Global Contagion of Non-Viral Information

Alon Bartal
Department of Software and
Information Systems Engineering
Ben-Gurion University
bartala@post.bgu.ac.il

Gilad Ravid
Department of Industrial and
Management Engineering
Ben-Gurion University
rgilad@bgu.ac.il

Oren Tsur
Department of Software and
Information Systems Engineering
Ben-Gurion University
orentsur@bgu.ac.il

Abstract

Contagion in Online Social Networks (OSN) is typically measured by the tendency of users to re-post information or to adopt a new behavior after exposure to that information/behavior. Most contagion research is bound by modeling: (i) only local neighbor-to-neighbor contagion (ii) the spread of viral information. However, most contagion events are non-viral and can also occur globally by non-neighbors through for example, exposure to information by exploratory browsing, or by content recommendation algorithms. This study is the first to address the phenomenon of both global and local contagion of non-viral information in a quantitative way. Analysis of Twitter networks reveals the prevailing nature of global contagion, the different temporal patterns between global and local contagion, and the ways it varies across topical categories. An interesting finding shows that users who retweeted due to global contagion have more Followers than those who retweeted due to local contagion.

1.   Introduction

Exposure to information is often modeled as a local neighbor-to-neighbor spreading process [1, 2, 3]. For example, the Following relationships of a Twitter user (accounts a user chooses to follow) largely determine the information she is exposed to [4]. Such exposure can lead to information contagion, e.g. re-posting a message (retweeting), or adopting the use of a a specific linguistic pattern or a hashtag [2].

A local exposure in an OSN occurs when a user reads content posted by a neighbor (e.g. a Twitter user whom she follows). However, exposure also occurs in other, global, ways [5], e.g. by exposure to promoted posts [3], external sources [6, 7] like mainstream media, exploratory browsing (users who seek and traverse information) [8], or the use of a recommender systems [9, 10]. We define this type of exposure as *global exposure* which can lead to *global contagion*.

Contagion is typically modeled through one of two approaches: structural and non-structural. Structural models [2, 11, 12], infer the spread of local contagion by users who expose network neighbors to their posted content. These models assume that information can reach users only within the network, but ignore the spread of information by non-neighbors. Two-step diffusion models [13, 11] argue that information first spreads through the media to opinion leaders and then propagates from node-to-node to a broader population. Non-structural models [14, 15] consider a contagion process based on the global state of the network, but largely ignore the nodal level. Understanding how contagion spread in OSNs, requires the consideration of both local and global effects.

Myers et al., [6] found evidence of global contagion due to exposure to external sources (main stream media) by detecting a sequence of external events that caused bursts of activity on Twitter. Albeit, indicating that their results are only valid for posts that were shared at least 50 times. However, most posts are shared a scant number of times [16, 17] as virality takes the form of a long-tailed distribution [18]. While most studies focus on the spread of viral content, research on non-viral content is sparse [19].

The contribution of this study is twofold: (i) Considering both local and global contagions, and (ii) Modeling the contagion and reach of non-viral content. This perspective yields significant insights into online human behavior.

We use Twitter data to learn the spread of different types of contagion by retweets across topics. Differentiating between local and global non-viral contagion, our findings complement the vast body of scholarship regarding the spread of viral content.

Organization.   Section 2 sets the theoretical background. Section 3 elaborates upon the methods used in this study, and describes the analyzed dastaset. Section 4 reports the empirical results. Finally, Section 5 covers the contributions, innovations, and the limitations of this study.

2. Related Work

Social media has revolutionized how people create and consume information. One way by which information can reach a user is through local exposure to content, posted by other network neighbors (i.e., nodes at a distance of 1-hop from the user). For example, Twitter exposes users to information through the Following-Followers lists [20]. Such exposure can lead to information contagion (infection, influence, adoption). A local contagion on Twitter can occur when a user retweets a message [2, 21], posted by a neighbor. Structural contagion models build upon the assumption that contagion spreads by a user-to-user mechanism.

2.1. Structural Contagion Models

Most studies about contagion spread focus on user-to-user exposure that leads to local contagion, causing the exposed user to engages in activities [22]. Leskovec et al., [23] studied a recommendation graph and measured the extent to which a user's activity in recommending a product is contagious, i.e., affecting her local neighbors' purchase decisions. Sun at al., [1] studied the contagion affect on users' participation in fan-pages after some number of their friends have done so. Bakshy et al., [24] examined contagion in adopting the use of "gestures" between friends in the Second Life platform. The Linear Threshold Model [25] focuses on contagion, defined by user transition from a non-active to an active state, following a transition in the participation of a network neighbor. The Independent Cascade Model [26], allows a user to activate its inactive neighbors, thus contagion is propagated by causing a user to participate.

As explained in Section 1, OSN users are increasingly exposed to varied content [3] by local neighbors, and by global sources. Global exposure can lead to global contagion [5]. For example, Facebook and Reddit allow global exposure via trending topics that appear on a user's front page [27]. Twitter exposes users to information posted by non-neighbors via exposure to hashtags [28] and via exposure to promoted content on a user's Timeline feed [4]. Thus, the need to consider global contagion beyond network structure arises.

2.2. Non-structural Contagion Models

Non-structural contagion models do not rely on the structure of the network to infer contagion. For example, the Susceptible-Infected-Resistant (SIR) and the Susceptible-Infectious-Susceptible (SIS) models [29] assume that every individual has the same probability to be infected, i.e., all individuals have the

same contact rate, which can be indicated by an edge formation in a network. Leskovec et al., [23] propose a SIS model where all users have the same probability to be infected. However, contagion in OSNs is not evenly distributed among users [30] and is likely to depend on exposure rates [6]. The Linear Influence Model assumes a static network structure were users are affected by the rate of contagious users in the network [14]. Wang et al., [31] predict contagion by temporal and topological dynamics but, regarding network structure, consider only the distance from a user to the contagious user.

Myers et al., [6] found evidence of global contagion in OSNs by exposure to external (out-of-network) sources such as mainstream media, albeit, indicated that their results are only valid for posts that were shared at least 50 times. However, many posts fail to attract attention [16] and do not become viral [17]. Detecting viral messages have been extensively studied [3, 2]. Since most posts are non-viral [18, 17], studying how non-viral content spreads can lead to important insights on human behavior online [2].

To summarize, an exposure event occurs when a user is exposed to information. Such exposure can lead to information contagion e.g. retweeting (RT) a message on Twitter [21]. Twitter users can pass interesting pieces of information to their followers by a RT activity which is the action of re-posting someone else's tweet [20]. A retweeted message indicates the ability of the user who created the message to generate contagious information.

Most contagion research is focused on modeling only local user-to-user contagion, ignoring non-neighbors, or contagion by all network users, largely ignoring the topology of the network. Detecting both local and global contagion while considering network structure is crucial to better understand human behavior online, as reflected by user interactions [32].

2.3. Types of User Interactions

OSNs provide users platforms to form interaction activities. These online activities create an *activity network* and include, for example, post-reply, and content sharing activities by which users expose others to information. Twitter, for example, supports two type of networks: a *social network* based on the Following-Follower relationships, and an activity network of who tweets whom [20]. In a Twitter RT network, nodes are authors of the tweets and edges represent the contagion spreading paths [33].

Dow et al., [17] focused on contagion spread by resharing a photo on Facebook at least 100 times while also considering the social network of friends. They found that viral cascades have a large depth (larger

than 20), many re-shares occurred within a single step from the source, and that a significant fraction of re-shares occurred deeper than one level into the cascade. Liben-Nowell et al., [34] measured cascades with hundreds of steps. These studies did not analyze the dynamics that govern non-viral information spread.

Some OSNs only provide users the ability to interact, without the ability to form an explicit relationship (social) network. Therefore, the relationships between users are implicit [35]. However, Shi et al., [36] found that the contagion spreading patterns in the activity networks are similar to those of explicit relationships networks in online spaces. Moreover, the Twitter social network was found to be less informative regarding contagion spread than the activity network, which exposes users by a RT, a mention, or a reply [7].

As discussed, most posts fail to attract attention. Lack of sufficient knowledge about the spread of non-viral content and sparse research about global contagion in OSNs are the motivation of this study.

The proposed modeling approach of local and global contagion, which provides multiple pathways in which contagion can spread, is the basis for asking the following two Research Questions (RQ):

1. Does global and local contagion of non-viral information have different spreading patterns in OSNs?

2. Does global and local contagion of non-viral information spread differently in OSNs depending on the discussed topics?

Our work differs significantly from other studies by considering both local contagion as well as global contagion that are not restricted to exposure by mainstream media, or user-to-user mechanisms, and can also result from exposure to content promoted by algorithms. Our findings regarding the spread on non-viral content complement the vast body of scholarship regarding the spread of viral content.

3. Methods and Data

3.1. Detecting Local and Global Contagion

This section is designed to cope with RQ #1. Consider a directed social network as a graph $G = (V, E)$ were nodes (V) represent Twitter users and edges (E) represent Following relationships among users. If a user v_j follows another user v_i then, there exists a directed edge $e_{ij} \in E$, connecting them. The direction of e_{ij} signals an exposure path where v_i post is (potentially) read by v_j who follows v_i. Such exposure might result in a spread of a local contagion.

More formally, let w denote an information nugget (e.g. an original tweet) which is posted at time t_0 by user $v_0 \in V$. The retweeting spread of w at times $t_1...t_k$ by users $v_1, ..., v_j \in V$, along with v_0 could be thought of as a temporal *activity network* $G_{Tw} = (V_{Tw}, E_{Tw})$, laid over the social network G. In G_{Tw}, nodes (V_{Tw}) represent users who retweet w at times t_i t_k, starting with v_0 and edges (E_{Tw}) are defined by retweet (RT) relations which are known from the dataset. If $v_i \in V_{Tw}$ has retweeted a tweet that was posted by $v_j \in V_{Tw}$ then, $e_{ij} \in E_{Tw}$. The edge e_{ij} points to the node who spread the contagion (i.e., the infecting node).

Next, we define the two types of contagion.

Local contagion. We define a contagion as local if $e_{ij} \in E_{Tw}$ and $e_{ji} \in E$. That is, v_i retweeted an original post w *after* one of the users she follows have retweeted or wrote it, hence the contagion is local.

Global contagion. We define a contagion as global if $e_{ij} \in E_{Tw}$ and $e_{ji} \notin E$. That is, v_i retweeted w *before* any of the users she follows have retweeted it, hence the contagion is global.

This contagion process is illustrated in Figure 1. A user v_0 posts a tweet w at t_0, locally exposing her followers (v_2, v_4). User v_2 retweeted w at time t_1, presenting a local contagion, since v_2 follows v_0. In addition, v_1 who does not follow v_0 has also retweeted w at time t_1. We say that, v_1 was exposed by some global source to w and was infected globally (i.e., global contagion). In Figure 1, the rest of the contagion events (v_3 at t_2, v_4 at t_3 and v_5 at t_4) are local since their opposite edges in E_{Tw} exist in E.

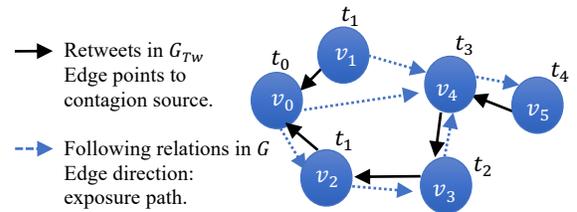

Figure 1: An activity network with a social network. Contagion time (t_i) of a node is indicated near the node.

To cope with RQ #2, we analyze user retweet activity by topics which requires detecting the discussed topics.

3.2. Topics Detection

To detect the topics of tweets, we follow two steps. First, detect the list of topics in the dataset by inspecting a random sample of tweets from the dataset.

From a statistical point of view the number of topics in a randomly sampled twitter data exhibits high positive

correlation with the number of topic in the dataset when the randomly selected sample is large enough [37].

Second, each tweet in the dataset was annotated by three independent annotators into one of the categories, identified in the First step. A Tweet's topic (category) was decided by majority vote. If each of the annotators assigned a tweet a different category, preventing a majority, the principal researchers assigned the category.

3.3. Dataset Description

Using the Twitter stream API, we collected into the *initial* dataset during March 2017 all public tweets either mentioning (i.e., @username) the username of the president of the United States (US) or the username of a member of the US Congress. The accounts of users who posted the tweets were also collected. The initial dataset was selected since it contains tweets on US politics with a potential to go viral, allowing us to select active users. Since the discussed topics in the initial dataset are somewhat biased toward politics, we counter such bias by analyzing tweets on any topic of users who posted a tweet in the initial dataset.

We used the Tweepy Python library to create a stream session and routed tweet messages to a StreamListener instance. We saved the tweets as json files for further processing with the R and Python software, as described next (Figure 2):

1. Select randomly a user who posted a tweet in the initial dataset. To avoid "celebrities" who tend to go viral, we include that user if the user's Followers and Friends-lists' sizes range between 100 and 1,000. To avoid inactive users, accounts with less than 100 Followers were omitted. We curated the analyzed dataset in Steps #2 to #5 during October 2018, starting with Step #2 which is designed to counter the bias in the initial dataset toward politics topics.

2. Following [38] we collected the 200 most recent tweets posted by each of the users selected in Step #1 without being limited to a specific topic like politics. The older a tweet, the more likely that its retweet list will be missing a user (blocked, deleted, suspended, or protected), which can negatively effect the analysis of contagion spreading paths. Selecting the most recent tweets minimizes the chance for a biased view. The next step is designed to select non-viral tweets from the tweets collected in step #2.

3. Obtain only tweets with 10 to 100 retweets (RTs) for each tweet collected in Step #2. The upper bound of 100 RTs was selected following [11, 17] who indicated that viral posts infect more than 100 nodes. One month after the collection date, we checked the RT-count of each tweet captured in this step to verify that a tweet did not become viral, discarding the viral ones.

4. Collect Twitter accounts of users who retweeted an original tweet in Step #3. Throughout this process we waved out suspected Bots (automated agents) with the botcheck R package which calls the Botometer API [39]. Only users with a score bellow 3 were selected, presenting an average score of 2.3 out of 5. The lower the score the less likely a user is a Bot. The removal of Bots is necessary to learn human behavior.

5. Collect the Followers list of users from Step #4.
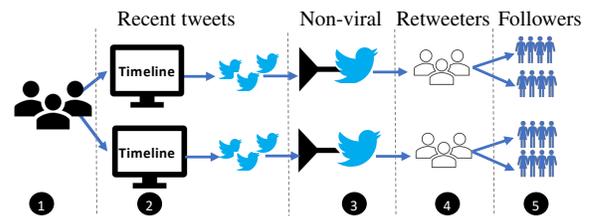

Figure 2: Data collection steps (# denoted in a circle).

4. Analysis and Results

Following Section 3.3, in Step #1 we selected a seed of 200 users from the initial dataset. During October 2018 we curated the analyzed dataset by following Steps #2 to #5. The 200 users allowed us to collect and analyze a large dataset (elaborated next) of retweet contagions events, similar to other studies on contagion [4, 7, 5, 6]. The set of selected users had Followers and Friends-list sizes between 100 and 1,000. In Step #2 we collected the most recent tweets of selected users. Step #3 has resulted in 3,901 tweets, out of which 745 tweets were discarded since their RT count grew over 100. Thus, the remaining 3,156 tweets represent non-viral content. Step #4 resulted in 25,607 users who retweeted 172,206 times. Finally, Step #5 resulted in 11,038,902 users.

To summarize, we analyzed 3,156 tweets that were retweeted by 25,607 users, 172,206 times. The retweeting users had a total of 11,038,902 Followers.

Given a network, a local contagion begins at some node (user) and then spreads from user-to-user over the edges of the network, creating a cascade [40]. Typically, cascades are measured on the activity network [40], ignoring information about users' social relationships

(i.e., social network), which can explain the observed spreading patterns of a contagion [17]. To cope with RQ #1, and better understand human behavior online in spreading non-viral content, we start by detecting a contagion using the activity network (G_{Tw}), and measure the depth of a cascade on the social network (G). The cascade depth is defined as the largest distance the information spreads from the user originating the information [40].

To detect the cascade depth of w, each user retweeting w was assigned a distance (d) on G from the user v_0, who originated w. An infinite distance (INF) was assigned in the absence of a path on G. Distances were calculated using the R igraph package [41]. Figure 3a presents a distribution of the distances from v_0. Since a cascade is defined as a user-to-user spread, the depth of a cascade for each w is $max(\{d|d < INF\})$.

A path might exists at larger distances (e.g. $d = 10$) on the Following circles that we collected in Step #5. However, such contagion is global and thus, does not affect the analyses that follow.

Dow et al., [17] found that contagion occurred deeper than one level into viral cascades. We find similar results regarding the spread of contagion with non-viral content (Figure 3a). The results show that non-viral content has a contagion cascade with depth $d \approx 9$. If a path exists on G, the more distant a user is from the source v_0, the less likely she is to retweet w (Figure 3a). As expected, most contagion events are local ($d = 1$).

To provide a view of the global contagion, we need to consider the temporal order of the retweeting events. For example, consider a slight variation in Figure 1 where v_0 is the source of w; v_3 retweets w at time t_1 before v_2 retweets w at time t_2. This order of events implies v_3 was infected globally.

To account for such cases, each infected node was assigned a contagion type (local/global) by analyzing the activity network and the social network as described in Section 3.1. Then, we computed the minimal distance on G from any user $v_i \in V_{Tw}$ who retweeted w at time t_k (V_{Tw} are adaptors), to any user who retweeted w before t_k, or to the user who originated w (Figure 3b). Returning to Figure 1, consider a case where the tweeting source is v_0, and nodes v_0, v_3 are the only users in V_{Tw} at time t_{k-1}. Next, at time t_k global contagion spread to v_5. The minimal distance from any user in V_{Tw} before t_k, to v_5 is $d(v_3, v_5) = 2$, or $d(v_0, v_5) = 2$.

In Figure 3b, $d = 1$ indicates a local contagion. As expected, this is the most frequent contagion mechanism. Whereas, $d > 1$ indicates a distance that was measured from the closest adopter to a user who was infected by a global contagion.

In Figure 3b, for $d \leq 6$, the closer a user to an
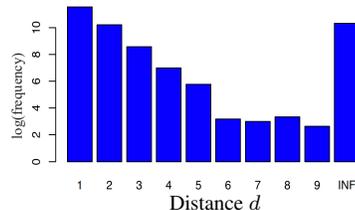
(a) Histogram of distances from the tweeting source.
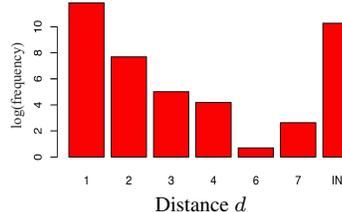
(b) Histogram of distances from the closest adaptor.

Figure 3: Histograms of measured distances.

adaptor, the more likely that user will retweet. This finding supports the six-degree of separation theory [42]. An interesting twist of a trend for $d > 6$: the larger the distance, the more likely a user will retweet.

Most users are aware of information up to a social distance of three hops [43]. Thus, the trend observed for $d \leq 6$ is in line with previous scholarship and it is likely that a user who is retweeting at $d \leq 6$ (Figure 3b) was exposed to content through locality, through manually crawling through the feeds of friends (and friends of friends), or through exposure to the Twitter Timeline containing tweets of accounts that a user follows.

On the other hand, contagion occurring with $d > 6$ is more likely to result from a global exposure by means of exposure to external sources (e.g. main stream media), or promoted content by the Twitter Timeline algorithm.

Recall that our dataset contains non-viral tweets which are less likely to be covered by mainstream media, and therefor we believe that the Twitter Timeline algorithm has exposed users and led to global contagion.

Twitter organises content on a user's Timeline in three main groups: 1) Tweets of accounts that a user follows, 2) Promoted tweets - purchased by advertisers, and 3) Tweets that are ranked by Twitter as having a large engagement potential [44]. Groups #2, and #3 can contain content by non-neighbors. Due to user low available time and limited attention, pushed information on the user's Timeline significantly increases contagion rate [21, 44]. Thus, global contagion is more likely to occur due to content promoted by Twitter's Timeline.

To better study the patterns by which contagion spread, we calculated the hourly time difference between every two consecutive retweets, i.e.,
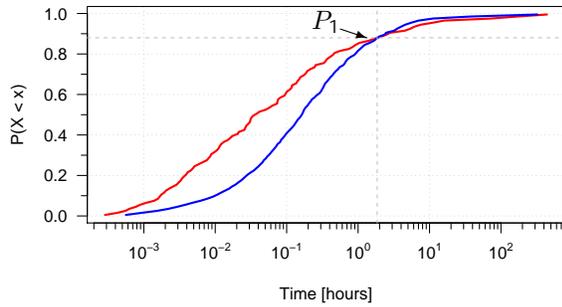
(a) CDFs of inter-retweet time (log time scale).
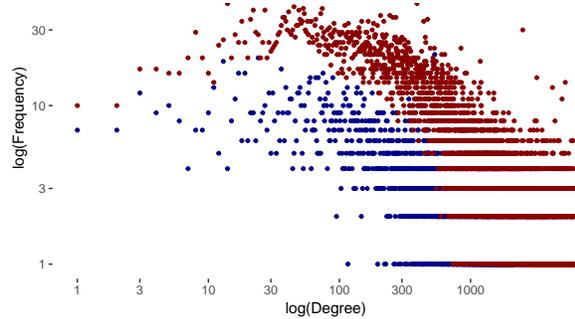
(b) Followers log-log distribution.

Figure 4: Patterns of global (red) and local (blue) contagion.

inter-retweet times. Figure 4a presents two cumulative distribution functions (CDFs) of inter-retweet times that correspond to each contagion type (local and global). The two CDFs show different trends, uncovering different patterns of user behavior.

For global contagions, 80% of inter-retweet times were less than 0.44 hours (27.35 minutes), while for 99% of them, the inter-tweet time was less than 305.96 hours (12.74 days). For local contagions, 80% of the inter-retweet times were less than 0.91 hours (54.6 minutes), whereas, for 99% of them, the inter-tweet time was less than 180.94 hours (7.54 days). This finding in inline with previous scholarship [45], and shows that Twitter members interact in a bursty nature, where members send several tweets in short periods, separated by long periods of reduced activity [46].

We observe a change in Figure 4a at point P_1. Up to point P_1 users retweeting due to global contagion respond quicker than those infected locally. At P_1, 88% of the inter-retweet times of users were less than 2 hours. After P_1 global contagion spreads slightly slower than the local contagion but lasts longer.

A possible explanation for this finding is that local contagion spreads slower than global contagion in the initial phase since: 1) In local contagion a user has to wait for the information to reach her through neighbors,

and 2) Users' inclination to retweet non-viral content is limited and governed by friendships.

While contagion is often perceived as spreading like an infectious disease, social contagion differs in that OSN users actively decide to spread it. On the other hand, users who are exposed by a global source are not restricted by a neighbor-to-neighbor mechanism.

We also study the effect of the number of Followers on the contagion types. Users who retweeted due to global contagion have more Followers than those who retweeted due to local contagion (Figure 4b). A user with many Followers acts as an information source. To maintain this reputation, a user has to actively consume diverse and innovative information beyond her local neighborhood [47].

To cope with RQ #2, we pursue a more fine-grained analysis between local and global contagion, by analyzing next, user retweet activity by topics.

4.1. Retweet Analysis by Topics

Following the First step in Section 3.2 we inspected a random sample of 1,000 tweets from the dataset which reflects the topic distribution within the dataset with 95% confidence level. We identified 19 topics in the sample as presented in Figure 5. In the Second step, each tweet in the dataset was annotated by three independent annotators into one of the 19 topics, identified in the First step (Table 1). Fleiss's Kappa [48] of inter-rater agreement score (0.68) indicates an adequate agreement.

Figure 5a shows that the size of global contagion varies across topical categories. The three most contagious categories are Politics, Sports, and Celebrity. Breaking contagious to types, we find that Politics has only 17% of global contagion (Figure 5b). Several studies [2, 49] found that political related topics spread the most. Our analysis uncovers similar findings for the spread of local contagions (Figure 5b). However, global contagion shows different results. A low global contagion in Politics may be the result of "echo chambers", i.e., restricted exposure to opinions of like minded users [3], tending to retweet each other. Technology, Community, and Sports related topics enjoy the highest percentage of global contagion. Since this study focuses on non-viral content, the analyzed tweets are less likely covered by mainstream media. We hypothesize that in the absence of exposure by mainstream media, and the existence of global contagion, individuals are exposed by other global exposure sources.

We further analyze the density distributions of the number of global (Figure 6a) and local (Figure 6b)
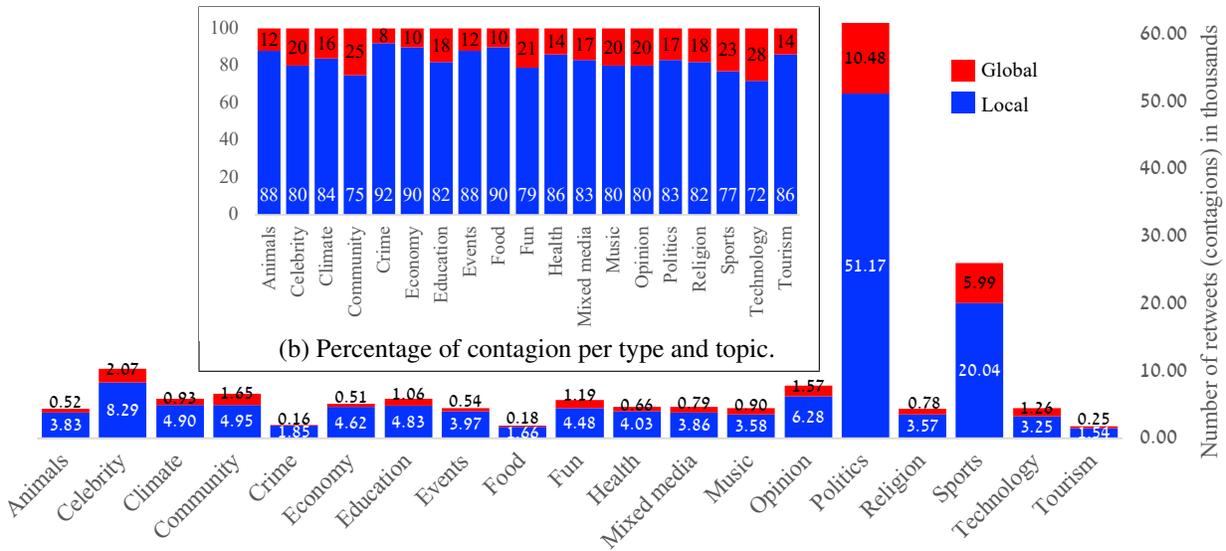
(a) Sum of contagions per type and topic.

Figure 5: Patterns of global (red) and local (blue) contagion.
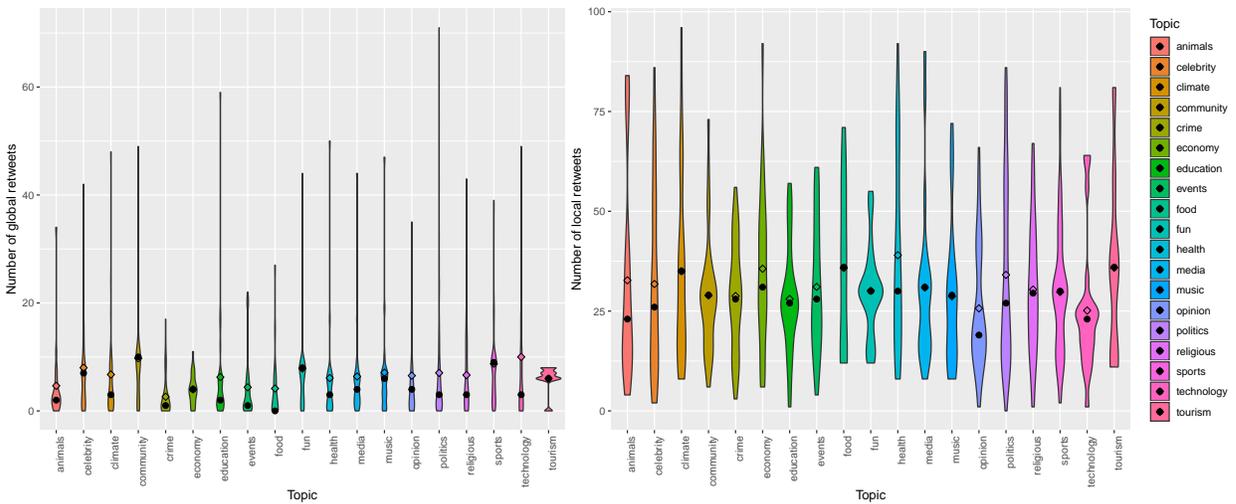
(a) Density distributions of # global contagions per topic.

(b) Density distributions of # local contagions per topic.

Figure 6: Density distributions of the number of contagions by type. Black circles: median; Diamonds: mean.

contagions (retweets) across different topics, illustrated by violin plots. A violin plot allows the comparison of multiple density distributions, similar to a box plot with a rotated kernel density plot on each side.

From Figure 6a, 6b it can be observed that within each topic, the local and global retweet distributions present different shapes. Using a Kruskal-Wallis rank (KW) test [50] we found significant differences in the density for both local ($P_{value} = 22 \quad 10^{16} < 0.05$), and global ($P_{value} = 3 \quad 10^{1} < 0.05$) contagions across topics. To detect where those differences lie,

we conducted two Conover-Iman tests [51] with the Benjamini-Hochberg method. Figure 7 presents the P_{value} of each test, one for the global contagion, and the other for the local contagion. Considering local contagion, most topic groups are not significantly different in their number of retweets (indicated by red and yellow squares in Figure 7). The low P_{value} in the KW test regarding local contagion was achieved due to significant differences between a small number of topical categories. For example, Technology differs from Tourism (Figure 7). Whereas for global contagion,

Table 1: Definitions of categories used for annotation with examples.

Category	Definition	Example
Animals	Requests to save animals, reports of animal abuse, photos of animals.	"#Blackfish be careful out there. They are taking our' fish."
Celebrity	The subject is a person or a group. Political figures were classified as "Politics". Sports teams were classified as "Sports".	"Had the pleasure of interviewing the fabulous @JoeCongelAuthor."
Climate	Whether, global warming, green environment.	"We broke down climate change by region."
Community	Public requests such as donations, giving advice, sharing feelings with others, erotic texts, complaints, empowerment.	"Ok dear friends, if you're thinking of an organization to support next year, pls consider Culture of Life Africa."
Crime	Posts on illegal actions, trials, and laws.	"Accuse an 82 grandmother of vandalism?"
Economy	Posts on business issues e.g. deals, trade, consumption of goods, and investments.	"The UN spends more than 80% of its annual peacekeeping budget in Africa."
Education (Edu)	Research results, books, schools, universities, historical facts, quotation of wisdom comments.	"Afghanistan museum unveils restored ancient Buddha."
Events	News, current and live events reported by users (e.g. earthquake).	"#TrafficAlert I15 NB. Approx 100 gallons of diesel across all lanes."
Food	Experiences regarding food and drinks (e.g. restaurants recommendations).	"I'm hungry, let's go grab food. How does Lazy Moon sound?"
Entertainment (Fun)	Humor, spare time, hobbies, jokes, short funny videos, fashion, lifestyle.	"In the ER, a guy wearing a jacket that says, "Invincible," he does not seem"
Health	Any health-related subject (e.g. a symptoms).	"436 hospitalized due to McDonald's salads."
Mixed Media (Media)	Texts about movies, main stream media, or social media.	"Star Wars: Clone Wars Revived for Surprise Season."
Music	Songs, albums, bands, music,and playing tools.	"Drop your favorite song '?' in the comments."
Opinion	Opinions on any subject not belonging to other categories, with inconclusive finding or facts.	"People in LA are the best."
Politics	A political figure, a party or movement, elections, a location that is the subject of controversial political discussion, or an issue that is the subject of a political discussion.	"More Americans would still vote for Trump again than Clinton".
Religion	Texts about users' beliefs, often containing the words: church, pray, God, Christian, and bible.	"God does not forsake the work which his own hands have begun."
Sports	Names of sports teams, leagues, sport events, fan groups, or references to news items on sports.	"IT'S DONE!! Welcome, Las Vegas Raiders!"
Technology	Names of Web sites, applications, devices, or events about any of these.	"change your password immediately" if you stayed at a Marriott Hotel."
Tourism	Holiday, and vacation recommendations.	"Travel guys... while you're able to."

Definition taken from [2]; In absent of a majority vote among annotators, the authors assigned the category.

most topic groups significantly differ from one another, as indicated by the absent of numbers in Figure 7.

5. Conclusions

Most contagion research is bound by modeling only local contagion, and the spread of viral information. This study focused on modeling contagion spread on Twitter by users' engagement in retweeting non-viral content. Contrary to the common assumption that contagion diffuses from node-to-node, we find that contagion in OSNs also spreads globally, beyond social network links. In addition, we find that the fractions of global and local contagion differ between topics. The percentage of global contagion on Twitter ranges from 8% in Crime to 28% in Technology related topics.

Significant differences in the spread of global and local contagion of non-viral information, imply on different mechanisms of contagion at work. Two behaviors are observed in the spread of global
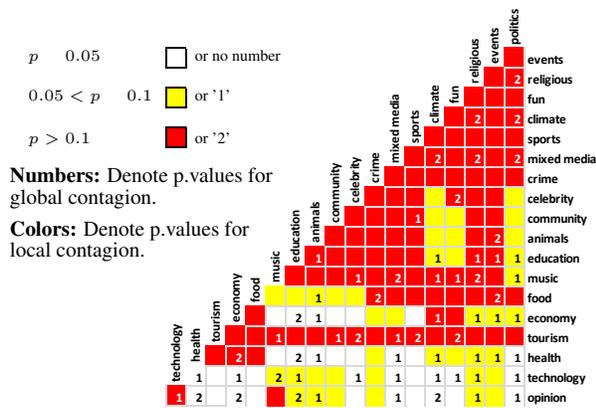
Figure 7: P_{values} of Conover-Iman test.

contagion. First, at distances $d \not\in [2,6]$ on the social network from a user who adopted the information: the larger the distance the less likely a user will be infected. Second, when $d > 6$ the trend is reversed. When $d > 6$ users are less likely to search information by a user-to-user page crawling. Thus, contagion more likely occurs due to content promoted by external sources such as algorithms, mass media, or searching information in the Twitter search box (i.e., jumping between pages).

We also analyzed the temporal retweeting activity of users and find that global contagion spreads faster than local contagion, indicated by short inter-retweet times up to point P_1. At point P_1, 88% of the inter-retweet times were less than 2 hours, indicating that global contagion is more bursty than local contagion. This might be due to a blocking mechanism of local contagion where retweeting non-viral content is less appealing. However, after point P_1 the trend is reversed.

Two main limitations of this study are 1) Human annotation has subjective judgments in topic assignments. Some categories definition can overlap, depending on the annotator's subjective reading. However, the level of agreement among annotators presented adequate results, and 2) There might have been retweets by deleted accounts which we cannot observe in a retweet list.

Whereas, most studies focus on user-to-user contagion, this study contributes to the knowledge on how non-viral contagion spread. Since most content is non-viral, analyzing it makes a valuable contribution to the OSN research. The novelty associated with this study is in better modeling user behavior in an OSN, accounting to local and global effects.

References

[1] E. Sun, I. Rosenn, C. A. Marlow, and T. M. Lento, "Gesundheit! modeling contagion through facebook news feed," in *Third international AAAI conference on weblogs and social media*, 2009.

[2] D. M. Romero, B. Meeder, and J. Kleinberg, "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter," in *Proceedings of the 20th international conference on World wide web*, pp. 695–704, ACM, 2011.

[3] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on facebook," *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015.

[4] S. Wakamiya, Y. Kawai, and E. Aramaki, "Twitter-based influenza detection after flu peak via tweets with indirect information: text mining study," *JMIR public health and surveillance*, vol. 4, no. 3, p. e65, 2018.

[5] A. Bartal, N. Pliskin, and G. Ravid, "Modeling influence on posting engagement in online social networks: Beyond neighborhood effects," *Social Networks*, vol. 59, no. 1, pp. 61–76, 2019.

[6] S. A. Myers, C. Zhu, and J. Leskovec, "Information diffusion and external influence in networks," pp. 33–41, 2012.

[7] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi, "Measuring user influence in twitter: The million follower fallacy," *Icwsm*, vol. 10, no. 10-17, p. 30, 2010.

[8] A. Jain, N. Lupfer, Y. Qu, R. Linder, A. Kerne, and S. M. Smith, "Evaluating tweetbubble with ideation metrics of exploratory browsing," in *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, pp. 53–62, ACM, 2015.

[9] E. Pariser, *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.

[10] R. Buettner, "A framework for recommender systems in online social network recruiting: An interdisciplinary call to arms," in *2014 47th Hawaii International Conference on System Sciences*, pp. 1415–1424, IEEE, 2014.

[11] S. Goel, A. Anderson, J. Hofman, and D. J. Watts, "The structural virality of online diffusion," *Management Science*, vol. 62, no. 1, pp. 180–196, 2015.

[12] A. D. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences*, vol. 111, no. 24, pp. 8788–8790, 2014.

[13] E. Katz, "The two-step flow of communication: An up-to-date report on an hypothesis," *Public opinion quarterly*, vol. 21, no. 1, pp. 61–78, 1957.

[14] J. Yang and J. Leskovec, "Modeling information diffusion in implicit networks," in *2010 IEEE International Conference on Data Mining*, pp. 599–608, IEEE, 2010.

[15] F. Wang, H. Wang, and K. Xu, "Diffusive logistic model towards predicting information diffusion in online social networks," in *2012 32nd International Conference on Distributed Computing Systems Workshops*, pp. 133–139, IEEE, 2012.

[16] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec, "Can cascades be predicted?," in *Proceedings of the 23rd international conference on World wide web*, pp. 925–936, ACM, 2014.

[17] P. A. Dow, L. A. Adamic, and A. Friggeri, "The anatomy of large facebook cascades," in *Seventh international AAAI conference on weblogs and social media*, 2013.

[18] D. R. Bild, Y. Liu, R. P. Dick, Z. M. Mao, and D. S. Wallach, "Aggregate characterization of user behavior in twitter and analysis of the retweet graph," *ACM Transactions on Internet Technology (TOIT)*, vol. 15, no. 1, p. 4, 2015.

[19] K. Subbian, B. A. Prakash, and L. Adamic, "Detecting large reshare cascades in social networks," in *Proceedings of the 26th International Conference on World Wide Web*, pp. 597–605, International World Wide Web Conferences Steering Committee, 2017.

[20] T. O'Reilly and S. Milstein, *The twitter book*. " O'Reilly Media, Inc.", 2011.

[21] N. O. Hodas and K. Lerman, "The simple rules of social contagion," *Scientific reports*, vol. 4, p. 4343, 2014.

[22] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in *Proceedings of the 21st international conference on World Wide Web*, pp. 519–528, ACM, 2012.

[23] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Transactions on the Web (TWEB)*, vol. 1, no. 1, p. 5, 2007.

[24] E. Bakshy, B. Karrer, and L. A. Adamic, "Social influence and the diffusion of user-created content," in *Proceedings of the 10th ACM conference on Electronic commerce*, pp. 325–334, ACM, 2009.

[25] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *10th International Conference on Data Mining (ICDM)*, pp. 88–97, IEEE, 2010.

[26] J. Kleinberg, "Cascading behavior in networks: Algorithmic and economic issues," *Algorithmic game theory*, vol. 24, pp. 613–632, 2007.

[27] A. Richterich, "karma, precious karma!karmawhoring on reddit and the front pages econometrisation," *Journal of Peer Production, 4(1).*, 2014.

[28] H. Chang, "A new perspective on twitter hashtag use: Diffusion of innovation theory," *Proceedings of the Association for Information Science and Technology*, vol. 47, no. 1, pp. 1–4, 2010.

[29] M. E. J. Newman, "The structure and function of complex networks," *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.

[30] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *ACM SIGMOD Record*, vol. 42, no. 2, pp. 17–28, 2013.

[31] W. Wang, H. Zhou, K. He, and J. E. Hopcroft, "Learning latent topics from the word co-occurrence network," in *National Conference of Theoretical Computer Science*, pp. 18–30, Springer, 2017.

[32] A. Bartal and G. Ravid, "Member behavior in dynamic online communities: Role affiliation frequency model," *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[33] S. Aral, L. Muchnik, and A. Sundararajan, "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks," *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, pp. 21544–21549, 2009.

[34] D. Liben-Nowell and J. Kleinberg, "Tracing information flow on a global scale using internet chain-letter data," *Proceedings of the national academy of sciences*, vol. 105, no. 12, pp. 4633–4638, 2008.

[35] V. Gómez, A. Kaltenbrunner, and V. López, "Statistical analysis of the social network and discussion threads in slashdot," in *Proceedings of the 17th international conference on World Wide Web*, pp. 645–654, ACM, 2008.

[36] X. Shi, J. Zhu, R. Cai, and L. Zhang, "User grouping behavior in online forums," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 777–786, ACM, 2009.

[37] F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley, "Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose," in *Seventh international AAAI conference on weblogs and social media*, 2013.

[38] J. Mahmud, J. Nichols, and C. Drews, "Where is this tweet from? inferring home locations of twitter users," in *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.

[39] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," *arXiv preprint arXiv:1703.03107*, 2017.

[40] Z.-K. Zhang, C. Liu, X.-X. Zhan, X. Lu, C.-X. Zhang, and Y.-C. Zhang, "Dynamics of information diffusion and its applications on complex networks," *Physics Reports*, vol. 651, pp. 1–34, 2016.

[41] G. Csardi, T. Nepusz, *et al.*, "The igraph software package for complex network research," *InterJournal, Complex Systems*, vol. 1695, no. 5, pp. 1–9, 2006.

[42] J. Guare, *Six degrees of separation: A play*. Vintage, 1990.

[43] R. T. A. Leenders, "Modeling social influence through network autocorrelation: constructing the weight matrix," *Social networks*, vol. 24, no. 1, pp. 21–47, 2002.

[44] N. Koumchatzky and A. Andryeyev, "Using deep learning at scale in twitters timelines." https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines.html, 2017. Accessed: 2019-05-17.

[45] Y. M. Kalman, G. Ravid, D. R. Raban, and S. Rafaeli, "Pauses and response latencies: A chronemic analysis of asynchronous cmc," *Journal of Computer-Mediated Communication*, vol. 12, no. 1, pp. 1–23, 2006.

[46] M. Karsai, K. Kaski, A.-L. Barabási, and J. Kertész, "Universal features of correlated bursty behaviour," *Scientific reports*, vol. 2, p. 397, 2012.

[47] M. Granovetter, "Threshold models of collective behavior," *American journal of sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.

[48] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica: Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[49] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, "Fake news on twitter during the 2016 us presidential election," *Science*, vol. 363, no. 6425, pp. 374–378, 2019.

[50] A. Vargha and H. D. Delaney, "The kruskal-wallis test and stochastic homogeneity," *Journal of Educational and Behavioral Statistics*, vol. 23, no. 2, pp. 170–192, 1998.

[51] W. J. Conover, "Practical nonparametric statistics," 1980.