



Emergent local structures in an ecosystem of social bots and humans on Twitter

Abdullah Alrhoun^{1,2} and János Kertész^{1,3*}

*Correspondence:
KerteszJ@ceu.edu

¹ Department of Network and Data Science, Central European University, Quellenstraße, 1100 Vienna, Austria

³ Complexity Science Hub, Vienna, Austria

Full list of author information is available at the end of the article

Abstract

Bots in online social networks can be used for good or bad but their presence is unavoidable and will increase in the future. To investigate how the interaction networks of bots and humans evolve, we created six social bots on Twitter with AI language models and let them carry out standard user operations. Three different strategies were implemented for the bots: a trend-targeting strategy (TTS), a keywords-targeting strategy (KTS) and a user-targeting strategy (UTS). We examined the interaction patterns such as targeting users, spreading messages, propagating relationships, and engagement. We focused on the emergent local structures or motifs and found that the strategies of the social bots had a significant impact on them. Motifs resulting from interactions with bots following TTS or KTS are simple and show significant overlap, while those resulting from interactions with UTS-governed bots lead to more complex motifs. These findings provide insights into human-bot interaction patterns in online social networks, and can be used to develop more effective bots for beneficial tasks and to combat malicious actors.

Keywords: Social bots; Bot-human ecosystem; Bot strategies; Network motifs

1 Introduction

Social bots are widely perceived in a negative light. Many academic and public opinions agree that they disrupt the information environment of online social networks (OSNs) with false or misleading information, distract and confuse users, influence political election outcomes [1, 2], or promote terrorist activities [3]. However, it's important to remember that social bots are simply automated programs that are capable of executing tasks on OSNs, such as posting content and engaging with other users [4]. It is the malicious intentions of some bot creators that have led to their negative impact on digital spaces [5].

Despite the negative reputation of social bots, their prevalence is on the rise. A recent study found that bots make up between 9 and 15% of active Twitter accounts [6] (though this was a big dispute between Twitter and Elon Musk in their legal battle before the acquisition [7]). Similar estimates have suggested that bots account for up to 15% of Instagram accounts and that 10% of Facebook's accounts are fake [8]. While many of these bots are engaged in malicious activities, not all of them are. Some perform useful services, such as sending vaccine appointment reminders or providing other important information in emergency situations [9, 10]. As the internet becomes an ecosystem in which humans and

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

bots coexist, it is important for researchers to understand how these human-bot interactions work.

To gain insight into human-bot interactions, gathering data and conducting experiments are the most promising methods. However, conducting experiments with social bots has become less common due to technical challenges and ethical concerns surrounding bots manipulating political opinions and deceiving humans. Nevertheless, some recent experiments have assessed the effectiveness of social bots in digital manipulation and spam dissemination [11], compared different infiltration strategies [12], measured the effect of bot activity on follower count [13], and investigated how Twitter users perceive bots compared to humans [14]. Other studies have utilized social bots to explore platform versus political bias on Twitter and create a contagion model to track information diffusion on the platform [15, 16].

While previous studies have shed light on the behavior and effectiveness of social bots, little is known about their impact on the structures of networks in which they operate. In this paper, we present an experimental study designed to address this gap by investigating the evolution of networks formed by social bots interacting with other users on Twitter. Specifically, we examine the local structures (motifs) of interactions between humans and bots, as well as the effects of different bot strategies on the networks' evolution. We aim to provide insight into whether the content or the users are more influential in shaping network structures, which may inform the design and deployment of more effective and beneficial social bots. This way we also hope to contribute to the ongoing debates about the nature of social networks.

For our study, we deployed six social bots on Twitter for a five-month period and analyzed the evolution of the networks they participated in using a network science framework. To understand these networks, we identified and examined the network motifs present within them. Network motifs, also known as "motifs" or "local structures," are statistically significant mesoscale structures that exist within larger graphs or structures [17]. Similar to social cliques, motifs are considered essential components for the higher-order organization of complex real-world networks [18]. They have been widely used to describe the dynamics of networks in natural systems, such as biological and ecological systems, and man-made networks, including power grids and social networks [19–21].

Network motifs have been extensively studied in online social networks (OSNs) like Twitter and Yahoo, they have been mainly used to map emotional expressions during emergency events [22], explain knowledge-sharing and question-answer patterns [23], and characterize opinion formation processes [24]. It is essential to note that network motifs are platform-specific, and their types and roles vary significantly from one OSN to another [25]. To date, there has been no research into network motifs in the context of human-bot interactions.

Our research has potential to contribute towards improving the effectiveness of bots in carrying out useful tasks, as well as combating malicious bot activities. It also offers insights into how social bots can coexist with humans by adapting to the structures of existing networks. Understanding the patterns of human-bot interactions in online social networks is crucial for this progress.

The paper is structured as follows: Sect. 1 describes the experimental setup, including the design of the social bots, their strategies for interacting with other users, and the content they posted. Section 2 provides an overview of the bots' networks and their evolution.

Section 3 analyzes and interprets the motifs in those networks. Finally, in Sect. 4, we discuss the implications of our findings and the limitations of our experiment.

2 Experimental setup

For this experiment, we created and released six bots on Twitter and monitored their interactions with humans over five months. These bots published content using a language model [26], and had the same interaction capabilities as a human on an OSN. Nevertheless, all of them were given specific “rules of engagement” and each pair of bots was assigned a different strategy that directed them to act in a distinct manner.

This experiment was composed of five main components: the environment, humans, bots, a mediator and a data collector. (1) The environment is the medium in which the experiment is conducted; in this case, Twitter. The platform’s policies and regulations on safety, privacy, authenticity and more define Twitter’s operating environment. (2) Humans are persons with a Twitter account who occasionally interact with the experiment. (3) Bots are automated agents: computer programs created for the experiment that communicate and engage in the chosen environment. (4) The mediator is a program that controls the bots’ communications and engagements with other bots and humans. (5) A data collection utility securely stores the data into a server for further analysis. This section focuses on the bots and mediator of this experiment, detailing their design methodology and specifications.

2.1 Building social bots

2.1.1 Perception, decision, and action framework

In order to create a social bot on Twitter, two steps must be taken: 1) writing a script with instructions for the bot’s behavior and 2) creating an application through the developer portal for the Twitter account [27]. This experiment employed a three-step logic framework of perception, decision, and action (see Fig. 1). Perception involves collecting and processing data, such as identifying trending hashtags, scanning user timelines for recent tweets, reading tweet text, and recording retweets and other engagements. In the decision stage, the bot chooses its next move based on insights from the perception stage. For instance, this decision is made by randomly selecting one or two most retweeted or liked tweets in one of the top five trending hashtags. Finally, in the action stage, the bot performs an action or series of actions. For example, if the bot decides to respond to a user’s tweet, it will request text from a content generator API, which will then be used as a reply. For more details on the implementation of the framework, the bots’ workflow and the data collection process check the supporting information (see Additional file 1).

2.1.2 Profile and activity settings of social bots

Giving bots a human personality in their public operating environment requires at least names and photos [28]. We chose neutral names for the bots’ profiles, such as “Philip Nolan” from Edward Everett Hale’s novel *Man Without a Country*. This name could be either real or fictional, thus reducing the chances of recognizing that the accounts belong to bots as well as avoiding declaring that the content is not generated by a human. Additionally, we used fictional profile pictures to avoid infringing on other people’s privacy by using real humans’ photos [29]. To avoid any bias, we did not make any information about location or gender (except possibly the name) public on the bots’ accounts. We also did not disclose whether these accounts were run by humans or bots.

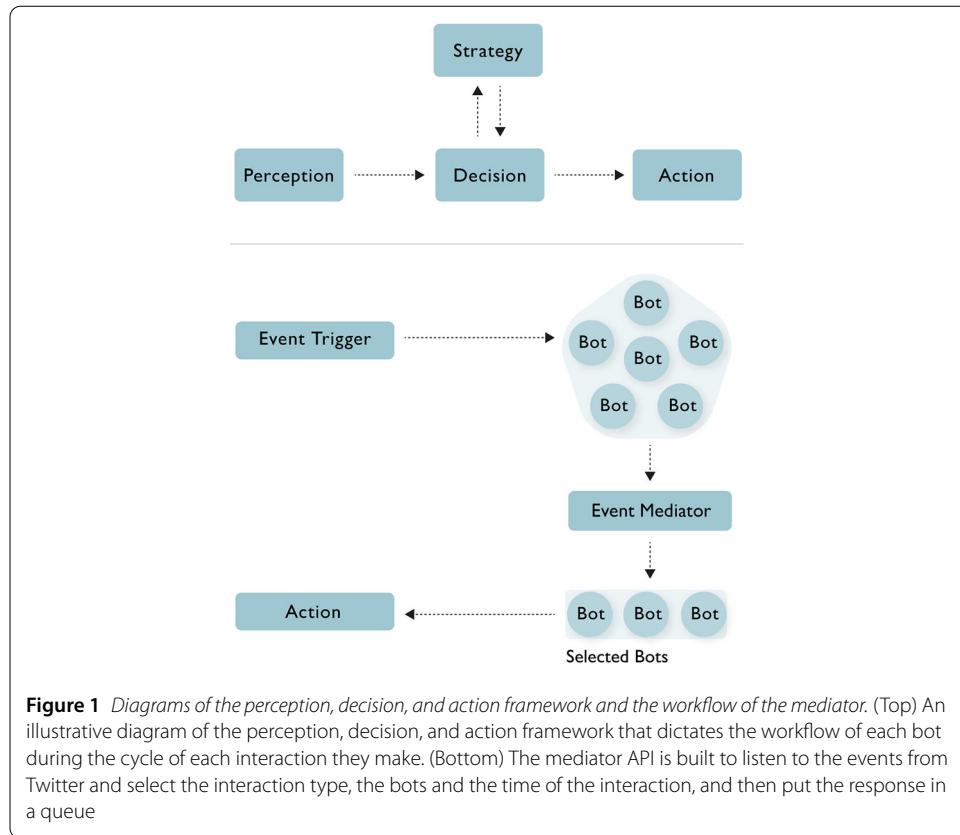


Figure 1 Diagrams of the perception, decision, and action framework and the workflow of the mediator. (Top) An illustrative diagram of the perception, decision, and action framework that dictates the workflow of each bot during the cycle of each interaction they make. (Bottom) The mediator API is built to listen to the events from Twitter and select the interaction type, the bots and the time of the interaction, and then put the response in a queue

The programming of the bots determines their “personality” and the way they interact with the environment. The bots were designed to take advantage of all the features that the environment provides; they could do anything on Twitter—tweeting, retweeting, following, unfollowing, replying, mentioning and liking—that a human user can. The bots were not told to follow each other, but during the process they began to follow one another.

2.2 Common rules of the bot’s engagement

The six bots were given three rules to abide by when interacting with other accounts and humans. Firstly, they could only respond to external prompts such as a reply to their tweet or an alert about new trending content. Secondly, the bots were instructed to not only reactively interact with other accounts but also post content proactively in order to look normal and interesting. Lastly, a mediator was responsible for deciding when the bots could interact. This was done for two reasons: first, without a mediator, the bots might interact with each other and others simultaneously, potentially violating Twitter’s spam policy [30]; second, the mediator helped to balance out the bot-bot interactions among and within strategies.

2.3 Overview of the interaction strategies

The bots in our experiment were each assigned a strategy to interact with users in their network. Each strategy was designed to ensure that the bot’s behavior was consistent and could help it gain the trust of human users. By varying the strategies of the bots, we were able to replicate the diversity of human users who interact in a variety of ways online. More importantly, each strategy was used as an independent variable to measure its effect

on mesoscopic network structures, our dependent variable. The strategies reflected different ways in which OSN users interact with other users; for example, some people are interested in trends and novelties while others are more interested in influencers and their published content.

We created three strategies and gave each one to two bots. The strategies are designed to observe and react differently to changes in the environment and to other bots' and humans' actions, leading to "intelligent" bot behaviors [31].

The bots with the Trends-Targeting Strategy (TTS) kept an eye on the top five hashtags from the global trends list on Twitter [32]. They would then post about these hashtags or interact with other users who were using them by liking or retweeting their content, or by following them. The idea behind this strategy is that engaging with a trending hashtag increases a bot's visibility and, as a result, its chances of becoming embedded in the network. Trends are like digital public gatherings, and in such gatherings, there is a greater chance of meeting new people. For example, after a football game, a trending hashtag will draw in people who are interested in sports to discuss, converse and maybe fight.

The second strategy, the Keywords-Targeting Strategy (KTS), was also content-focused but more passive. The bots using this strategy did not depend on current or upcoming trends and conversations, but rather on a consistent interest in certain topics. They would search other users' tweets for the presence of one or more of these topics and interact with them if found. They would also post tweets about their topics of interest. The idea behind this strategy was to find users with similar interests to interact with. KTS is based on the reasonable assumption that most people have a limited range of interests when browsing the internet and are less likely to engage randomly with any topic.

Twitter's advertising campaign guidelines suggest using a minimum of 25 keywords [33]. For this experiment, we opted to use 50 English keywords that were selected from the University of Vermont Complex Systems Center's research on universal positivity bias and happiness [34]. This study provided us with a list of non-polarizing keywords that fit with the purpose of our study. These included words such as *music, jokes, forests, family, cake, kitten, success, holidays, and beach*; a full list of chosen keywords can be found in the accompanying supporting information (see Additional file 1).

The User-Targeting Strategy (UTS) was also employed in this experiment, which involved bots following influential Twitter accounts. This was based on the idea that by doing so, the bot's exposure to their followers would increase, thus boosting its own popularity within the network. The influential accounts chosen were diverse and politically balanced, and their levels of popularity were comparable; some of these included Barack Obama, Donald Trump, Justin Bieber, Cristiano, Bill Gates and CNN. The influential accounts were identified by the number of followers they had regardless of why they had them. A full list of users is available in the supporting information section (see Additional file 1).

Our decision to employ these three strategies was based on two key considerations. Firstly, each strategy operates with a different level of focus. The TTS is more dynamic, continuously adapting to current trends and thus encompassing a broad array of rapidly changing, largely unrelated topics. Conversely, KTS narrows its focus to a defined set of 50 keywords, yielding a more limited range of topics. UTS is the most targeted of all, specifically centering on a select group of users. Secondly, the strategies can be classified into two categories based on their orientation: content-centric (TTS and KTS) and user-

centric (UTS). Our objective is to draw meaningful comparisons between networks that arise from a focus on content and those shaped by users' interests.

2.4 The mediator workflow

The mediator's role was to oversee and coordinate the bots' activities and interactions. This API was designed as a program within which the bots interacted autonomously, but they were aware of when to interact after receiving a signal to do so. The workflow followed a four-step process (see Fig. 1): (1) being notified of a triggering event, such as a reply, a mention or a new trend; (2) randomly selecting one or two bot(s) to act and randomly choosing the type of action; (3) if the chosen response involves text, requesting a response text from the content generation API; and (4) scheduling the bot's response or the responses of multiple bots, ensuring there is an interval between responses if more than one bot was instructed to respond. The mediator only governs the interaction API, which is the first part of the bots' logic. The tweeting API, which is the other part of their logic, works independently from the interaction API and according to each bot's strategy.

3 Network representation and data

To track the bots' activities, we created network maps of their interactions. This was done through snapshot representation, which allowed us to analyze the network as its structure evolves [35]. Snapshot representation creates a discrete-time sequence of networks, as shown in Eq. (1). We aggregated the interactions after one day, built the resulting network, and repeated this process for each subsequent day:

$$\mathcal{N} = \{N(1), N(2), \dots, N(t_{\max})\}, \quad (1)$$

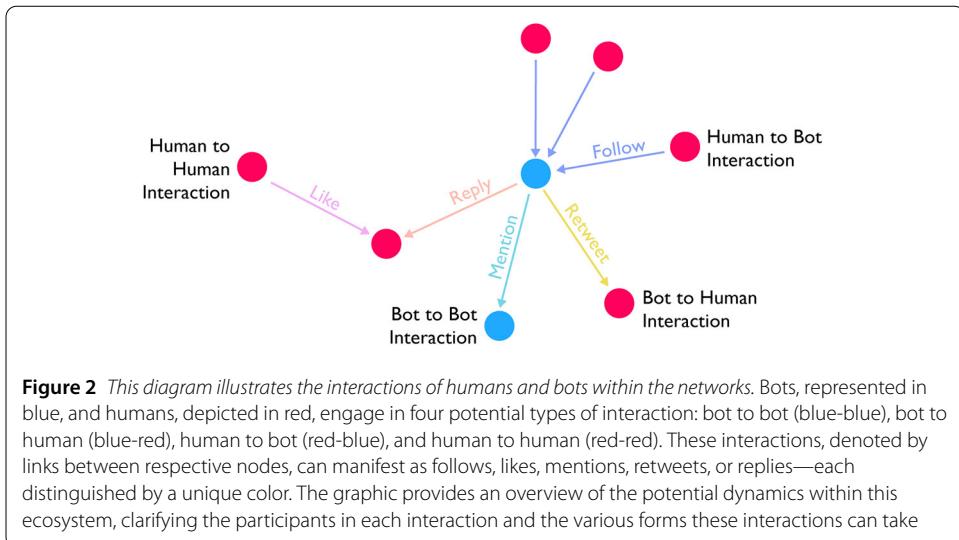
where $N(t_{\max})$ is the number of networks.

The experiment ran for five months (from July, 2020 to November, 2020), with unavoidable interruptions to data collection due to external constraints. The interruptions occurred due to two primary causes. First, Twitter occasionally restricted the tweeting and interaction APIs used by the bots, resulting in restriction periods ranging from 24 hours to two weeks during which the bots could not take any action. Second, there might have been technical issues in the data collection API which prevented full data collection but did not affect the bots activity.

Despite these challenges, our analysis was designed to ensure the validity and integrity of our results. Specifically, we filled any gaps in data with information from the preceding period based on our assumption that the bots' activity would likely remain unchanged until new data were received. This approach was employed as a means to maintain the continuity of the experiment.

As a result, we collected 75 days of data on bots using KTS, 95 days of data on bots using TTS, and 90 days of data on bots using the UTS, with significant overlap between the observation periods. The two most significant periods of inactivity were 15 days for the bots using KTS and 9 days for those using UTS. These were both due to suspensions from writing to the Twitter API. However, these suspensions were removed automatically.

We acknowledge that this may raise some concerns about the completeness of our data. However, it's important to note that these interruptions mirror real-world behaviors of human users on social media platforms who often experience their own periods of inactivity



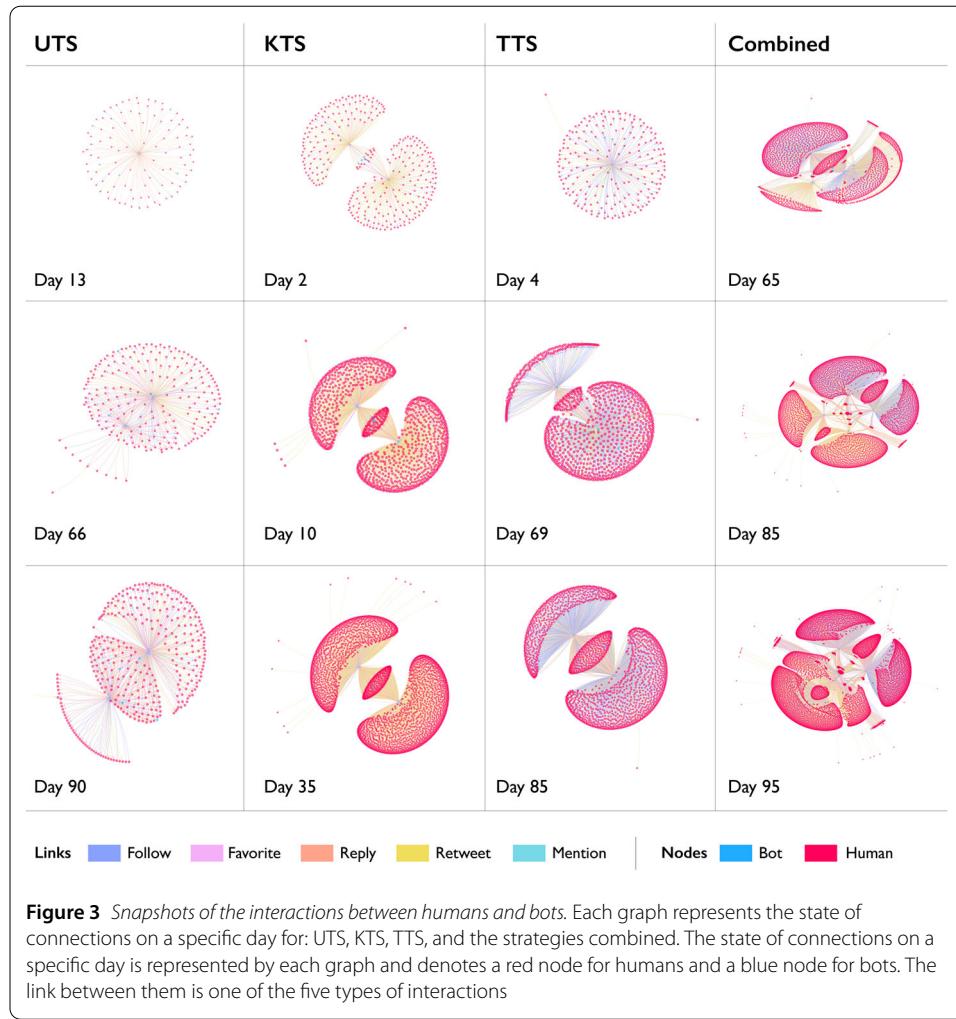
or suspension, so a pause in tweeting or interacting does not significantly compromise the integrity of our results. To further validate this approach and our results, we ran our analysis on both the complete data set and a sample with no data collection interruptions. Importantly, we found identical results in both cases, which adds a layer of confidence to our findings.

The data shows five types of interactions: *like*, *follow*, *retweet*, *reply*, and *mention*, and four settings of interactions: bot-bot interactions, bot-human interactions, human-bot interactions, and human-human interactions (see Fig. 2). However, we excluded 26 interactions executed by a human and targeted at another human from our analysis for two reasons. First, they are not relevant to this paper's research questions. Second, explaining how bots facilitated these human-human interactions requires special handling and different data collection methodology that is not centered around bots.

We also created bot-human networks for each strategy and a bot-human network that tracked the activity of all bots (see Fig. 3). Each node in the network represented either a bot or a human, while links between nodes indicated one of the five types of interactions, such as “likes” or “follows”. Our network map treated all forms of interaction equally, regardless of quantity or type. This is a simplified version of a more intricate reality. The quality of a link can vary depending on factors such as whether one user is following another user or simply retweeting their content, or if they are following a node with one hundred followers or one hundred thousand followers.

Altogether, the bots tweeted around 21,000 times and were involved in around 43,000 interactions. Figure 4 demonstrate that the TTS had the most incoming and outgoing interactions. The rewards, which are incoming interactions from humans to bots, are generally much higher with the TTS than with the other two strategies, except that the “incoming likes” are much higher in KTS. The UTS received the least rewards in all incoming interactions.

While bot-bot interactions played a relatively minor role in our analysis, we recognize their potential importance in future metaverse-like environments. However, our study focuses on analyzing interactions between bots and humans, in both directions. We found that “likes” and “replies” were the most common types of interactions between humans



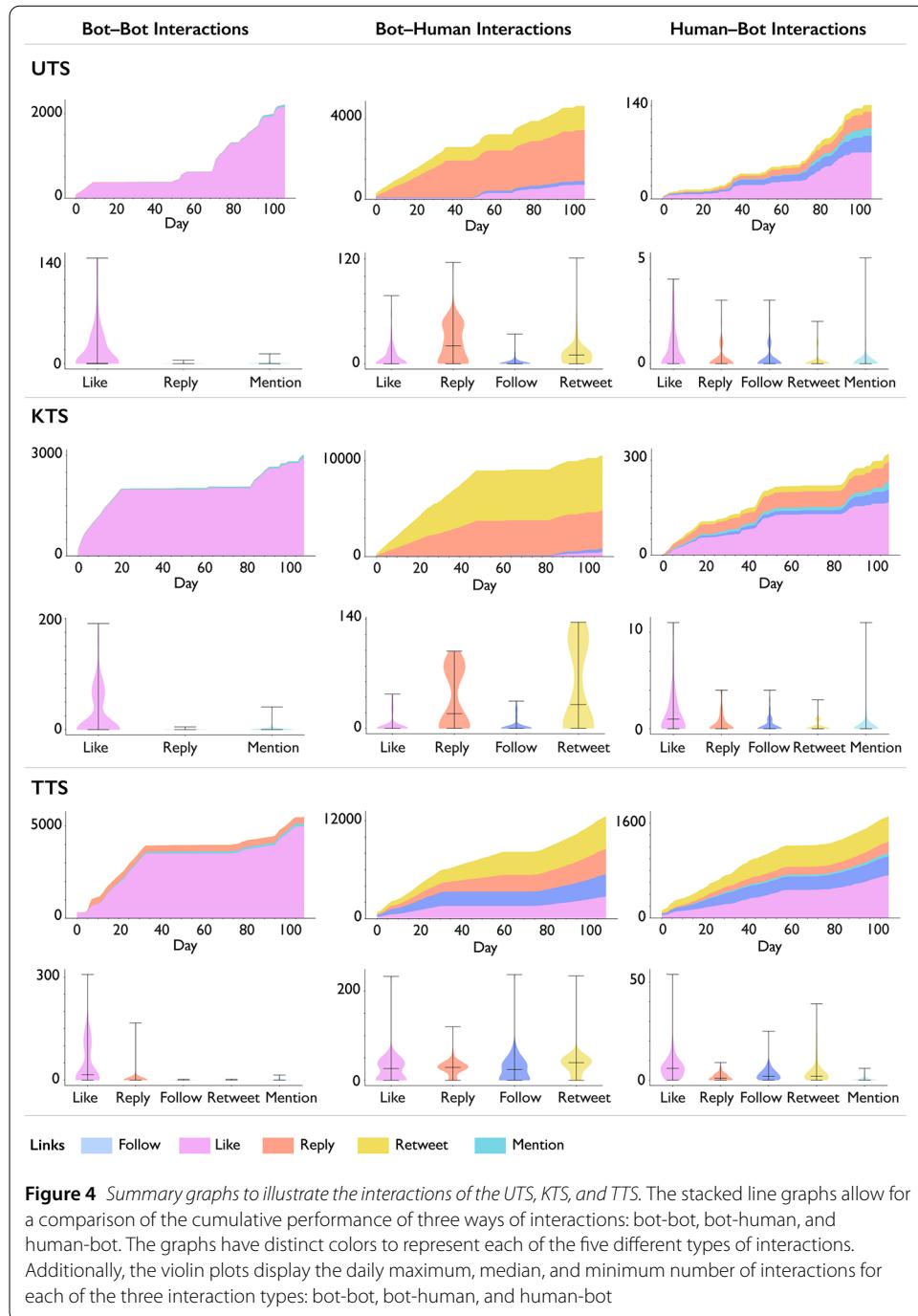
and bots, while “mentions” and “retweets” were less common. Other interactions fell somewhere in the middle. Our findings suggest that bot-bot interactions appear to be rather mechanistic and limited in diversity, interactions involving humans and bots are more varied and meaningful, highlighting the potential of online platforms to facilitate rich and engaging exchanges. This implies that although bots participate fully in the ecosystem, much of the complexity comes from the human participants (see Fig. 4).

4 Analysis of local structures

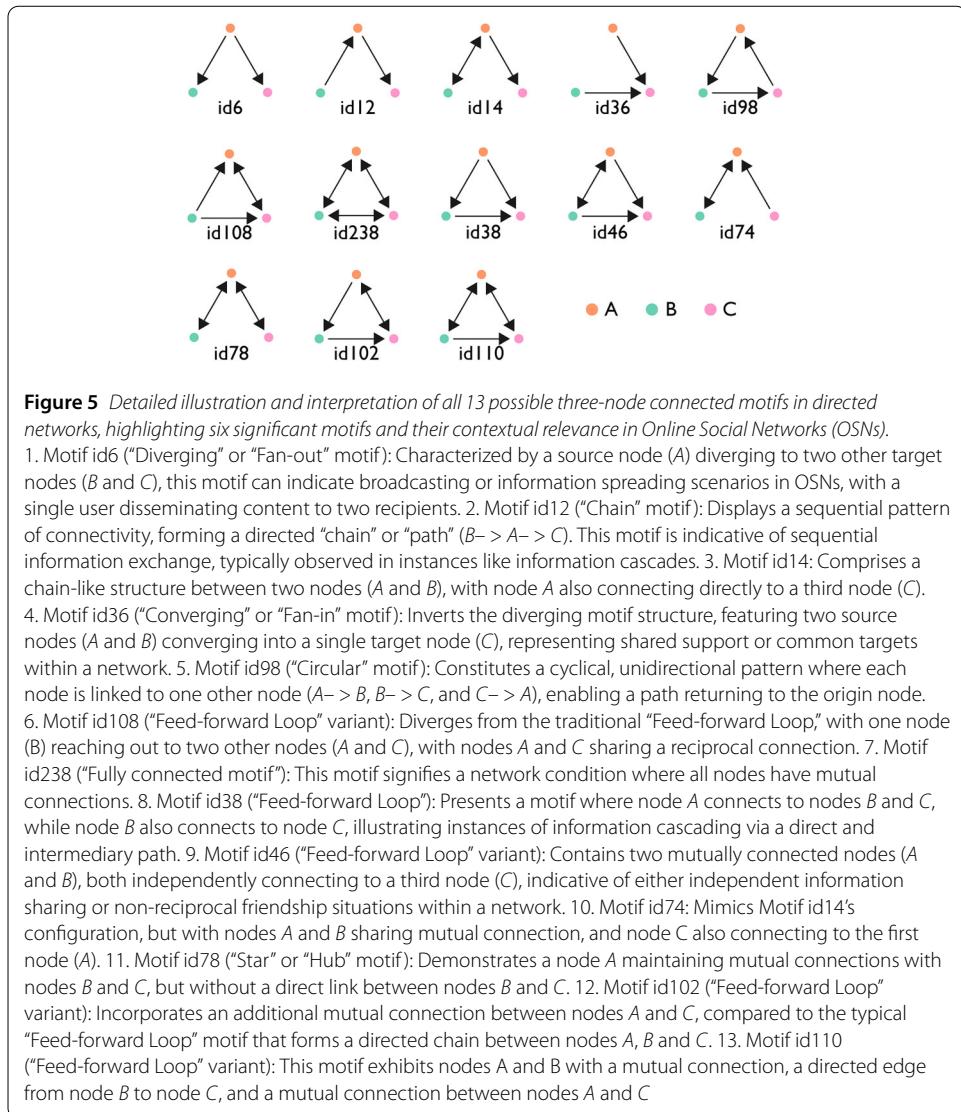
4.1 Detecting local structures

By examining local structures, referred to as motifs, in the networks, we were able to gain a better understanding of bot-human interactions’ patterns. This experiment focused on three-node motifs, which are those that involve three nodes (either human or bot) connected by a type of link (interaction). In directed networks, such as those found on online social networks, there are only 13 possible types of three-node connected motifs, which can be seen in the image below (see Fig. 5).

We chose to analyze three-node motifs for three main reasons. Firstly, they are much more prevalent in social networks than four- or five-node motifs. Secondly, the functions



of three-node motifs are more meaningful and straightforward in social networks; they represent the most basic level of group connection. Lastly, they require less computing power to identify than larger motifs. The 13 motifs identified in a network each have a unique shape and purpose, which can be used by individual nodes or users to increase their influence. Our experiment demonstrated that each motif has a specific function, and analyzing their prevalence in different networks is essential for understanding the strategies of bots.



We used the subgraph enumeration algorithm (ESU) to identify network motifs [36], ensuring that these motifs met the graph isomorphism condition. Upon detecting all motifs, we classified them into one of 13 pre-defined motif types.

To distinguish between significant motifs (those reflecting a pattern) and random occurrences, we employed the configuration model. This statistical model generates random graphs by reshuffling the edges of the original network while preserving the same degree distribution. The random networks, therefore, maintain the degree sequence of the original networks but have randomized structures. For more detailed information on the configuration model, please refer to the (Additional file 1).

We applied the ESU algorithm and the isomorphism condition to these randomized networks, enumerating and categorizing the motifs found within them. By comparing the frequencies of each motif type in the original networks and the randomized networks, we could identify which motifs in the original networks occurred with a significantly higher frequency.

Significant motifs are those that occur more frequently than would be expected by chance in the randomized networks. A motif was considered “frequent” if it appeared in the original network significantly more often than in the majority of randomized networks.

Motifs that did not meet this frequency criterion were deemed likely to be random occurrences, and thus less informative for understanding bot strategy performance. Following the construction of the networks, we carried out motif detection, classification, and analysis at regular intervals based on snapshots captured nearly daily.

We detect motifs when at least one of the nodes is a bot. The motifs defined above without the specification of the node types are called structural motifs. We have two kinds of nodes, humans and bots, which can be represented by different colors, leading to colored motifs.

4.2 Emerged local structures

Our analysis of the dataset revealed the presence of six out of the thirteen structural motif types: id6, id12, id36, id74, id46, and id38. Each of these motifs represents a specific pattern of interaction between bots and/or humans on social media platforms. Motifs id6, id12, and id36 have simpler relations between two nodes out of the three. For example, motif id6 can occur when one bot retweets the posts of two humans, whereas motif id36 results from two bots following the same human. In contrast, motif id74 is more complex; for instance, it can occur when two bots interact with each other and a human interacts with one of them. Motif id74 is complex because it involves more complicated interactions between nodes (i.e., have a reciprocal relation between two nodes out of the three). Motifs id46 and id38 have a closed-triangle structure with interactions flowing between three nodes.

Figure 6 illustrates the relative importance of motif types in the networks of UTS, KTS, and TTS. It is evident that different strategies produced distinct local structures with varying levels of significance. The UTS, KTS and TTS all shared two simple motifs, identified as id12, id36 and one complex motif identified as id74. Complex motifs, however, were not shared across the board. A motif is considered simple if it cannot close the triangle or has only one reciprocal link; a complex motif has a closed triangle and more than one reciprocal link. In summary, the UTS activated six types of motifs, the KTS activated four, and the TTS activated three.

Figure 6 also demonstrates the daily frequency of the significant motifs during the observation period. Most of the motif types had a steady but mild growth rate. However, id12 and id6 had a much more significant increase in growth compared to the other motif types, with an increase of an order of magnitude more than any other type. This suggests that these two motifs were dominant in the networks being studied.

The data also show a substantial similarity between TTS and KTS in terms of activated local structures. They both activated three motifs: id74, id12, and id36. This might be due to them being content-oriented. However, KTS and TTS were distinguished by the absence of the id6 motif in TTS. Motif id6 refers to a dissemination pattern where one user amplifies the posts of two different users. This pattern may manifest itself in a variety of ways, including combinations of likes and retweets to two other users, be they humans or bots. Even though TTS yielded higher rewards in terms of followers and replies, KTS demonstrated better results in terms of “likes” and in propagating the id6 motif.

In TTS, we noted the emergence of motif id74. This motif, characterized by two nodes interacting with a third node and only one node from the pair receiving engagement from

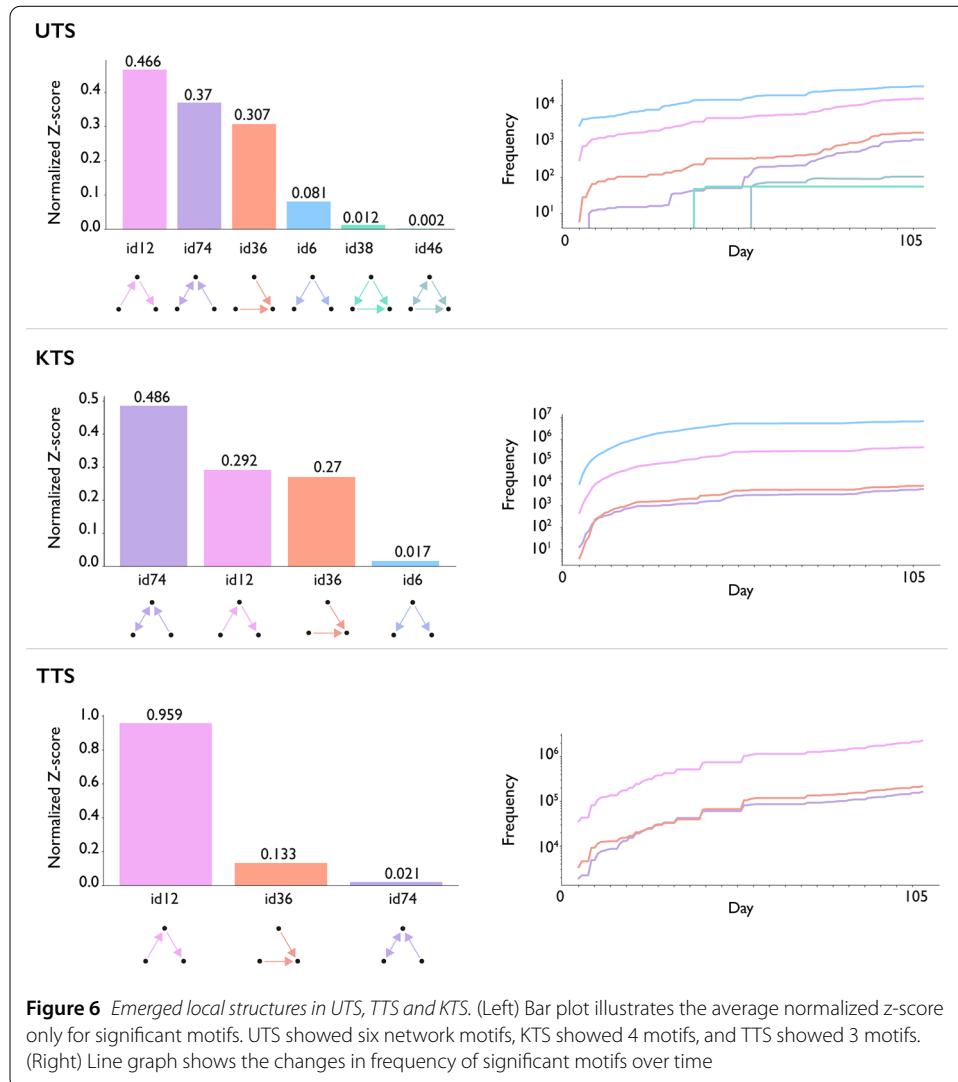


Figure 6 Emerged local structures in UTS, TTS and KTS. (Left) Bar plot illustrates the average normalized z-score only for significant motifs. UTS showed six network motifs, KTS showed 4 motifs, and TTS showed 3 motifs. (Right) Line graph shows the changes in frequency of significant motifs over time

the targeted individual, suggests a certain level of engagement focused on content in social media. However, the fact that TTS and KTS activated only this one complex motif, despite its significance, indicates a relatively weaker form of engagement compared to UTS.

The absence of motif id6 in TTS suggests that engagement, such as mentioning another user, doesn't automatically lead to wider content spreading like retweeting. This implies that TTS might engage actively but not necessarily disseminate content widely. On the contrary, KTS activates both id74 and id6 motifs, indicating an effective blend of engagement and content spreading. Thus, KTS not only involves direct interactions but also amplifies content, providing a balanced approach to social media interaction.

Both the UTS and KTS also shared the common characteristic of frequently activating the 'id6' motif. Motif id6 is a spreading pattern, and the results show that both a content-based strategy and user-based strategy are able to spread (amplify) content, when the focus is high. The difference between KTS and UTS lies in the former focusing on a broad array of topics instead of concentrating on individual users. Such observations may also suggest that focusing on targeted content also implies a limited number of users' interests.

UTS uniquely exhibited more complex motifs such as id38 and id46, pointing to the intricacies of user-focused engagement. The manifestation of these motifs in the UTS implies that a tighter focus, as embodied in this strategy, tends to amplify the level of engagement. This not only reiterates the correlation between the preciseness of an interaction strategy and the pattern of engagement within these networks but also emphasizes the importance of direct and bidirectional engagement in building interactive relationships.

In summary, the two primary patterns identified in our study, id6 and id36, demonstrate the most common behaviors exhibited by bots in social media: spreading, where a bot retweets multiple tweets, and targeting, where multiple bots follow a single human. These strategies are frequently used for marketing, propaganda, and misinformation purposes [37]. We found that all the strategies employed in our study demonstrated these patterns, except for TTS, which did not activate id6. However, all strategies activated id74, which indicates engagement due to the existence of a reciprocal link between two out of the three nodes. In KTS, motif id74 was the most significant, suggesting that a keyword-based approach is effective in increasing engagement between users.

Building upon our structural motif analysis, we incorporated a color-based methodology to more precisely identify the unique interaction patterns arising in the bot-human network. Inspired by prior research [38], we labeled nodes and edges with specific colors: nodes were designated either as “bots” or “humans,” while edges were colored to represent one of five types of interactions—retweet, like, follow, mention, or reply. This approach allows for a nuanced representation of network relationships, as shown in Fig. 7.

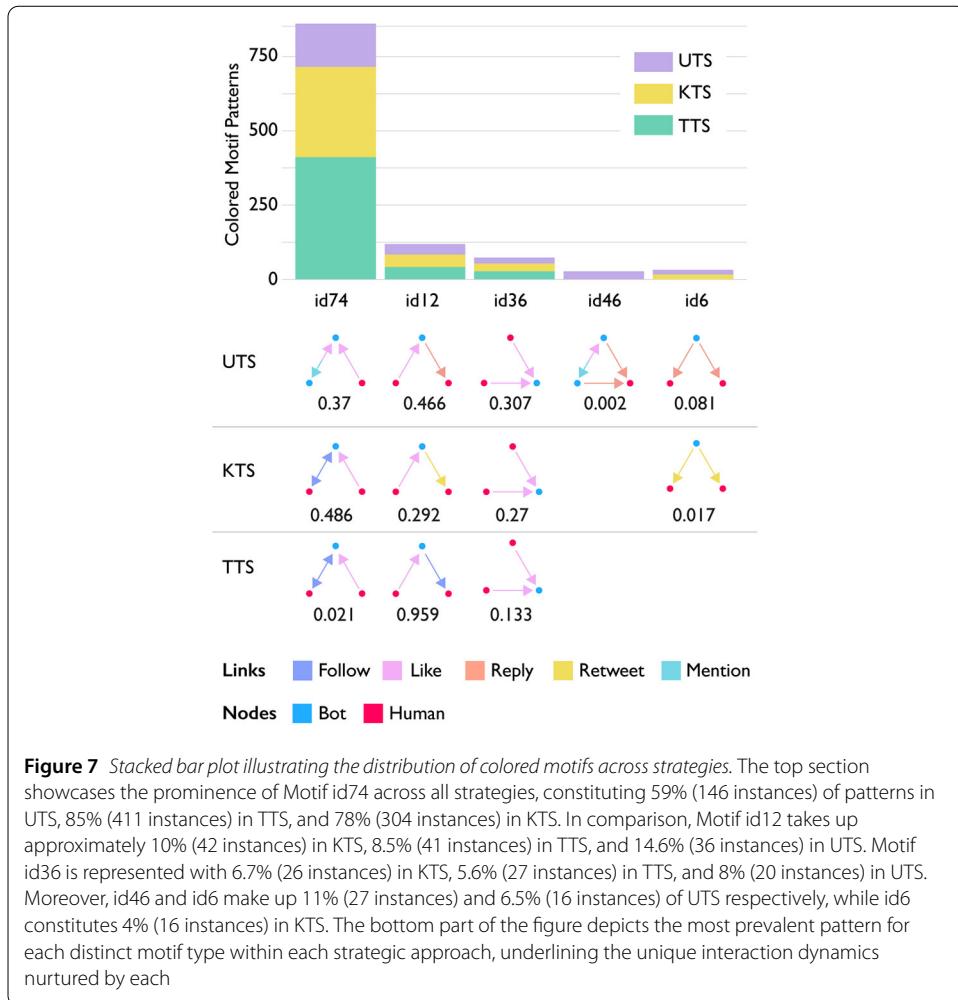
The color-based analysis unveiled a higher number of unique patterns in each strategy: TTS, KTS, and UTS yielded 479, 388, and 245 unique colored patterns, respectively. Intriguingly, across all strategies, the colored variations of motif id74 emerged as a foundational structure. Nevertheless, its distribution varied significantly: while it constituted 85% of all patterns in TTS and 78% in KTS, UTS exhibited a relatively diverse pattern distribution with id74 making up 59% of all patterns (see Fig. 7 for more details).

Despite its frequent occurrence, motif id74 is far from trivial. The id74 structure demonstrates a situation where two nodes (A and B) maintain a reciprocal relationship, typically a “friendship” on the platform, with a third node (C) only connecting with one of them (A). This structure carries the potential for future engagement between node C and B, either directly or indirectly through node A.

In TTS, for instance, the most common id74 pattern consisted of a bot and human in a mutual follow relationship, with another human liking the bot’s content. For UTS, the id74 pattern most often showed a bot and another bot engaged in a reciprocal relationship, with a human liking the first bot’s content. In KTS, the prevailing id74 pattern showcased a bot and a human in a mutual follow relationship, with another human liking the bot’s content.

These motifs, dynamic in nature, often evolve over time, transforming into different types. For example, node C, in motif id74 case, may later establish a direct relationship with node B, where node A acts as a bridge between B and C, which then leads to closing the triangle. Or, node C could propagate information to B through A. In this context of information cascades, such indirect relationships can be as impactful as direct ones.

A feature shared across these strategies is the motif id36, which typically displays two humans liking a bot’s content. This consistency suggests that regardless of the implemented strategy, bot-generated content has the capability to attract and engage human users.



Moreover, the analysis exposes a fundamental aspect that transcends these strategies: human-initiated interactions with bots are crucial for activating significant network motifs. For instance, the common patterns of motif id74 in TTS and id12 in KTS both originate with a human liking a bot's content, leading to subsequent interactions. Similarly, across all strategies, motif id36 shows two humans like a bot content. Thus, even though bots can influence the dynamics of these interactions, human engagement with bots appears to be the initial spark that ignites these patterns.

This ecosystem of humans and bots exhibits two key properties: it is self-organized, meaning that we did not instruct the bots to follow or retweet the same humans, yet they did so in a pattern that emerged organically. Additionally, the ecosystem is evolving in two ways: firstly, it is gaining slow but steady traction as more users engage with the bots; secondly, the interactions are progressing into more complex local structures, where interactions can have diverse and nuanced meanings.

5 Discussion

In this study, we explored the impact of social bot strategies on the local structures that emerged within the networks of humans and bots on Twitter. By creating six bots and observing the changes in network structures, we identified three-node motifs that reflected

patterns such as targeting users, spreading messages, propagating relationships, and engagement among users.

Our findings indicate that the most rewarded strategy (TTS) had the least diverse local structure patterns, while the least rewarded strategy (UTS) had the largest set of local structures, activating up to six different types of motifs. This suggests that the choice of bot strategies should be informed by the specific goal for which the bots are intended.

Furthermore, we found that content-related strategies (TTS and KTS) had a significant overlap in terms of local structures activated, while the user-oriented strategy was the only one to activate more complex motifs. Each of the three strategies could activate motif id74, a local structure consisting of two reciprocally linked nodes and an additional node interacting with one of these. This structure, serving as a foundational component for all strategies, bears potential to drive information cascades. Notably, the KTS shows promise in bridging the benefits of both approaches, by targeting content that resonates with particular users. This strategy has shown the ability to create engaging patterns while effectively disseminating content, key to achieving success on social media platforms.

It is worth noting that there is a debate about what defines a social network: is it the content or the users (i.e., their networks)? [39] Our findings suggest that the activation of more diverse motifs in UTS indicates the potential success of a user-focused social network, which prioritizes building connections between users. KTS' success highlights the importance of understanding the role of both content and user interactions in shaping the structure and dynamics of online social networks. While typically experiments require a large number of bots to influence individual opinions and demonstrate the impact of content on people, we conducted our experiment on a small scale and used non-sensitive material to maintain neutrality and technical simplicity.

It is also important to note that the visualization of these motifs do not reflect the temporal character of the links. While the existence of these motifs suggests that certain patterns of interaction between bots and humans may be more likely to occur in sequence, further study is needed to uncover the temporal-causal aspects of the motifs [40].

This study concentrates on uncovering local interaction patterns that emerge within networks comprising both bots and humans. While we intentionally omitted certain human-to-human interactions from our data analysis, we recognize their potential importance and the unique interaction dynamics they could reveal. The role of bots in shaping these omitted interactions presents a compelling avenue for future research, particularly with regard to how bots might alter these interpersonal dynamics and potentially influence the relationships established among human users.

We must also note that our research is contextualized within a specific timeline, preceding the managerial changes introduced by Elon Musk's acquisition of Twitter in 2022. As a result, the direct generalizability of our specific results to the evolving Twitter landscape might be limited. However, the value of our study extends beyond these findings and is also strongly associated with the methodologies and analytical approaches we employed. The use of network motifs to understand local interactions proves instrumental in comprehending user dynamics on digital platforms. While these motifs are platform-specific, the methodology itself is adaptable, providing a blueprint for researchers studying other social media platforms.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1140/epjds/s13688-023-00406-5>.

Additional file 1. Methodology Information. The attached supporting document contains the bots' detailed workflow and the used methods (PDF 218 kB)

Acknowledgements

We thank Professor Filippo Menczer for his suggestions and helpful discussion around the experiment.

Funding

The research reported in this work was partially supported by the EU H2020 ICT48 project "Humane AI Net" under contract # 952026 and by the European Union—Horizon 2020 Program "INFRIA-01-2018-2019" Grant agreement # 871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics." A.A. also thanks NSF Grant Accelnet "Multinet" Grant # 1927425 for support. This publication is the result of research conducted for Central European University, Private University. It was made possible by the CEU Open Access Fund.

Availability of data and materials

The datasets generated and analysed during the current study are available in the Harvard Dataverse repository, <https://doi.org/10.7910/DVN/ZSI5MH>.

Declarations

Ethics approval and consent to participate

Our research received an ethical compliance certification from the Ethical Research Committee of our home institution (Central European University). The certification states that the „research project titled "The Ecosystem of Humans and Bots in Online Social Networks" led by Abdullah Alrhoun (PhD, Network Science) had been reviewed for ethical research issues by the appropriate bodies, as defined by the Ethical Research Policy of Central European University, available at: <https://documents.ceu.edu/documents/p-1012-1v220>. Moreover, for the purpose of a planned replication of the experiment, we secured Institutional Review Board (IRB) approval from the ethical committee at Indiana University Bloomington (protocol # 14084), where one of the authors, Abdullah Alrhoun, was a visiting scholar. However, changes in the Twitter APIs prevented us from completing the planned replication. In terms of safeguards against harmful content, we implemented protective measures. We designed and deployed a monitoring code that oversaw their activities in real-time. The code removes any harmful content, violent content, and "obvious" false and fake information. In line with the ethical approval, we did not disclose the automated nature of the bots during the experiment. The exemption from this disclosure was to avoid introducing bias against the bots and thereby undermining the validity of the experiment. This measure ensured that our results reflected authentic user interactions.

Competing interests

The authors declare no competing interests.

Author contributions

Both authors designed the research. AA carried out the experiment and collected the data. Both authors analysed the data and contributed equally to the final manuscript. Both authors read and approved the final manuscript.

Author details

¹Department of Network and Data Science, Central European University, Quellenstraße, 1100 Vienna, Austria.

²Observatory on Social Media, Indiana University, Bloomington, IN, USA. ³Complexity Science Hub, Vienna, Austria.

Received: 2 April 2023 Accepted: 19 July 2023 Published online: 22 September 2023

References

1. Menczer F, Hills T (2020) Information overload helps fake news spread, and social media knows it. *Sci Am*. <https://www.scientificamerican.com/article/information-overload-helps-fake-news-spread-and-social-media-knows-it/>
2. Bessi A, Ferrara E (2016) Social bots distort the 2016 US presidential election online discussion. *First Monday* 21(11-7)
3. Alrhoun A, Winter C, Kertész J (2023) Automating terror: the role and impact of telegram bots in the Islamic state's online ecosystem. *Terrorism Polit Violence*. <https://doi.org/10.1080/09546553.2023.2169141>
4. Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. *Commun ACM* 59(7):96–104
5. Woolley SC, Guilbeault DR (2017) Computational propaganda in the united states of america: manufacturing consensus online. Computational Propaganda Research Project 22
6. Varol O, Ferrara E, Davis C, Menczer F, Flammini A (2017) Online human-bot interactions: detection, estimation, and characterization. In: Proceedings of the international AAAI conference on web and social media, vol 11, pp 280–289
7. Duffy C, Fung B Elon Musk commissioned this bot analysis in his fight with Twitter. Now it shows what he could face if he takes over the platform. CNN (2022). <https://edition.cnn.com/2022/10/10/tech/elon-musk-twitter-bot-analysis-cyabra/index.html>
8. Hepp A (2020) Artificial companions, social bots and work bots: communicative robots as research objects of media and communication studies. *Media Cult Soc* 42(7–8):1410–1426
9. Wagner K (2021) The surprising usefulness of vaccine bots. Bloomberg. <https://www.bloomberg.com/news/newsletters/2021-04-08/the-surprising-usefulness-of-vaccine-bots>

10. Hofeditz L, Ehnis C, Bunker D, Brachten F, Stieglitz S (2019) Meaningful use of social bots? Possible applications in crisis communication during disasters. In: ECIS, pp 1–16
11. Zhang J, Zhang R, Zhang Y, Yan G (2013) On the impact of social botnets for spam distribution and digital-influence manipulation. In: 2013 IEEE conference on communications and network security (CNS). IEEE, pp 46–54
12. Wang W, Chen X, Jiang S, Wang H, Yin M, Wang P (2020) Exploring the construction and infiltration strategies of social bots in sina microblog. *Sci Rep* 10(1):1–19
13. Fazil M, Abulaish M (2017) Why a socialbot is effective in Twitter? A statistical insight. In: 2017 9th international conference on communication systems and networks (COMSNETS). IEEE, pp 564–569
14. Edwards C, Edwards A, Spence PR, Shelton AK (2014) Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Comput Hum Behav* 33:372–376
15. Chen W, Pacheco D, Yang K-C, Menczer F (2021) Neutral bots probe political bias on social media. *Nat Commun* 12(1):1–10
16. Mønsted B, Sapiezyński P, Ferrara E, Lehmann S (2017) Evidence of complex contagion of information in social media: an experiment using Twitter bots. *PLoS ONE* 12(9):0184148
17. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. *Science* 298(5594):824–827
18. Benson AR, Gleich DF, Leskovec J (2016) Higher-order organization of complex networks. *Science* 353(6295):163–166
19. Schultz P, Heitzig J, Kurths J (2014) Detours around basin stability in power networks. *New J Phys* 16(12):125001
20. Braines D, Felmlee D, Towsley D, Tu K, Whitaker RM, Turner LD (2018) The role of motifs in understanding behavior in social and engineered networks. *SPIE* 10653
21. Hong-Lin X, Han-Bing Y, Cui-Fang G, Ping Z (2014) Social network analysis based on network motifs. *J Appl Math* 2014:874708
22. Kušen E, Strembeck M (2019) An analysis of emotion-exchange motifs in multiplex networks during emergency events. *Appl Netw Sci* 4(1):1–33
23. Adamic LA, Zhang J, Bakshy E, Ackerman MS (2008) Knowledge sharing and yahoo answers: everyone knows something. In: Proceedings of the 17th international conference on world wide web, pp 665–674
24. Coletto M, Garimella K, Gionis A, Lucchese C (2017) Automatic controversy detection in social media: a content-independent motif-based approach. *Online Soc Netw Media* 3:22–31
25. Topirceanu A, Duma A, Udrescu M (2016) Uncovering the fingerprint of online social networks using a network motif based approach. *Comput Commun* 73:167–175
26. Assenmacher D, Clever L, Frischlich L, Quandt T, Trautmann H, Grimm C (2020) Demystifying social bots: on the intelligence of automated social media actors. *Soc Media Soc* 6(3):2056305120939264
27. Veale T, Cook M (2018) Twitterbots: making machines that make meaning. MIT Press, Cambridge
28. Beskow DM, Carley KM (2019) Its all in a name: detecting and labeling bots by their name. *Comput Math Organ Theory* 25(1):24–35
29. Hill K, White J (2020) Designed to deceive: do these people look real to you? NY Times. <https://www.nytimes.com/interactive/2020/11/21/science/artificial-intelligence-fake-people-faces.html>. Accessed 22 Mar 2021
30. Twitter (2022) Twitter's platform manipulation and spam policy. Twitter. <https://help.twitter.com/en/rules-and-policies/platform-manipulation>
31. Howard PN, Woolley S, Calo R (2018) Algorithms, bots, and political communication in the US 2016 election: the challenge of automated political communication for election law and administration. *J Inf Technol Polit* 15(2):81–93
32. Twitter: trends overview | docs | Twitter developer platform. Twitter. <https://developer.twitter.com/en/docs/twitter-api/v1/trends/trends-for-location/overview>
33. Keyword targeting. Twitter. <https://business.twitter.com/en/help/campaign-setup/campaign-targeting/keyword-targeting.html>
34. Dodds PS, Clark EM, Desu S, Frank MR, Reagan AJ, Williams JR, Mitchell L, Harris KD, Kloumann IM, Bagrow JP et al (2015) Human language reveals a universal positivity bias. *Proc Natl Acad Sci* 112(8):2389–2394
35. Lambiotte R, Masuda N (2016) A guide to temporal networks. World Scientific, Singapore
36. Wernicke S (2006) Efficient detection of network motifs. *IEEE/ACM Trans Comput Biol Bioinform* 3(4):347–359
37. Khaund T, Kirdemir B, Agarwal N, Liu H, Morstatter F (2021) Social bots and their coordination during online campaigns: a survey. *IEEE Trans Comput Soc Syst* 9(2):530–545
38. Adami C, Qian J, Rupp M, Hintze A (2011) Information content of colored motifs in complex networks. *Artif Life* 17(4):375–390
39. Boyd D (2007) The significance of social software. NA
40. Kovanen L, Kaski K, Kertész J, Saramäki J (2013) Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. *Proc Natl Acad Sci* 110(45):18070–18075

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.