behavioral influence propagation.

Two empirical studies are present in this dissertation. The first study presents a novel method for tracking these influence relationships over time, which we call influence cascades, and presents a visualization technique to understand these cascades better. These influence patterns are investigated within and across online social media platforms using empirical data and comparing to a scale-free network as a null model. Our results show that characteristics of influence cascades and patterns of influence are, in fact, affected by the platform and the community of the users. The second study applies the same framework to re-construct interconnected social networks and explores the significance of cross-platform influence on social media users in the influence process. In particular, we explore the social dynamics of users with a higher number of social influence relationships across platforms, which we call interface users, and those with fewer social influence relationships across platforms, which we call core users. Our results find that interface users are more vulnerable to being influenced and influential than core users. Further, our results show that the interface users who are influenced to do initiation action exert significantly more influence on others than those who are influenced to contribute.

ACKNOWLEDGMENTS

This research would not have been possible to complete without the funding and data we received through the DARPA SocialSim program (DARPA program HR001117S0018 (FA8650-18-C-7823)) and the computational resources provided through the AWS research grant.

I would like to thank my advisor, Dr. Ivan Garibay, for providing me with constant support, guidance, and advice to grow as an independent researcher and complete this dissertation successfully. I appreciate it he allowed me to enter a new field that I was not exposed to before by accepting me as his one of graduate students. Furthermore, he supported me in all the possible ways throughout the past years, whether it was providing funding to participate in conferences, having the discussion and providing feedback to formulate the research question and improve the methodology, or providing all the computational resources that were needed to implement the experiments. I also would like to thank my co-advisor, Dr. William Rand, for providing me with constant support to improve my research work throughout these years. He always supported me in whatever possible ways, including helping formulate research questions, discussing the methodology and experiments performed in this dissertation, providing feedback, and reviewing my research drafts. Further, I would like to thank the rest of the committee, Dr. Vladimir Boginski, Dr. Luis Rabelo, and Dr. Alexander Mantzaris, for their guidance and advice throughout this dissertation process.

I want to thank all the lab mates in Complex Adaptive Systems Labs. Their encouragement and all the help provide me with great support. A big thank you to Chathika Gunaratne, previously my research group fellow and now Dr. Chathika Gunarathna, for mentoring me all these years with your experience. His constant support helps me significantly improve my skills and knowledge and complete this dissertation successfully. Also, I am grateful to other fellow Ph.D. students in my research group, Chathura Jayalath and Nisha Baral, for working alongside me on the DARPA

SocialSim challenges and encouraging and helping me throughout this journey.

In addition, I am most thankful to my parents and my brothers and sister, and my second brother's family back in Kandy, Sri Lanka, for providing me with constant support, encouragement, and unconditional love all these years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

C Contribution to an existing conversation or a post

CVE Cyber-vulnerabilities

GH GitHub

I Initiation of a conversation or a post

JS Jensen–Shannon

OSM Online Social Media

S Sharing of an existing post between conversations without changing the content

TW Twitter

CHAPTER 1: INTRODUCTION

Online social media (OSM) platforms allow people to interact and maintain their existing social networks as well as connect and exchange information with strangers based on their interests such as politics, entertainment, academic field, shopping, etc. [1]. With the rapid increase of online social media usage, social platforms now represent a large portion of daily communication and play a major role in information diffusion throughout society. The major driving force of this information diffusion can be identified as the social influence exerted or experienced by users in the network. Social influence in OSM can be defined as the ability of a user's action to affect the actions of other users. We refer to such occurrences as social influence relationships. However, in most cases, these relationships are asymmetric. A person who influences other users is referred to as an influencer and the person being influenced is referred to as an influencee. Social influence has been widely studied in many fields including marketing [2, 3, 4, 5, 6], political science [7], human and animal behavior [8, 9, 10, 11], and communication [12, 13].

In OSM, we can classify user actions into three types: (1) initiation of a conversation or a post (I), (2) contribution to an existing conversation or a post (C), or (3) sharing of an existing post between conversations without changing the content (S). Since we will use these three actions and this framework throughout this work, we will refer to it as the ICS classification. However, most existing studies on social influence in OSM assume an implicit monolithic notion of influence, i.e., that a user's influence is the same across all action types. In particular, in most of the previous work, social influence is measured using traditional influence measurements such as centrality measures, link topological measures, and coreness-based measures [14, 15], which were focused on the notion of centrality, or structural influence. For instance, centrality measurements such as the degree, closeness, betweenness, eigenvector, Katz centrality measurements, and their variations measure the importance of a given node's position in the network to propagate information, and the notion of importance varies for each measurement [16, 14, 17, 18, 19, 20, 21]. However, except for eigenvector centrality, these measurements treat the contribution of nodes to the measurement equally. Even in the case of eigenvector centrality, the only factor taken into account is the centrality of the connected nodes to a focal node. Further, the link topological ranking measurements including Hyperlink-Introduced Topic Search (HITS) [22] and PageRank algorithm [23] and PageRank-like algorithms such as TunkRank [24], InfluenceRank [25], SpreadRank [26], TURank [27], TwitterRank [28], InfRank [29] assume that the value of a node depends on the number of connections they have. Furthermore, the coreness-based measures such as the k-shell method and its variants [9, 30, 31, 32] assume coreness of a node is more important than their connectivity or centrality to predict the best spreaders accurately. Therefore, link topological ranking, and coreness-based measures are also structural-based measurements. In addition, some recent work used entropy-based measurements such as graph entropy [33], friend entropy, and interaction frequency entropy [34] to measure the social influence in OSM, as entropy captures the uncertainty and complexity of the social influence effectively [14]. However, these measurements are also defined based upon the network structure. As a result, these measurements cannot fully capture the behavioral influence.

However, in reality, there are differences in how users influence others through initiation, contribution, and sharing actions. Disregarding these differences in behavioral influence may hinder a comprehensive understanding of the real role of social influence in a wide variety of scenarios, including (1) information propagation and influence maximization, (2) knowledge transfer in a community and development of projects, such as in GitHub and Stack Overflow, (3) online influence campaigns, or (4) online brand engagement at different stages of the consumer purchase funnel. As an example, in online marketing campaigns, some users may create original content, some users may contribute to others' created content, and still other users may spread the content of others by sharing. If a marketing firm is interested in interacting with this information spread, it may want to identify different users based upon the role they play and how those users affect other users. Similarly, in the learning communities and knowledge-sharing communities such as GitHub, Stack Overflow, professional learning communities on Twitter and LinkedIn, it is essential to have user engagement in all three types of actions: I, C, and S, to sustain the community as well as to achieve the user and community goals.

Furthermore, though some measurements such as Twitter-Rank, InfluenceRank, InfRank, and SpreadRank are integrated with the behavioral influence by adding the number of tweets and retweets that users have to the measurement, these measurements are platform-specific. This limits the generalization capabilities of the results obtained from those state-of-art methods over multiple platforms. Therefore, moving beyond the monolithic notion of social influence and a single-platform focused study will enhance the state-of-art methods.

Moreover, while there are hundreds of online social networks out there supporting a broad spectrum of interests and practices, nowadays, many users use more than one social media based on their interests and needs resulting in a complex social media ecosystem [35]. The latest social media usage statistics show that today's online social media ecosystem is more interconnected than a few years ago. In 2018, around three quarters (73%) of the American public used at least two social media platforms [36], whereas, in October 2021, an internet user uses or visits averagely around 7.5 different social media platforms each month, globally [37].

This interconnected social media ecosystem allows online social media users to expand their social ties, gain exposure to diverse information through different platforms, and exchange information across platforms [38, 35, 39]. In turn, their actions on a specific platform are influenced by not only the interactions from the focal platform but also the interaction from other platforms they interact with. Not only that, being exposed to diverse information allows users to bring novel information to the focal platform. Having access to more diverse sources of information gives users social currency that will enable them to exert more influence on the focal platform. Therefore, the users who have social ties with other platforms might exert more influence compared to the rest of the users [40, 41]. As a result, a user on multiple platforms can unknowingly generate a cascade of actions and spread information across several platforms. For example, a Twitter user can share a YouTube video about a product, political campaign, or a piece of celebrity news on Twitter to initiate a tweet or reply to another tweet. Then it can follow a cascade of retweets. As a result, a Reddit user can start a discussion in Reddit about that YouTube video. Another example is a Twitter / Reddit discussion of technology such as cryptocurrency or software vulnerabilities can influence a GitHub user to contribute to the codebase development in a GitHub repository.

As users can engage with their networks through I, C, and S actions, the overall effect of the information spread can be positive or negative based upon their collective behavior. On top of that, as mentioned above, the cross-platform influence can amplify the effect of spread of information by spreading information across platforms. Therefore, as the online social media ecosystem becomes more interconnected, empirical studies to understand the significance of users who have social influence relationships across platforms are much more needed than ever before. It will contribute to potential strategies to maximize the positive outcomes in the scenarios such as online marketing campaigns, knowledge transfer in communities such as LinkedIn, Twitter professional educational communities, StackOverflow learning communities, and GitHub open-source code development communities; and to reduce the effect of the adverse outcomes in scenarios such as miss-information spread, cyberbullying.

However, most of the work that has been done on online social influence is focused on a single platform, and less attention has been paid to the cross-platform influence. Moreover, the existing few studies on cross-platform influence mainly focused on the topics such as user identity linkage across social media [42, 43], exploring sharing behavior across platforms [39, 44] macro-level information transfer across platforms [45, 46]. There has not been an empirical study to investigate the significance of the users who have behavioral social influence relationships across platforms.

Purpose of the Study

The prior work on social influence analysis is mainly lacking in four aspects, which are (1) assuming the monolithic notion of influence; i.e., all influence is measured using one number, (2) lack of generalizability of the proposed algorithms or measures of influence, (3) assuming the influence as a property of the user instead of the property of the relationship, (4) lack of studies on the significance of the cross-platform influence on users in the influence process. Hence, the work presented in this dissertation is focused on enhancing the literature on online social influence while filling these literature gaps. In particular, in this study, rather than modeling the influence as a single entity, I model the influence as multiple entities and examine social networks through the perspective of behavioral influence propagation.

This dissertation presents two empirical studies. In the first study, I introduce the concept of Influence cascades to track the multiple behavioral influence relationships over time and characterize the different social media and communities. An influence cascade can be defined as all of the actions in a chain that start from an initial user, who was prompted by an external (outside the social network) stimulus or intrinsic motivation to act, and the actions that the initial user then influences other users to take, and, in turn, the actions those users influence others to take and so forth until a user's action no longer influence any other users to act. In other words, an influence cascade is all of the users and events that were socially motivated and can be tracked back to an initial user that was not motivated socially, but due to an influence outside the social network. The presence of influence cascades indicates an underlying organizational structure. In the case of a highly distributed community, such as those that exist on OSM, such organizational structure is not explicitly expressed but is implicit in the users' actions. Analyzing influence cascades allows us to infer these underlying organizational structures. In this study, I extract influence cascades in a variety of scenarios over multiple platforms and visualize the underlying organizational structures.

Then, I explore the characteristics of influence cascades of OSM, and contrast the extracted OSM influence cascades against those from scale-free networks as well.

As we are interested in behavioral influence, with the premise that any I, C, or S action that a user can take can influence other users to do any I, C, or S actions, we defined nine types of influence relationships that can exist between any pair of users. We use transfer entropy to quantify these nine types of influence relationships [47] as it gives us the ability to model social influence as multidimensional while capturing the direction and causality of the influence.

Because of the action classification used, our model is abstracted from platform event types. As a result, we can compare influence cascades on different platforms using the same ICS classification. Different social media platforms enable different affordances for interaction. Though the actions on these platforms can still be characterized under the ICS classification, the algorithms and exact implementations may alter how users utilize these different actions. Hence, this gives us the ability to study human behavior on different platforms and determine if the affordances of the platform affect influence cascades. This is the first study to compare influence cascades between platforms.

Furthermore, in the second study, I investigate the interconnected social media to study the significance of cross-platform influence on social media users and the effect of the user actions that they are influenced to perform on cross-platform influence. In particular, I explore the social dynamics of the influential users who have a relatively higher number of social influence relationships across platforms, which I call "Interface users", compared to the influential users who have a relatively less number of social influence relationships across platforms, which I call the "Core users" to understand the significance of the interface users empirically. We can identify two events when interacting with social networks: (a) users experience influence by other users, and (b) users exert influence on other users. When users interact across platforms, these two events can occur through users of the focal platform and users of the other platform. Therefore, in this study, I explore the social dynamics of the interface users and core users based on (1) influence experienced from focal platform influence relationships (2) influence experienced from cross-platform influence relationships (3) influence exerted on focal platform users. Furthermore, we investigate the consistency of these social dynamics in different communities. Previous work showed that different actions of online social networks users influence other users in the network on further actions differently [48]. We could expect that the interface users who are influenced to perform initiation action will experience more influence, and they might exert more influence on the focal platform because they have a higher chance to expose to novel information through the cross-platform influence and, in turn, bring novel information to the focal platform. However, it has not been investigated empirically before. Therefore, we further analyzed interface users' social dynamics based on the actions they are influenced to do.

As our *ICS* action classification is platform-independent and transfer entropy could capture the causal influence relationships without needing any explicit underlying link structure, our framework in the first study gives us the ability to capture the influence relationships across platforms as well. Hence, the same framework was used to quantify the influence and re-construct the interconnected OSM.

I consider cryptocurrency (crypto), common vulnerable exposure (CVE), interest communities on GitHub (GH) and Twitter (TW) for the experiments in both studies. The first study results show that the depth and the structure of influence cascades depend on the platform and community of users. As a result of these observations, we can characterize the underlying organizational structures of these online communities. Moreover, the second study results show that users with a relatively higher number of social influence relationships are more vulnerable to being influenced and influential than others. Also, our results show that the interface users who are influenced to take the initiation actions are more influential than those who are influenced to take contribution actions.

Research Questions

In this dissertation the following research questions will be addressed.

1. How can we create a method to track the multitude of possible influence relationships caused by users' actions in online social networks at once?
2. How can that method help us characterize different social media and communities in understanding the structure of influence in social media platforms or communities?
3. What is the effect of cross-platform social influence relationships in the influence process?
4. How does the cross-platform influence affect the users based on the actions that they are influenced to do, in the influence process?

Scope of the Study

Though our introduction of influence cascades is novel, previous work has examined information cascades. However, information cascades differ in that the focus of the analysis is on the transmission of a particular piece of information and not the users influencing each other to transmit the information. Typically, information cascades are extracted by tracing a piece of information such as content, URL, or an image through the explicit link structures such as parent-child relationships [49, 50, 2, 51, 10]. However, such explicit link structures are not available in many data sets or may be incomplete [49, 52]. These studies focus on analyzing characteristics such as size, depth, degree distributions, or the growth of such information cascades, as opposed to understanding how one user directly influences another user.

Therefore, it should be noted that in this study, I am not following any retweet chains or reply chains of specific content. Influence cascades are not direct interaction chains, i.e., retweeting chains or reply chains. Instead, influence cascades are observed from looking at the time series of all users in the data set and observing how likely a user's particular event causes another event of another user. This means influence cascades do not always begin with an I action because a root user's I action may not be the action that influences other users, but instead, a root user's I, C, or S actions could all create influence chains. Also, there is a possibility to observe a C action influencing an I relationship in the cascade since that means that we observed that when a certain user performs contribution events, another user is likely to initiate a new thread.

Statement of Contributions

The prior work on social influence analysis is mainly lacking in four aspects; 1) assuming the monolithic notion of influence; i.e., all influence is measured using one number, 2) lack of generalizability of the proposed algorithms or measures of influence, 3) assuming the influence as a property of the user instead of the property of the relationship 4) lack of studies on the significance of the cross-platform influence on users in the influence process. Hence, this dissertation makes multiple contributions while addressing the above issues. Primarily, this dissertation presents one of the first general methods of tracking behavioral influence relationships caused by actions on social media, which we call influence cascades, using a platform-independent action classification and measuring the transfer entropy between the time-series of these actions. Secondly, this work presents a method to investigate the significance of the cross-platform influence in the influence process, using the same framework.

Thus, in overall, the following contributions are made in this research.

1. Presenting a novel and generalizable method to track influence relationships caused by ac-

tions of OSM.

2. Providing new insights to improve state-of-art methods that assume a monolithic notion of influence and homogeneous populations in the social influence analysis field.

3. Presenting the evidence to show that the platform and community determine depth and structure of influence cascades.

4. Presenting a method to investigate the significance of cross-platform influence in the influence process.

5. Presenting the evidence to show that the users with a relatively higher number of social influence relationships across platforms are more vulnerable to being influenced and influential than the others.

6. Providing provide insights for marketing firms, online community leaders, and policymakers to make proper intervention strategies to control the spread of information or misinformation.

Statement of Originality

Parts of this dissertation have been included in conference presentations, a journal publication. Other than the work discussed in the following list, the rest of this dissertation has not been published publicly at the time of writing.

1. Senevirathna, C., Gunaratne, C., Rand, W., Jayalath, C., Garibay, I. (2021). Influence Cascades: Entropy-Based Characterization of Behavioral Influence Patterns in Social Media. Entropy, 23(2), 160. doi:10.3390/e23020160

2. Senevirathna, C., Gunaratne, C., Jayalath, C.,Baral, N., and Rand, W., and Garibay, I. (2019b). Hidden Patterns in Influence Hierarchies in GitHub's Cryptocurrency Community. Santa Fe, NM. Computational SocialScience Society of the Americas.

3. Senevirathna, C., Gunaratne, C., Jayalath, C.,Baral, N., and Garibay, I. (2019a). Evidence of influence hierarchies ingithub's cryptocurrency community. Amsterdam, NL. International Conference on Computational Social Science.

CHAPTER 2: LITERATURE REVIEW

Social influence has long been studied in many areas such as information diffusion and influence maximization [12, 9], viral marketing [2, 3], influential blogger finding [10], health applications [53], spread of opinions and news [54, 13, 8], and so on. In these studies, social influence is measured in many different ways. Among these methods, most of the work has focused on the notion of centrality, or structural influence. Centrality measurements such as degree, closeness, betweenness, eigenvector, Katz, and their variations are used widely in studies of social influence [18, 19, 20, 21]. However, in most of these measurements except eigenvector centrality, there is no distinction of the contribution of individual nodes to the measurement [14], and even in the case of eigenvector centrality, the only difference is a structural difference, not behavior-based. The number of followers, which is related to degree centrality, is used by [3, 25] to measure influence in microblogs. However, in [3, 55, 56], the authors show that there is a weak correlation between behavioral influence and the number of followers. Hence, these measurements are not fully able to capture behavioral influence and state-of-art methods related to these measurements cannot comprehensibly address the scenarios where an organization is interested in different types of influence, as discussed in the Introduction.

In addition, some recent studies use deep learning models to capture social influence. The DeepInf developed by Qiu et al. [57] is able to predict the binary status (active/inactive) of a user, given the user's underlying local network structure and the status of the near neighbors of the user. Leung et al. [58] proposed the HPPNP model by integrating a feature from a page rank domain to the DeepInf model and improved the performance of the DeepInf model. These models use historical interactions to predict social influence. However, the accuracy of the prediction depends on the underlying social network that the model uses because of the assumption that only near neighbors influence users' actions. In [57, 58], the authors use underlying user networks such as follower/followee or friendship networks for their study. Hence, these studies fail to address users' actions that may occur when they identified posts using hashtags or keywords [59].

Another way to measure influence is based on entropy and information theory. Peng et al. [34] use node entropy based upon the degree of a user and interaction frequency entropy to evaluate social influence in mobile social networks. Sun and Ng [33] use graph entropy based upon the centrality of users to measure the influence of connectors on social networks. Chen et al. [60] consider network topology and proposed a method to rank the influential nodes by considering the Tsallis entropy of the users and their neighbors. Transfer entropy is another entropy-based measurement that is used to quantify influence. Transfer entropy is introduced by Schreiber [47] to capture the cause and effect in an interaction between two coupled systems effectively. It is an information-theoretic approach based on Shannon entropy [61] and it measures the uncertainty reduced by the prediction of the future of a system from the past of the system by knowing the past of another system. If two random processes are $X = \{X_t\}_{t \in N}$ and $Y = \{Y_t\}_{t \in N}$ then the transfer entropy can be defined as

$$TE_{X \to Y} = \sum_{x,y \in \Omega} P(Y_{t+1} = y, Y_t = y, X_t = x) \log \frac{P(Y_{t+1} = y | Y_t = y, X_t = x)}{P(Y_{t+1} = y | Y_t = y)}, \tag{2.1}$$

where Ω is the sample space that includes all realizations.

VerSteeg and Galstyan [62] use this approach to quantify the influence of content on users in social media and show that transfer entropy is able to capture some of the relationships that cannot be captured by the follower network or mention network successfully. Moreover, He et al. [63] use the same approach to reconstruct the underlying network structure of online social media and use transfer entropy to measure peer influence in OSM.

Information cascades have provided us with insight into how these social networks operate. For example, Adar and Zhang [49] study the sharing of URLs in the blog-space by inferring their explicit link structure and implicit link structure. Explicit link structure is constructed by tracing the provided information on the data. Implicit link structure is constructed by using a classifier that depends on the blog similarity measures. Gruhl et al. [64] propose a model to study the propagation of information in the form of topics throughout the blog space using a derived form of the independent cascade model on a network induced by the time series of the topics and the blog which posts that topic at that time. Further, Leskovec et al. [50] study the propagation of posts in the blog space to discover the patterns of information propagation. The authors analyze the cascades of blog posts by measuring the overall out-degree, in-degree, and in-degree distribution of nodes at level L of the collection of cascades. Further, they quantify the cascades by the number of nodes in the cascades and analyze the distribution of the cascade size over the collection of cascades they extracted. Their results show that blog posts have weekly periodicity but they do not have a bursty behavior. Moreover, Leskovec et al. [2] trace the diffusion of product recommendations using emails and show that product recommendation cascades do not grow very large. Kumar et al. [51] study the information cascades in yahoo!, Twitter, and Usenet groups by reconstructing the information cascades using the parent-child relationships that exist in the data and explore the distributions of size, depth, and degree of the information cascades. They show that degree distributions of information cascades are close to a power law. Bakshy et al. [10] study the information diffusion by studying the cascades of URL's sharing on Twitter and show that information mainly spreads through small cascades that are started by ordinary individuals while long cascades are rare. Dow et al. [65] study the cascade of image sharing on Facebook and explore them in terms of evaluation time and the distributions of the depth of the cascades. Further, they quantify the predictability of sub-cascades sizes. Cadena et al. [66] show that activity cascades in Twitter are predictive of civil unrest.

Moreover, with the variety of OSM today, people engage with multiple social media platforms giving them the opportunity to discuss and share their interests on multiple platforms. Hence, researchers have become interested in studying how human behavior differs on different platforms. Xiong et al. [67] propose a new approach to link GitHub and Stack Overflow accounts using a CART decision tree and explore developer behavior on these two platforms. Waterloo et al. [68] study how users express their emotions on WhatsApp, Facebook, Twitter, and Instagram and find that there are differences in the patterns of emotional expression based on the platform. Furthermore, Kim et al. [69] propose a method to estimate the information transfer across mainstream news, social networking sites, and blogs using transfer entropy. Also, a similar study from Bhattacharjee [46] analyzes information transfer across social media, in particular Twitter, Reddit, and GitHub. Bhattacharjee uses symbolic transfer entropy to measure the influence from one platform to the other. Our work extends this past work into the realm of influence cascades, so we can understand not only how the same user operates on different platforms, but also whether users on one platform influence users on other platforms.

Furthermore, as users tend to use more than one online social media, identifying the same user across multiple platforms provide many advantages to link prediction studies and recommendation studies. In [42], the authors present a review of key achievements of studies of user identity linkage across social media and state-of-art algorithms that use to identify the user identities. Further, they present a framework consisting of feature extraction such as profile features, content features, and model construction which can be done using supervised or unsupervised ways to identify user identities. Also, Zhou et al. [43] propose a method, which calls ACCount eMbedding (ACCM), to identify similar users across different online social media platforms using the semantics of networks structures. They show that their method outperforms several state-of-art baseline methods using real-world and synthetic data.

Moreover, as different platforms offer the features to share the content across social media, Ham et al. [39] study the motivations of users to share the content in the social media ecosystem. They show that users' preference for sharing their thoughts and feelings with others plays the most significant role in sharing behavior. Finally, as the sharing of content across platforms may impact both positively and negatively on society, Cody et al. [44] study the effect of recent implementation to exclude the potentially harmful content from video recommendations on YouTube while they did not remove the content from the platform. Their results show that this implementation reduces the probability of sharing harmful videos across platforms.

CHAPTER 3: INFLUENCE CASCADES: ENTROPY-BASED CHARACTERIZATION OF BEHAVIORAL INFLUENCE PATTERNS IN SOCIAL MEDIA

This chapter addresses research questions one and two; 1) How can we create a method to track the multitude of possible influence relationships caused by users' actions in online social networks at once? 2) How can that method help us characterize different social media and communities in understanding the structure of influence in social media platforms or communities? First, we present a method to trace the multitude of possible influence relationships caused by the users' actions on OSM, which we call influence cascades. We then explore the characteristics of extracted influence cascades from cryptocurrency and cyber-vulnerability communities on GitHub and Twitter and compare within and across platforms. Further, we compare the influence cascades of OSM against those from the scale-free network as well.

Methodology

First, we examine the basic concept of influence cascades, in this section. In particular, we start by examining users who are not socially influenced themselves but exert influence on others, and how different actions contribute to the accumulation of social influence as it progresses through the network via influencer-influenced relationships. We use the ICS classification in order to replicate our findings across two social media platforms and two different communities. We extract social influence cascades observed in four online user communities: (1) GitHub users working on cryptocurrency, (2) Twitter users discussing cryptocurrency, (3) GitHub users working on cyber-vulnerabilities (CVEs), and (4) Twitter users discussing CVEs.

We performed this 2 $\times$ 2 comparison to give us the ability to analyze both platform and subject community differences. In regrading to answer the research questions 2, we compared the extracted social influence cascades against those expected on an artificially generated scale-free network. By using this scale-free model as a null model, we provide a basis of a comparison that is independent of any of the intrinsic properties of underlying networks, to compare and contrast our results. Therefore, we can identify what aspects are related to the particular circumstances of the platform and community and what aspects are present in any network.

Defining Influence Relationships

In this study, we built our framework based on ICS classification. We let the set of actions a user can perform be denoted as $A = \{I, C, S\}$. Once a user, u, performs an action, a, there is a chance that his action influences another user, v, to preform another action, b, which we describe as a social influence relationship of type $u_a \rightarrow v_b$, where $a, b \in A$ and u, v are users in the network W. Hence, we can define nine influence relationships as follows: $u_I \rightarrow v_I$, $u_I \rightarrow v_C$, $u_I \rightarrow v_S$, $u_C \rightarrow v_I$, $u_C \rightarrow v_C$, $u_C \rightarrow v_S$, $u_S \rightarrow v_I$, $u_S \rightarrow v_C$, and $u_S \rightarrow v_S$. As an example, we can use $u_I \rightarrow v_I$ to symbolize an influence relationship where u's initiation of a conversation influenced the initiation of another conversation by v. We use transfer entropy to quantify these influences and infer causal relationships [47]. Transfer entropy has been shown to capture influence better than other commonly used measures such as centrality and number of followers [62]. Also, by using transfer entropy, we are not restricted to limitations in the follower network that may occur if a user is influenced by, but does not follow, another user [3, 56, 59].

Extraction of Influence Cascades

We first quantify the magnitude of influence for each relationship $u_a \rightarrow v_b$ by calculating the transfer entropy, from a time series of action type a of user u to the time series of action type b of user v [70]. Next, we extract influence cascades from the pruned influence network and visualize them as follows.

Constructing the Influence Network

Since each directed user pair $(\vec{u,v})$ can have nine types of influence relationships $u_a \rightarrow v_b$, we define the total social influence from user u to v as a vector $\overrightarrow{\gamma}_{u,v}$ with the corresponding influence measurement values $\gamma_{uv(ab)}$ as its vector components. If at least one influence relationship exists, i.e., at least one non-zero influence vector component exists from u to v, then we can say that u influences v. Accordingly, we define the influence network $G(V,E)$ according to Equations (4.1) and (4.2).

$$V = \{\forall u, v \in W | \sum_{a,b \in A} \gamma_{uv(ab)} > 0\}, \tag{3.1}$$

$$E = \{\{u,v\} | \sum_{a,b \in A} \gamma_{uv(ab)} > 0\}. \tag{3.2}$$

It must be noted that G is a directed graph. Furthermore, we attribute the influence vector components to edge weights of G. In this manner, W is pruned of edges that have no social influence from one user to another, forming G.

Extracting Influence Cascades

Next, to study the characteristics and reach of the quantified influences, we extract the influence cascades of the users as follows. Externally motivated but not socially influenced users R is defined according to Equation (3.3).

$$R = \{u \in V | in - degree(u) = 0\} \tag{3.3}$$

In order to extract the influence cascades from any $u \in R$, we first extract all the outgoing neighbors of u, $N_o(u)$, and their corresponding edges from u. We then extract all the outgoing neighbors of users in $N_o(u)$ and their corresponding edges and repeat this process until there are no more identifiable outgoing edges. The initial user, at the top of the cascade, is called the root user and their node level is 0. Level 0 users are chosen as those who have no incoming edges, i.e., have no influencing users, but exert influence on other users. The node level of other users in the cascade is labeled based on the hop distance from the level 0 user to them. Figure 3 shows an example influence cascade using this process.

Visualization of Influence Cascades

As Figure 3 shows, the extracted cascades help analyze basic characteristics, such as the size and length of the cascades. However, this representation does not identify whether the influencing action was a I, C, or S. Hence, we propose a visualization technique that can integrate the information of influence cascades as follows:

Let $L_{i,i+1}$ represent the set of influence vectors flow from the ith level to $(i+1)$th level, where $i \in \{0, 1, , \ldots, n-1\}$ and n is the depth of the cascade. The normalized vector component of the total social influence an action a has on an action b, $\gamma_{i,i+1(ab)}$ is calculated as shown in Equation

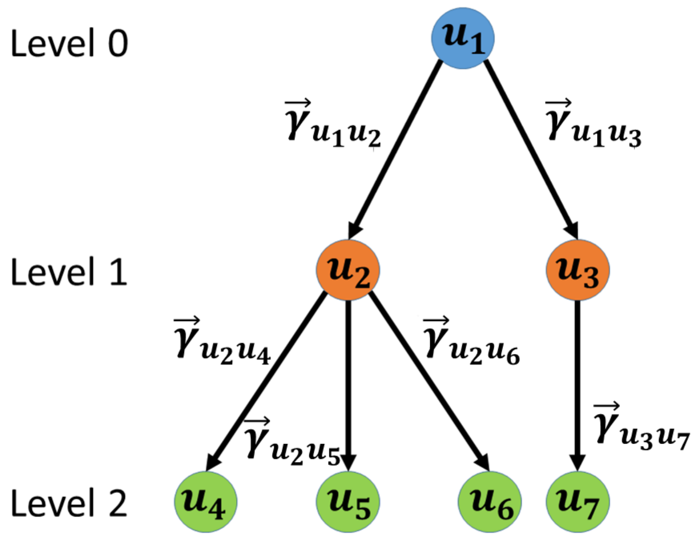

Figure 3.1: Example of an Influence Cascade. User u_1 is selected as a root user as it has a zero in-degree, i.e. it is not socially influenced. User u_1 socially influences users $\{u_2, u_3\}$ at level 1. Users $\{u_2, u_3\}$ influence users $\{u_4, u_5, u_6, u_7\}$ at level 2. $\vec{\gamma}_{u_i,u_j}; i, j = 1, 2, \ldots, 7$ represent the total influence vector from user u_i to user u_j.

(3.4).

$$\gamma_{i,i+1(ab)} = \frac{\sum_{l \in L_{i,i+1}} l(ab)}{\sum_{i=0}^{i=n-1} \sum_{l \in L_{i,i+1}} l(ab)} \tag{3.4}$$

We visualize the influence cascade through a Sankey diagram [71]. In the Sankey diagram, nodes represent the influencing actions ($a \in A$), while flows represent the total magnitude of influence exerted by this action on users at the next level of the cascade, normalized across the cascade. Figure 3.2 shows an example of a Sankey diagram produced by the proposed method.

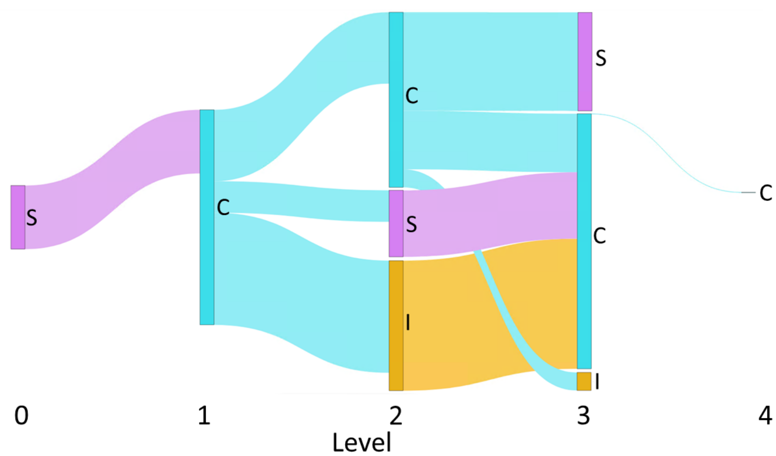

Figure 3.2: Example of a Sankey diagram produced by the proposed method. The diagram visualizes the normalized flow of total influence, categorized by influence relationships, along the length of a influence cascade. The nodes represent the different activity types, *I*: Initiation (yellow), *C*: Contribution (blue), and *S*: Sharing (pink), and their heights represent the relative magnitude of influence each level exerts on the next. The thickness of the blue, pink, yellow flow lines are proportionate to the magnitude of the normalized total influence value that *C*, *S*, and *I* events have on corresponding actions at the next level, respectively.

Experiments

Data

We considered two OSM platforms, Twitter and GitHub for our experiments. Twitter is a popular social networking site that allows users to post and interact with comments. Though GitHub may not appear to be an OSM on its surface, it provides powerful tools for interaction and commenting, allowing users to socially interact in a fashion similar to other OSM [72].

The empirical data consisted of temporal user activity related to discussions and project development of selected cryptocurrencies (Crypto) and cyber-vulnerabilities (CVE) on both Twitter (TW) and GitHub (GH). The data is gathered as follows:

- GH-Crypto data was collected by extracting events related to more than 20 target coins' official repositories, repositories labeled with target coin names, and repositories that mentioned the target coin names in their descriptions.
- TW-Crypto data was collected by extracting all tweets from official websites related to more than 20 target coins and by matching the target coin names, code, hashtags, etc. with the full Twitter firehose. Extraction was limited to English language tweets and users from either unknown countries or the UK, India, Canada, Russia, and the Netherlands.
- GH-CVE data was collected by extracting events related to any repositories that were related to CVE at a certain point in their life cycle, found by matching CVE textual patterns against repository descriptions and texts related to events.
- TW-CVE data was collected by matching the CVE textual patterns against the collection of tweets extracted through the public Twitter API.

The TW-Crypto data was extracted from the 1st of August 2018 to the 30th of November 2018 while GH-Crypto, GH-CVE, and TW-CVE data were extracted from the 1st of January 2017 to the 31st of March 2017. The raw data sets contained 111821, 19166, 11875, and 3278 unique users respectively. As low activity users have less impact on influencing others over time we only considered active users who had an average monthly activity greater than five events within these time periods. The filtered data sets contained 4170, 1784, 1989, and 92 unique users respectively. Table 3.1 shows the categorizations of 14 different GitHub events and 4 different Twitter events into initiation, contribution, and sharing action classes.

Table 3.1: Classification of GitHub and Twitter actions.

	Initiation	Contribution	Sharing
GitHub	CreateEvent	CommitCommentEvent, GollumEvent, IssueCommentEvent, IssuesEvent, PullRequestEvent, PullRequestReviewCommentEvent, PushEvent, DeleteEvent	ForkEvent, WatchEvent, MemberEvent, PublicEvent, ReleaseEvent
Twitter	Tweet	Reply, Quote	Retweet

The extracted influence networks of GH-Crypto, TW-Crypto, GH-CVE, and TW-CVE had 1406, 3365, 151, and 80 nodes (users), respectively. Each of these influence networks consisted of 568, 2385, 111, and 45 users who were not socially influenced but influenced others (root nodes).

Experimental Setup

We began our experiment by exploring the influence cascades in our empirical networks. For comparison, we constructed generic scale-free networks that were similar in size as null models. As an example, we constructed a scale-free network with 1406 nodes as a null model of GH-Crypto net-

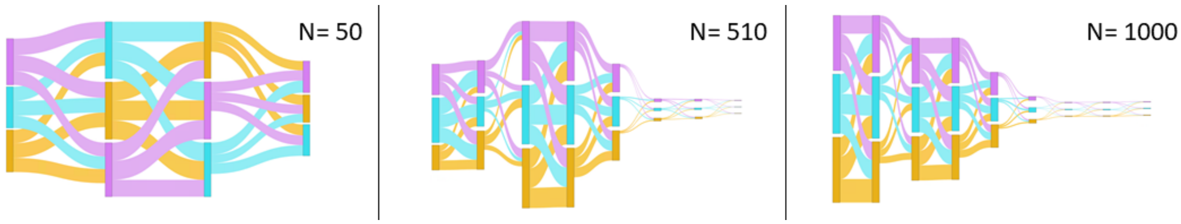

Figure 3.3: Examples of uniformly distributed influence cascades over scale-free networks of varying network sizes.

work which has 1406 users in its influence network. Python 3 and the NetworkX scale_free_graph library [73] were used to generate directed scale-free networks. Except for the number of nodes, the other parameter values were kept constant while producing the scale-free networks. Any loops and multi-edges that resulted were removed. Next, for each resultant edge, nine random values from $U[0,1]$ were assigned as the magnitude of the influence of the nine influence relationships. Given this network, we extracted the influence cascades from root nodes by identifying those nodes that had zero in-degree, i.e., no influencing nodes. For each network, we aligned all the influence cascades by level and aggregated the normalized total influence vector components (Equation 3.4) by their median. For some examples of these cascades see Figure 3.3.

We explored the user distribution of influenced cascades by comparing the mean number of users as well as the cumulative mean number of users per cascade level by platform and community. The Jensen–Shannon (JS) Divergence test was performed to measure the similarity of user distributions between influence cascades extracted from empirical networks and their null models as well as between the platforms/communities.

In order to study the similarities of the structure of influence cascades in terms of the distribution of influence from different extracted networks, we explored the residual differences between the median normalized total influence values extracted from influence cascades and those from influence cascades generated by the corresponding null model, both within and across platforms, by influence relationship. A Spearman's correlation test was performed on these residuals by influence relationship, grouping by platform and community, in order to infer the statistical significance of the observations. The null hypothesis H_0 tested, was that there is no correlation between the residuals in the magnitude of influence of two platform-communities. In other words, if the comparison was significant that means that two platforms or communities are significantly similar in terms of the distribution of the magnitude of influence. As we have multiple comparisons, we applied a Bonferroni correction to minimize the error rate. Therefore, we used a significance level of $0.05/9 = 0.0055$ in order to consider an individual test as significant.

The computer code for extracting influence cascades, visualizations, and conducted experiments was developed in a Jupyter notebook which is publicly available. The influence data extracted from the OSM and code is available at `https://github.com/Csenevirathna/InfluenceCascades`. The versions of the software and packages which are used are as follows: Python 3.6.3, pandas 1.0.1, NumPy 1.19.1, NetworkX 2.4, seaborn 0.9.0, Plotly 3.6.0 and, statsmodels 0.12.0.

Results

Comparisons of the mean number of users per influence cascade level by platform and community are shown in Figure 3.4. Similarly, comparisons of the cumulative mean number of users per cascade level by platforms and community are shown in Figure 3.5. For both of these sets of measurements, the measurement of the corresponding scale-free null-model, matched by network size, has been included as a control. We observed that the user distributions for the CVE community closely followed that of their corresponding scale-free null-model, in contrast to that of the cryptocurrency community, where a larger deviation from the scale-free null-model was observed. This result is confirmed in Table 3.2, where the JS-divergences for each platform-community from their corresponding scale-free networks are shown. The JS Divergences for the CVE community networks is a magnitude smaller than that of the cryptocurrency community networks, regardless of platform.

Instead, we found that the distributions of users across levels were similar for the cryptocurrency community, regardless of platform. In particular, we observed that on average the user distributions culminate at level 4 for both cryptocurrency networks, producing influence cascades that are much shorter than are expected based on their comparison against the corresponding scale-free null-models. In other words, the mean user distributions over influence cascades for the cryptocurrency community were robust across platforms, while those for the CVE community were more platform-sensitive. This result is further confirmed in Table 3.3, where the Jensen–Shannon divergence between each platform-community is displayed. According to this comparison, the JS divergence is lowest within communities rather than within platforms. Furthermore, the JS divergence is lower when comparing across platforms within the cryptocurrency community, rather than the JS divergence when comparing across platforms within the CVE community. Also, we see that the JS divergence between the cryptocurrency and CVE communities on Twitter is much lower than that on GitHub. In other words, the influence structures within the cryptocurrency community are more robust across platforms than the CVE community, and the influence structures on Twitter are more robust across communities in comparison to those on GitHub.

Table 3.2: Jensen-Shannon Divergence test statistics for each empirical network from its corresponding scale-free null model. JS divergences are least within CVE communities and their scale-free null models, in comparison to that of cryptocurrency communities.

Community	Platform	JS-Divergence
CVE	GitHub	0.0964
Crypto	GitHub	0.1765
CVE	Twitter	0.0858
Crypto	Twitter	0.2138

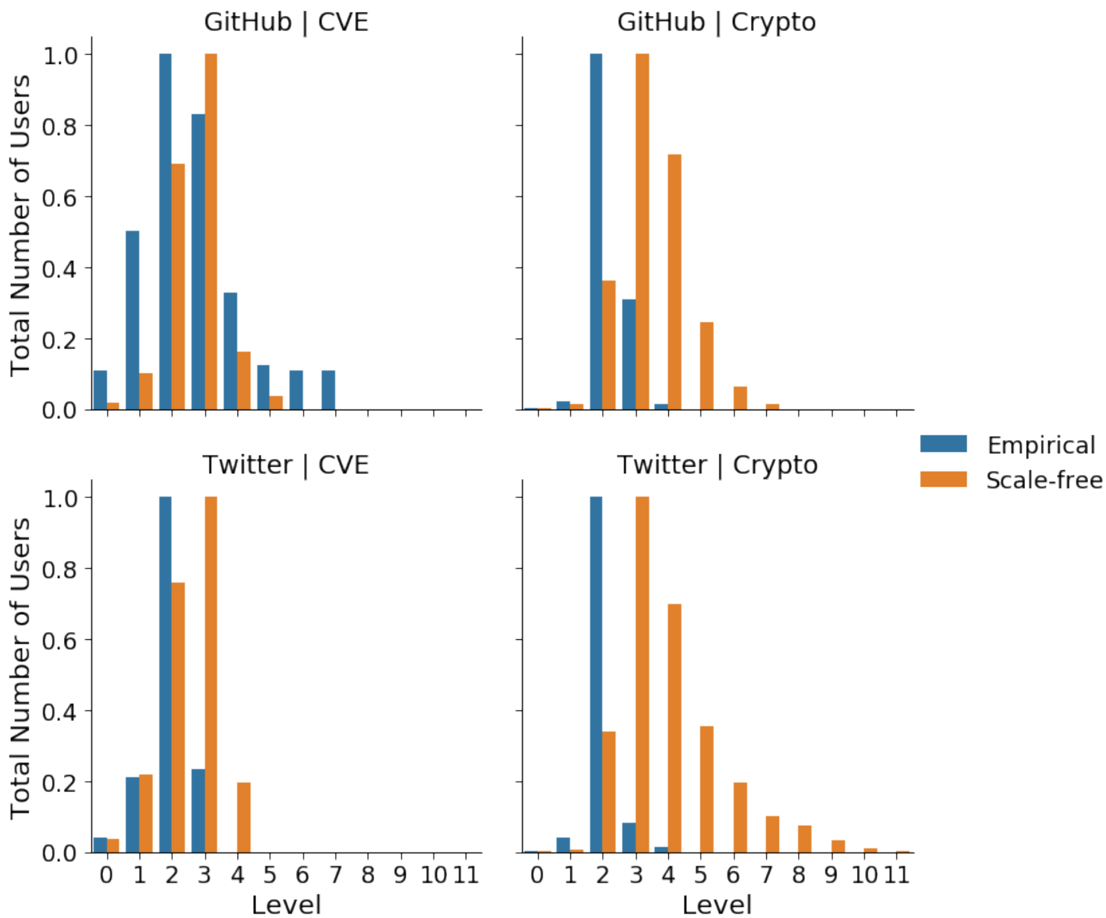

Figure 3.4: The mean total number of users at each level by platforms and communities for empirical networks and their scale-free networks. The user distribution of the common vulnerable exposure (CVE) community follows the scale-free null model closer than the cryptocurrency community.

We then compare the distributions of influence over cascade level by action for the empirical networks against their corresponding scale-free null models matched by network size. Figure 3.6 displays the median normalized total influence exerted from lower to higher levels by action (I, C, and S) for the four empirical networks. The same measurements for their corresponding scale-free null models are shown in Figure 3.7. Figure 3.6 shows a clear distinction between how different influence relationships are distributed along the cascades within platforms and across

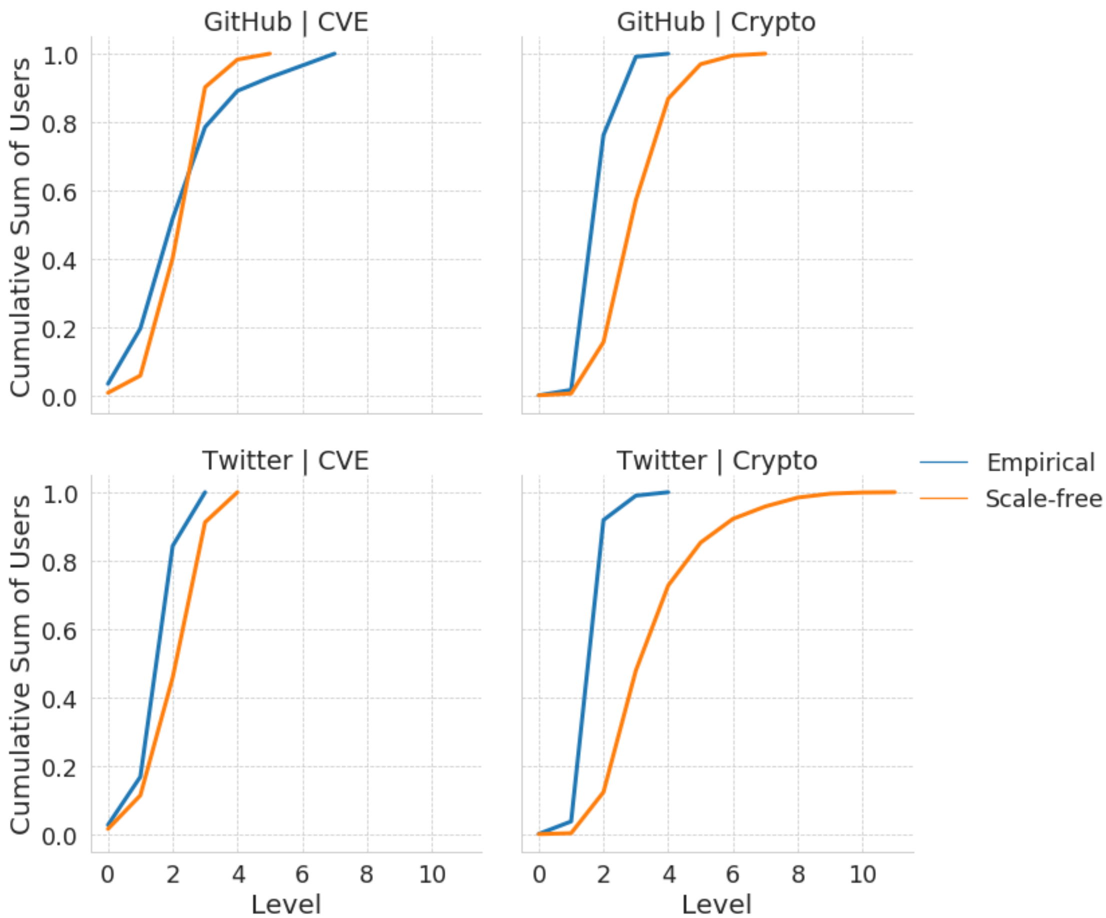

Figure 3.5: The cumulative sum of the mean total number of users at each level by platforms and communities for empirical networks and their scale-free networks. The cumulative user distribution of CVE community follows the scale-free null model closer than cryptocurrency community.

Table 3.3: Jensen-Shannon Divergence test statistics between each empirical network. JS divergences are least within communities and across varying platforms, in comparison to within platforms across varying communities.

Community 1	Platform 1	Community 2	Platform 2	JS-Divergence
Crypto	GitHub	Crypto	Twitter	0.1634
CVE	GitHub	CVE	Twitter	0.1944
Crypto	Twitter	CVE	Twitter	0.1983
Crypto	GitHub	CVE	GitHub	0.3414

platforms. Despite the closeness of user distributions of the CVE networks to their corresponding scale-free null models, we observe that how influence is distributed among this community differs from that expected through the scale-free null models. We observe a similar difference in influence distribution from the scale-free null models for the cryptocurrency community. We observed that GH-Crypto, TW-Crypto, and TW-CVE have a common shape to their influence cascade, with the highest fraction of influence flow for most relationships in these platform-communities happening towards the middle of the cascade. However, for GH-CVE this happens at the head of the cascade. Interestingly, the distribution of influence seen in GH-CVE, which has a smaller influence network size (151 users), is similar to that seen in the larger networks of GH-Crypto and TW-Crypto scale-free null models (1406 and 3365 users respectively).

These results indicate that root nodes in GH-CVE are more influential compared to all of the other users in the cascade, whereas root nodes in the cryptocurrency community and the TW-CVE community are not very different from the other users in the cascades in terms of the amount of influence they exert on others. This can be explained by the popularity of and interest towards cryptocurrencies among all the users regardless of platform and difference in the interest of users on CVE's in different platforms. Moreover, these results indicate that influence cascades of empirical networks have less similarity with the scale-free null models by further confirming the effect of communities and platforms on influence cascades.

Furthermore, it was observed that not all nine influence relationships existed between every consecutive level of the influence cascades for any of the social networks, unlike that observed in the scale-free null models. This means that influence exerted by users is not uniform and depends on the type of action they are more inclined to perform given their platform and community, and also that the preference for certain actions is heterogeneous among users of a particular platform and community. Specifically, we observed that influence cascades of Twitter have a more equal distribution of influence through all three actions. Instead, we observe that contribution actions

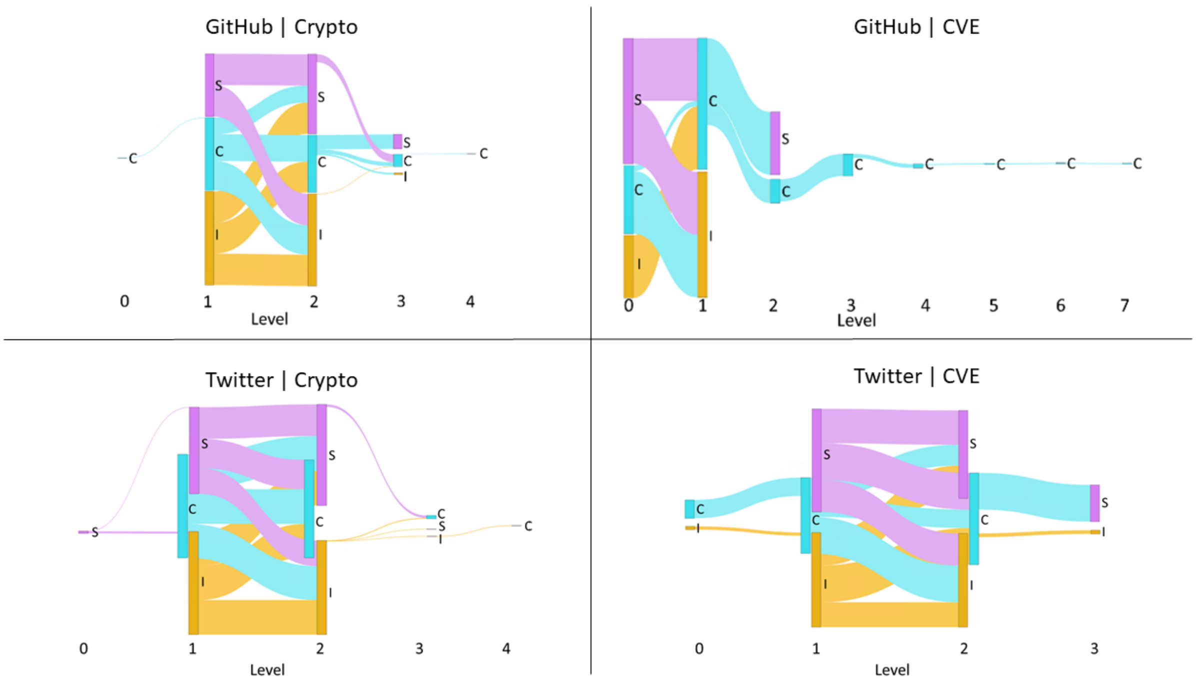

Figure 3.6: The median normalized total influence, by activity types I,C and S, along with the levels of the influence cascades of the GitHub (GH)-Crypto, GH-common vulnerable exposure (CVE), Twitter (TW)-Crypto, and TW-CVE empirical networks. The typical influence cascade in GH-CVE is much longer than the other platform-communities and is dominated by contribution actions influenced by contribution actions.

have more influence throughout the cascades observed on GitHub. This result can be explained by the differences in the nature of GitHub and Twitter. That is, GitHub is a platform for developers that are intensively involved in open source software development, but Twitter serves as a platform to share and post short discussions.

The residuals between the median normalized total influence values by relationship, over cascade level, of the empirical networks when compared to those of their corresponding scale-free null models across both platforms and communities are shown in Figure 3.8. Again we observe that the distribution of influence on GH-CVE is very different compared to that of the other three networks

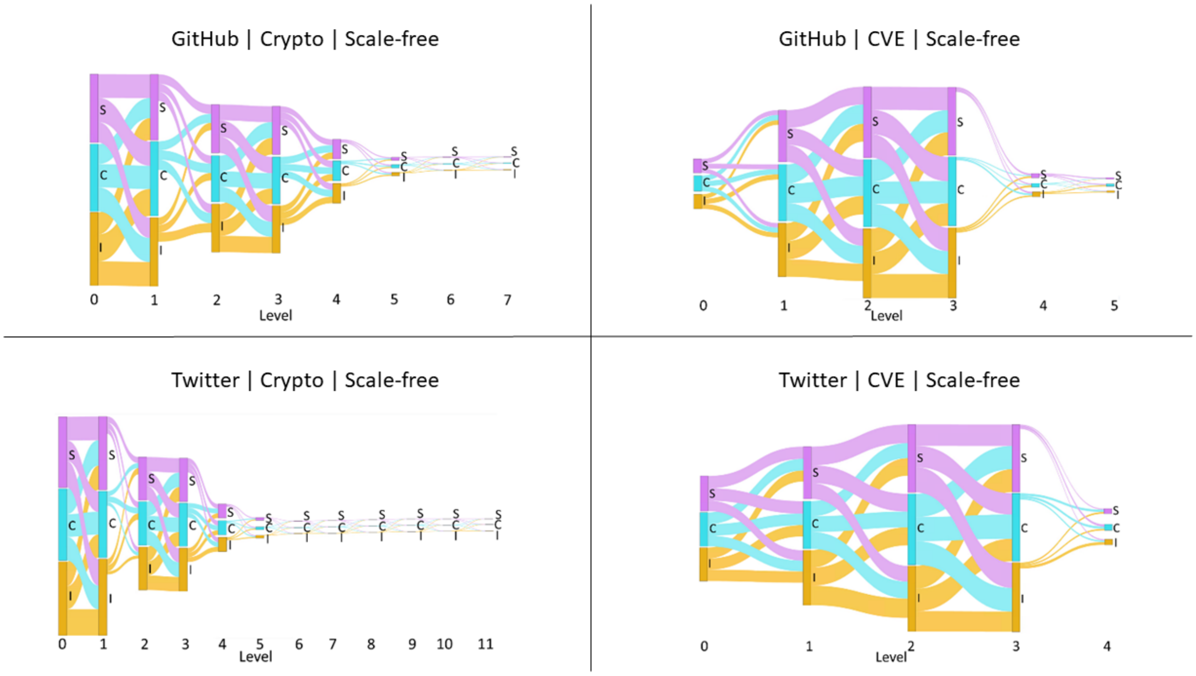

Figure 3.7: The median normalized total influence, by activity types I,C and S, along with the levels of the uniformly distributed influence cascades over the scale-free null models corresponding to GH-Crypto, GH-CVE, TW-Crypto, and TW-CVE influence networks by equal network size.

for almost all relationship types, except for $C \rightarrow S$. Furthermore, we see that the differences between cryptocurrency community influence cascades on GitHub and Twitter occur through $S \rightarrow C$ and $C \rightarrow S$ relationships.

In the case of Spearman's correlation tests, the null hypothesis for our experiments is that there is no correlation between the residuals in the magnitude of influence of two platform-communities when examined by the nine action-action relationships. The results of this test at original significance = 0.05 (Bonferroni-corrected significance = 0.0055) are shown in Table 3.4. The only significant correlations were observed between GitHub and Twitter within the cryptocurrency community for most influence relationship types, with the exception of $C \rightarrow S$ and $S \rightarrow C$. In other words, how influence was propagating within the cryptocurrency community over both Twitter and GitHub were similar with the exception of contribution and sharing events. This result can be explained by the higher importance that contribution events (such as commits, and commit comments) have within GitHub, compared against the popularity that sharing (or retweeting) has on Twitter. In contrast, we can state that how influence is propagated within the CVE community differs based on the platform, Twitter or GitHub, upon which the users interact. We can also state that there are no similarities in how influence is propagated when comparing between the two communities on either platform. Appendix A provides further visual validation via scatter graphs for each test above. It must be noted that an ANOVA could not be applied in place of the above correlation test as the residuals of the influence relationships failed to satisfy the normality assumption (Further information in Appendix A).

Discussion

We examine social networks through the perspective of influence propagation based on user actions, and compare four social networks, the cryptocurrency and CVE communities of Twitter and GitHub. In order to facilitate cross-platform comparison, we categorized actions into three abstract types that existed on multiple social media platforms: initiation, contribution, and sharing. The influence of these actions by users on further actions by other users was measured for all nine resulting relationships. We propose a novel method to measure and visualize the social influence exerted by users through these actions over time. We illustrate how transfer entropy can be used as the measurement of influence to estimate the degree to which causal relationships existed between user actions. User pairs that had at least one influence relationship of non-zero magnitude formed the basis of a network of influence. Users of this network that were not influenced by others, but did exert influence on others were selected as the roots of influence cascades. The users influenced by these root influencers were identified recursively, extracting cascades of influence, propagated via all nine relationships. The extracted empirical influence cascades were compared against uniform influence cascades on scale-free networks of equal node count, as null models.

Our results indicate that the manner in which influence cascades through online social media is affected by the social media platform and the online community. In particular, we find that the cryptocurrency community exhibits influence structures that are similar across both GitHub and Twitter, while this is not true for the CVE interest community. More specifically, within the cryptocurrency community, we notice that the only significant difference in influence cascades exist between relationships where contribution actions influence sharing actions, and vice versa. In other words, the influence relationships that exist between users engaged in cryptocurrency related development on GitHub and the influence relationships that exist between users engaged in cryptocurrency discussions are similar, with the exception of contribution actions influencing sharing actions (or sharing influencing contribution). The fact that code-development on GitHub is driven primarily through contribution actions, such as commits and pull-requests, while Twitter is driven by sharing actions, specifically retweets, offers an explanation for this exception. This technique and visualization enable the automatic identification and analysis of these differences.

In contrast, there is no similarity between the influence relationships on GitHub and Twitter within the CVE community. Additionally, we see that the influence cascades of the CVE developers on GitHub are longer than those of the CVE discussions on Twitter. This leads us to conclude that CVE developers on GitHub are generally more responsive to social influence than users discussing CVE related topics on Twitter. Further, we observe generally longer cascades of contribution actions influencing contributions actions within the CVE community on GitHub. In other words, individuals of the CVE community are more likely to engage in contributions to GitHub projects in the CVE domain than engage in CVE related discussions on Twitter.

Finally, we find evidence that the influence structures of Twitter show higher similarity across communities, compared to those of GitHub. However, we do not find any individual influence relationships across the two communities on Twitter that show significantly similar progressions of the magnitude of influence over cascade level.

Some of these differences in platforms versus communities may have to do with the nature of the communities themselves. Cryptocurrencies have been a growing topic since Bitcoin was introduced to the financial market as a medium of exchange. Hence, we could explain the similar organizational structure in the Crypto community as a fact of the popularity of the cryptocurrencies in both Twitter and GitHub. However, the structural differences in the relationships where contribution actions influence sharing actions and sharing actions influence contribution actions can be explained as a result of the different nature of the contribution and sharing actions in GitHub and Twitter.

However, unlike cryptocurrencies, discussions of cyber-vulnerabilities maybe very different on Twitter and GitHub. On Twitter, CVE discussions may be interesting to one group of users who are interested in the news around CVEs, while on Github, the most active users may be individuals who are actively trying to develop solutions to CVEs. This disparity between the types of users engaged on Twitter versus Github is greater for the CVE community than the Cryptocurrency community.

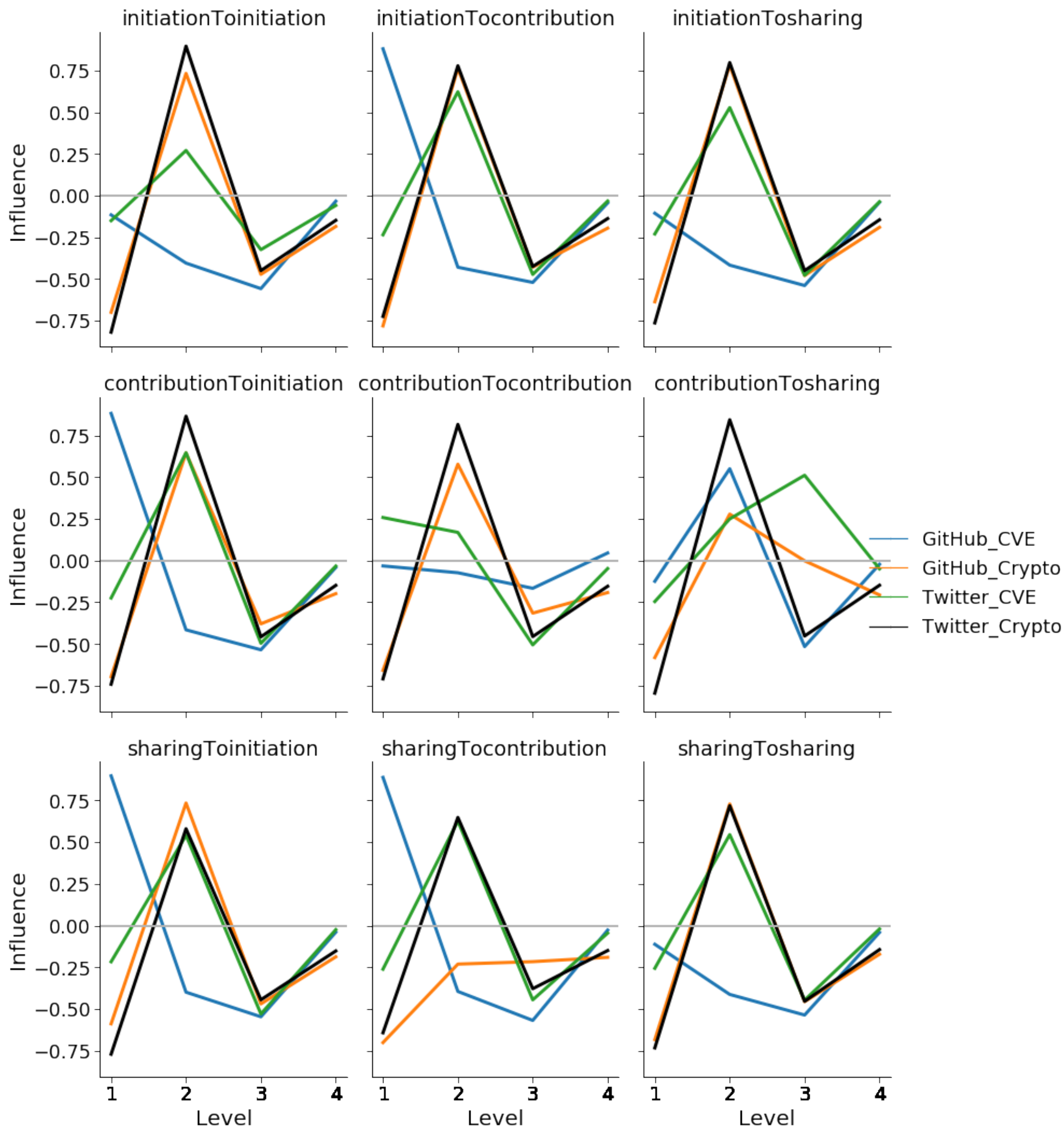

Figure 3.8: The residuals between the median normalized total influence values by relationship, over cascade level, of the empirical networks when compared to those of their corresponding scale-free null models. Most relationships in the cryptocurrency community seem correlated despite the difference in platform.

Table 3.4: Spearman's correlation for the H_0: there is no correlation between the residuals in magnitude of influence of two platform-communities by relationship, at original significance = 0.05 (Bonferroni-corrected significance = 0.0055). For each platform-community, the results for the influence relationships compared are sorted in descending order of correlation coefficient. The only significant correlations are observed between the influence relationships of the cryptocurrency community on Twitter and GitHub, with the exception of $C \rightarrow S$ and $S \rightarrow C$.

Community 1	Platform 1	Community 2	Platform 2	Influence Relationship	ρ	p-value
Crypto	GitHub	Crypto	Twitter	$I \rightarrow I$	1	0
				$I \rightarrow C$	1	0
				$I \rightarrow S$	1	0
				$C \rightarrow I$	1	0
				$C \rightarrow C$	1	0
				$S \rightarrow I$	1	0
				$S \rightarrow S$	1	0
				$C \rightarrow S$	0.8	0.2
				$S \rightarrow C$	0.4	0.6
CVE	GitHub	CVE	Twitter	$I \rightarrow I$	0.4	0.6
				$I \rightarrow S$	0.4	0.6
				$C \rightarrow C$	0.4	0.6
				$S \rightarrow S$	0.4	0.6
				$I \rightarrow C$	0.2	0.8
				$C \rightarrow I$	0.2	0.8
				$S \rightarrow I$	0.2	0.8
				$S \rightarrow C$	0.2	0.8
				$C \rightarrow S$	-0.2	0.8
Crypto	Twitter	CVE	Twitter	$I \rightarrow I$	0.8	0.2
				$I \rightarrow C$	0.8	0.2
				$I \rightarrow S$	0.8	0.2
				$C \rightarrow I$	0.8	0.2
				$S \rightarrow I$	0.8	0.2
				$S \rightarrow C$	0.8	0.2
				$S \rightarrow S$	0.8	0.2
				$C \rightarrow S$	0.4	0.6
				$C \rightarrow C$	-0.2	0.8
Crypto	GitHub	CVE	GitHub	$C \rightarrow S$	0.4	0.6
				$I \rightarrow I$	0	1
				$I \rightarrow S$	0	1
				$C \rightarrow C$	0	1
				$S \rightarrow S$	0	1
				$I \rightarrow C$	-0.4	0.6
				$C \rightarrow I$	-0.4	0.6
				$S \rightarrow I$	-0.4	0.6
				$S \rightarrow C$	-0.4	0.6

CHAPTER 4: ENTROPY-BASED ANALYSIS OF CROSS-PLATFORM INFLUENCE ON SOCIAL MEDIA USERS.

This chapter addresses research questions three and four; 3) What is the effect of cross-platform social influence relationships in the influence process? 4) How does the cross-platform influence affect the users based on the actions that they are influenced to do, in the influence process? First, we present a method to reconstruct the underlying influence relationships in interconnected OSM platforms and a method to investigate the cross-platform influence in the influence process. We then compare the cross-platform influence patterns in cryptocurrency and cyber-vulnerability communities on GitHub and Twitter.

Methodology

This section presents the method that we used to construct the underlying influence network structure of interconnected online social networks by identifying the possible influence relationships between users within focal platforms and across platforms. We used the *ICS* classification to define the influence relationships, and it allows us to model the influence platform independently. We used transfer entropy to capture the influence relationships and quantify the influence without needing any explicit underlying link structure. We then identified the interface users and core users in each platform. We explored their social dynamics based on (1) influence experienced from focal platform influence relationships, (2) influence experienced from cross-platform influence relationships and (3) influence exerted on focal platform users. Finally, we did a similar investigation on the groups of interface users, which were categorized based on the actions they were influenced to perform.

We performed our experiments on two social media platforms and two different communities. In particular, we consider four online user communities: (1) GitHub users working on cryptocurrency (2) Twitter users discussing cryptocurrency (3) GitHub users working on cyber-vulnerabilities (4) Twitter users discussing cyber-vulnerabilities This allows us to explore the similarities and differences in social dynamics of interface users compared to the core users in different platforms and communities.

Influence Relationships and Influence Network

An influence relationship occur when a user's action result in another user performing an action. As users' different actions exert influence differently on other users to perform actions [48], we define influence relationships based on the *ICS* classification. Let the set of actions be defined by $A = \{I, C, S\}$ and the interconnected social networks defined by the Z. If a user u do an action a then it can influence user v to perform an action b by creating a social influence relationship of type $u_a \to v_b$, where u, v are users in the network Z and $a, b \in A$. Hence, these three actions result in nine types of influence relationships in one direction as follows: $u_I \to v_I$, $u_I \to v_C$, $u_I \to v_S$, $u_C \to v_I$, $u_C \to v_C$, $u_C \to v_S$, $u_S \to v_I$, $u_S \to v_C$, and $u_S \to v_S$. Therefore, the social influence from user u to user v is defined as a vector $\overrightarrow{\gamma}_{u,v} =< \gamma_{uv(ab)} >$, where $\gamma_{uv(ab)}$ are corresponding influence measurements.

We used transfer entropy to identify these causal influence relationships within and across platforms from the empirical online social network data and quantify the influence. Transfer entropy is an information-theoretic technique based on Shannon entropy [61] and the Schreiber introduces it to capture the casual interactions between two random processes [47]. It measures the uncertainty reduced by the future prediction of a random process given its past behavior and knowing another random process's past behavior. It has been used to measure the influence of content on users in social media and [62] to reconstruct the underlying network structure of online social media successfully [63, 70, 48]. This approach is independent of the explicit underlying link structures such as follower networks or network constructs using parent-child relationships. Therefore, it can capture the cross-platform influence relationships and some of the relationships within platforms that such networks cannot capture. For example, associations that exist when a user can reply to a tweet that is found by using a keyword or hashtag without following the user who posts it [59] or relationships that are unable to detect because of incomplete data.

We calculate the transfer entropy between the time series of a action type a of user u to the time series of action type b of user v to quantify the magnitude of the influence of $u_a \rightarrow v_b$, which is same as $\gamma_{uv(ab)}$. If there exist an at least one type of influence relationship with $\gamma_{uv(ab)} > 0$ from users u to user v then it indicates that user u influence user v. Thus, we remove the non-influential edges from Z and define the directed influence network $Z^{'}(V,E)$ of Z according to the equations (4.1) and (4.2).

$$V = \{\forall u, v \in Z | \sum_{a,b \in A} \gamma_{uv(ab)} > 0\}, \tag{4.1}$$

$$E = \{\{u,v\} | \sum_{a,b \in A} \gamma_{uv(ab)} > 0\}. \tag{4.2}$$

Experiments

Data

We used cryptocurrency and cyber-vulnerability interest community data on GitHub and Twitter online social networks for our experiments. The data was extracted from 01^{st} of January 2017 to 31^{st} of March 2017, and they consisted of user activities related to discussions and project development of the selected cryptocurrencies (Crypto) and cyber-vulnerabilities (CVE) on Twitter

(TW) and GitHub (GH). GH-Crypto data was collected by querying events related to more than 20 target coins' official repositories, repositories labeled with coin names, and the repositories that mentioned the target coin names in their description. TW-Crypto data was collected by querying tweets related to more than 20 target coins, and extraction was limited to English tweets. GH-CVE data was collected by matching the CVE textual patterns with the repository descriptions and text-related events. TW-CVE data was collected by matching the CVE textual patterns with the collection of tweets extracted through the public API. The collected data contained 14 different GitHub events and four different Twitter events. They were classified into initiation, contribution, and sharing classes, same as in chapter 3, Table 3.1.

The raw data sets of GH-Crypto, TW-Crypto, GH-CVE, and TW-CVE contained 10679, 95161, 200761, and 841 unique users, respectively. We only consider the users who had more than 15 average monthly activities because less active users have less probability of influencing others over time. The filtered data sets contained 544, 2836, 7689, and 29 unique users, respectively, and the respective influence networks had 453, 1698, 6639, and 20 unique users.

Experimental Setup

Different platforms afford different functions in the social network ecosystem and work under different algorithms. Also, communities on the same platform may focus on various topics. Therefore, the social dynamics resulting from the interaction between users in a platform but for different communities and between users across platforms for the same community might depend on both the platform and the community. Thus, we first performed an exploratory analysis of the relative percentage of influence relationships that users have from the other platform to the focal platform for each platform-community.

Secondly, in order to study the social dynamics of interface users and core users, we grouped the users of each platform-community in the order of priority as follows: (1) sorted the users in the descending order of the number of influence relationships coming from the other platform (equal to the number of influencers from the other platform), (2) if there are users who have the same number of influence relationships from the other platform then, sorted those users in ascending order of the number of influential relationships coming from the focal platform. For example, if we consider the GH-CVE platform-community, we first sorted the GH-CVE users in descending order of the number of influence relationships coming from TW-CVE platform-community (i.e., the number of in-edges that GH-CVE users have from TW-CVE users). For example, suppose two GH-CVE users have ten influence relationships coming from different TW-CVE users. Then we sorted those users in ascending order of the number of influence relationships coming from GH-CVE users. In this manner, we can guarantee that the top users prioritize the influence coming from the other platform while the bottom users least prioritize the influence coming from the other platform.

As we focus on two measurements of each user, namely: (a) average influence experienced by other users and (b) average influence exerted on other users, for our exploration, we defined and quantified them as follows.

If u_i, where $i = 1, 2, ..., n$ are the influencers of a user x, the influence experienced by the user x is defined as in the equation (4.3).

$$\text{Influence Experienced by a user } x = \sum_{i=1}^{i=n} \sum_{a,b \in A} \gamma_{u_i x(ab)} \tag{4.3}$$

Next, if v_j, where $j = 1, 2, ..., k$ are influencees of user the x then the influence exerted by the user x is defined as in the equation (4.4).

$$\text{Influence Exerted by a user } x = \sum_{j=1}^{j=k} \sum_{a,b \in A} \gamma_{xv_j(ab)} \tag{4.4}$$

Next, in order to investigate the significance of inference users in the process of influence compared to core users, we grouped the users based on their focal platform-community and identified their influencers and influencees from the focal platform-community and the other platform-community separately. Then we calculated the magnitude of the: (1) influence experienced by focal platform influence relationships, (2) influence experienced by cross-platform influence relationships, and (3) influence exerted on focal platform users, using the above defined measurements and explored whether there are significant differences between the distribution of the resulting data for interface users and core users within and across platforms for two communities. The Kruskal-Wallis H test (for the comparison of more than two groups) and Mann-Whitney U test (for the comparison of two groups) were performed (significance level =0.05) accordingly to infer the statistical significance of the observations. For multiple comparisons, Bonferroni corrected significant values were used.

Finally, we grouped the interface users as follows to investigate the interface users based on the actions they are influenced to take. If a user u exists such that $\gamma_{uv(ab)} > 0$ for interface user v, then user v is a interface user who influenced to do action type b. We denote interface users who are influenced to do action type b as "Interface Users_b." Then we followed the same experiments above on these groups.

Results

The histogram of the relative percentage of influence relationships from the other platform compared to the influence relationships from the focal platform is shown in Figure 4.1. We observed that except in GH-CVE, all the users in the other three platform-communities have at least one link

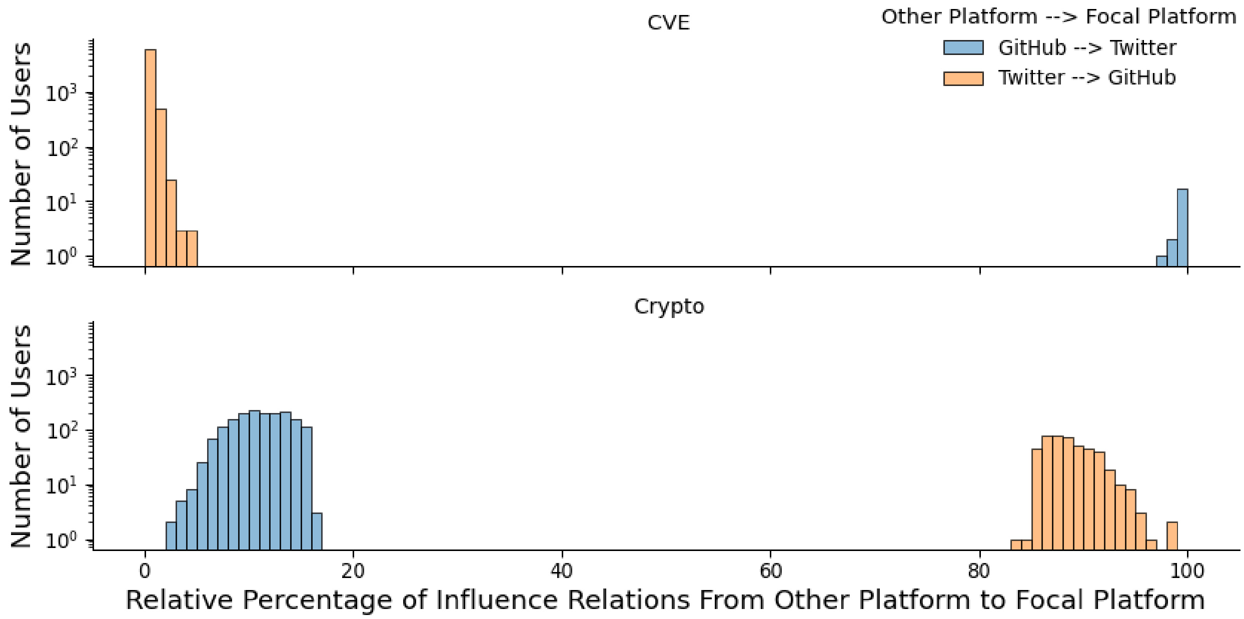

Figure 4.1: The relative percentage of influence relationships coming to GH-CVE, TW-CVE, GH-Crypto, and TW-Crypto from TW-CVE, GH-CVE, TW-Crypto, and GH-Crypto, respectively. Only GH-CVE contained users who are only influenced by the influence relationships within the focal platform-community (0 % cross-platform influence relationships). All the users in GH-CVE and TW-Crypto have more than 50% influence relations coming from TW-CVE and GH-Crypto, respectively. All the users in TW-CVE and GH-Crypto, have more than 50% influence relationships coming from their focal platform-communities.

from the corresponding other platform. This shows that most of the users who participate in the influence process do not limit their influence network to the focal platform but also expand their influence network to other platforms and bring information from other platforms to the focal platform. Also, this observation re-enforces the importance of studying the cross-platform influence. Moreover, in the CVE community, all Twitter users had more influencers from GitHub than from Twitter. At the same time, all GitHub users had more influencers from GitHub than from Twitter. However, in the crypto community, all Twitter users had more influencers from within Twitter than from GitHub. At the same time, all GitHub users had more influencers from Twitter than from within GitHub. These results can be explained by the social dynamics of how these communities operate. Previous work showed that the cyber-vulnerabilities started to be discussed on GitHub not only before they were discussed on Twitter but also even before they were officially published on NVD [74]. This implies that the CVE community is a development-driven community. Therefore, our results align with this previous finding by showing that more people tend to work on the development related to CVEs in GitHub, which drives online discussions on Twitter. Similarly, we can infer from our results that the crypto community is a discussion-driven community. It can be explained by the popularity of cryptocurrencies as digital investment currencies. Because of that, more people are sensitive to the topics related to cryptocurrencies and tend to discuss them frequently, which drives development in GitHub.

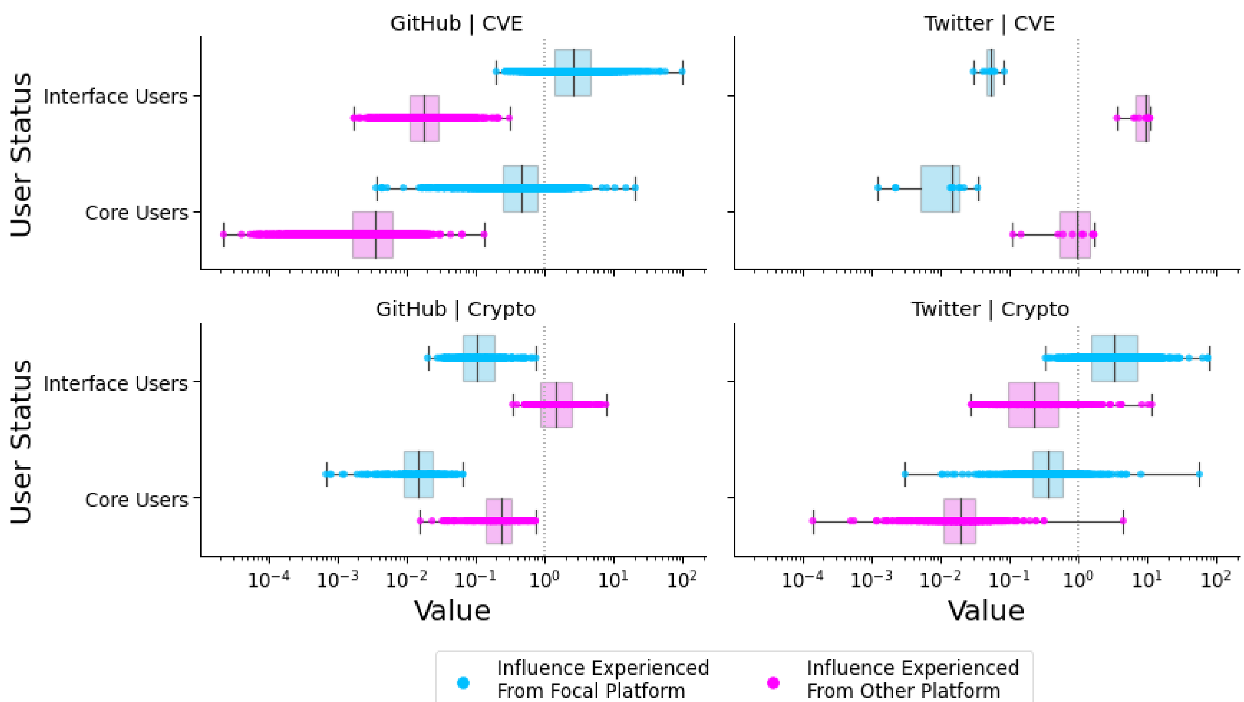

Figure 4.2: The distributions of influence experienced by the interface users and core users from the focal platform and the corresponding other platform. Interface users experienced more influence than core users. The subplot titles represent the focal platform | community.

Comparison of influence experienced by focal platform and other platform for interface users and core users are shown in Figure 4.2. We observed that both interface users and core users on GH-CVE and TW-Crypto platform-communities experienced more influence by the focal platform-community than by the corresponding other platform-community. Whereas both interfaces users and core users in GH-Crypto and TW-CVE experienced more influence by the corresponding other platform-community than by the focal platform-community. As we saw how the influence relationships distributed within the focal platform and cross-platform for both communities in Figure 4.1, these observations are to be expected. Hence, these results also can be explained as a result of the social dynamics of how the communities operate. Moreover, when we consider the same source of influence, i.e., the focal platform or the other platform, we observed clearly that influence experienced by the interface users is higher than that of the core users in all the platform-communities except TW-CVE.

These results are confirmed in the Table 4.1 and Table 4.2. Table 4.1 shows the Kruskal-Wallis H test statistics and mean rank of the influence experienced for each group, and Table 4.2 shows test statistics for pairwise comparison between each group. According to the Kruskal-Wallis H test statistics, mean experienced influence values were statistically significantly different between the four groups (groups are defined by the user status, i.e., interface users or core users; and the source of the influence, i.e., focal platform or other platform) for all the platform-communities (asymp.p <0.001). Further, when we consider the groups of the same source of influence, the mean influence experienced for interface uses higher than the core users. The pairwise comparisons revealed that there is a statistically significant difference in mean experienced influence values between all the groups in GH-CVE, GH-Crypto, and TW-Crypto (adj.p <0.001 for all the pairwise comparisons). In TW-CVE, the mean influence experienced between the two groups under interface users as well as between two groups under core users are significantly different, but not between the groups of interface users and core users who experienced the influence from the same source.

Table 4.1: Kruskal-Wallis H test statistics for the alternative hypothesis that the distributions of influence experienced by the interface users and core users from the focal platform and the corresponding other platform are significantly different. The group numbers represent the following groups: 1- Influence experienced by the interface users from the corresponding other platform , 2- Influence experienced by the interface users from the focal platform , 3- Influence experienced by core users from the corresponding other platform , 4- Influence experienced by core users from the focal platform. The alternative hypothesis is accepted for each platform-community.

Platform-Community	Group_num	N	Mean Rank	Kruskal-Wallis H	Asym.p
GH-CVE	1	3319	4529.89	11554.92	<0.001
	2	3319	11097.05		
	3	3013	1693.29		
	4	3320	8181.38		
TW-CVE	1	10	35.50	36.44	<0.001
	2	10	15.40		
	3	10	25.50		
	4	10	5.60		
GH-Crypto	1	226	790.84	766.23	<0.001
	2	226	400.28		
	3	227	505.74		
	4	227	118.39		
TW-Crypto	1	849	1597.43	2720.07	<0.001
	2	849	2922.48		
	3	849	451.04		
	4	849	1823.05		

Comparison of the influence exerted on the focal platform by the Interface users, and core users are shown in Figure 4.3. We observed that interface users exert more influence on the focal platform than the core users for all the platform-communities. This result is further confirmed in Table 4.3, where the Mann-Whitney U test statistics between interface users and core users are shown. As mean rank values for interface users are higher than the core users and p-values < 0.05, for each platform-community, we can conclude that influence exerted by the interface users on focal platform-community users is statistically significantly higher than for core users.

Next, the comparison of the distributions of influence experienced from the focal platform and the

Table 4.2: The test statistics of the pairwise comparison between groups in Table 4.1 for the alternative hypothesis that the group i and Group j are significantly different. The group numbers represent the following groups: 1- Influence experienced by the interface users from the corresponding other platform , 2- Influence experienced by the interface users from the focal platform , 3- Influence experienced by core users from the corresponding other platform , 4- Influence experienced by core users from the focal platform. Bonferroni adjusted p values are compared with the significance level of 0.05 to test the alternative hypothesis.

Platform-Community	Group i-Group j	Test Statistic	p	Adj. p
GH-CVE	3-1	2836.604	<0.001	<0.001
	3-4	-6488.089	<0.001	<0.001
	3-2	9403.762	<0.001	<0.001
	1-4	-3651.485	<0.001	<0.001
	1-2	-6567.158	<0.001	<0.001
	4-2	2915.673	<0.001	<0.001
TW-CVE	4-2	9.800	0.061	0.365
	4-3	19.900	<0.001	0.001
	4-1	29.900	<0.001	<0.001
	2-3	-10.100	0.053	0.320
	2-1	20.100	<0.001	0.001
	3-1	10.000	0.056	0.335
GH-Crypto	4-2	281.891	<0.001	<0.001
	4-3	387.357	<0.001	<0.001
	4-1	672.453	<0.001	<0.001
	2-3	-105.466	<0.001	<0.001
	2-1	390.562	<0.001	<0.001
	3-1	285.096	<0.001	<0.001
TW-Crypto	3-1	1146.392	<0.001	<0.001
	3-4	-1372.015	<0.001	<0.001
	3-2	2471.446	<0.001	<0.001
	1-4	-225.623	<0.001	<0.001
	1-2	-1325.054	<0.001	<0.001
	4-2	1099.431	<0.001	<0.001

corresponding other platform between the group of interface users who are grouped according to their behavioral role, i.e., being influenced to do the S, C, or I action, shows in Figure 4.4. We observed that the pattern we observed for the interface users in Figure 4.2 is still preserved for each group in all the platform-communities. i.e., influence experienced from the focal platform is

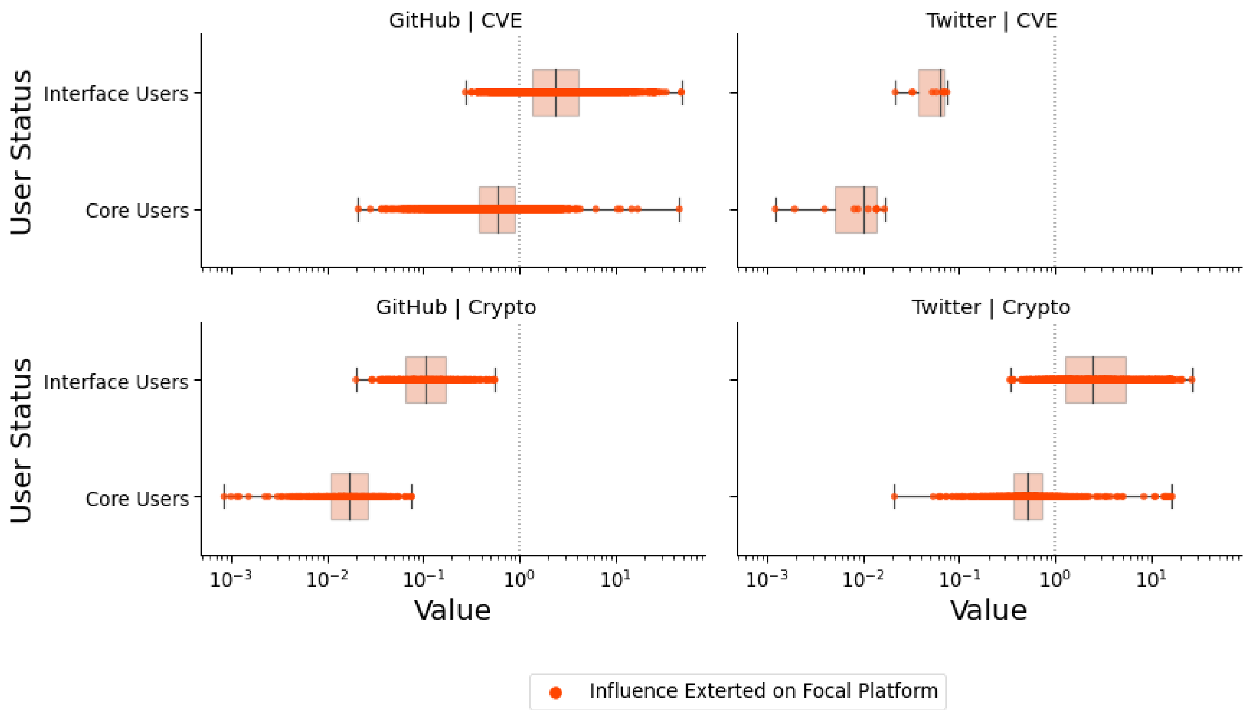

Figure 4.3: The distributions of influence exerted on the focal platform-community by interface users and core users. Interface users exert more influence on the focal platform-community than core users in each platform-community. The subplot titles represent the focal platform | community.

higher than that of the corresponding platform for each group in GH-CVE and TW-Crypto, and influence experienced from the corresponding other platform is greater than that from the focal platform for TW-CVE and GH-Crypto. However, it was difficult to observe a clear difference between the distributions when we considered the same source of influence (i.e., whether from the focal platform or the other platform).

The Kruskal-Wallis H test was performed to infer the statistical significance of the observations, and the results are shown in Table 4.4, Table 4.5, Table 4.6, and Table 4.7. Because of the lack of data points (< 5) in some groups in TW-CVE, we didn't consider TW-CVE for this statistical anal-

Table 4.3: Mann-Whitney U test statistics for the alternative hypothesis that the influence exerted by interface users and core users on the focal platform is significantly different. (*a*: Asymptotic significance value, *b*: exact significance value.) Group numbers represent the following groups: 1 – Influence exerted by the interface users on the focal platform-community , 2 – Influence exerted by core users on the focal platform-community . The influence exerted by interface users is significantly higher than that of core users.

Pltform-Community	Group_num	N	Mean Rank	Mann-Whitney U	p
GH-CVE	1	3319	4714.94	879746.00	$< 0.001^a$
	2	3320	1925.48		
TW-CVE	1	10	15.50	0.00	$< 0.001^b$
	2	10	5.50		
GH-Crypto	1	226	336.66	867.00	$< 0.001^a$
	2	227	117.82		
TW-Crypto	1	849	1216.52	48798.00	$< 0.001^a$
	2	849	482.48		

ysis. Table 4.4 shows the Kruskal-Wallis H test statistics and mean value ranks of the experienced influence corresponding to each group. The data revealed that at least one distribution differs significantly from the other distributions for each considered platform-communities (asym. p-value < 0.001). Subsequently, the test statistics of the pairwise comparison of the distributions of each group which is shown in Table 4.5, Table 4.6, and Table 4.7 revealed the mean experienced influence from the focal platform is statically significantly higher than that from the corresponding other platform for GH-CVE and TW-Crypto. But, the mean experienced influence from corresponding other platform is statically significantly higher than that from the focal platform for GH-Crypto. As we discussed earlier, this could be explained by the social dynamics of the communities.

Moreover, when comparing the groups within the same source of influence, (1) in GH-CVE, the mean influence experienced by Interface Users_C > Interface Users_S > Interface Users_I and the values are statistically significant only between the Interface Users$_C$ and the Interface Users$_I$ (adj. p-value < 0.05) regardless of whether influence coming from the focal platform or from the TW-CVE, (2) in GH-Crypto, mean influence experienced by Interface Users_C > Interface Users_I >

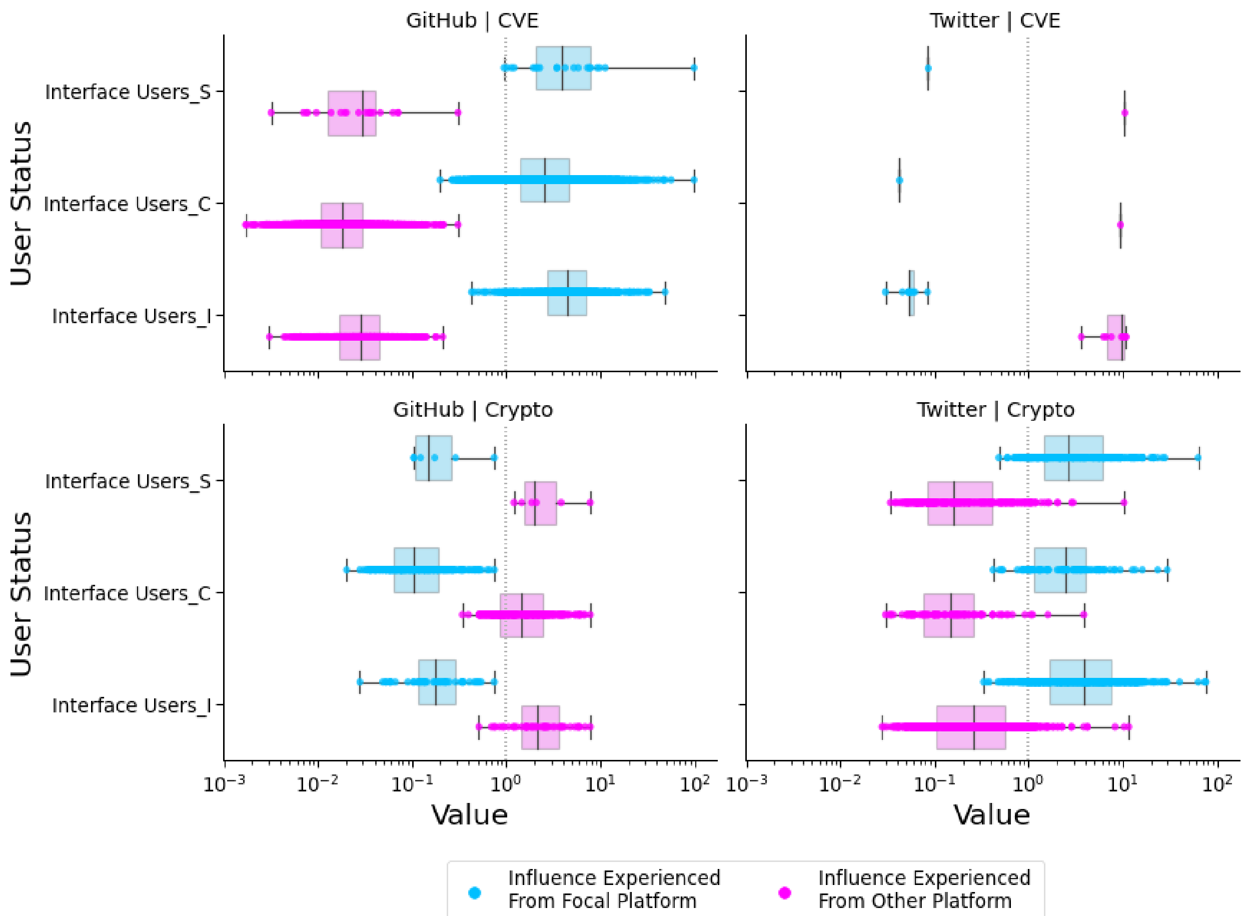

Figure 4.4: The distributions of influence experienced from the focal platform-community and the corresponding other platform-community by the groups of interface users, which are categorized based on the action they are influenced to do. The subplot titles represent the focal platform | community.

Interface Users_*S* but the values are not statistically significant (adj. p-value < 0.05) for any group combination regardless of whether influence coming from the focal platform or from the TW-Crypto, (3) in TW-Crypto, mean influence experienced by Interface Users_*C* > Interface Users_*S* > Interface Users_*I* and the values are statistically significant only between the Interface Users_*C* and the Interface Users_*I* (adj. p-value < 0.05) when the influence is coming from GH-Crypto. These results suggest that the influence experienced by the interface users can be significantly different based on the actions that they are influenced to do. Still, the variations depend on the platform and the community they work on.

Table 4.4: Kruskal-Wallis H test statistics for the alternative hypothesis that the influence experienced by the interface users from the focal platform-community and from the corresponding other platform-community are significantly different based on the action they are influenced to do, *I*, *C*, or *S*. Group numbers represent the following groups: 1- Influence experienced by Interface Users_*I* from the focal platform 2- Influence experienced by Interface Users_*I* from the other platform 3- Influence experienced by Interface Users_*C* from the focal platform 4- Influence experienced by Interface Users_*C* from the other platform 5- Influence experienced by Interface Users_*S* from the focal platform 6- Influence experienced by Interface Users_*S* from the other platform. The alternate hypothesis is accepted for each platform-community.

Platform-Community	Group_num	N	Mean Rank	Asym. p
GH-CVE	1	20	6273.35	<0.001
	2	20	2299.20	
	3	3312	5767.90	
	4	3312	1860.05	
	5	589	6510.28	
	6	589	2517.21	
GH-Crypto	1	6	190.83	<0.001
	2	6	459.67	
	3	226	131.42	
	4	226	406.75	
	5	45	175.39	
	6	45	451.37	
TW-Crypto	1	256	1490.01	<0.001
	2	256	491.82	
	3	85	1413.68	
	4	85	415.98	
	5	697	1560.94	
	6	697	581.18	

Finally, the distributions of exerted influence on the focal platform by the interface users based on the action that they are influenced to do are shown in Figure 4.5. TW-CVE was not considered for the analysis due to a small number of data points (<5) in some groups. We observed slight variations in the amount of influence exerted by the interface users based on the actions they are

Table 4.5: The test statistics of the pairwise comparison of the GH-CVE groups in Table 4.4 for the alternative hypothesis that the group i and Group j are significantly different. Group numbers represent the following groups: 1- Influence experienced by Interface Users_*I* from the focal platform 2- Influence experienced by Interface Users_*I* from the other platform 3- Influence experienced by Interface Users_*C* from the focal platform 4- Influence experienced by Interface Users_*C* from the other platform 5- Influence experienced by Interface Users_*S* from the focal platform 6- Influence experienced by Interface Users_*S* from the other platform. Bonferroni adjusted p values are compared with the significance level of 0.05 to test the alternative hypothesis.

Pltform-Community	Group i-Group j	Test Statistic	p	Adj. p
GH-CVE	4-2	439.151	0.387	1.000
	4-6	-657.166	<0.001	<0.001
	4-3	3907.849	<0.001	<0.001
	4-1	4413.301	<0.001	<0.001
	4-5	-4650.227	<0.001	<0.001
	2-6	-218.015	0.672	1.000
	2-3	-3468.698	<0.001	<0.001
	2-1	3974.150	<0.001	<0.001
	2-5	-4211.076	<0.001	<0.001
	6-3	3250.683	<0.001	<0.001
	6-1	3756.135	<0.001	<0.001
	6-5	3993.061	0.000	<0.001
	3-1	505.452	0.320	1.000
	3-5	-742.378	<0.001	<0.001
	1-5	-236.926	0.645	1.000

influenced to do. The Kruskal-Wallis H test statistics which are shown in Table 4.8 and the test statistics of pairwise comparisons of each group which hare shown in Table 4.9 revealed that the (1) in GH-CVE, mean influence exerted on focal platform by Interface Users_*C* > Interface Users_*S* >Interface Users_*I* and the mean values are statistically significant only between Interface Users_*C* and Interface Users_*I* (adj. p-value <0.05), (2) in GH-Crypto, mean influence exerted on focal platform by Interface Users_*C* > Interface Users_*I* > Interface Users_*S* and the mean values are statistically significant only between Interface Users_*C* and Interface Users_*I* (adj. p-value <0.05) and , (3) in TW-Crypto, mean influence exerted on focal platform by Interface Users_*C* > Interface Users_*S* >Interface Users_*I* and the mean values are statistically significant between

Table 4.6: The test statistics of the pairwise comparison of the GH-Crypto groups in Table 4.4 for the alternative hypothesis that the group *i* and Group *j* are significantly different. Group numbers represent the following groups: 1- Influence experienced by Interface Users_*I* from the focal platform 2- Influence experienced by Interface Users_*I* from the other platform 3- Influence experienced by Interface Users_*C* from the focal platform 4- Influence experienced by Interface Users_*C* from the other platform 5- Influence experienced by Interface Users_*S* from the focal platform 6- Influence experienced by Interface Users_*S* from the other platform. Bonferroni adjusted *p* values are compared with the significance level of 0.05 to test the alternative hypothesis.

Pltform-Community	Group *i*-Group *j*	Test Statistic	*p*	Adj. *p*
GH-Crypto	3-5	-43.966	0.092	1.000
	3-1	59.411	0.370	1.000
	3-4	-275.332	<0.001	<0.001
	3-6	-319.944	<0.001	<0.001
	3-2	328.244	<0.001	<0.001
	3-1	15.444	0.824	1.000
	3-4	231.366	<0.001	<0.001
	5-6	-275.978	<0.001	<0.001
	5-2	284.278	<0.001	0.001
	1-4	-215.921	0.001	0.017
	1-6	-260.533	<0.001	0.003
	1-2	-268.833	0.004	0.054
	4-6	-44.612	0.088	1.000
	4-2	52.912	0.424	1.000
	6-2	8.300	0.905	1.000

Interface Users_*C* and Interface Users_*I* (adj. *p*-value <0.05) as well as between Interface Users_*C* and Interface Users_*I* (adj. *p*-value <0.05). From these results, we can conclude that there is a significant difference in the influence exerted on the focal platform by the interface users based on the action that they are influenced to do. Though these variations depend on the platform and the community that they are worked on, the results show that the influence exerted by the interface users who are influenced to do initiation action is significantly higher than that from the interface users who are influenced to do the contribution action, in all three platform-communities that we consider. This might result from the novelty of the information and where it is presented during a conversation. The influence relationships across platforms allow users to bring novel information

Table 4.7: The test statistics of the pairwise comparison of the TW-Crypto groups in Table 4.4 for the alternative hypothesis that the group i and Group j are significantly different. Group numbers represent the following groups: 1- Influence experienced by Interface Users_*I* from the focal platform 2- Influence experienced by Interface Users_*I* from the other platform 3- Influence experienced by Interface Users_*C* from the focal platform 4- Influence experienced by Interface Users_*C* from the other platform 5- Influence experienced by Interface Users_*S* from the focal platform 6- Influence experienced by Interface Users_*S* from the other platform. Bonferroni adjusted p values are compared with the significance level of 0.05 to test the alternative hypothesis.

Pltform-Community	Group i-Group j	Test Statistic	p	Adj. p
TW-Crypto	4-2	75.836	0.312	1.000
	4-6	-165.197	0.016	0.247
	4-3	997.694	<0.001	<0.001
	4-1	1074.031	<0.001	<0.001
	4-5	-1144.956	<0.001	<0.001
	2-6	-89.361	0.041	0.620
	2-3	-921.858	<0.001	<0.001
	2-1	998.195	<0.001	<0.001
	2-5	-1069.120	<0.001	<0.001
	6-3	832.497	<0.001	<0.001
	6-1	908.834	<0.001	<0.001
	6-5	979.759	<0.001	<0.001
	3-1	76.337	0.309	1.000
	3-5	-147.262	0.032	0.487
	1-5	-70.925	0.105	1.000

to the focal platform. Suppose the novel information they obtain is used to initiate a post rather than contributing to a post in the middle of an existing conversation. In that case, that information could gain more attraction from the other users.

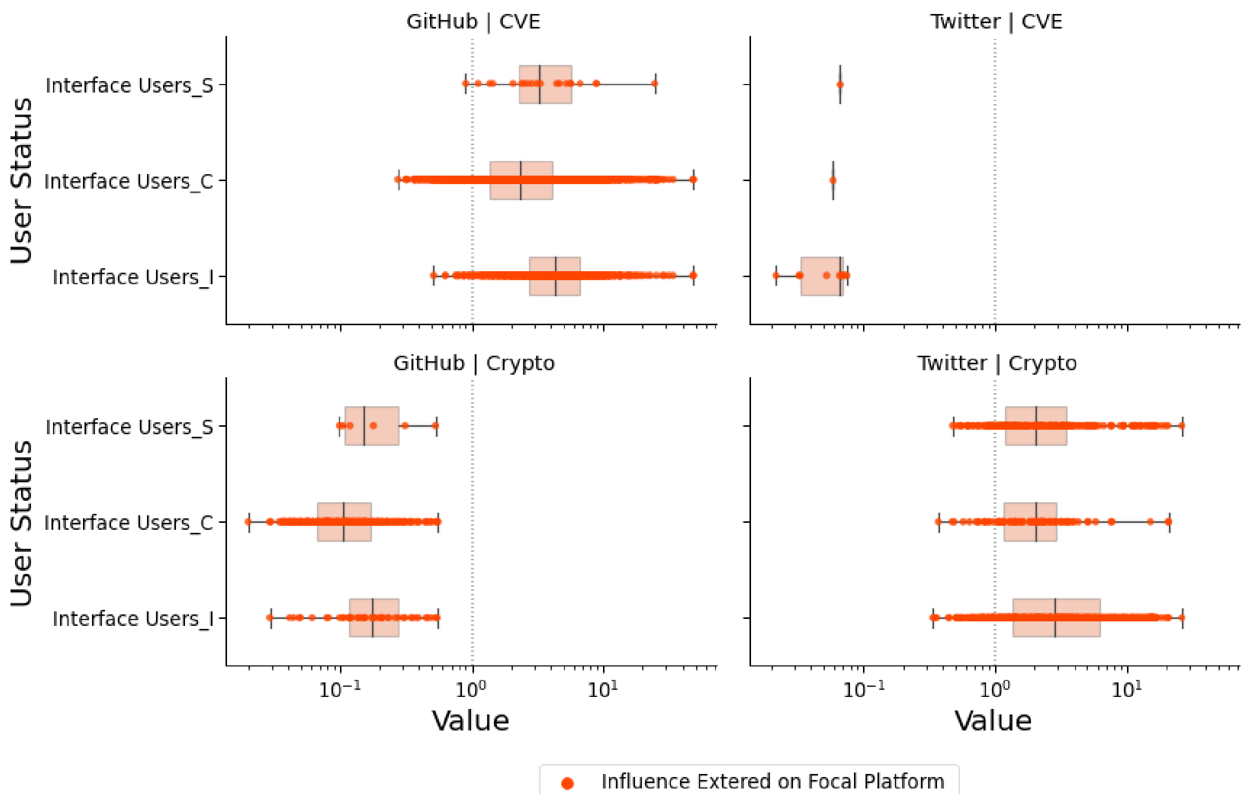

Figure 4.5: The distributions of influence exerted on the focal platform-community by the groups of interface users, which are categorized based on the action they are influenced to do. The subplot titles represent the focal platform | community.

Table 4.8: Kruskal-Wallis H test statistics for the alternate hypothesis that the amount of Influence Exerted on the focal platform by the interface users based on the action they are influenced to do, *I*, *C*, or *S* are significantly different. The group numbers represent the following groups: 1- Influence exerted by the Interface Users_*S* on the focal platform 2- Influence exerted by the Interface Users_*C* on the focal platform 3- Influence exerted by the Interface Users_*I* on the focal platform .

Platform-Community	Group_*num*	N	Mean Rank	Asym. *p*
GH-CVE	1	20	2344.85	<0.001
	2	3312	1836.78	
	3	589	2646.49	
GH-Crypto	1	6	185.67	<0.001
	2	226	130.04	
	3	45	177.79	
TW-Crypto	1	256	465.11	<0.001
	2	85	423.42	
	3	697	551.19	

Table 4.9: The test statistics of the pairwise comparison between the groups in Table 4.8 for the alternate hypothesis that the group *i* and Group *j* are significantly different. The group numbers represent the following groups: 1- Influence exerted by the Interface Users_*S* on the focal platform 2- Influence exerted by the Interface Users_*C* on the focal platform 3- Influence exerted by the Interface Users_*I* on the focal platform.

Platform-Community	Group *i*-Group *j*	Test Statistic	*p*	Adj.*p*
GH-CVE	2-1	508.074	0.045	0.136
	2-3	-809.710	0.000	0.000
	1-3	-301.636	0.241	0.724
GH-Crypto	2-3	-47.751	0.000	0.001
	2-1	55.629	0.093	0.280
	3-1	7.878	0.821	1.000
TW-Crypto	2-1	41.694	0.267	0.800
	2-3	-127.776	0.000	0.001
	1-3	-86.082	0.000	0.000

Discussion

We investigated the cross-platform influence in the influence process of social media users by comparing four social networks, the cryptocurrency and CVE communities of GitHub and Twitter. First, we used platform-independent action classification, which are initiation, contribution, and sharing, to model influence as a multidimensional entity. Then we used transfer entropy to capture the nine types of casual influence relationships resulting from those three actions and quantify the influence. Finally, with the premise that the existence of at least one type of influence relationship between a user pair implies the social influence relationship between them, we reconstruct the influence network of the interconnected social media platforms using the identified influence relationships.

In order to explore the cross-platform influence in the influence process, we analyzed the influence experienced from the focal platform and cross-platform and the influence exerted on the focal platform by two groups of users: interface users and core users. Interface users are users in the focal platform with a relatively higher number of influence relationships coming from the other platform. Conversely, core users are users in the focal platform with relatively fewer influence relationships coming from the other platform. Further, we execute the same analysis between the groups of interface users categorized by the action they are influenced to perform.

Our results from an exploratory analysis of the distribution of the number of influence relationships within the platform and across the platforms for two communities find evidence that users do not always tend to form more ties within the focal platform. Instead, their primary source of influence, i.e., the focal platform or the other platform, depends on the social dynamics of how the community operates. In particular, we find that GitHub users have more influence relationships from within GitHub than from Twitter in the CVE community, which is development-driven. But, Twitter users have more influence relationships from GitHub than from within Twitter. In contrast, we observe the opposite pattern in the discussion-driven cryptocurrency community. Subsequently, our results indicate that the interface users and core users of the GitHub CVE community and Twitter cryptocurrency community experienced more influence from the focal platform than from Twitter and GitHub, respectively. In contrast, GitHub cryptocurrency and Twitter CVE communities experienced more influence from Twitter and GitHub, respectively, than from their focal platform.

Nevertheless, the comparison between interface users and core users revealed that interface users experienced significantly greater influence than core users. Moreover, it shows that interface users are more influential than core users. The fact that interface users have a higher chance than the core users to expose to novel information and bring them to the focal platform explains this result.

Moreover, we find significant differences in influence experienced by the interface users based on the actions they are influenced to do in GitHub CVE and Twitter cryptocurrency communities. In particular, we find those interface users who are influenced to do initiation actions are significantly more vulnerable to being influenced than those who are influenced to make contributions. This indicates that users interact across platforms more to initiate posts in the focal platform than contribute to the existing content. Furthermore, our results from GitHub CVE, GitHub, and Twitter cryptocurrency communities reveal significant differences between the influence exerted on the focal platform based on the action that interface users influence to do. However, our results show that the interface users who are influenced to do initiation are more influential than those influenced to contribute across all platforms and communities. As contribution actions happen in the middle of an existing conversation/post, they might attract fewer users than initial posts resulting in the difference between the effect of initiation and contribution action on further actions by other users. These results lead us to conclude that there are significant differences in the influence experienced by interface users and the influence exerted by them on others based on the actions they are influenced to do. However, these differences will vary depending on the platform and the community.

CHAPTER 5: CONCLUSIONS

Even though online social influence has long been studied, this research area still has a lot of space to grow because of the growing complexity of the online social media ecosystem. In this dissertation, I present a method to examine social networks through the perspective of behavioral influence propagation by addressing four drawbacks in the previous literature; 1) assuming the monolithic notion of influence; i.e., all influence is measured using one number, 2) lack of generalizability of the proposed algorithms or measures of influence, 3) assuming the influence as a property of the user instead of the property of the relationship, and 4) lack of studies on the significance of the cross-platform influence on users in the influence process. As influence can depend on many factors, users' actions can be one such factor, and different actions may influence users differently. However, assuming a monolithic notion of influence hinders the comprehensive understanding of such behavioral influences. Further, the lack of generalizability of proposed algorithms limits comparing and contrasting the results over different platforms. Further, as both users play a role in an influence relationship, assuming the influence as a user's property will reduce the accuracy of results. Moreover, as the number of different OSM used by OSM users is growing, treating OSM platforms as independent entities might overlook the advantages and disadvantages of information spread across platforms.

Therefore, to address these issues, first, I produce a transfer-entropy-based method to measure behavioral influence between online social media users. As transfer entropy captures the relationships between two random processes, it helps to model the influence as a property of the relationship. This method abstracts user actions into initiation, contribution, and sharing actions, allowing us to analyze users' social influence on different platforms and across platforms. We use this action classification to define influence relationships. Then, I examine how the strength of influence relationships vary between interest cryptocurrency and CVE communities and across GitHub and

Twitter platforms. Finally, we compare the empirical results against social influence propagation patterns expected by scale-free null models. Our findings show that characteristics of influence cascades and the patterns of influence are determined by the platform and community of the users.

Secondly, I use the same framework to reconstruct the interconnected OSM and investigate the cross-platform influence in the influence process of OSM users, as it does not limit us from analyzing the cross-platform influence relationships. Then I examine the social dynamics between the users with a relatively higher number of influence relationships coming from the corresponding cross-platform and those with the relatively fewer influence relationships coming from the corresponding cross-platform. The experiments consider the empirical data of Cryptocurrency and the CVE community on GitHub and Twitter platforms. Our results indicate that the interface users experience greater influence than core users. Also, results show that the interface users exert more influence on the focal platform than the core users. Moreover, our results show that interface users who are influenced to do initiation action exert more influence than those who are influenced to do contribution action.

Therefore, through this research, we extend the existing literature by discarding the traditional monolithic notion of influence and by providing new insights into the differences and similarities of how social influence propagates within and across different communities and platforms. Overall, our study contributes to science by (1) presenting a novel and generalized method to track influence relationships caused by actions of OSM, (2) providing new insights to improve state-of-art methods that assume a monolithic notion of influence and homogeneous populations in the social influence analysis field, (3) characterizing influence cascades caused by actions of social network media across platforms and communities and presenting the evidence to show that the depth and structure of influence cascades are determined by the platform and community, (4) presenting a method to investigate the significance of cross-platform influence in the influence process, (5) presenting the evidence to show that the users with a relatively higher number of social influence relationships across platforms are more vulnerable to being influenced and influential than the others, and (6) providing insights for marketing firms, online community leaders, and policymakers to develop intervention strategies to control the spread of information or misinformation.

Future Work

Although we have analyzed networks within the confines of two platforms and two communities in this study, it is exciting to explore the influence cascades and cross-platform influence of other communities and platforms in the future. For example, both communities we focused on in this study are related to technology. However, the communities related to entertainment or information operations might show utterly different influence cascade structures and cross-platform influence patterns. Also, even if we consider misinformation related to health and politics, we might see different behavioral structural properties. We believe that this would be beneficial to the research community.

APPENDIX A: SUPPLEMENTARY MATERIALS - STATISTICAL TESTS

Spearman's Correlation Test

We present here scatter plots corresponding to each Spearman's correlation test which we perform. We observe that except Crypto community (Figure A.1) neither CVE community (Figure A.2) nor GitHub (Figure A.3), Twitter (Figure A.4) platforms shown to have clear monotonic relation. Further, we observe that contribution to sharing and sharing to contribution influence relationships in the Crypto community not showing a clear monotonic relation.

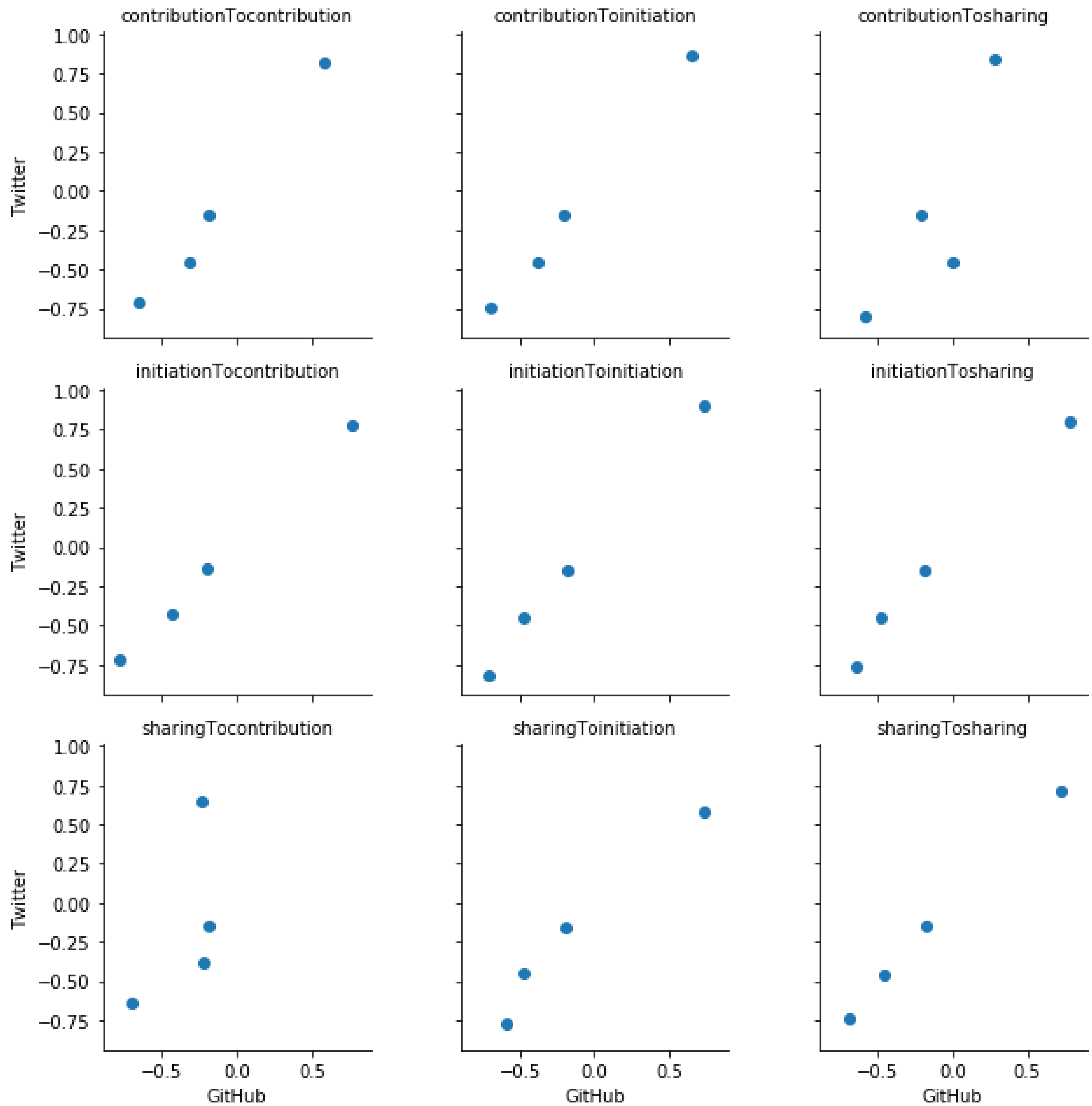

Figure A.1: The scatter plot of residuals of median total influence values of GitHub and Twitter platforms in Crypto community.

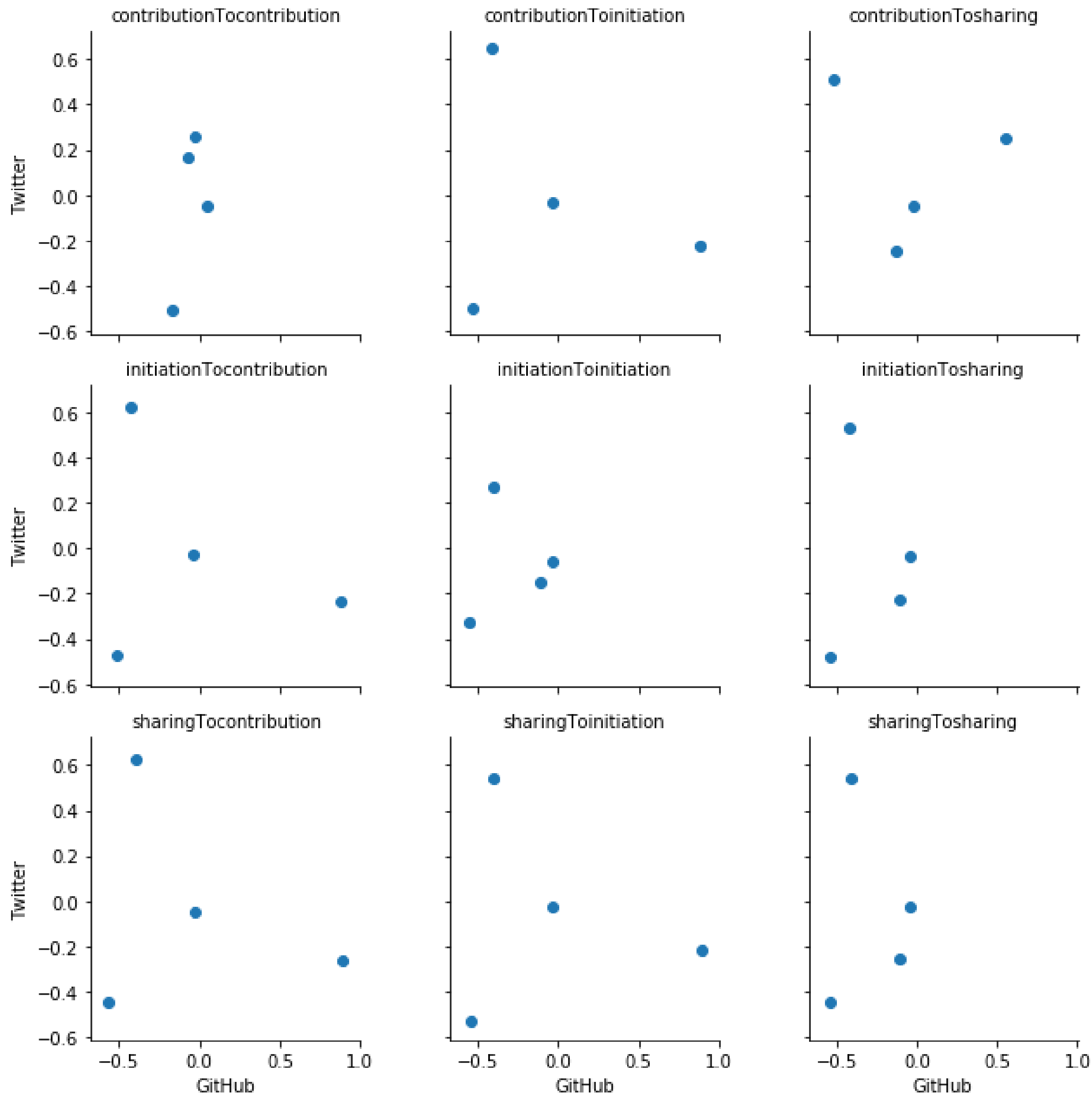

Figure A.2: The scatter plot of residuals of median total influence values of GitHub and Twitter platforms in CVE community.

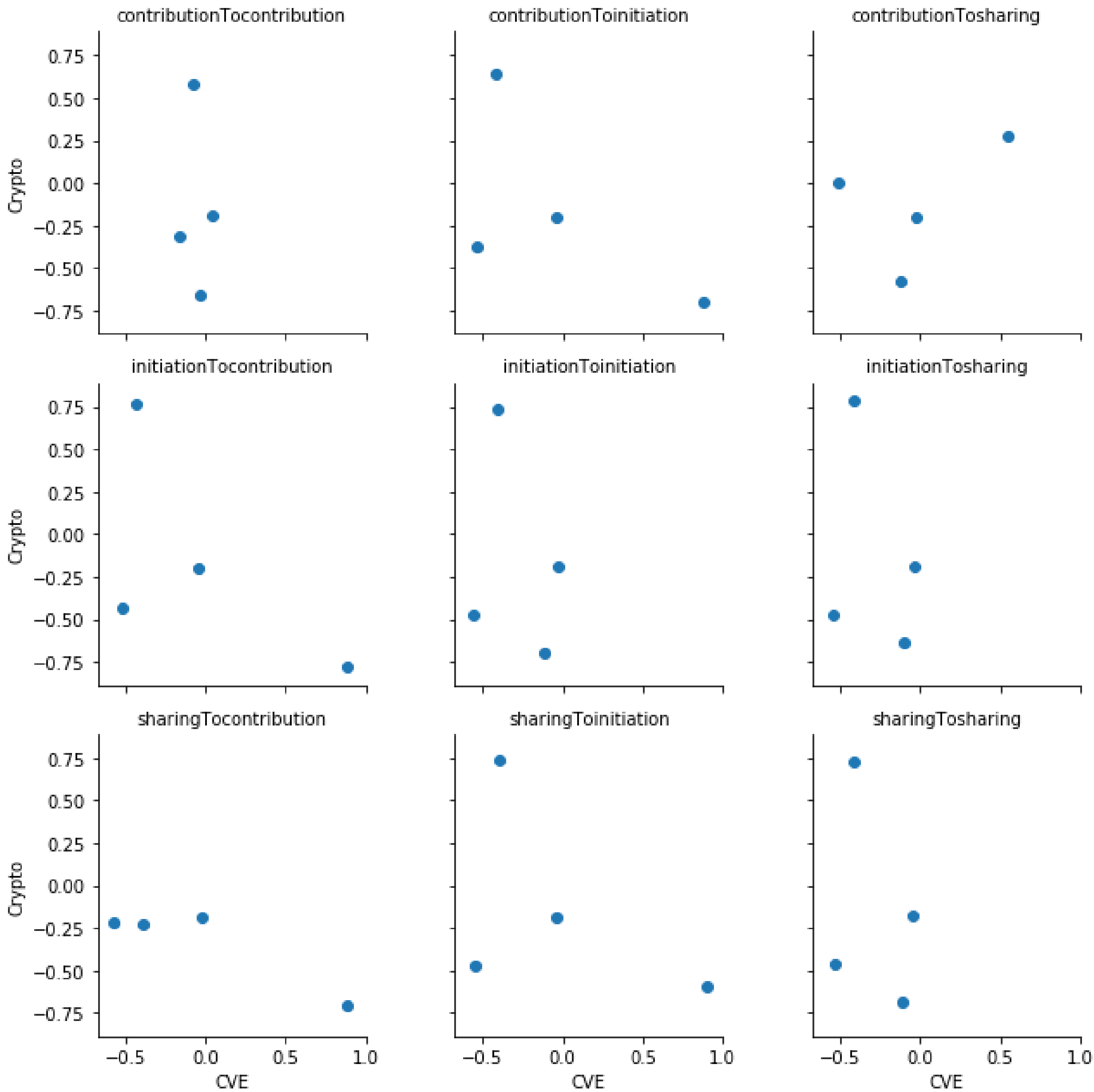

Figure A.3: The scatter plot of residuals of median total influence values of Crypto and CVE community in GitHub.

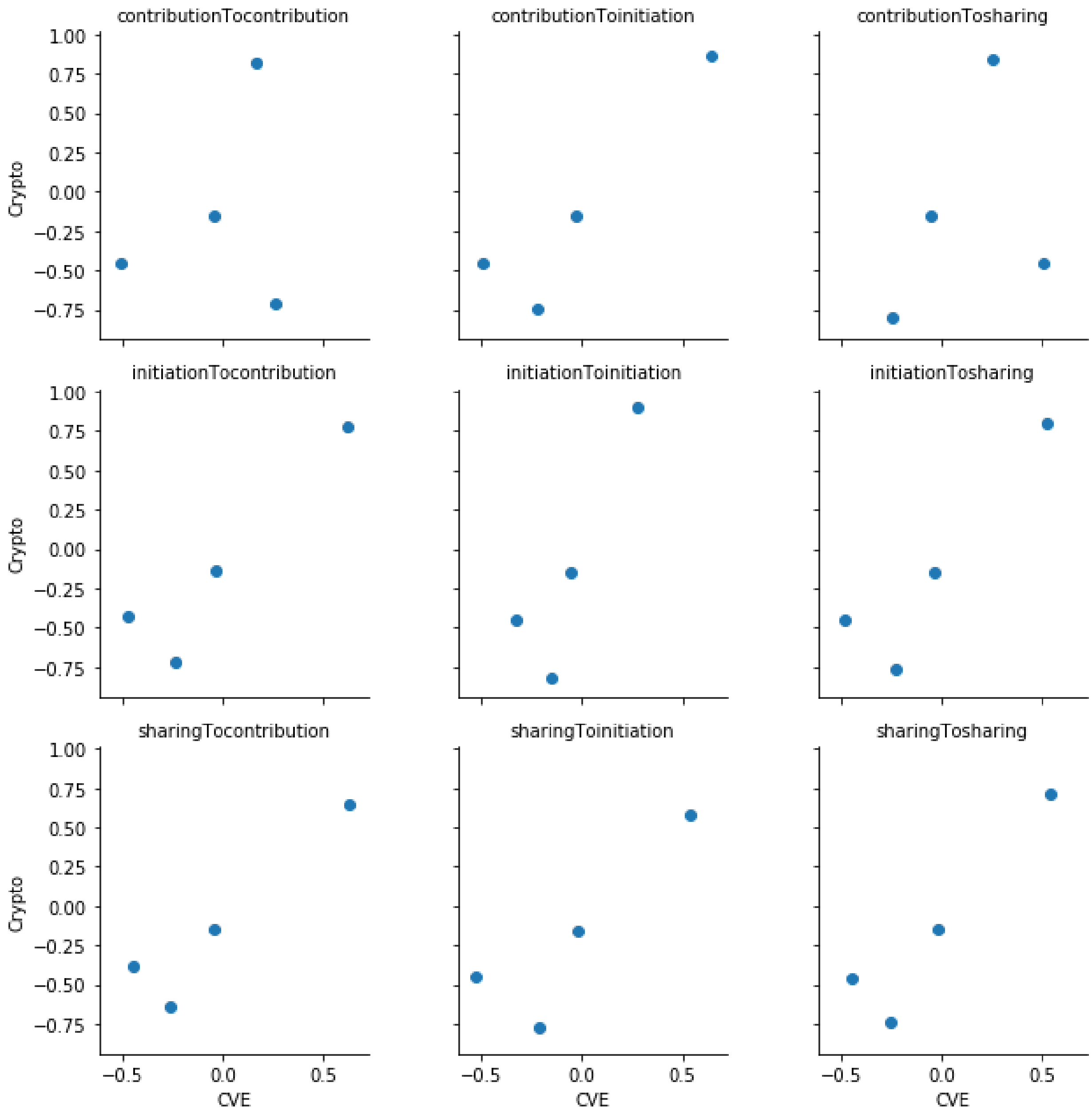

Figure A.4: The scatter plot of residuals of median total influence values of Crypto and CVE community in Twitter.

Factorial ANOVA Test

We examine the use of the factorial ANOVA test as a statistical significance test to infer the similarities of the structure of the influence cascades across platforms and communities. Forty influence cascades are chosen randomly from the set of influence cascades extracted from each empirical influence networks. The residual values between total normalized influence vector components and the median total normalized influence vector components of the influence cascades from the corresponding scale-free network are calculated. Data is grouped by platform-community, influence relationships, and level, and the normality of each data set was examined using the Shapiro-Wilk test. The p-values are compared with 0.05. Not all the data sets were able to satisfy the normality assumption as shown in Tables A.1, A.2, A.3, and A.4.

Table A.1: Test statistics of factorial ANOVA (Part I).

Platform-Community	Influence Relationship	Level	Statistics	p-value	H_0
GitHub-CVE	contributionTocontribution	1	0.78	2.72E-06	rejected
GitHub-CVE	contributionTocontribution	2	0.9	1.79E-03	rejected
GitHub-CVE	contributionTocontribution	3	0.93	1.53E-02	rejected
GitHub-CVE	contributionTocontribution	4	0.8	8.28E-06	rejected
GitHub-CVE	contributionToinitiation	1	0.39	1.04E-11	rejected
GitHub-CVE	contributionToinitiation	2	1	1.00E+00	not rejected
GitHub-CVE	contributionToinitiation	3	1	1.00E+00	not rejected
GitHub-CVE	contributionToinitiation	4	1	1.00E+00	not rejected
GitHub-CVE	contributionTosharing	1	0.23	3.25E-13	rejected
GitHub-CVE	contributionTosharing	2	0.64	1.05E-08	rejected
GitHub-CVE	contributionTosharing	3	0.61	4.91E-09	rejected
GitHub-CVE	contributionTosharing	4	1	1.00E+00	not rejected
GitHub-CVE	initiationTocontribution	1	0.39	1.04E-11	rejected
GitHub-CVE	initiationTocontribution	2	1	1.00E+00	not rejected
GitHub-CVE	initiationTocontribution	3	1	1.00E+00	not rejected
GitHub-CVE	initiationTocontribution	4	1	1.00E+00	not rejected
GitHub-CVE	initiationToinitiation	1	1	1.00E+00	not rejected
GitHub-CVE	initiationToinitiation	2	1	1.00E+00	not rejected
GitHub-CVE	initiationToinitiation	3	1	1.00E+00	not rejected
GitHub-CVE	initiationToinitiation	4	1	1.00E+00	not rejected
GitHub-CVE	initiationTosharing	1	1	1.00E+00	not rejected
GitHub-CVE	initiationTosharing	2	1	1.00E+00	not rejected
GitHub-CVE	initiationTosharing	3	1	1.00E+00	not rejected
GitHub-CVE	initiationTosharing	4	1	1.00E+00	not rejected
GitHub-CVE	sharingTocontribution	1	0.23	3.25E-13	rejected
GitHub-CVE	sharingTocontribution	2	1	1.00E+00	not rejected
GitHub-CVE	sharingTocontribution	3	1	1.00E+00	not rejected
GitHub-CVE	sharingTocontribution	4	1	1.00E+00	not rejected
GitHub-CVE	sharingToinitiation	1	1	1.00E+00	not rejected
GitHub-CVE	sharingToinitiation	2	1	1.00E+00	not rejected
GitHub-CVE	sharingToinitiation	3	1	1.00E+00	not rejected
GitHub-CVE	sharingToinitiation	4	1	1.00E+00	not rejected
GitHub-CVE	sharingTosharing	1	1	1.00E+00	not rejected
GitHub-CVE	sharingTosharing	2	1	1.00E+00	not rejected
GitHub-CVE	sharingTosharing	3	1	1.00E+00	not rejected
GitHub-CVE	sharingTosharing	4	1	1.00E+00	not rejected

Table A.2: Test statistics of factorial ANOVA (Part II).

Platform-Community	Influence Relationship	Level	Statistics	p-value	H_0
GitHub-Crypto	contributionTocontribution	1	0.73	3.02E-07	rejected
GitHub-Crypto	contributionTocontribution	2	0.89	1.32E-03	rejected
GitHub-Crypto	contributionTocontribution	3	0.91	3.21E-03	rejected
GitHub-Crypto	contributionTocontribution	4	0.3	1.36E-12	rejected
GitHub-Crypto	contributionToinitiation	1	0.42	2.43E-11	rejected
GitHub-Crypto	contributionToinitiation	2	0.77	2.07E-06	rejected
GitHub-Crypto	contributionToinitiation	3	0.77	2.11E-06	rejected
GitHub-Crypto	contributionToinitiation	4	1	1.00E+00	not rejected
GitHub-Crypto	contributionTosharing	1	1	1.00E+00	not rejected
GitHub-Crypto	contributionTosharing	2	0.95	5.75E-02	not rejected
GitHub-Crypto	contributionTosharing	3	0.95	5.75E-02	not rejected
GitHub-Crypto	contributionTosharing	4	1	1.00E+00	not rejected
GitHub-Crypto	initiationTocontribution	1	0.15	6.64E-14	rejected
GitHub-Crypto	initiationTocontribution	2	0.6	3.49E-09	rejected
GitHub-Crypto	initiationTocontribution	3	0.57	1.13E-09	rejected
GitHub-Crypto	initiationTocontribution	4	1	1.00E+00	not rejected
GitHub-Crypto	initiationToinitiation	1	1	1.00E+00	not rejected
GitHub-Crypto	initiationToinitiation	2	0.63	7.16E-09	rejected
GitHub-Crypto	initiationToinitiation	3	0.23	3.25E-13	rejected
GitHub-Crypto	initiationToinitiation	4	1	1.00E+00	not rejected
GitHub-Crypto	initiationTosharing	1	1	1.00E+00	not rejected
GitHub-Crypto	initiationTosharing	2	0.6	3.26E-09	rejected
GitHub-Crypto	initiationTosharing	3	0.15	6.64E-14	rejected
GitHub-Crypto	initiationTosharing	4	1	1.00E+00	not rejected
GitHub-Crypto	sharingTocontribution	1	0.65	1.72E-08	rejected
GitHub-Crypto	sharingTocontribution	2	0.65	1.68E-08	rejected
GitHub-Crypto	sharingTocontribution	3	0.77	1.70E-06	rejected
GitHub-Crypto	sharingTocontribution	4	1	1.00E+00	not rejected
GitHub-Crypto	sharingToinitiation	1	0.15	6.64E-14	rejected
GitHub-Crypto	sharingToinitiation	2	0.29	1.19E-12	rejected
GitHub-Crypto	sharingToinitiation	3	1	1.00E+00	not rejected
GitHub-Crypto	sharingToinitiation	4	1	1.00E+00	not rejected
GitHub-Crypto	sharingTosharing	1	1	1.00E+00	not rejected
GitHub-Crypto	sharingTosharing	2	0.15	6.64E-14	rejected
GitHub-Crypto	sharingTosharing	3	1	1.00E+00	not rejected
GitHub-Crypto	sharingTosharing	4	1	1.00E+00	not rejected

Table A.3: Test statistics of factorial ANOVA (Part III).

Platform-Community	Influence Relationship	Level	Statistics	p-value	H_0
Twitter-CVE	contributionTocontribution	1	0.15	6.64E-14	rejected
Twitter-CVE	contributionTocontribution	2	1	1.00E+00	not rejected
Twitter-CVE	contributionTocontribution	3	1	1.00E+00	not rejected
Twitter-CVE	contributionTocontribution	4	1	1.00E+00	not rejected
Twitter-CVE	contributionToinitiation	1	0.15	6.64E-14	rejected
Twitter-CVE	contributionToinitiation	2	0.62	5.59E-09	rejected
Twitter-CVE	contributionToinitiation	3	0.15	6.64E-14	rejected
Twitter-CVE	contributionToinitiation	4	1	1.00E+00	not rejected
Twitter-CVE	contributionTosharing	1	0.22	2.81E-13	rejected
Twitter-CVE	contributionTosharing	2	0.63	7.17E-09	rejected
Twitter-CVE	contributionTosharing	3	0.61	4.53E-09	rejected
Twitter-CVE	contributionTosharing	4	1	1.00E+00	not rejected
Twitter-CVE	initiationTocontribution	1	0.59	2.07E-09	rejected
Twitter-CVE	initiationTocontribution	2	0.65	1.67E-08	rejected
Twitter-CVE	initiationTocontribution	3	1	1.00E+00	not rejected
Twitter-CVE	initiationTocontribution	4	1	1.00E+00	not rejected
Twitter-CVE	initiationToinitiation	1	0.79	4.87E-06	rejected
Twitter-CVE	initiationToinitiation	2	0.85	9.62E-05	rejected
Twitter-CVE	initiationToinitiation	3	0.74	4.97E-07	rejected
Twitter-CVE	initiationToinitiation	4	1	1.00E+00	not rejected
Twitter-CVE	initiationTosharing	1	0.62	6.69E-09	rejected
Twitter-CVE	initiationTosharing	2	0.8	7.55E-06	rejected
Twitter-CVE	initiationTosharing	3	0.64	1.21E-08	rejected
Twitter-CVE	initiationTosharing	4	1	1.00E+00	not rejected
Twitter-CVE	sharingTocontribution	1	0.15	6.64E-14	rejected
Twitter-CVE	sharingTocontribution	2	0.54	5.09E-10	rejected
Twitter-CVE	sharingTocontribution	3	0.23	3.25E-13	rejected
Twitter-CVE	sharingTocontribution	4	1	1.00E+00	not rejected
Twitter-CVE	sharingToinitiation	1	0.45	4.74E-11	rejected
Twitter-CVE	sharingToinitiation	2	0.7	1.17E-07	rejected
Twitter-CVE	sharingToinitiation	3	0.44	3.48E-11	rejected
Twitter-CVE	sharingToinitiation	4	1	1.00E+00	not rejected
Twitter-CVE	sharingTosharing	1	0.33	2.45E-12	rejected
Twitter-CVE	sharingTosharing	2	0.71	1.30E-07	rejected
Twitter-CVE	sharingTosharing	3	0.46	6.44E-11	rejected
Twitter-CVE	sharingTosharing	4	1	1.00E+00	not rejected

Table A.4: Test statistics of factorial ANOVA (Part IV).

Platform-Community	Influence Relationship	Level	Statistics	p-value	H_0
Twitter-Crypto	contributionTocontribution	1	0.36	5.16E-12	rejected
Twitter-Crypto	contributionTocontribution	2	0.65	1.40E-08	rejected
Twitter-Crypto	contributionTocontribution	3	0.56	8.47E-10	rejected
Twitter-Crypto	contributionTocontribution	4	1	1.00E+00	not rejected
Twitter-Crypto	contributionToinitiation	1	0.53	3.80E-10	rejected
Twitter-Crypto	contributionToinitiation	2	0.69	7.75E-08	rejected
Twitter-Crypto	contributionToinitiation	3	0.52	2.98E-10	rejected
Twitter-Crypto	contributionToinitiation	4	1	1.00E+00	not rejected
Twitter-Crypto	contributionTosharing	1	0.26	6.03E-13	rejected
Twitter-Crypto	contributionTosharing	2	0.61	4.73E-09	rejected
Twitter-Crypto	contributionTosharing	3	0.55	6.01E-10	rejected
Twitter-Crypto	contributionTosharing	4	1	1.00E+00	not rejected
Twitter-Crypto	initiationTocontribution	1	0.42	1.96E-11	rejected
Twitter-Crypto	initiationTocontribution	2	0.36	5.14E-12	rejected
Twitter-Crypto	initiationTocontribution	3	0.36	5.42E-12	rejected
Twitter-Crypto	initiationTocontribution	4	0.15	6.64E-14	rejected
Twitter-Crypto	initiationToinitiation	1	0.45	4.32E-11	rejected
Twitter-Crypto	initiationToinitiation	2	0.36	5.45E-12	rejected
Twitter-Crypto	initiationToinitiation	3	0.22	2.70E-13	rejected
Twitter-Crypto	initiationToinitiation	4	0.15	6.64E-14	rejected
Twitter-Crypto	initiationTosharing	1	0.31	1.83E-12	rejected
Twitter-Crypto	initiationTosharing	2	0.25	5.11E-13	rejected
Twitter-Crypto	initiationTosharing	3	0.24	3.90E-13	rejected
Twitter-Crypto	initiationTosharing	4	0.15	6.64E-14	rejected
Twitter-Crypto	sharingTocontribution	1	0.67	2.85E-08	rejected
Twitter-Crypto	sharingTocontribution	2	0.64	1.25E-08	rejected
Twitter-Crypto	sharingTocontribution	3	0.58	1.52E-09	rejected
Twitter-Crypto	sharingTocontribution	4	0.15	6.64E-14	rejected
Twitter-Crypto	sharingToinitiation	1	0.75	7.77E-07	rejected
Twitter-Crypto	sharingToinitiation	2	0.79	4.20E-06	rejected
Twitter-Crypto	sharingToinitiation	3	0.49	1.37E-10	rejected
Twitter-Crypto	sharingToinitiation	4	1	1.00E+00	not rejected
Twitter-Crypto	sharingTosharing	1	0.71	1.30E-07	rejected
Twitter-Crypto	sharingTosharing	2	0.58	1.55E-09	rejected
Twitter-Crypto	sharingTosharing	3	0.48	1.06E-10	rejected
Twitter-Crypto	sharingTosharing	4	0.15	6.64E-14	rejected

APPENDIX B: IRB OUTCOME LETTER

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901, 407-882-2012 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

NOT HUMAN RESEARCH DETERMINATION

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Ivan I Garibay, Alexander Mantzaris, Gita Reese Sukthankar, Stephen M Fiore**

Date: **February 21, 2018**

Dear Researcher:

On 02/21/2018, the IRB determined that the following proposed activity is not human research as defined by DHHS regulations at 45 CFR 46 or FDA regulations at 21 CFR 50/56:

Type of Review:	Not Human Research Determination
Project Title:	Deep Agent: A Framework for Information Spread and Evolution in Social Networks
Investigator:	Ivan I Garibay
IRB ID:	SBE-18-13732
Funding Agency:	DARPA
Grant Title:	Deep Agent: A Framework for Information Spread and Evolution in Social Networks
Research ID:	1062483

University of Central Florida IRB review and approval is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are to be made and there are questions about whether these activities are research involving human subjects, please contact the IRB office to discuss the proposed changes.

This letter is signed by:

Signature applied by Jennifer Neal-Jimenez on 02/21/2018 04:32:37 PM EST

Designated Reviewer

APPENDIX C: ENTROPY COPYRIGHT PERMISSION TO REUSE THE AUTHOR PUBLISHED PAPER

CS

Chathurani Senevirathna
Thu 2/25/2021 11:27 AM
To: entropy@mdpi.com
Cc: Ivan Garibay

Dear editors,

I would like to reuse this article ([Entropy] Manuscript ID: entropy-1054302; doi: 10.3390/e23020160) in my dissertation and the dissertation will be published electronically by the University of Central Florida. May I know if it allows or is there any procedure that I can request permission for that?

Thank you.

Best Regards,

Chathurani Senevirathna
Graduate Research Assistant
Complex Adaptive Systems Lab
University of Central Florida

Room 314, Engineering II,
12800 Pegasus Dr.,
Orlando, FL 32816-2993

[Entropy] Manuscript ID: entropy-1054302; doi: 10.3390/e23020160. Paper has been published.

Entropy <entropy@mdpi.com>
Thu 2/25/2021 11:21 PM
To: Chathurani Senevirathna
Cc: Ivan Garibay

Dear Chathurani,

Thank you for your message. This paper is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Therefore when publishing your article, please cite the original article as follows:

Senevirathna, C.; Gunaratne, C.; Rand, W.; Jayalath, C.; Garibay, I.
Influence Cascades: Entropy-Based Characterization of Behavioral Influence Patterns in Social Media. Entropy 2021, 23, 160.
https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdoi.org%2F10.3390%2Fe23020160&data=04%7C01%7Ccsenevirathna%40Knights.ucf.edu%7C9b012e6979654efe414108d8da0e0935%7C5b16e18278b3412c919668342689eeb7%7C0%7C1%7C637499101123249854%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=2xaJh0k8D5CqL2NLbqqyGCyOEbp69EHDR9DyfcggXA0%3D&reserved=0

Any question, please feel free to let me know.

Kind regards,

Connie Xiong
Managing Editor
Entropy (https://nam02.safelinks.protection.outlook.com/?url=http%3A%2F%2Fwww.mdpi.com%2Fjournal%2Fentropy&data=04%7C01%7Ccsenevirathna%40Knights.ucf.edu%7C9b012e6979654efe414108d8da0e0935%7C5b16e18278b3412c919668342689eeb7%7C0%7C1%7C637499101123249854%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=08yv8no3UNej3hMCOs3KPPK3viuOQs7dMKVHM8nvxnw%3D&reserved=0)

LIST OF REFERENCES

[1] danah m. boyd and Nicole B. Ellison. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 10 2007.

[2] Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1(1):5–es, May 2007.

[3] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna Gummadi. Measuring user influence in twitter: The million follower fallacy. *Proceedings of the International AAAI Conference on Web and Social Media*, 4(1), May 2010.

[4] Haichuan Zhao. Information quality or entities' interactivity? understanding the determinants of social network-based brand community participation. *Future Internet*, 11(4):87, Apr 2019.

[5] Salvatore Carta, Alessandro Sebastian Podda, Diego Reforgiato Recupero, Roberto Saia, and Giovanni Usai. Popularity prediction of instagram posts. *Information*, 11(9):453, Sep 2020.

[6] Ludovico Boratto, Salvatore Carta, Gianni Fenu, and Roberto Saia. Using neural word embeddings to model user behavior and detect user segments. *Knowledge-Based Systems*, 108:5 – 14, 2016. New Avenues in Knowledge Bases for Natural Language Processing.

[7] Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D.I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, sep 2012.

[8] Duncan J. Watts and Peter Sheridan Dodds. Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*, 34(4):441–458, 05 2007.

[9] Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, 2010.

[10] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. Everyone's an influencer: Quantifying influence on twitter. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, page 65–74, New York, NY, USA, 2011. Association for Computing Machinery.

[11] Shaokai Wang, Xutao Li, Yunming Ye, Shanshan Feng, Raymond Lau, Xiaohui Huang, and Xiaolin Du. Anchor link prediction across attributed networks via network embedding. *Entropy*, 21(3):254, Mar 2019.

[12] Sinan Aral and Paramveer S. Dhillon. Social influence maximization under empirical influence models. *Nature Human Behaviour*, 2(6):375–382, jun 2018.

[13] Zi-Ke Zhang, Chuang Liu, Xiu-Xiu Zhan, Xin Lu, Chu-Xu Zhang, and Yi-Cheng Zhang. Dynamics of information diffusion and its applications on complex networks. *Physics Reports*, 651:1 – 34, 2016. Dynamics of information diffusion and its applications on complex networks.

[14] Sancheng Peng, Yongmei Zhou, Lihong Cao, Shui Yu, Jianwei Niu, and Weijia Jia. Influence analysis in social networks: A survey. *Journal of Network and Computer Applications*, 106:17 – 32, 2018.

[15] Mohammed Ali Al-Garadi, Kasturi Dewi Varathan, Sri Devi Ravana, Ejaz Ahmed, Ghulam Mujtaba, Muhammad Usman Shahid Khan, and Samee U. Khan. Analysis of online social network connections for identification of influential users: Survey and open research issues. *ACM Comput. Surv.*, 51(1), January 2018.

[16] Kousik Das, Sovan Samanta, and Madhumangal Pal. Study on centrality measures in social networks: a survey. *Social network analysis and mining*, 8(1):1–11, 2018.

[17] Roshan Rabade, Nishchol Mishra, and Sanjeev Sharma. Survey of influential user identification techniques in online social networks. In *Recent advances in intelligent informatics*, pages 359–370. Springer, 2014.

[18] Evelien Otte and Ronald Rousseau. Social network analysis: a powerful strategy, also for the information sciences. *Journal of information Science*, 28(6):441–453, 2002.

[19] Pei Li, Jeffrey Xu Yu, Hongyan Liu, Jun He, and Xiaoyong Du. Ranking individuals and groups by influence propagation. In Joshua Zhexue Huang, Longbing Cao, and Jaideep Srivastava, editors, *Advances in Knowledge Discovery and Data Mining*, pages 407–419, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[20] W. Wang and W. N. Street. A novel algorithm for community detection and influence ranking in social networks. In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, pages 555–560, Aug 2014.

[21] Xujun Li, Yezheng Liu, Yuanchun Jiang, and Xiao Liu. Identifying social influence in complex networks: A novel conductance eigenvector centrality model. *Neurocomputing*, 210:141 – 154, 2016.

[22] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, September 1999.

[23] Sergey Brin and Lawrence Page. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 56(18):3825–3833, 2012. The WEB we live in.

[24] Daniel Tunkelang. A twitter analog to pagerank, 2009.

[25] W. Chen, S. Cheng, X. He, and F. Jiang. Influencerank: An efficient social influence measurement for millions of users in microblog. In *2012 Second International Conference on Cloud and Green Computing*, pages 563–570, Nov 2012.

[26] Zhao-yun Ding, Yan Jia, Bin Zhou, Yi Han, Li He, and Jian-feng Zhang. Measuring the spreadability of users in microblogs. *Journal of Zhejiang University SCIENCE C*, 14(9):701–710, 2013.

[27] Yuto Yamaguchi, Tsubasa Takahashi, Toshiyuki Amagasa, and Hiroyuki Kitagawa. Turank: Twitter user ranking based on user-tweet graph analysis. In Lei Chen, Peter Triantafillou, and Torsten Suel, editors, *Web Information Systems Engineering – WISE 2010*, pages 240–253, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[28] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. Twitterrank: Finding topic-sensitive influential twitterers. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, page 261–270, New York, NY, USA, 2010. Association for Computing Machinery.

[29] Lamjed Ben Jabeur, Lynda Tamine, and Mohand Boughanem. Active microbloggers: Identifying influencers, leaders and discussers in microblogging networks. In Liliana Calderón-Benavides, Cristina González-Caro, Edgar Chávez, and Nivio Ziviani, editors, *String Processing and Information Retrieval*, pages 111–117, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[30] An Zeng and Cheng-Jun Zhang. Ranking spreaders by decomposing complex networks. *Physics Letters A*, 377(14):1031–1035, 2013.

[31] Mohammed Ali Al-garadi, Kasturi Dewi Varathan, and Sri Devi Ravana. Identification of influential spreaders in online social networks using interaction weighted k-core decomposition method. *Physica A: Statistical Mechanics and its Applications*, 468:278–288, 2017.

[32] Phil Brown and Junlan Feng. Measuring user influence on twitter using modified k-shell decomposition. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1), Jul. 2011.

[33] Beiming Sun and Vincent TY Ng. Identifying Influential Users by Their Postings in Social Networks. pages 128–151. Springer, Berlin, Heidelberg, 2013.

[34] Sancheng Peng, Jian Li, and Aimin Yang. Entropy-Based Social Influence Evaluation in Mobile Social Networks. pages 637–647. Springer, Cham, nov 2015.

[35] Richard Hanna, Andrew Rohm, and Victoria L. Crittenden. We're all connected: The power of the social media ecosystem. *Business Horizons*, 54(3):265–273, 2011. SPECIAL ISSUE: SOCIAL MEDIA.

[36] A. Smith and M. Anderson. Social media use in 2018. *PEW RESEARCH CENTER* `https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/`, 2018.

[37] Kepios. Global social media stats. *DataReportal* `https://datareportal.com/social-media-users`, 2021.

[38] Goulet L. S. Rainie H. Purcell K. Hampton, K. N. Social networking sites and our lives: Pew internet american life project. *PEW RESEARCH CENTER* `https://www.pewresearch.org/internet/2011/06/16/social-networking-sites-and-our-lives/`, 2011.

[39] Chang-Dae Ham, Joonghwa Lee, Jameson L Hayes, and Young Han Bae. Exploring sharing behaviors across social media platforms. *International Journal of Market Research*, 61(2):157–177, 2019.

[40] R Burt. *Structural Holes: The Social Structure of Competition.* Cambridge: Harvard University Press, 1995.

[41] Sinan Aral and Marshall Van Alstyne. The diversity-bandwidth trade-off. *American Journal of Sociology*, 117(1):90–171, 2011.

[42] Kai Shu, Suhang Wang, Jiliang Tang, Reza Zafarani, and Huan Liu. User identity linkage across online social networks: A review. *Acm Sigkdd Explorations Newsletter*, 18(2):5–17, 2017.

[43] Fan Zhou, Kunpeng Zhang, Shuying Xie, and Xucheng Luo. Learning to correlate accounts across online social networks: An embedding-based approach. *INFORMS Journal on Computing*, 32(3):714–729, 2020.

[44] Cody Buntain, Richard Bonneau, Jonathan Nagler, and Joshua A. Tucker. Youtube recommendations and effects on sharing across online social platforms. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1), apr 2021.

[45] Minkyoung Kim, David Newth, and Peter Christen. Macro-level information transfer across social networks. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 321–322, 2014.

[46] Abhishek Bhattacharjee. Measuring influence across social media platforms: Empirical analysis using symbolic transfer entropy. 2019.

[47] Thomas Schreiber. Measuring information transfer. *Physical Review Letters*, 85:461–464, Jul 2000.

[48] Chathurani Senevirathna, Chathika Gunaratne, William Rand, Chathura Jayalath, and Ivan Garibay. Influence cascades: Entropy-based characterization of behavioral influence patterns in social media. *Entropy*, 23(2), 2021.

[49] Eytan Adar, Li Zhang, Lada A Adamic, and Rajan M Lukose. Implicit+ structure+ and+ the+ dynamics+ of+ blogspace. 2004.

[50] Jure Leskovec, Mary McGlohon, Christos Faloutsos, Natalie Glance, and Matthew Hurst. Patterns of cascading behavior in large blog graphs. In *Proceedings of the 2007 SIAM international conference on data mining*, pages 551–556. SIAM, 2007.

[51] Ravi Kumar, Mohammad Mahdian, and Mary McGlohon. Dynamics of conversations. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, page 553–562, New York, NY, USA, 2010. Association for Computing Machinery.

[52] Greg Ver Steeg and Aram Galstyan. Information transfer in social media. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, page 509–518, New York, NY, USA, 2012. Association for Computing Machinery.

[53] Nhathai Phan, Javid Ebrahimi, Dave Kil, Brigitte Piniewski, and Dejing Dou. Topic-Aware Physical Activity Propagation in a Health Social Network. *IEEE Intelligent Systems*, 31(1):5–14, jan 2016.

[54] Mengdie Hu, Shixia Liu, Furu Wei, Yingcai Wu, John Stasko, and Kwan-Liu Ma. Breaking News on Twitter. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, New York, New York, USA, 2012. ACM Press.

[55] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, page 591–600, New York, NY, USA, 2010. Association for Computing Machinery.

[56] Daniel M. Romero, Wojciech Galuba, Sitaram Asur, and Bernardo A. Huberman. Influence and passivity in social media. In Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, and Michalis Vazirgiannis, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 18–33, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[57] Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, and Jie Tang. Deepinf: Social influence prediction with deep learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining*, KDD '18, page 2110–2119, New York, NY, USA, 2018. Association for Computing Machinery.

[58] C. K. Leung, A. Cuzzocrea, J. J. Mai, D. Deng, and F. Jiang. Personalized deepinf: Enhanced social influence prediction with deep learning and transfer learning. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 2871–2880, 2019.

[59] Xin Zhang, Ding-Ding Han, Ruiqi Yang, and Ziqiao Zhang. Users' participation and social influence during information spreading on twitter. *PLOS ONE*, 12(9):1–17, 09 2017.

[60] Xuegong Chen, Jie Zhou, Zhifang Liao, Shengzong Liu, and Yan Zhang. A novel method to rank influential nodes in complex networks based on tsallis entropy. *Entropy*, 22(8):848, Jul 2020.

[61] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948.

[62] Greg Ver Steeg and Aram Galstyan. Information-theoretic measures of influence based on content dynamics. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, WSDM '13, page 3–12, New York, NY, USA, 2013. Association for Computing Machinery.

[63] Saike He, Xiaolong Zheng, Daniel Zeng, Kainan Cui, Zhu Zhang, and Chuan Luo. Identifying peer influence in online social networks using transfer entropy. In G. Alan Wang, Xiaolong Zheng, Michael Chau, and Hsinchun Chen, editors, *Intelligence and Security Informatics*, pages 47–61, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[64] Daniel Gruhl, R. Guha, David Liben-Nowell, and Andrew Tomkins. Information diffusion through blogspace. In *Proceedings of the 13th International Conference on World Wide Web*, WWW '04, page 491–501, New York, NY, USA, 2004. Association for Computing Machinery.

[65] P. Alex Dow, Lada Adamic, and Adrien Friggeri. The anatomy of large facebook cascades. *Proceedings of the International AAAI Conference on Web and Social Media*, 7(1), Jun. 2013.

[66] Jose Cadena, Gizem Korkmaz, Chris J. Kuhlman, Achla Marathe, Naren Ramakrishnan, and Anil Vullikanti. Forecasting social unrest using activity cascades. *PLOS ONE*, 10(6):1–27, 06 2015.

[67] Yunxiang Xiong, Zhangyuan Meng, Beijun Shen, and Wei Yin. Mining developer behavior across github and stackoverflow. In *SEKE*, pages 578–583, 2017.

[68] Sophie F Waterloo, Susanne E Baumgartner, Jochen Peter, and Patti M Valkenburg. Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New media & society*, 20(5):1813–1831, may 2018.

[69] Minkyoung Kim, David Newth, and Peter Christen. Macro-level information transfer in social media: Reflections of crowd phenomena. *Neurocomputing*, 172:84 – 99, 2016.

[70] Chathika Gunaratne, Chathurani Senevirathna, Chathura Jayalath, Nisha Baral, William Rand, and Ivan Garibay. A multi-action cascade model of conversation. In *5th International Conference on Computational Social Science*, 2019.

[71] R.C. Lupton and J.M. Allwood. Hybrid sankey diagrams: Visual analysis of multidimensional data for understanding resource use. *Resources, Conservation and Recycling*, 124:141 – 151, 2017.

[72] Tomek Strzalkowski, Teresa Harrison, Ning Sa, Gregorios Katsios, and Ellisa Khoja. Github as a social network. In Tareq Z. Ahram, editor, *Advances in Artificial Intelligence, Software and Systems Engineering*, pages 379–390, Cham, 2019. Springer International Publishing.

[73] Béla Bollobás, Christian Borgs, Jennifer T Chayes, and Oliver Riordan. Directed scale-free graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, volume 3, pages 132–139. Society for Industrial and Applied Mathematics, 2003.

[74] Prasha Shrestha, Arun Sathanur, Suraj Maharjan, Emily Saldanha, Dustin Arendt, and Svitlana Volkova. Multiple social platforms reveal actionable signals for software vulnerability awareness: A study of github, twitter and reddit. *Plos one*, 15(3):e0230250, 2020.